INFECTION AND IMMUNITY, Dec. 1987, p. 3228–3232
0019-9567/87/123228-05$02.00/0

Vol. 55, No. 12

Nucleotide Sequence of the Streptolysin O (SLO) Gene: Structural Homologies between SLO and Other Membrane-Damaging, Thiol-Activated Toxins

MICHAEL A. KEHOE,[1]* LORNA MILLER,[1] JOHN A. WALKER,[2] AND GRAHAM J. BOULNOIS[2]

Department of Microbiology, Medical School, University of Newcastle upon Tyne, Newcastle upon Tyne, NE2 4HH,[1] and Department of Microbiology, University of Leicester, Leicester, LE1 7RH,[2] United Kingdom

Received 4 May 1987/Accepted 8 September 1987

The complete nucleotide sequence of a cloned streptolysin O (SLO) gene and the amino acid sequence of SLO, predicted from the DNA sequence, are reported. SLO contains a single cysteine residue located close to the C terminus of the molecule and shares extensive structural homologies with other thiol-activated toxins, which allow us to predict functionally important features.

A variety of proteins from both procaryotes and eucaryotes lyse cells by forming hydrophilic channels in the target cell membrane (2, 4–6, 11). Even though the primary sequences of a number of pore-forming cytolysins are known (7, 9, 15), the molecular mechanisms involved in pore formation are not understood, partly because the functional domains in these very different sequences have not been defined. The thiol-activated toxins, in addition to being of interest as important bacterial virulence factors, are ideal models for studies on the mechanisms of pore formation. Although they are produced by taxonomically diverse species of pathogenic bacteria, they share biological and immunological properties, suggestive of a similar mechanism of action and common evolutionary origin (1, 5, 12, 24). All are lytic for a wide range of eucaryotic cells, reversibly sensitive to oxidation (thiol activated), irreversibly inhibited by cholesterol and related sterols, and immunologically cross-reactive. However, hybridization experiments using cloned thiol-activated toxin gene probes have failed to detect homology with DNA from species producing other thiol-activated toxins (17). This suggests that thiol-activated toxin genes may have undergone a considerable degree of divergence and that functionally important structures might therefore be identifiable as conserved regions in the primary amino acid sequences of toxins from different species. To facilitate detailed studies on structure-function relationships among thiol-activated toxins, the genes for streptolysin O (*slo*) and pneumolysin (*ply*) have been cloned and the *ply* gene sequence has been reported (17, 25). In this paper, we describe the complete nucleotide sequence of the *slo* gene, the predicted amino acid sequence of streptolysin O (SLO), and its relationship with the predicted pneumolysin sequence (25). This comparison reveals common structural features, which may be important in function.

The complete nucleotide sequence of both strands of the *slo* gene was determined by the dideoxy chain termination method (20), from a series of M13mp10 or M13mp11 clones containing specific and overlapping restriction endonuclease-generated fragments which were subcloned from the *slo* encoding region of plasmid pMK157 (17). Sequences were analyzed with the programs described previously (18, 19, 21, 22, 24). The *slo* gene sequence (Fig. 1) consists of an open reading frame extending from the ATG codon at base position 199 to 201 to the TAG codon at base position 1912 to 1914. Upstream from this open reading frame, there are sequences which resemble consensus ribosome-binding and promoter sequences (14, 16), including those compiled from other sequenced streptococcal genes (10), but in the absence of direct supporting evidence, caution should be exercised in assuming that these sequences are involved in the expression of the *slo* gene. Nine bases upstream from the designated *slo* open reading frame, there is a second, in-frame, ATG codon (position 190 to 193). At present, we do not know which of these adjacent ATG triplets is used as the initiation codon, but from the positions of the triplets with respect to potential ribosome-binding sites, we currently assume that it is the second.

The sequence of the primary *slo* gene product, predicted from the DNA sequence and starting at the second in-frame ATG codon, consists of 571 amino acid residues and has a molecular weight of 63,645. The 33 N-terminal residues possess all the consensus features of a gram-positive bacterial signal peptide (22), suggesting that they may be removed during secretion of SLO by *Streptococcus pyogenes*, to produce a 538-residue secreted form of the toxin with a molecular weight of 60,151. Native SLO appears to undergo a proteolytic cleavage subsequent to secretion, removing a segment with an estimated M_r of 7,000, to produce a predominant low-molecular-weight form of the toxin in culture supernatants (3, 17). The N-terminal amino acid sequence of the high-molecular-weight form of secreted SLO has not been determined, because it has not been possible to purify sufficient quantities of this form for sequencing studies. Therefore, the prediction of a signal peptide based on sequence data awaits direct confirmation. Although SLO is very difficult to purify to homogeneity, sufficient quantities of the highly purified low-molecular-weight form have recently been obtained to allow its four N-terminal residues to be tentatively identified as Ser-Asp-Glu-Asp- (P. Falmagne and J. Alouf, personal communication). The only similar sequence (Ser-Glu-Glu-Asp-) in our predicted SLO sequence occurs between residues 101 and 104 (Fig. 1). Ser is preceded by Lys, suggesting that a protease-sensitive site and the proteolytic removal of the 67 residues between the end of

* Corresponding author.

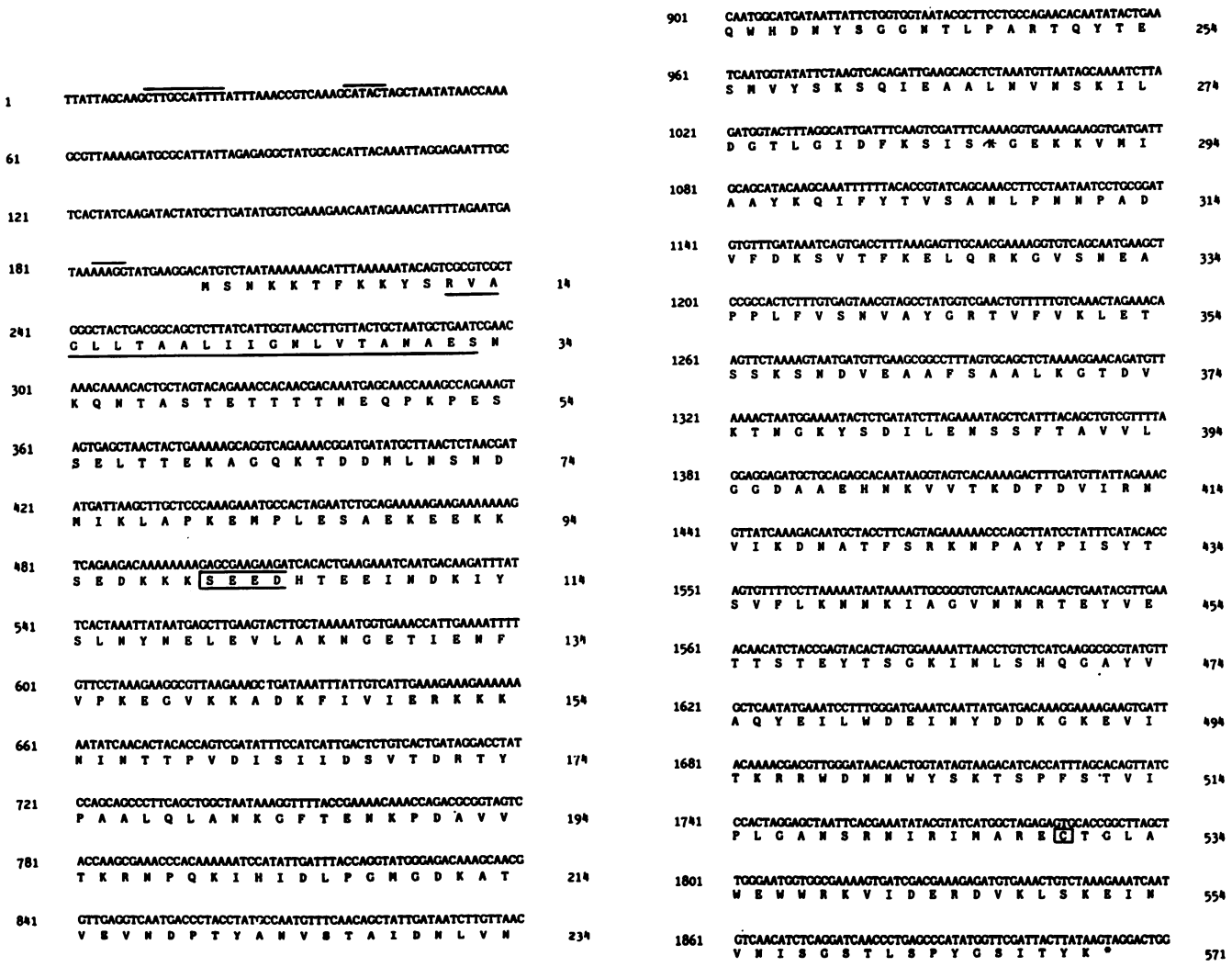

FIG. 1. Nucleotide sequence of the *slo* gene and predicted primary sequence of SLO. The nucleotide sequence is numbered at the left of the diagram, and the locations of possible promoter and ribosome-binding recognition sites (13, 15) are denoted by lines above the sequence. The predicted amino acid sequence is numbered on the right of the diagram. The N-terminal residues which possess the consensus features of a signal peptide involved in secretion are underlined. The probable N terminus of the low-molecular-weight form of SLO is denoted by the open box. The location of the single Cys residue, at position 530, is highlighted by a closed box.

the predicted signal sequence (see above) and this Ser would be consistent with the observed difference between the sizes of the high- and low-molecular-weight forms of the toxin. This, together with the fact that the alignment between the predicted SLO and pneumolysin sequences begins in this region (see below), strongly suggests that Ser-101 represents the N terminus of the low-molecular-weight form of SLO, but more extensive N-terminal amino acid sequencing would be desirable to confirm this.

Like other thiol-activated toxins, SLO and pneumolysin activity in crude culture supernatants is rapidly lost on exposure to air (1). This has lead to the assumption that loss of activity is due to the formation of an intramolecular disulfide bridge. However, only one Cys residue has been detected in the predicted SLO and pneumolysin sequences (Fig. 1; also see Fig. 3), indicating that intramolecular disulfide bridges cannot be formed. Loss of activity may result from reactions between the single Cys residue and the Cys of a second toxin molecule or a Cys in other proteins present in culture supernatants. The observation that the activity of purified SLO or pneumolysin is stable on exposure to air suggests that the latter is the case (1). Biochemical studies with thiol-blocking agents have shown that thiol-activated toxins contain at least one essential Cys residue (13). This allows us to predict that the regions in SLO and pneumolysin containing the single Cys residues correspond to functionally important structures. This is supported by the observations described below.

Hybridization probes failed to detect homology between *slo* and *ply* gene sequences (17). However, a dot matrix comparison of the complete nucleotide sequences suggests that these genes have originated from a common evolutionary source but reveals that they have undergone a considerable degree of divergence (Fig. 2). Despite this divergence at the DNA level, extensive structural homology exists between the two toxins at the amino acid sequence level (Fig. 3 and 4). The predicted sequence of the low-molecular-weight form of SLO, starting with Ser-101, is 471 residues long. This is remarkably similar in size to the predicted pneumolysin sequence, which, minus the N-terminal Met, is

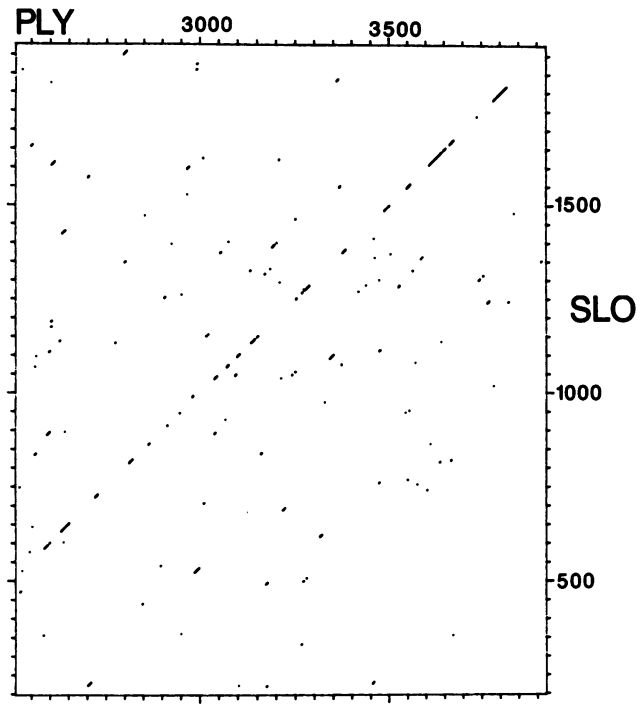

FIG. 2. Dot matrix comparison of the nucleotide sequences of the *slo* (*x* axis) and *ply* (*y* axis) genes. Each dot represents a match of 14 nucleotides in a window of 21. At higher stringencies, the number of matches decreases rapidly. Comparisons were performed by using a program described by Maizel and Lenk (19).

470 residues in length (pneumolysin is not secreted by the producing organism and, after removal of the N-terminal Met, does not undergo proteolytic processing). These two sequences can be aligned, by using a best-fit program (23), such that there are no extensive gaps and 42% of the residues match (Fig. 3a). Interestingly, this alignment places the single Cys residues at identical positions in both toxins. The extent of the structural homology between the two toxins is even more striking when the sequences are aligned by matching structurally similar, in addition to identical, residues (Fig. 3b). On this basis, over 60% of the residues match. An examination of the hydropathicity profiles of the molecules, in which the average hydropathicity at each position is calculated in the context of six flanking residues (Fig. 4), suggests that these amino acid sequence homologies are reflected by an overall structural homology between the two toxins. Superimposition of the SLO and pneumolysin profiles (Fig. 4, center) shows that they are almost indistinguishable.

The longest continuous identical sequence in SLO and pneumolysin consists of 12 residues encompassing the single Cys residue, close to the C termini of the molecules. This supports our suggestion that this is a functionally important region. Interestingly, preliminary sequencing studies on the 3′ end of a third thiol-activated toxin gene, encoding listeriolysin from *Listeria monocytogenes*, have revealed significant amino acid sequence homologies with the corresponding regions of SLO and pneumolysin (P. Cossart, personal communication). Thus, thiol-activated toxins from different species, and indeed different genera, may share a strongly conserved, functionally important structure close to the C-terminal ends of the molecules.

Except for the predicted signal sequence of SLO, neither toxin contains long highly hydrophobic regions. A similar situation has been observed for other cytolytic toxins for which primary sequence data is available. For example, the *Staphylococcus aureus* alpha-hemolysin contains only three short hydrophobic regions, separated in its primary sequence (15). It is possible that only a limited region of these toxins interacts with the hydrophobic region of the membrane. Alternatively, toxin folding or conformational changes induced by binding may bring short hydrophobic segments, which are separated in the primary sequences, into juxtaposition to form a hydrophobic face in the membrane. It seems likely that the region of the toxin involved in

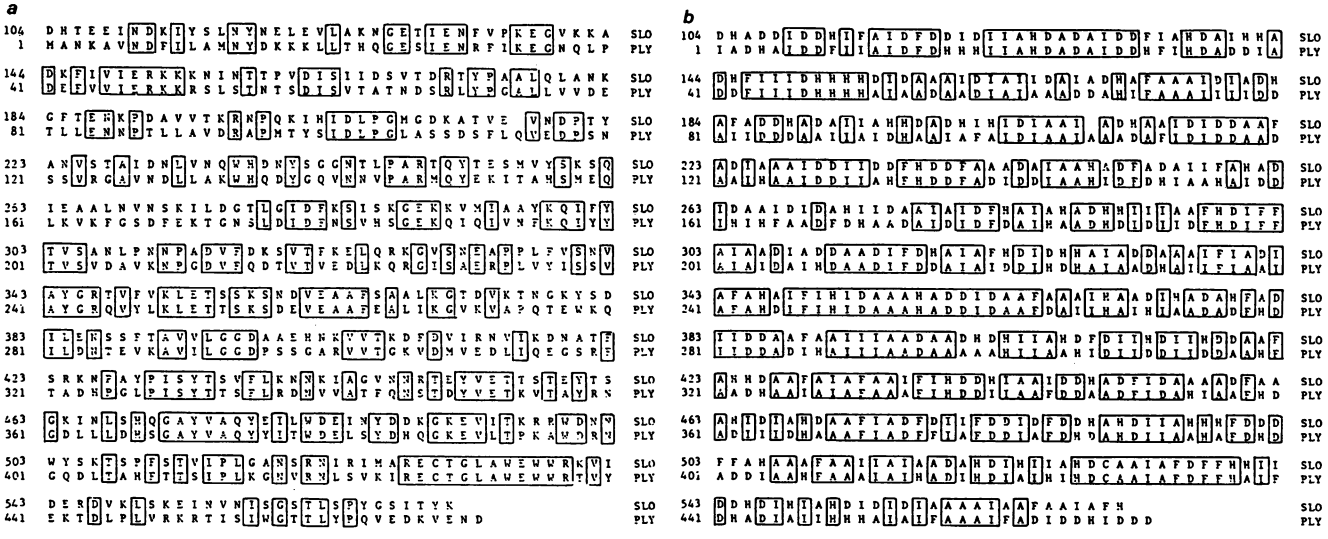

FIG. 3. Alignment of the predicted amino acid sequences of SLO and pneumolysin. Boxes denote common residues. (a) Alignment based on identical residues, represented by single-letter code. (b) Alignment based on similar residues. Abbreviations: A, neutral, weakly hydrophobic (P, A, G, S, T); D, hydrophilic, acid, amine (Q, N, E, D); H, hydrophilic, basic (H, K, R); I, hydrophobic (L, I, V, M); F, hydrophobic, aromatic (F, Y, W); and C, cysteine. The comparisons were made using a best-fit program, described by Smith and Waterman (23).

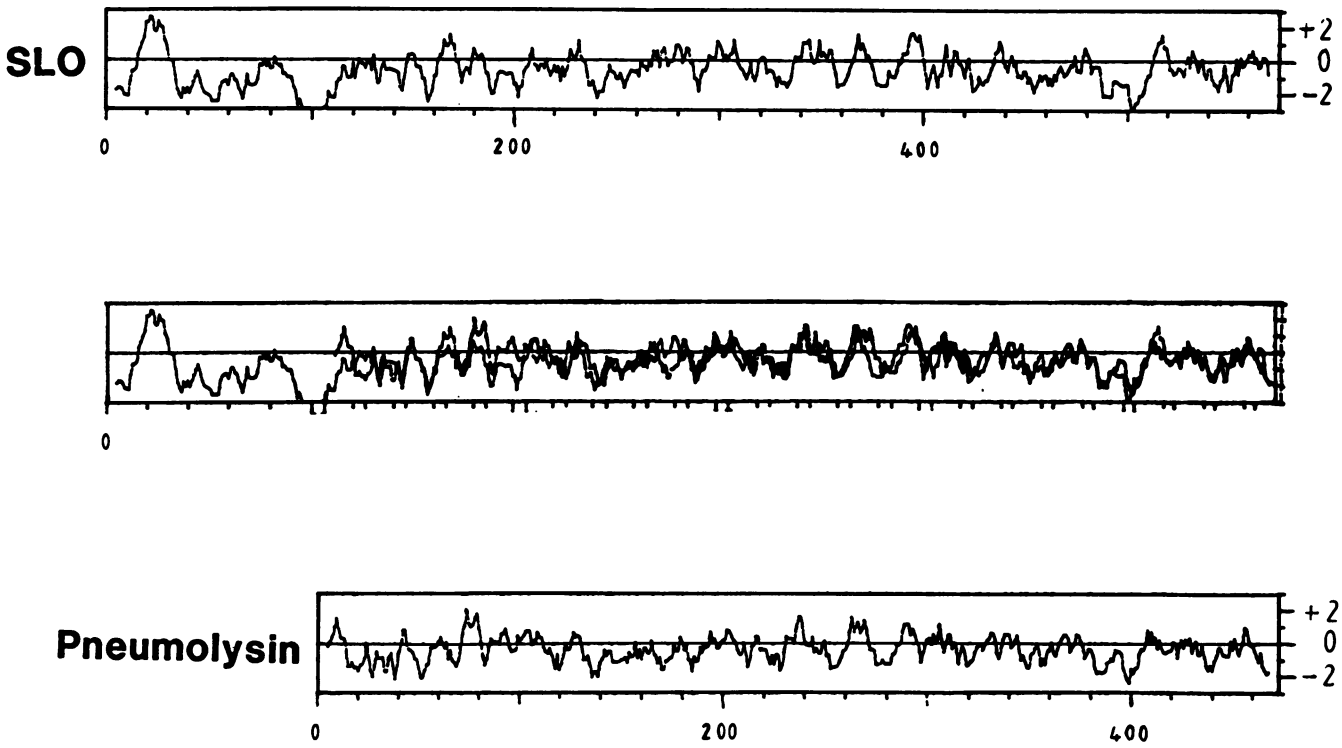

FIG. 4. Hydropathic profiles of SLO (top) and pneumolysin (bottom). The center displays the superimposition of the SLO and pneumolysin profiles, aligned at the single Cys residue in both molecules. The hydropathic indices were determined by the rules of Kyte and Dolittle (18). Positive values denote hydrophobicity, and negative values denote hydrophilicity. The lower scales denote the amino acid sequences.

binding to the cholesterol receptor in the membrane (1, 5, 24) would be composed of strongly conserved, specific residues. It is possible that the 12 residues encompassing Cys in SLO and pneumolysin form part of this structure. After binding, it may be the overall structure (folding) of the toxin in the membrane, rather than specific active sites, which mediates cell lysis. This would be consistent with the strong conservation of structure throughout the whole lengths of the SLO and pneumolysin molecules. Folding must produce a hydrophobic face to interact with membrane lipids, with at least two toxin-toxin faces which join noncovalently to form the large oligomeric structures which have been visualized by electron microscopy in lysed erythrocyte membranes (5, 8), and probably a hydrophilic pore-forming face. The availability of the cloned and sequenced *slo* and *ply* genes will allow us to investigate the role of individual structures in function by constructing mutant and hybrid toxins and altering individual residues by site-directed mutagenesis. In addition, the prediction based on sequence data of the SLO signal peptide can now be tested directly by constructing *slo-ply* gene fusions. The striking structural similarities between SLO and pneumolysin described here suggest that the failure of pneumolysin to be secreted might be due solely to the absence of a signal sequence, and it would be interesting to determine whether the 33 N-terminal residues of SLO could direct the secretion of pneumolysin. The construction of mutant and hybrid toxins could also contribute to our understanding of the role of the residues (probably 67) which are removed after secretion to generate the low-molecular-weight form of SLO. It has been reported that their removal is not required for activation of cytolytic activity, at least in vitro (3), but the possibility that they play a role in SLO activity or stability in vivo, in sublytic SLO activities against host cells (1), or in toxin secretion by group A streptococci cannot be ruled out.

M.K. thanks K.N. Timmis for his continued interest in his work on SLO, which was initiated during his stay in the laboratory of K.N.T. G.B. is a Lister Institute Jenner Research Fellow.

The work described here was supported by grants from the Nuffield Foundation (to M.K.) and the United Kingdom Medical Research Council (G846211CB to M.K. and G8409870CB to G.B.).

LITERATURE CITED

1. **Alouf, J. E.** 1980. Streptococcal toxins (streptolysin O, streptolysin S, erythrogenic toxin). Pharmacol. Ther. **11**:661–717.
2. **Bhakdi, S., N. Mackman, J.-M. Nicaud, and I. B. Holland.** 1986. *Escherichia coli* hemolysin may damage target cell membranes by generating transmembrane pores. Infect. Immun. **52**:63–69.
3. **Bhakdi, S., M. Roth, A. Sziegoleit, and J. Tranum-Jensen.** 1984. Isolation and identification of two hemolytic forms of streptolysin O. Infect. Immun. **46**:394–400.
4. **Bhakdi, S., and J. Tranum-Jensen.** 1978. Molecular nature of the complement lesion. Proc. Natl. Acad. Sci. USA **75**:5655–5659.
5. **Bhakdi, S., and J. Tranum-Jensen.** 1986. Membrane damage by pore-forming bacterial cytolysins. Microb. Pathogen. **1**:5–14.
6. **Bhakdi, S., J. Tranum-Jensen, and A. Sziegoleit.** 1985. Mechanism of membrane damage by streptolysin-O. Infect. Immun. **47**:52–60.
7. **DiScipio, R. G., M. R. Gehring, E. R. Podack, C. C. Kan, T. E. Hugli, and G. H. Fey.** 1984. Nucleotide sequence of cDNA and derived amino acid sequence of human complement component C9. Proc. Natl. Acad. Sci. USA **81**:7298–7302.
8. **Duncan, J. L., and R. Schlegel.** 1975. Effect of streptolysin O on erythrocyte membranes, liposomes and lipid dispersions. J. Cell Biol. **67**:160–173.

9. **Felmlee, T., S. Pellett, and R. A. Welch.** 1985. Nucleotide sequence of an *Escherichia coli* chromosomal hemolysin. J. Bacteriol. **163**:94–105.
10. **Ferretti, J. J., and R. Curtiss III (ed.).** 1987. Streptococcal genetics, p. 293. American Society for Microbiology, Washington, D.C.
11. **Fussle, R., S. Bhakdi, A. Sziegoleit, J. Tranum-Jensen, T. Kranz, and H. J. Wellensiek.** 1981. On the mechanism of membrane damage by Staphylococcus aureus α-toxin. J. Cell Biol. **91**: 83–94.
12. **Geoffroy, C., and J. E. Alouf.** 1984. Antigenic relationships between sulfhydryl-activated toxins, p. 241–243. *In* J. E. Alouf, F. J. Fehrenbach, J. H. Freer, and J. Jeljaszewicz (ed.), Bacterial protein toxins. Academic Press, Inc. (London), Ltd., London.
13. **Geoffroy, C., A.-M. Gilles, and J. E. Alouf.** 1981. The sulfhydryl groups of the thiol-dependent cytolytic toxin from *Bacillus alvei*. Evidence for one essential sulfhydryl group. Biochem. Biophys. Res. Commun. **99**:781–788.
14. **Gold, L., D. Pribnow, T. Schneider, S. Shinedling, B. Swefilius-Singer, and G. Stormo.** 1981. Translational initiation in procaryotes. Annu. Rev. Microbiol. **35**:365–403.
15. **Gray, G. S., and M. Kehoe.** 1984. Primary sequence of the α-toxin gene from *Staphylococcus aureus* Wood 46. Infect. Immun. **46**:615–618.
16. **Hawley, D. K., and W. R. McClure.** 1983. Compilation and analysis of *Escherichia coli* promoter DNA sequences. Nucleic Acids Res. **11**:2237–2254.
17. **Kehoe, M., and K. N. Timmis.** 1984. Cloning and expression in *Escherichia coli* of the streptolysin O determinant from *Streptococcus pyogenes*: characterization of the cloned streptolysin O determinant and demonstration of the absence of substantial homology with determinants of other thiol-activated toxins. Infect. Immun. **43**:804–810.
18. **Kyte, I., and R. F. Dolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**:105–132.
19. **Maizel, J. V., and R. P. Link.** 1981. Enhanced graphic matrix analysis of nucleic acid and protein sequences. Proc. Natl. Acad. Sci. USA **78**:7665–7669.
20. **Messing, J.** 1983. New M13 vectors for cloning. Methods Enzymol. **101**:20–78.
21. **Queen, C. L., and L. J. Korn.** 1984. A comprehensive sequence analysis program for the IBM personal computer. Nucleic Acids Res. **12**:581–599.
22. **Sarvas, M.** 1986. Protein secretion in bacilli. Curr. Top. Microbiol. Immunol. **125**:103–125.
23. **Smith, T. F., and M. S. Waterman.** 1981. Comparison of biosequences. Adv. Appl. Math. **2**:482–489.
24. **Smyth, C. J., and J. L. Duncan.** 1978. Thiol-activated (oxygen-labile) cytolysins, p. 129–183. *In* J. Jeljaszewicz and T. Wadstrom (ed.), Bacterial toxins and cell membranes. Academic Press, Inc. (London), Ltd., London.
25. **Walker, J. A., R. A. Allen, P. Falmagne, M. K. Johnson, and G. J. Boulnois.** 1987. Molecular cloning, characterization, and complete nucleotide sequence of the gene for pneumolysin, the sulfhydryl-activated toxin of *Streptococcus pneumoniae*. Infect. Immun. **55**:1184–1189.