The nucleotide sequence of toxic shock syndrome toxin-1 (TSST-1) has been determined. In addition, one-third of the predicted amino acid sequence was confirmed by amino acid sequence analysis of cyanogen bromide-generated TSST-1 protein fragments. The DNA sequencing results identified a 708-base pair molecule; kb, kilobase pairs. DNA sequencing results identified a 708-base pair fragment containing the first 40 amino acids and had characteristic structural similarities with other bacterial signal peptides. The coding sequence of the mature protein was 585 base pairs (194 amino acids) in length, and the molecular weight of the predicted protein was 22,049. This is in good agreement with the previously reported molecular weight of TSST-1 (22,000), as determined by sodium dodecyl sulfate-polyacrylamide gel electrophoresis. NH₂-terminal amino acid sequence analysis performed on isolated TSST-1 CNBr fragments determined the position of the peptides in the TSST-1 sequence and verified the predicted amino acid sequence in those positions. Computer analyses of the amino acid sequence showed that TSST-1 has little or no sequence homology with biologically related toxins, streptococcal pyrogenic exotoxin A, and staphylococcal enterotoxins B and C.

Toxic shock syndrome (TSS) is a multisystem illness characterized by the acute onset of high fever, hypotension or dizziness, rash, desquamation of skin upon recovery, and variable multisystem involvement (1-4). Staphylococcus aureus producing TSST-1 has been isolated from nearly 100% of menstrual-associated TSS patients (5, 6). Since bacteremia rarely occurs in TSS, it was proposed that a staphylococcal exotoxin was causing the widespread systemic effects. TSST-1 has been cited by several investigators as a major toxin most likely responsible for the symptoms of TSS (5-8).

The biological activities of TSST-1 have been examined extensively to build a relationship between the toxin's effects and the symptoms of TSS. The biological activities include: the capacity to induce fever, enhancement of host susceptibility to lethal shock by endotoxin, nonspecific T lymphocyte mitogenicity, suppression of immunoglobulin M synthesis against sheep erythrocytes, enhancement of delayed type hypersensitivity reactions, and induction of immunological tolerance in certain rabbits (9). The TSS model proposed by Schlievert (10) correlates these biological functions to the pathology seen in TSS patients.

To date, TSST-1 has been purified and biochemically characterized (11). In addition, the TSST-1 structural gene has been cloned from the bacterial chromosome (12). The toxin was purified to homogeneity by differential precipitation with ethanol and resolubilization in water followed by successive electrophoresis in pH gradients of 3-10 and 6-8 (11). TSST-1, thus isolated, migrated as a single band in SDS-polyacrylamide gel electrophoresis gels to a molecular weight of 22,000. The toxin gene was localized within a 10.6-kb chromosomal insert contained in plasmid pRN6100 and was previously shown to produce TSST-1. Analysis of subclones of the 10.6-kb insert in pBR322 assigned the structural gene to the leftmost 3.2-kb ClaI fragment (12).

Here we report the nucleotide and partial amino acid sequence analysis of TSST-1. These experiments were conducted to determine the complete amino acid sequence of TSST-1 as an approach to understanding the multiple structure-function relationships of this toxin molecule.

MATERIALS AND METHODS

M13 and DNA Sequencing—DNA sequencing was performed by the dideoxynucleotide chain termination method of Sanger et al. (13), following the subcloning of isolated DNA fragments from pRN6101 into M13 vectors mp10 and mp11 (12). A universal M13 sequencing primer (17 nucleotides, Boehringer Mannheim) was used in the annealing reaction.

Preparation of TSST-1 CNBr Fragments—TSST-1 was purified as previously described (11). The CNBr reactions were performed as reported by Gross (14) using a 2000:1 molar ratio of CNBr:TSST-1. The formic acid and excess CNBr reagent were removed by lyophilization. As a control, toxin was incubated with formic acid alone and then analyzed by gel electrophoresis. No significant protein degradation occurred due to the acidic reaction conditions.

Separation of CNBr Fragments by Gel Filtration—Lyophilized TSST-1 fragments (5-10 mg) were solubilized in 30% acetic acid and then loaded onto a Sephadex G-75 column (1.7 × 120 cm) equilibrated with 10% acetic acid. Fractions (2.0 ml each) were eluted with 10% acetic acid at a flow rate of 12 ml/h. The acetic acid was removed from each fraction by lyophilization. After lyophilization, column fractions containing a visible amount of material were suspended in 0.2 ml of water. The protein content of each fraction was then analyzed on a 10-18% linear gradient SDS gel (1-5 μg/lane) to
signal sequence and the blackened region of mature TSST-1. The cartridge by means of Polyhrene-impregnated glass fiber filters. The region, were separately cloned into coliphage M13 and seeded. Peptide and intact toxin samples were immobilized in the reaction cartridge by means of Polyhrene-impregnated glass fiber filters. The amount of peptide loaded in each run was in the range of 500 pmol. An aliquot (10%) of the phenylthiohydantoins was taken for the injection directly into a high performance liquid chromatograph (Model 120A, Applied Biosystems, Foster City, CA). Resolution of the phenylthiohydantoins was accomplished on a reverse-phase column (Brownlee “Spheri-5” C18, 0.21 x 22 cm) with a complex gradient composed of 114 mM acetate, pH 4.0, in 5% tetrahydrofuran (solvent A) and acetonitrile (solvent B), as specified in the technical bulletin of Applied Biosystems.

Computer Analysis—The amino acid sequences of TSST-1, staph-ylococcal enterotoxins B and C (19, 20), and streptococcal pyrogenic exotoxin type A (21, 22) were compared using the computer program Fast Protein Data Base (FASTP) written by Lipman and Pearson (23; sequence analysis programs distributed by the National Institutes of Health). This program is based on modifications of the algorithm of Wilbur and Lipman (23). To determine the statistical significance of sequence similarities, we employed Monte Carlo analysis using another algorithm written by Lipman and Pearson.

RESULTS AND DISCUSSION

Nucleotide Sequence Analysis of the TSST-1 Gene—Previously, subclones of pRN6101 were isolated which failed to express the TSST-1 protein (12). These results indicated that a 200-bp BamHI-HindII fragment was required for expression and established the sequencing strategy (Fig. 1) since it could be assumed that this region was within or very close to the TSST-1 structural gene. Using the deoxy cDNA chain termination method of Sanger (13), the overlapping 1.1-kb HindII and the 1.0-kb BamHI fragments, which bracket the gene-specific region, were separately cloned into colipage M13 and sequenced. The completion of the entire nucleotide sequence of the tsf gene and its controlling regions required additional cloning and sequencing of subfragments depicted in Fig. 1.

The final percentage of the sequence determined from both strands was 70%.

The toxin sequence, presented in the 5’ to 3’ orientation, starts in the AluI fragment and extends to the 3’ end in the HindII fragment (Fig. 2). The sequencing data identified a 705-bp open reading frame starting with an ATG at nucleotide position 478 and terminating at TAA nucleotide number 1180. A good consensus Shine-Dalgarno site lies 7 bp up-stream of the ATG start position (dotted line in Fig. 2). Amino acid sequence analysis of intact TSST-1 identified the NH₂ terminus of the mature TSST-1 protein as Ser-Thr-Asp-Asp-Ile-Lys-Asp-Leu, thus confirming previous studies of Igarashi et al. (24). In addition, these data defined the cleavage point for the signal peptide at an Ala/Ser sequence and determined the length of the signal peptide to be 40 amino acids. The signal peptide is indicated by a solid line in Fig. 2. The TSST-1 signal peptide contained the characteristic structural homologies found in other bacterial signal peptides: (a) 1–3 basic amino acids at the amino terminus, (b) a hydrophobic region of approximately 15 residues, (c) a Pro or Gly in the hydrophobic core, (d) a Ser or Thr near the carboxyl terminus of the core, and (e) an Ala or Gly at the cleavage site (25).

The coding sequence of the mature protein was 585 bp in length (194 amino acid residues), and the calculated molecular weight was 22,049, which is in complete agreement with the molecular weight of TSST-1 determined by SDS-polyacrylamide gel electrophoresis, 22,000 (11).

Generation and Isolation of CNBr Peptides—The TSST-1 protein fragments generated by CNBr cleavage are shown in Fig. 3. CNBr cleaved the toxin at its 2 methionine residues, producing five peptide fragments. The estimated masses of the fragments were: CN1, 18 kDa; CN2, 17 kDa; CN3, 14 kDa; CN4, 6–8 kDa; and CN5, 4 kDa.

The mixture of CNBr-generated toxin fragments was then applied to a Sephadex G-75 gel filtration column. The elution profile of the column fractions showed one broad protein peak. Therefore, each column fraction containing a visible

---

**Fig. 1.** Restriction map and sequencing strategy for the tsf gene. The top line depicts the nucleotide positions of the restriction sites on a pRN6101 subclone. The cross-hatched line indicates the signal sequence and the blackened line designates the coding sequence of mature TSST-1. The solid lines below show the subclones generated for sequencing and the restriction enzymes used. A, AluI; B, BamHI; D, DdeI; H, HindII. The arrows indicate the direction and distance of the sequence analysis.

**Fig. 2.** The nucleotide sequence of TSST-1 and the deduced amino acid sequence. The boxed area indicates those predicted residues which were verified by amino acid sequence analysis of the intact TSST-1 and the CNBr-generated fragments. The solid line identifies the signal sequence. The dotted line designates the Shine-Dalgarno sequence.
Amino Acid Sequence Analysis of the CNBr Peptides—

Purified protein fragments were analyzed by automated Edman degradation to determine the amino acid sequence of each fragment. These data were used to determine the position of each CNBr fragment in the TSST-1 molecule, as well as verifying the inferred amino acid sequence in those positions. Furthermore, defining the position of each fragment is critical for structure-function studies; localizing a specific biological activity to a mapped fragment will ultimately determine the functional domains of the intact protein.

Amino acid sequence analysis of purified CN3 and CN4 mapped their positions on the TSST-1 protein as shown in Fig. 4 (CN3, 35 cycles; CN4, 26 cycles). The repetitive yields in the CN3 and CN4 analyses were 90% and 93%, respectively. The solid area of each box indicates the sequenced region of each peptide, and the open area shows the remaining unsequenced protein. A sample containing CN1 and CN2 generated two unambiguous peaks at each amino acid position (24 cycles, repetitive yield 91%). These two signals were easily resolved by examining the inferred amino acid sequence; one signal depicted a peptide originating at the NH₂ terminus, the second signal described a fragment starting at methionine residue 33. The positions of CN1 and CN2 were determined by comparing the SDS gel estimated molecular weights and the molecular weights calculated from the predicted amino acid sequence (Table I). Peptide CN5 was not retrieved from the gel filtration column and, therefore, was not sequenced (indicated by the completely open box in Fig. 4). The position of CN5 is proposed based on the SDS gel estimated molecular weight and the calculated molecular weight of the predicted sequence. All of the SDS gel estimated molecular weights are in good agreement with the predicted amino acid sequence values except for CN4 (Table I). CN4 is an acidic fragment, and it is known that SDS-polyacrylamide gel electrophoresis overestimates the molecular weight of several other acidic proteins (26, 27).

Each of the samples described above was analyzed once, but analysis of the CN1, CN2 preparation verified the earlier CN3 results, and intact toxin analysis confirmed and extended the CN2 assignments.

The boxed sequences in Fig. 2 identify the inferred amino acid sequence which was confirmed by amino acid sequence analysis of intact TSST-1 and CNBr-generated fragments. Approximately one-third of the nucleotide sequence was con-
firmed. Note, amino acid residues 25–33 were determined by sequence analysis of intact TSST-1 (38 cycles, repetitive yield 97%).

Analysis of TSST-1 Amino Acid Sequence—Previously, we reported the amino acid composition of TSST-1 predicted from the nucleotide sequence (11). The subsequent CNBr fragment amino acid sequence analysis showed a few changes in the predicted amino acid sequence. The changes occurred in those areas where there was some ambiguity in the reading of the gel sequences. The amino acid sequence analysis of the CN1, CN2 preparation and of intact TSST-1 identified three excess nucleotides between amino acid numbers 31–32, 39–40, and 43–44. The corrected predicted amino acid composition of the gel sequences. The amino acid sequence analysis of the excess nucleotides between amino acid numbers 31–32, 39–40, and 43–44. The corrected predicted amino acid composition is listed in Table II. As previously noted, this composition correlates closely with other TSST-1 protein compositions (5, 24), excluding the differences in cysteine residues reported to be present by Reiser et al. (28).

The most interesting features of the toxin's amino acid sequence are the abundance of hydrophobic residues, the clusters of proline residues, and the two predicted β-turns. Approximately 25% of the total amino acids in TSST-1 are hydrophobic residues. Also, the TSST-1 amino acid sequence had four different areas containing clusters of proline residues: amino acids 48–56, 95–101, 112–117, and 179–180. Evaluation of the secondary structure of TSST-1 by the Chou-Fasman method suggested the presence of two β-turns at residues 35–39 and 47–50 (29).

TSST-1 Amino Acid Sequence Homology with Related Toxins—Previously, studies have shown that staphylococcal enterotoxins B and C, and streptococcal pyrogenic exotoxin type A have highly significant protein sequence homology (19–22). Since TSST-1 belongs to the same general family, based upon shared biological activities, analyses were performed to compare the toxin protein sequences. No homology was observed between TSST-1 and enterotoxins B and C. Minimal homology was seen between TSST-1 and streptococcal pyrogenic exotoxin type A. However, Monte Carlo analysis indicated that this minimal homology was not significant. Furthermore, streptococcal pyrogenic exotoxin A and staphylococcal enterotoxins B and C showed some serological cross-reactivity, as detected by Western blot analysis using polyclonal antisera against each toxin (data not shown). In support of the sequence homology results, none of the related toxins showed any cross-reactivity with TSST-1 antiserum, nor did the TSST-1 band cross-react with antisera against the other toxins.

In summary, knowing the complete amino acid sequence for TSST-1 and its CNBr fragments now allows for careful examination of the structure-function relationships of this immunoregulatory toxin. We plan to localize the biological activities retained in the CNBr fragments, as well as using DNA technology to further specify those particular amino acids involved.

Acknowledgments—We gratefully thank Yvonne Guttill for typing the manuscript, Dr. Robert Wolhuter for the amino acid sequence analysis, and Timothy Leonard for artwork and photography.

REFERENCES

1. Todd, J., Fishman, M., Kapral, F., and Welch, T. (1978) Lancet 2, 1116–1118
2. Dohrmann, J. P., Chesney, P. J., Wand, P. J., LaVenture, M., and The Investigation and Laboratory Team (1980) N. Engl. J. Med. 303, 1429–1435
3. Shands, K. N., Schmid, G. P., Dan, B. B., Blum, D., Guidotti, R. J., Harpsect, J. T., Anderson, R. L., Hill, D. L., Broome, C. V., Band, J. D., and Frazer, D. W. (1980) N. Engl. J. Med. 303, 1436–1442
4. Toffe, R. W., and Williams, D. N. (1981) Ann. Intern. Med. 94, 149–156
5. Schlievert, P. M., Shands, K. N., Dan, B. B., Schmid, G. P., and Dohrmann, J. D. (1981) J. Infect. Dis. 143, 509–516
6. Bergdoll, M. S., Crass, B. A., Reiser, R. F., Robbins, R. N., and Davis, J. P. (1981) Lancet 1, 1017–1021
7. de Azaledo, J. P., and Arbuthnott, J. P. (1984) Infect. Immun. 46, 314–317
8. Reiser, R. F., Arko, R. J., Feeley, J. C., Chandler, F. W., Thornberry, C., Gibson, R. J., Cohen, M. L., Jeffries, C. F., and Broome, C. V. (1985) Infect. Immun. 47, 598–604
9. Schlievert, P. M. (1984) Surv. Synth. Pathol. Res. 3, 54–62
10. Schlievert, P. M. (1983) J. Infect. Dis. 147, 391–398
11. Blomster-Hautamaa, D. A,, Kreiswirth, B. N., Novick, R. P., and Schlievert, P. M. (1986) Biochemistry 25, 54–59
12. Kreiswirth, B. N., Lofsdahl, S., Betley, M. J., O'Reilly, M., Schlievert, P. M., Bergdoll, M. S., and Novick, R. P. (1983) Nature 306, 709–712
13. Sanger, F., Coulson, A. R., Barrel, B. G., Smith, A. J. H., and Roe, B. A. (1980) J. Mol. Biol. 143, 161–178
14. Gross, E. (1967) Methods Enzymol. 11, 253–255
15. Swank, R. T., and Munkres, K. D. (1971) Anal. Biochem. 39, 77
16. Hashimoto, F., Horigome, T., Kanbayashi, M., Yoshida, K., and Sugano, H. (1983) Anal. Biochem. 129, 192–199
17. Oakley, R. B., Kirsch, D. R., and Morris, N. R. (1980) Anal. Biochem. 105, 361–363
18. Hovnan, W. M., Hunkapiller, M. W., Hood, L. E., and Dreyer, W. J. (1981) J. Biol. Chem. 256, 7990–7997
19. Huang, L.-Y., and Bergdoll, M. S. (1970) J. Biol. Chem. 245, 3518–3525
20. Schmidt, J. J., and Spero, L. (1983) J. Biol. Chem. 258, 6300–6302
21. Weeks, C. R., and Ferretti, J. J. (1986) Infect. Immun. 52, 144–150
22. Johnson, L. P., L’Italien, J. J., and Schlievert, P. M. (1986) Mol. Gen. Genet. 203, 354–356
23. Wilbur, W. J., and Lipman, D. J. (1983) Proc. Natl. Acad. Sci. U. S. A. 80, 726–730
24. Igarashi, H., Fujikawa, H., Usami, H., Kawabata, W., and Morita, T. (1984) Infect. Immun. 44, 175–181
25. Vlasuk, G. P., Inouye, S., Ito, H., Itakura, K., and Inouye, M. (1983) J. Biol. Chem. 258, 7141–7148
26. Burton, Z., Burgess, R. R., Lin, J., Moore, D., Holder, S., and Gros, C. A. (1981) Nucleic Acids Res. 9, 2889–2903
27. Kaufmann, E., Geisler, N., and Weber, K. (1984) FEBS Lett. 170, 81–84
28. Reiser, R. F., Robbins, R. N., Khoe, G. P., and Bergdoll, M. S. (1983) Biochemistry 22, 3907–3912
29. Chou, P. Y., and Fasman, G. D. (1974) Biochemistry 13, 222–245