Sequence of the Cloned *Escherichia coli* K1 CMP-N-acetylneuraminic Acid Synthetase Gene*

Gerardo Zapata‡, Willie F. Vann§, Wendy Aaronson¶, Marc S. Lewis†, and Malcolm Moos**

From the ‡Laboratories of Bacterial Polysaccharides, §Bacterial Toxins, and ¶Biochemical Pharmacology, Center for Biologics Evaluation and Research and the †Biomedical Engineering and Instrumentation Branch, Division of Research Services, National Institutes of Health, Bethesda, Maryland 20892

(Received for publication, March 16, 1989)

The *Escherichia coli* CMP-N-acetylneuraminic acid (CMP-NeuAc) synthetase gene has been cloned into the vector pKK223-3 containing the tac promoter. The complete nucleotide sequence of the gene encoding CMP-NeuAc synthetase has been determined by progressive deletion of M13 clones containing the 2.7-kb fragment. CMP-NeuAc synthetase is the 49,000-dalton polypeptide, has been located on a 3.3-kb HindIII fragment of this plasmid (15). The enzyme has been purified to homogeneity and a partial amino-terminal amino acid sequence determined (15).

The present study reports the cloning of the DNA fragment encoding CMP-NeuAc synthetase into the expression vector pKK223-3, the nucleotide sequence of the fragment that encodes CMP-NeuAc synthetase, and confirms the DNA sequence using protein sequencing methods.

MATERIALS AND METHODS

**Plasmids and Strains**—The plasmid pSR35 containing the gene encoding CMP-NeuAc synthetase has been described (14, 15). The plasmid vectors m13mp18, m13mp19, and pKK223-3 and the host JM105 were obtained from Pharmacia LKB Biotechnology Inc.

**Plasmid DNA Techniques**—Plasmid DNA was purified by cesium chloride-ethidium bromide density gradient centrifugation as described by Maniatis (11). Restraint endonuclease digestions were performed with premixed reagents according to protocols supplied by Bethesda Research Laboratories for EcoRI, BamHI, HindIII, and HincII, Boehringer Mannheim for KpnI and United States Biochemical Corp. for PstI. Electrophoresis of DNA was carried out on vertical 0.7% agarose gels as described by Davis et al. (17), using bacteriophage λ DNA standards as molecular weight markers. DNA transformation into *E. coli* strain JM105 was carried out as described by Dager and Ehrlich (18). Promega’s Erase-A-Base system was used for construction of M13 subclones as described by Henikoff (19).

**Cloning Strategy for Unidirectional Deletions with Exonuclease III**—The plasmid pSR35 (Fig. 1) was digested with EcoRI and the overhang at the cleavage site filled by treatment with Klenow DNA polymerase and deoxynucleotide triphosphates (16, 17). The resulting DNA was digested with HindIII to yield a 2.7-kb fragment with a 5′ overhang at the cleavage site. The 2.7-kb fragment was isolated by electrophoresis in low melting point agarose (SeaPlaque, FMC Bio).
products, Rockland, ME) and ligated into HindIII-HincII-digested M13mp19. The resulting clone pWG4 contains a 2.7-kb fragment oriented in such a way that a collection of unidirectional deletions can be created with exonuclease III in the DNA region coding for CMP-NeuAc synthetase (Fig. 1). A similar strategy was used to create the vector pWG5. pSR35 was enzymatically digested to the restriction sites are: BamHI (B); EcoRI (E); HincII (Hc); HindIII (H); PstI (P); PvuII (Pv). Diagonal digestion as described (18) was used for the preparation of pWAl.

Preparation of Unidirectional Exonuclease Clones—The plasmid pWG4 was ethanol-precipitated, dried, redissolved, and digested with KpnI. The reaction mixture was phenol-extracted and purified by gel filtration. A second digestion was performed with BamHI, followed by phenol extraction and ethanol precipitation. The double-digested protein was then subjected to exonuclease III digestion as described in Promega's Erase-a-Base System. A similar procedure was used to digest the pWG5 plasmid but PstI and HindIII were the enzymes used to prepare pWG5 for EcoRI digestion.

DNA Sequencing and Peptide Sequencing—Nucleotide sequencing was performed by the procedure of Sanger et al. (20) using Sequenase (United States Biochemical Corp.). The polyacrylamide-urea gel electrophoresis system described by Biggen et al. (18) was used for the separation of DNA fragments. Data were analyzed by the DNA analysis programs of Staden on a MicroVax minicomputer. The sequence of the peptide was determined by using an Applied Biosystems, Inc. automated gas phase sequencer according to the instructions of the manufacturer.

Construction of pWAl—Plasmid pSR35 was enzymatically digested with EcoRI and HindIII. The resulting 2.7-kb fragment containing CMP-NeuAc synthetase and P7 was purified by electrophoresis in low melting point agarose. The expression vector pKK223-3 was also digested with EcoRI and HindIII. The 2.7-kb fragment and the pKK223-3 plasmid were ligated overnight and the new plasmid pWAl was used to transform E. coli JM105 cells.

Enzyme Purification—CMP-NeuAc synthetase was purified by ion exchange and affinity chromatography techniques as described previously (15). Isolation of enzyme from reaction mixtures and exchange of buffers was achieved by HPLC using a Superose 12 gel permeation column (Pharmacia), equilibrated with 0.05 M ammonium bicarbonate, 0.7 M NaCl, 10% glycerol, pH 7.6, and (b) 0.1 M Tris, 20 mM MgCl₂, pH 9.0. The column was equilibrated for 1 h at a flow rate of 1 ml/min and calibrated with commercial gel permeation standards (Bio-Rad) equilibrated in the appropriate buffer. CMP-NeuAc synthetase was concentrated and equilibrated in buffer using Centricron-10 filters.

Molecular Weight Determination by Ultracentrifugation—Samples of CMP-NeuAc synthetase were dialyzed against (a) 0.05 M Tris, 0.3 M NaCl, 10% glycerol, pH 7.6, and (b) 0.1 M Tris, 20 mM MgCl₂, 1 mM DTT, pH 9.0, and (c) same as (b) plus 1 M ammonium sulfate. Enzyme solutions (0.3 mg/ml) were ultracentrifuged at 4°C and 17,000 rpm until equilibrium was attained and the molecular weight calculated as described (23).

CMP-NeuAc Synthetase Assay—The enzyme was assayed for the formation of CMP-NeuAc by the thioarbituric acid assay as described (15).

RESULTS AND DISCUSSION

The gene encoding CMP-NeuAc synthetase is located on the plasmid pSR35 (14, 15). In order to determine its nucleotide sequence, a 2.7-kb HindIII-EcoRI fragment of pSR35 containing the CMP-NeuAc synthetase gene was subcloned into the multiple cloning sites of the sequencing vectors M13mp18 and M13mp19. These plasmids pWG4 and pWG5 (Fig. 1) permitted the use of Henikoff's method (19) to create progressive unidirectional deletions in the CMP-NeuAc synthetase gene. Plasmids pWG4 and pWG5 contain the CMP-NeuAc synthetase gene in opposite directions with respect to the DNA sequencing primer, thus permitting sequencing of both strands of DNA.

We sequenced a total of 1300 nucleotides covering the entire DNA sequence of the CMP-NeuAc synthetase gene. It has been shown that the Tn5 insertion mutation kps4 in the p7 maps 1.3 kb from the EcoRI site on pSR35 (14). The strategy used for deducing the nucleotide sequence is illustrated in Fig. 2. Sixteen subclones of the plasmid pWG4 were sequenced from the amino terminus and 12 subclones of the plasmid pWG5 were used to obtain the sequence in the opposite direction. Difficulty was encountered in obtaining overlap between consensus sequences in subclones of pWG4 in the

Fig. 1. Plasmids used for nucleotide sequencing and gene expression. The location of CMP-NeuAc synthetase is designated by P6, the restriction sites are: BamHI (B); EcoRI (E); HindIII (Hc); HindIII (H); KpnI (K); PstI (P), and pTAC denotes the tac promoter.
The peptides were fractionated by HPLC and sequenced. The sequences of the peptides are shown in Table I. Confirmation from trypsin- and cyanogen bromide-generated peptides. The agreement of the sequence determined for peptides T2, T3, T4, and CN2 and the predicted amino acid sequence confirms a significant portion of the nucleotide sequence. Comparison of the sequence of the tryptic peptides T1-T4 with the predicted amino acid sequence agrees with the presence of arginine at positions 2, 193, 242, and 358. The peptide CN2 co-purified with a minor component, CN1. Automated Edman degradation revealed that CN1 originated from the carboxyl terminus predicted by the nucleotide sequence and is also predicted to have a molecular weight and a pI similar to those of CN2. The methionine residues at positions 25, 200, and 393 were located by comparison of the sequence of peptides CN1, CN2, and T2 (Table I) with the predicted amino acid sequence.

In a separate experiment the carboxyl-terminal sequence of CMP-NeuAc synthetase was confirmed using hydroxylamine to chemically cleave asparaginyl-glycine bonds. There are only two potential cleavage sites for hydroxylamine cleavage in CMP-NeuAc synthetase as predicted by the nucleotide derived amino acid sequence (arrows in Fig. 3). Cleavage at these two sites (Asn-405 near the carboxyl terminus and Asn-173 near the middle of the sequence) should yield a 15-amino acid carboxyl-terminal peptide in addition to two larger peptides. CMP-NeuAc synthetase was treated with hydroxylamine and the resulting peptides fractionated by HPLC. The sequence of the low M, peptide was determined by automated Edman degradation to be GYTTLNEIEAIEIVK. This sequence is identical to residues 406-418 in the predicted amino acid sequence in Fig. 3 and confirms the presence of a lysine as the carboxyl-terminal amino acid in CMP-NeuAc synthetase.

Comparison of Predicted Amino Acid Sequences of CMP-NeuAc Synthetase and CMP-KDO Synthetase—The amino acid and nucleotide sequences determined above (Fig. 3) were used to search the September 1988 National Biomedical Research Foundation and November 1988 GenBank data bases for homologous sequences. The best fit obtained by the Lipman-Pearson algorithm (25) was with the translated nucleotide sequence of the E. coli kdsB gene product (26), CMP-KDO synthetase with 19.2% homology. An alignment was made using the program GAP and suggests a 40% similarity of amino acids allowing five gaps. Both CMP-NeuAc and CMP-KDO synthetase catalyze the transfer of an α-keto acid to form a cytidine-monophosphate sugar nucleotide. As illustrated in Fig. 4, several regions of homology were observed. The regions of strongest homology occurred at residues 8-12 (HIPAR) as reported previously and at residues 46-55 (EKIVYTTDSE) at the amino terminus of both proteins. The region of homology observed with CMP-KDO synthetase and elongation factor Tu (26) was not observed with the sequence of CMP-NeuAc synthetase. No significant homology was detected between the protein sequences of E. coli CMP-NeuAc synthetase and rat liver β-galactosyl α2,6-sialyltransferase (27). Calculations as described by Chou and Fasman (28) and Garnier et al. (29) were used to predict secondary structure of the two proteins. The CMP-NeuAc synthetase is predicted to have 47% α helix and 36% β structures; and the CMP-KDO synthetase is predicted to have 54% α helix and 31% β structures. Both proteins are predicted to have β structures between residues 3 and 12 and α helix near residue 40. These data suggest that common functional residues may occur in the amino-terminal regions of these proteins.

Native Molecular Weight—The polypeptide molecular
weight is known from the amino acid sequence. The native molecular weight of the CMP-NeuAc synthetase was therefore estimated in order to determine the enzyme subunit structure. The apparent molecular mass of the major activity fraction of 65,000-80,000 on the FPLC column Superose 12 in non-denaturing buffers (data not shown). Sedimentation equilibrium was used as an independent method for determining molecular weight (Table II). The partial specific volume was estimated from the amino acid composition predicted by the sequence (30). The enzyme was dialyzed against nondenaturing buffers and centrifuged. A major species with a molecular mass of 49,000 was observed in 20 mM M$^+$ at pH 9.0. Enzyme activity and protein was observed with a molecular mass of 70,000-80,000 daltons (Table II). Varying amounts of high monomer and may also exist as a dimer or a higher aggregate in some buffers. The cause of this aggregation was not investigated.

**Overproduction of CMP-NeuAc Synthetase**—The gene cod-

---

**Table I**

Sequence of peptides isolated from digests of CMP-NeuAc synthetase

Peptides were isolated from digests of a purified sample of CMP-NeuAc synthetase using two reversed-phase columns as described under "Materials and Methods." The isolated peptides were then sequenced using an Applied Biosystems, Inc. gas sequencer.
phenomena. Changes in sialylated oligosaccharides have been implicated as a marker for malignant transformation (34). Although CMP-NeuAc is commercially available, it is expensive. Because of the difficulty of synthesis of CMP glycosides and their potential utility in the preparation of oligosaccharide receptors, inhibitors, and specific oligosaccharide labels (35) there is interest in the enzymatic synthesis of the sugar nucleotide. Analogs and O-acetylated derivatives of NeuAc (11, 36) have been activated to CMP-NeuAc by preparations of bovine CMP-NeuAc synthetase. Research on the biological roles of sialylation should be facilitated by the readily available source such as E. coli.

REFERENCES
1. DeWitt, C. W., and Rowe, J. A. (1961) J. Bacteriol. 82, 838-848
2. Robbins, J. B., Schneerson, R., Egan, W. B., Vann, W., and Liu, D. T. (1980) The Molecular Basis of Microbial Pathogenicity (Smith, H., Skehel, J. J., and Turner, M. J., eds) pp. 115-132, Dahlem Konferenzen 1980, Verlag Chemie GmbH, Weinheim/Bergst, Federal Republic of Germany
3. Kajiser, B., Hanson, L. A., Jodal, V., Linden-Johnson, G., and Robbins, J. B. (1977) Lancet 1, 664-666
4. Robbins, J. B., McCracken, G. H., Jr., Gotschlich, E. C., Orslov, F., Orslov, I., and Hanson, L. A. (1974) N. Engl. J. Med. 290, 1216-1220
5. Roth, J., Zuber, C., Wagner, P., Tastjes, D. J., Weisberger, C., Heitz, F. U., Gordis, C., and Bitter-Sustermann, D. (1988) Proc. Natl. Acad. Sci. U. S. A. 85, 2999-3003
6. Hoffman, S., Sorkin, B. C., White, P. C., Brackenbury, R., Mailhammer, R., Rutishauser, U., Cunningham, B. A., and Edelman, G. M. (1982) J. Biol. Chem. 257, 7720-7729
7. Troy, F. A., Vijay, I. K., McCloskey, M. A., and Rohr, T. E. A. (1982) Methods Enzymol. 89, 540-548
8. Blacklow, R. S., and Warren, L. (1962) J. Biol. Chem. 237, 3520-3526
9. Kean, E. L., and Roseman, S. (1966) J. Biol. Chem. 241, 5643-5650
10. Schauer, R. (1989) Adv. Carbohydr. Chem. Biochem. 40, 131-234
11. Higa, H., and Paulson, J. C. (1985) J. Biol. Chem. 260, 8838-8849
12. Gross, H. J., and Brossmer, R. (1987) Glycoconjugate J. 4, 145-156
13. Gross, H. J., and Brossmer, R. (1988) Eur. J. Biochem. 177, 543-589
14. Silver, R. P., Vann, W. F., and Aaronson, W. (1984) J. Bacteriol. 157, 568-575
15. Vann, W. F., Silver, R. P., Abeijon, C., Chang, K., Aaronson, W., Sutton, A., Finn, C. W., Lindner, W., and Kotsatos, M. (1987)
CMP-N-acetylneuraminic Acid Synthetase Sequence

16. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
17. Davis, L. G., Dibner, M. D., and Batty, J. F. (1986) Basic Methods in Molecular Biology, Elsevier Scientific Publishing Co., Inc., New York
18. Dagert, M., and Ehrlich, S. D. (1979) Gene (Amst.) 6, 23–28
19. Henikoff, S. (1984) Gene (Amst.) 28, 351–359
20. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 5463–5467
21. Biggen, M. D., Gibson, T. J., and Hong, G. F. (1983) Proc. Natl. Acad. Sci. U. S. A. 80, 3963–3965
22. Moore, S. (1963) J. Biol. Chem. 238, 235–237
23. Gladner, J. A., Lewis, M. S., and Chung, S. I. (1981) J. Biol. Chem. 256, 1772–1781
24. Shine, J., and Dalgarno, L. (1974) Proc. Natl. Acad. Sci. U. S. A. 71, 1342–1348
25. Lipman, D. J., and Pearson, W. R. (1985) Science 227, 1434–1441
26. Goldman, R. C., Bolling, T. J., Kohlbrenner, W. E., Kim, Y., and Fox, J. L. (1986) J. Biol. Chem. 261, 15831–15835
27. Weinstein, J., Lee, E. U., McEntee, K., Lai, P.-H., and Paulson, J. C. (1987) J. Biol. Chem. 262, 17735–17743
28. Chou, P. Y., and Fasman, G. D. (1978) Annu. Rev. Biochem. 47, 251–276
29. Garnier, J., Osguthorpe, D. J., and Robeson, B. (1978) J. Mol. Biol. 120, 97–120
30. Zamyatnin, A. A. (1984) Annu. Rev. Biophys. Bioeng. 13, 145–165
31. Prakash, V., and Timasheff, S. N. (1985) Methods Enzymol. 117, 53–80
32. Arakawa, T., and Timasheff, S. N. (1985) Methods Enzymol. 117, 60–65
33. Icho, T. (1988) J. Bacteriol. 170, 5117–5124
34. Hakomori, S. (1984) Annu. Rev. Immunol. 2, 103–126
35. Kanie, O., Nakamura, J., Kiso, M., and Hasegawa, A. (1987) J. Carbohydr. Chem. 6, 105–115
36. Gross, H. J., Buensch, A., Paulson, J. C., and Brossmer, R. (1987) Eur. J. Biochem. 168, 595–602