

ing is found (2, 12). This study shows that Hoogsteen base pairs can coexist with Watson-Crick base pairs and it reinforces the possibility that such DNA polymorphism may be involved in various biological phenomena.

REFERENCES AND NOTES

1. A. H.-J. Wang *et al.*, *Nature (London)* **282**, 680 (1979).
2. D. E. Pulleyblank, D. B. Haniford, A. R. Morgan, *Cell* **42**, 271 (1985).
3. A. H.-J. Wang *et al.*, *Science* **225**, 1115 (1984).
4. G. Ughetto *et al.*, *Nucleic Acids Res.* **13**, 2305 (1985).
5. M. M. Van Dyke and P. B. Dervan, *Science* **225**, 1122 (1984).
6. C. M. L. Low, H. R. Drew, M. J. Waring, *Nucleic Acids Res.* **12**, 4865 (1984).
7. C. M. L. Low, R. K. Olsen, M. J. Waring, *FEBS Lett.* **176**, 414 (1984).
8. W. A. Hendrickson and J. Konnert, in *Biomolecular Structure: Conformation, Function and Evolution*, R. Srinivasan, Ed. (Pergamon, Oxford, 1979), p. 43.
9. M. Windholz, Ed., *The Merck Index* (Merck and Co., Inc., Rahway, NJ, ed. 10, 1983).
10. K. A. Hartman, Jr., and A. Rich, *J. Am. Chem. Soc.* **87**, 2033 (1965).
11. A. Rich, A. Nordheim, A. H.-J. Wang, *Annu. Rev. Biochem.* **53**, 791 (1984).
12. H. R. Drew and R. E. Dickerson, *EMBO J.* **6**, 663 (1982).
13. Supported by grants from the National Institutes of Health, National Science Foundation, American Cancer Society, Office of Naval Research, National Aeronautics and Space Administration, and the Netherlands Organization for the Advancement of Pure Research. G.U. acknowledges support from NATO and Instituto Strutturistica Chimica, Consiglio Nazionale Delle Ricerche (Italy). We thank T. Yoshida of Shionogi Co., Osaka, Japan, for the triostin A.

2 December 1985; accepted 18 March 1986

Pertussis Toxin Gene: Nucleotide Sequence and Genetic Organization

CAMILLE LOCHT AND JERRY M. KEITH

The current pertussis vaccines, although efficacious, in some instances produce undesirable side effects. Molecular engineering of pertussis toxin, the major protective antigen, could provide a safer, new generation of vaccines against whooping cough. As a first critical step in the development of such a vaccine, the complete nucleotide sequence of the pertussis toxin gene was determined and the amino acid sequences of the individual subunits were deduced. All five subunits are coded by closely linked cistrons. A promoter-like structure was found in the 5'-flanking region, suggesting that the toxin is expressed through a polycistronic messenger RNA. The order of the cistrons is S1, S2, S4, S5, and S3. All subunits contain signal peptides of variable length. The calculated molecular weights of the mature subunits are 26,024 for S1, 21,924 for S2, 21,873 for S3, 12,058 for S4, and 11,013 for S5. Subunits S2 and S3 share 70% amino acid homology and 75% nucleotide homology. Subunit S1 contains two regions of eight amino acids homologous to analogous regions in the A subunit of both cholera and *Escherichia coli* heat labile toxins.

PERTUSSIS TOXIN (1) IS ONE OF THE various toxic components produced by virulent *Bordetella pertussis*, the microorganism that causes whooping cough. A wide variety of biological activities, such as histamine sensitization, insulin secretion, lymphocytosis promotion, and immunopotentiating effects can be attributed to this toxin (2). In addition, the toxin protects mice challenged intracerebrally (3, 4) or by aerosol (4) with virulent *B. pertussis*. Pertussis toxin is, therefore, an important constituent in the vaccine against whooping cough and is included in the acellular component vaccines being tested and used in several countries (5). However, the toxin may also be the cause of the harmful side effects associated with the current vaccines (6). It may be possible to develop a new vaccine with reduced side effects by genetic manipulation of the toxin gene.

The toxin is structured in a hexamer composed of five dissimilar subunits, designated S1 through S5 relative to their electrophoretic migration in denaturing gels (7). Subunit S1 contains an enzymatic adenosine diphosphate (ADP)-ribosylation activity

(8) and subunits S2 through S5 contain a target cell receptor binding activity (9). Thus, by analogy to other bacterial ADP-ribosylating toxins, pertussis toxin is structured in an A-B model (7, 10) in which the A moiety contains the enzymatic activity and the B moiety the binding activity (11).

The isolation of a 4.5-kb DNA fragment containing the pertussis toxin S4 subunit gene was described previously (12). Because of the polycistronic nature of prokaryotic genes, and by analogy to other A-B structured bacterial toxins that have similar enzymatic activities (13, 14), it seemed likely that the cistrons coding for all the pertussis toxin subunits would be linked. We therefore determined the nucleotide sequence of the 4.5-kb cloned fragment. Here, we present this sequence and demonstrate that it indeed codes for the entire pertussis toxin structural gene.

Digests with a variety of six base pair (bp)-specific restriction enzymes and DNA sequence analysis were used to establish the restriction map shown in Fig. 1a. DNA was sequenced by the base-specific chemical cleavage method (15) starting from the re-

striction sites Eco RI, Sal I, Sma I, and Bgl II and by the dideoxy chain termination method (16) after subcloning overlapping regions into vectors M13 mp18 and M13 mp19 (17). Because of the high C+G content of *B. pertussis* DNA, it was necessary to use both of these methods with a combination of 8% and 20% polyacrylamide-8M urea gels for sequence analysis. Each nucleotide has been sequenced in both directions an average of 4.13 times. The final consensus sequence of the sense strand is presented in Fig. 2. In agreement with the overall high C+G content of *B. pertussis* DNA (18), the entire sequence contains 62.2% C+G with 19.6% A, 33.8% C, 28.4% G, and 18.2% T in the sense strand.

The DNA sequence was translated in all six reading frames. On the basis of our previous data (12), the open reading frame (ORF) corresponding to the S4 subunit was identified and is shown in Fig. 1b. The assignment of the other subunits to their respective ORF's, as shown in Fig. 1, b and c, is based on the following evidence: size of ORF's, high coding probability, deduced amino acid composition, predicted molecular weights, ratios of acidic to basic amino acids, amino acid homology to other bacterial toxins, mapping of Tn5-induced mutations, and partial amino acid sequence.

Those ORF's long enough to code for any of the five toxin subunits were analyzed by the statistical TESTCODE algorithm designed to differentiate between real protein coding sequences and fortuitous open reading frames (19). The amino acid composition of each ORF with a high protein coding probability was calculated, starting from either the predicted amino terminus of the mature proteins or from the first amino acid of the mature protein determined by amino acid sequencing of subunits purified by high-performance liquid chromatography. These data were then compared with

Department of Health and Human Services, Public Health Service, National Institutes of Health, National Institute of Allergy and Infectious Diseases, Laboratory of Pathobiology, Molecular Pathobiology Section, Rocky Mountain Laboratories, Hamilton, MT 59840.

the experimentally determined compositions of the individual subunits (7). On the basis of the similarity of the amino acid compositions shown in Table 1, all five subunits were identified and assigned to the ORF regions shown in Fig. 1b, and the gene order of S1, S2, S4, S5, and S3 from the 5'-end to the 3'-end was established (Fig. 1d).

Table 1 shows that the deduced amino acid composition from all five assigned subunits are in good agreement with the experimentally determined compositions, with two significant exceptions. First, the S1 subunit contains no lysine residues in the deduced amino acid sequence, whereas a 2.2% lysine content was determined experimentally. Second, in subunits S2, S3, S4, and S5 the proportion of cysteines was substantially underestimated in the experimentally observed compositions. These discrepancies, as well as the remaining minor differences observed for all subunits, including the previously assigned S4 subunit, can only be attributed to experimental error during amino acid analysis (20) or to the individual subunits, which are purified by liquid column chromatography, being slightly impure. The amino acid composition of significant ORF's (Fig. 1, b and c) not assigned to any subunit show no similarity to any of the experimentally determined amino acid compositions, although some of these ORF's are quite long and have a high coding potential. It is possible that these

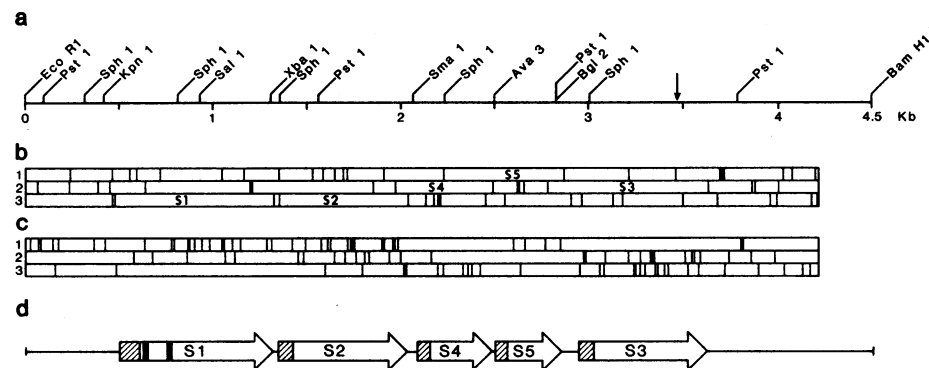


Fig. 1. Physical map and genetic organization of the pertussis toxin gene. (a) Restriction map of the 4.5-kb Eco RI-Bam HI fragment from pPTX42 containing the pertussis toxin gene cloned from *B. pertussis* strain 3779 (12). The arrow indicates the position of the Tn5 DNA insertion in pertussis toxin-negative Tn5-induced mutant strains BP356 and BP357 (12, 23). (b) Open reading frames in the forward direction. (c) Open reading frames in the backward direction. The vertical lines indicate termination codons. (d) Organizational map of the pertussis toxin gene. The arrows show the translational direction and length of the protein coding regions for the individual subunits. The hatched boxes represent the signal peptides. The solid bars in S1 represent the regions homologous to the A subunits in cholera and *E. coli* heat labile toxins (see Fig. 3).

regions code for other proteins that may be involved in the assembly or transport of pertussis toxin.

The experimentally estimated molecular weights and isoelectric points of the individual subunits were compared to the calculated molecular weights and ratios of acidic to basic amino acids of the putative proteins encoded by the ORF's. Table 1 shows that differences in these ratios reflect corresponding differences in the observed isoelectric

points for each subunit, that is, the higher the acidic content, the lower the isoelectric point. The calculated molecular weights also correspond well to the experimentally determined values, with slight differences for the S1 (less than 10%) and the S5 (about 15%) subunits. These small differences are within acceptable limits for protein molecular weights determined by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE).

Table 1. Comparison of the observed amino acid composition with the calculated composition from DNA sequences for mature pertussis toxin subunits. The data for the observed values are from Tamura *et al.* (7); see also (20). ND, not determined.

| Item | S1 | | S2 | | S3 | | S4 | | | S5 | |
|-------|-----------------|-------------------|-----------------|-------------------|-----------------|-------------------|------------------------|------------------------|-------------------|-----------------|-------------------|
| | Observed values | Calculated values | Observed values | Calculated values | Observed values | Calculated values | Exp. 1 observed values | Exp. 2 observed values | Calculated values | Observed values | Calculated values |
| M_r | 28 kD | 26.0 kD | 23 kD | 21.9 kD | 22 kD | 21.9 kD | 11.7 kD | | 12.1 kD | 9.3 kD | 11.0 kD |
| A/B* | | 1.3 | | 0.89 | | 0.83 | | | 0.65 | | 1.4 |
| pI† | 5.8 | | 8.5 | | 8.8 | | 10.0 | 10.0 | | 5.0 | |
| Ala | 10.6 | 11.5 | 6.5 | 6.0 | 11.7 | 11.1 | 9.4 | 9.8 | 8.2 | 9.8 | 9.0 |
| Arg | 5.9 | 9.0 | 6.2 | 6.0 | 6.1 | 6.5 | 5.1 | 5.4 | 5.5 | 3.3 | 3.0 |
| Asn‡ | 9.3 | 5.6 | 6.3 | 2.5 | 6.3 | 2.0 | 5.3 | 5.0 | 0.9 | 8.2 | 3.0 |
| Asp | | 4.3 | | 4.0 | | 4.0 | | | 3.6 | | 5.0 |
| Cys | 1.0 | 0.9 | 1.3 | 3.0 | 1.1 | 3.0 | 0.9 | 0.7 | 3.6 | 1.6 | 4.0 |
| Gln§ | 10.6 | 3.0 | 8.7 | 3.5 | 9.0 | 4.5 | 9.5 | 9.1 | 3.6 | 9.3 | 3.0 |
| Glu | | 7.3 | | 4.0 | | 3.5 | | | 4.5 | | 6.0 |
| Gly | 11.2 | 7.7 | 13.0 | 10.6 | 11.9 | 10.1 | 9.6 | 8.9 | 6.4 | 8.7 | 8.0 |
| His | 1.7 | 2.6 | 2.4 | 2.0 | 1.0 | 1.0 | 0.5 | 0.5 | 0.9 | 3.0 | 3.0 |
| Ile | 3.2 | 3.4 | 4.2 | 5.5 | 5.0 | 6.5 | 2.0 | 1.8 | 1.8 | 3.4 | 3.0 |
| Leu | 5.5 | 3.4 | 7.3 | 7.5 | 8.1 | 8.0 | 8.4 | 8.7 | 9.1 | 13.8 | 15.0 |
| Lys | 2.2 | 0 | 3.4 | 3.0 | 2.7 | 2.5 | 6.9 | 7.6 | 7.3 | 4.7 | 5.0 |
| Met | 1.6 | 1.7 | 1.4 | 1.5 | 1.1 | 1.5 | 5.1 | 4.3 | 7.3 | 1.6 | 2.0 |
| Phe | 3.5 | 3.0 | 3.2 | 2.5 | 3.2 | 2.5 | 3.6 | 4.5 | 4.5 | 4.9 | 5.0 |
| Pro | 4.4 | 3.4 | 4.6 | 4.5 | 5.7 | 5.0 | 9.1 | 9.9 | 10.0 | 5.6 | 5.0 |
| Ser | 10.6 | 9.8 | 8.5 | 8.5 | 6.3 | 5.0 | 8.0 | 7.3 | 5.5 | 6.9 | 6.0 |
| Thr | 7.4 | 7.3 | 10.4 | 10.1 | 8.2 | 8.0 | 5.0 | 5.1 | 4.5 | 6.9 | 7.0 |
| Trp | ND | 0.9 | ND | 1.0 | ND | 0.5 | ND | ND | 0 | ND | 1.0 |
| Tyr | 4.6 | 8.1 | 7.6 | 8.0 | 7.9 | 9.5 | 2.2 | 2.0 | 1.8 | 4.3 | 4.0 |
| Val | 6.7 | 7.3 | 4.9 | 6.0 | 4.7 | 5.0 | 9.4 | 9.4 | 10.9 | 4.0 | 3.0 |

*A/B is the number of acidic amino acids (Glu + Asp) divided by number of basic amino acids (Arg + Lys).
 †pI, isoelectric pH.
 ‡Observed values are equal to Asn + Asp.
 §Observed values are equal to Gln + Glu.

EcoRI
GAATTCCGCCTCGCCTGGTTCGCCGTCATGGCCCCAAGGAACCGACCCCAAGATA
100
ATCGTCTGCTCAACGCCACATCAACGAGGCGCTGAGTCCAAAGCGGTCTCGAGGCC
TTTCCGCCCAAGGCGCCACGCGGTCATGCCACGCCGATCAGACCCGCGGCTTCATC
200
GCAGACGAGATCCAGCGCTGGGCCGCGCTGTGCGCGAAACCGGCCCAAGCTGAAGTAG
300
CAGCGCAGCCTCCAAACGCCCATCCCCGTCCGCCGCGCACCATCCCGCATACGTGTGG
CAACGCCCAACCGCATCGGTGACGATTCTGCTACAAAACCTCGATTCTTCGTATCAT
400
CCCGTACTGCAATCCAACCGCATGAACGCTCTCTCGCGCAAAAGTCGCGCATGGTA
-35
CCGGTCACCGTCCCGACCGTCTGACGCCCTGCCATGGTGTGATCCGTAAATAGGCAC
-10
CATAAAACGCGAGGGGAAGACGGGATCGCTTGCACTCGGGCAATTCCGCAACCGCAA
500
GAACAGGCTGGCTGACGTGGCTGGCGATTCTTCCGCTACGGCGCCGCTGACTTCGCCG
R T G W L T W L A I L A V T A P V T S P
CATGGCCGACGATCTCCGCCACCGTATACCGTATGACTCCCGCCCGCGGAGGAGC
A W A *D D P P A T V Y R Y D S R P P E D
700
TTTTCCAGAACGGATTACCGCGTGGGGAACCAACGACAATGTGCTGCACATCTGACCG
V F Q N G F T A W G N N D N V L D H L T
GACGTTCTCCAGGTCGCGACGACCAACGCGCTTTCGCTCCACGACGACGACCGGC
G R S C Q V G S S N S A F V S T S S S R
800
GCTATACCGAGGTCTATCTCAACATCGCATGAGGAGCGGTGAGGCGCAACGCGCG
R Y T E V Y L E H R M Q E A V E A E R A
900
GCAGGGGACCGCCACTTCTCGGCTACATCTACGAAGTCGCGCGCAACAACTTCT
G R G T G H F I G Y I Y E V R A D N N F
ACGGCGCGCGCTCGTACTTCGAATACGTCGACACTTATGGCGCAATCGCGCGCTGA
Y G A A S S Y V D T Y G D N A G R
1000
TCTCGCGCGCGCTGGCCCTACCGAGGCAATATCTGGCACCGCGCGCTTCCG
I L A G A L A T Y Q S E Y L A H R R I P
CCGAAACATCCGAGGGTAAACGGGTCTATCAACGCGCATCCGCGCGAGACCGCA
P E N I R R V T Y G H G I T Y G E T T
1100
CCACGGAGTATTCACCGCTCGCTACGTCAGCCAGCACTCGCGCAATCCCAACCGCT
T T E Y S N A R Y V S Q Q T R A N P N P
1200
ACACATCGCGAAGGTCGCTAGCGTCGATCGTCGGCACATGGTGCATGGCGCGGTGATA
Y T S R R S V A S I V G T L V H G A G D
GCGCTGTCATGGCGCGGAGCGCAAGCTCCGAGGCGCATGGCAGCCTGGTCCGAACCG
S A C M A R Q A E S S E A M A A W S E R
1300
CCGGCGAGGCGATGGTCTCTGTACTACGAAGCATCGCGTATTCGTTCTAGACCTGGC
A G E A M V L Y Y E S
1400
CCAGCCCCGCCCAACTCCGGTAATTGAACAGCATGCCGATCGACCGCAAGACGCTCTGCC
FM P I D R K T L C
ATCTCTGTCTCGTTCGCTTGGCCCTCTCGGATCTCACTGGCGCGGCGCTCCACGC
H L L S V L L A L L G S H V A R A *S T
1500
CAGGCATCGTTCATTCGCGCGAGGAACAGATTACCCAGCATGGCAGCCCTATGGACGCT
P G I V I P P Q E Q I T Q H G S P Y G R
GCGCAACAGACCGCTGCCCTGACCGTGCGGAATTGCGCGGACGCGGCTATCGCAGG
C A N K T R A V A E L R G S G D L Q
1600
AGTACCTGCGTATGTGACGCGCGCTGGTCAATTTGCGCTCTACGATGGCACCTATC
E Y L R H V T R G W S I F A L Y D G T Y
TCGGCGCGAATATGGCGCGTGTATCAAGGACGGAACACCGCGCGGCTTCGACCTGA
L G G E Y G V I K D G T P G G A F D L
1700
AAACGAGCTTCTGCATCAGCACCGCAATACGGGTCAACCGCGCAACGATCACTACT
K T T F C I M T T R N T G Q P A T D H Y
1800
ACAGCAACGTCACCGCACTCGCTGCTCTCCAGCACCAACAGCAGGCTATGCGCGGTCT
Y S N V T A T R L L S S T N S R L C A V
TCGTGAGAAGCGGCAACCGTCAATTGGCGCTGACCAACCGCGTATGACGCGCAAGTACT
F V R S G Q P V I G A C T S P Y D G K A Y
1900
GGAGCATGTACAGCGGCTCGGGAATGCTTACCTGATCTACGTGGCGCGCATCTCCG
W S M Y S R L R K M L Y L I Y V A G I S
TACGCGTCCATGTACAGGAAGAACAGTATTACGACTATGAGGACGCAACCTTCGAGA
V R G H V S K E E Q T T Y D Y E D A T F E
2000
CTTACGCCCTTACCGCATCTCCATCTGCAATCTGGATATCTTATGCTGAGACGCTT
T Y A L T G I S I C N P G S S L C U
2100
CCCCACTCGAACCACCGCCCGGACAGGGCGGCGCCGCGGTGCGCGGTGCGCGCCT
FM R A L

GGCGTGGTGTGCGCATCGGCGCGATGACGATCTTTCCCGCGCCTGGCGGACGTTCC
A W L L A S G A M T H L S P A L A *D V P
2200
TTATGTGCTGGTGAAGACCAATATGGTGGTCAACGCGTAGCCATGAAGCGGTGAAGT
Y V L V K T N M V V T S V A M K P Y E V
CACCVCAGCGCATGCTGGTCTGCGGCATCGCCGCAAACTGGGCGCGCGGCGAGCAG
T P T R M L V C G I A A K L G A A A S S
2300
CCCGGACGCGACGTGCGCTTCTGCTTGGCAAGGATCTCAAGCGTCCCGGACGAGTCC
P D A H V P F C F G K D L K R P G S S P
2400
CATGGAAGTATGTTGCGCGCGCTTTCATGCAACACGCGCGTGGCATGTTTCTGGG
M E V M L R A V F M Q Q R P L R M F L G
TCCAAGCAACTCACTTTGAAAGGAAGCGCGCTCGAAGTATCGGATGGTGAATG
P K Q L T F E G K P A L E L I R M V E C
2500
CAGCGCAAGCAGGATTGCGCTTGAAGGCAACCCCATACCATCGATCCATCCCTG
S G K Q D C P U
FM H T I A S I L
TTGTCGCTGCTCGCATATACAGCCCGGCTGACGTGCGCGGCTTGGCGACCATCTGTAC
L S V L G I Y S P A D V *A G L P T H L Y
2600
AAGAAGTCTACGTGTCAGGAGTGGCTTGAAGTGAAGGCGAAGATCAGGAGTCTGTC
K N F T V Q E L A L K L K G K N Q E F C
2700
CTGACCGCTTCATGTCGGGCAAGCGTGGTCCGGGCGCTGCTGTCGACGCGGACAC
L T A F M S G R S L V R A T C T P G D L H
GAGCAGCACGCTGGTTCGACCATGCTTGGCTTGGCATATCGCGTATGCGCTCAAG
E H D T W F D T M L G F A I S A Y A L K
2800
AGCGGATCGCGTACGCTGGAAGACTGCGGTATCGGGGCACTCCCGGATCTGCTC
S R I A L T V E D S P Y P G T P G D L H
GAACTGACAGTCTGCGCGCTCAACGGAATTTGCGAATGAACCTTCCGAGGTTTCGACG
E L Q I C P L N G Y C E U
2900
TTTCCGCGCAATCCGCTTGAGACGATCTTCCGCGCTGGTTCATTCCGGGAACACGCA
3000
CATGCTGATCAACAACGAAGCTGCTTATCATCATCTGCGCATCTGGTCTCGCGCT
FM L I N N K K L L H H I L P I L V L A L
GCTGGGATGCGCAGCGCGCGGCTGCGCAGGATCTGTCATCCGCGGAAGGCACT
L G M R T A Q A *V A P G I V I P P K A L
3100
GTTCCACCAACAGGCGCGCTATGGACGCTGCGGCAACCGGCGCTTACCGT
F T A Q Q G G A Y G R C P N G T R A L T V
GGCGGAAGTCCGCGGCAACCGCAATTCGACGATTTGCGCAGATACCGCGCGCTG
A E L R G N A E L Q T R I T P G W
3200
GTCCATATACGGTCTCTATGACGCTACCTAGGCGGCGGCGTACGCGGCGCATCA
S I Y G L Y D G T Y L G Q A Y G G I I K
3300
GGACGCGCGCGGCGGCGGCTTATTTATCGGCAAACTTTGTCATACGACCATAT
D A P P G A G F I Y R E T F C I T I Y
CAAGACCGGCAACCGGCTGCGGATCACTACTACGCAAGGTACGCGGCAACGCGCTGCT
K T G Q P A A D H Y Y S K V T A T R L L
3400
CGCCAGCACCAACAGCAGGCTGTGCGCGGTATTGTCAGGCGGCGCAATCGGTATCGG
A S T N S R L C A V F V R D G Q S V I G
AGCGTCCGCGCGCTATGAAGGCGGTACAGACATGTACGACGCGCTGCGCGCGCT
A C A S P Y E G R Y R D M T A L R L
3500
GCTGTACATGATCTATATGTCGCGCTTGGCGTACGCGTACGTCAGCAAGGAAGACA
L Y M I Y M S G L A V R V H V S K E E Q
3600
GTATTACGACTACGAGGACGCGACATTCAGACCTATGCGCTACCGGCAATTCCTCTG
Y D Y E D A T F Q T Y A L T G I S L C
CAACCGGCGCGTGTATGCTGAGCGCGGCTCGGATCTGTTGCGCTGTCATGTTT
N P A A S I C U
3700
TTCCTTGACGATACCGGGAATGAATCCCTTGAAGACTTGAGAGCATGCTACCGCGCC
TGCCCTTCATGGCAGCTGACCGCTGTTGTCGCGACGCTGCGGACCTGCGGAGGCGG
3800
GCGCGGCGTGCAGCGCTGTCAACCACTTCATGGCGAGCATGCTGGTCTACTGCGCGG
CGGTGAGTGGCCACGCTGACCATCGCCATAATCTGGCGGCGTACAAGCTGCTGTTCCG
3900
CACGCGGATGCTGAGCGTGGTGGTGGTGGTGGCGGAGCTGCTGATCGCGCATC
GGCGAAATCGCTGTTATCTGCTGACCTGAATCTGGACATGCAACATCGTGATCC
4000
GCTTTTCAAGGCTGCAACCGCGCGGCTGCTGATGGCTACCGCGGCGGAGGCGG
4100
TGTGACGCGGCGACCATCTGCTGGGCTATCTGGTTCAGCATCGCTTCTGCGCTT
4200
GTTTCCGCTGCGATGCTGGCGATGCGGATCATGATCGGCGGATGACGAGCAGTTCG
Sau3A
CCTGATC

Fig. 2. Complete nucleotide sequence of the pertussis toxin gene. The insert DNA from pPTX42, that is, the 4.5-kb Eco RI–Bam HI fragment shown in Fig. 1a, was digested with restriction enzymes and subcloned by standard procedures (43) with the cloning vectors M13 mp18 and M13 mp19 and *E. coli* strain JM101 (17). Both strands of the DNA were sequenced by either the Maxam and Gilbert base-specific chemical cleavage method (15) or the dideoxy chain termination method (16) with the universal 17-base primer, or both. The DNA sequence and the derived amino acid sequence were analyzed with the use of MicroGenie computer software. The deduced

amino acid sequence of the individual subunits are shown in the single letter code below the nucleotide sequence. The proposed signal peptide cleavage sites are indicated by asterisks. The start of the protein coding region for each subunit is indicated by the box and arrow over the initiation codon. Putative ribosomal binding sites are underlined. The promoter-like sequence is shown in the –35 and –10 boxes. Proposed transcriptional start site is indicated by the arrow in the CAT box. Inverted repeats are indicated by the arrows in the flanking regions.

The assignment of S1 is further supported by a significant homology of two regions in the amino acid sequence with two analogous regions in the A subunits of both cholera and *Escherichia coli* heat labile toxins (Fig. 3). These regions may be part of functional domains responsible for the similar catalytic activities of all three toxins (21). Furthermore, the assignment of S1, as well as the correct prediction of the signal peptide cleavage site, is supported by preliminary amino acid sequence data for the mature protein (22). The absence of lysine residues in S1 may explain why lysine-specific chemical modifications do not affect the biological and enzymatic activities of S1 (10).

Subunits S2 and S3 share 70% amino acid homology, which makes the correct assignment of these subunits to their ORF's difficult if it is based only on the amino acid composition and the molecular weight. Nevertheless, we could determine the gene order as shown in Fig. 1d with the help of a Tn5-induced *B. pertussis* mutant. Although unable to produce active pertussis toxin in the culture supernatant (23), this mutant still produces the S2 subunit (24), thus the mutation is not located in the cistron for S2. We previously mapped this Tn5-insertion 1.3 kb downstream from the start site for the S4 subunit cistron (12), which positions it in the ORF for the S3 subunit (Fig. 1).

The amino acid sequence for each subunit was deduced from the nucleotide sequence and is presented in Fig. 2. The mature proteins contain 234 amino acids for S1, 199 amino acids for S2 and S3, 110 amino acids for S4, and 100 amino acids for S5. The calculated molecular weights from the deduced sequence of the mature peptides are 26,024 for S1, 21,924 for S2, 21,873 for S3, 12,058 for S4, and 11,013 for S5. Since S4 is present in two copies per toxin molecule (7, 10), the total molecular weight for the holotoxin should be 104,950. This is in agreement with the apparent molecular weight estimated by nondenaturing PAGE.

It is likely that all subunits contain signal peptides, as expected for secretory proteins. The length of the putative signal peptides was estimated after analysis of the hydrophobicity plot (25), the predicted secondary structure (26), and the application of von Heijne's rule for the prediction of the most probable signal peptide cleavage site (27). The predicted cleavage site for each subunit is indicated in Fig. 2 by the asterisks and has been confirmed for S1, S3, and S4 by sequencing the amino termini of the purified mature subunits (12, 22). The length of the signal peptides varies from 34 residues for S1, 28 residues for S3, and 27 residues for S2, to 21 residues for S4, and 20 resi-

Region 1

| | | | | | | | | | | |
|------------------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
| Pertussis S1 subunit | (8) | Tyr | Arg | Tyr | Asp | Ser | Arg | Pro | Pro | (15) |
| Cholera A subunit | (6) | Tyr | Arg | Ala | Asp | Ser | Arg | Pro | Pro | (13) |
| <i>E. coli</i> HLT A subunit | (6) | Tyr | Arg | Ala | Asp | Ser | Arg | Pro | Pro | (13) |

Region 2

| | | | | | | | | | | |
|------------------------------|------|-----|-----|-----|-----|-----|-----|-----|-----|------|
| Pertussis S1 subunit | (51) | Val | Ser | Thr | Ser | Ser | Ser | Arg | Arg | (58) |
| Cholera A subunit | (60) | Val | Ser | Thr | Ser | Ile | Ser | Leu | Arg | (67) |
| <i>E. coli</i> HLT A subunit | (60) | Val | Ser | Thr | Ser | Leu | Ser | Leu | Arg | (67) |

Fig. 3. Comparison of two homologous regions in ADP-ribosylating subunits of pertussis, cholera, and *E. coli* heat labile toxins. The numbers in parentheses refer to the amino acid position in the mature proteins. *Escherichia coli* heat labile and cholera toxin sequences are from Yamamoto *et al.* (21). Common residues in all three toxins are boxed. HLT, heat labile toxin.

dues for S5. All of the signal peptides contain a positively charged amino terminal region of variable length, followed by a sequence of hydrophobic amino acids, usually in α -helical or partially α -helical, partially β -pleated conformation. A less hydrophobic carboxyl terminal region follows, usually ending in a β -turn conformation at the signal peptide cleavage site. All subunits except S5 follow the -1, -3 rule (28), which positions the cleavage site after Ala-X-Ala. The amino terminal charge for the subunit signal peptides varies between +4 for S1 and +1 for S4 and S5. All described properties correspond very well to the general characteristics for bacterial signal peptides (29).

An interesting feature of the sequence is the overall, relatively high amount of cysteines compared to *E. coli* proteins. Cysteines are not involved in inter-subunit links in the quaternary structure of the toxin, since all subunits can be separated by SDS-PAGE in the absence of reducing agents. The cysteines are probably involved in intra-chain bonds, since reducing agents significantly change the electrophoretic mobility of all subunits except S4 (30). Serines, threonines, and tyrosines also are represented more frequently than in average *E. coli* proteins. The hydroxyl groups of these residues may be involved in the quaternary structure through hydrogen bonding.

The most striking feature of the predicted peptide sequences is the high degree of homology between S2 and S3. The two peptides are 70% homologous in their amino acid sequences and 75% homologous in their nucleotide sequences. This suggests that both cistrons were generated through a duplication of an ancestral cistron followed by mutations that result in functionally different peptides. Despite their high homology, also reflected in the predicted secondary structures and hydrophilicities, S2 and S3 subunits cannot substitute for each other in the functionally active pertussis toxin (7). The comparison between the two subunits may be useful in localizing their functional domains in relation to their primary, secondary, and tertiary structures. The differ-

ences between S2 and S3 are scattered throughout the whole sequence and are slightly more frequent in the amino terminal half of the peptides. Each of the subunits binds to one S4 subunit (7). This function could be located in the more conserved carboxyl terminal domains of S2 and S3. The two resulting dimers are thought to bind to the same S5 subunit. This function could be assigned to the more divergent amino terminal domains of S2 and S3. Alternatively, it is possible that the dimers bind to the S5 subunit through S4 and that the amino terminal domains of S2 and S3 are involved in some other function, possibly the interaction of the binding moiety (S2 through S5) with the enzymatically active moiety (S1).

The S1 subunit was compared to the A subunits of other bacterial toxins. Two regions with significant homology to cholera and *E. coli* heat labile toxins were found (Fig. 3). They are located in analogous regions of all three toxins. However, the three amino acid differences found in these regions cannot be explained by single base pair changes in the DNA. Furthermore, in most cases the homologous amino acids use quite different codons in pertussis toxin compared to cholera and *E. coli* heat labile toxins. This, together with the fact that no other significant homology in the primary structure could be found and that the amino acid sequences of the other subunits are completely different from the sequence of any other ADP-ribosylating toxin, strongly suggests that pertussis toxin is not evolutionarily related to any of the other known bacterial toxins. The nicotinamide adenine dinucleotide (NAD)-binding site for the two enterotoxins has been identified at the carboxyl terminal region of their A₁ subunit (13, 31). No significant homology could be found between the carboxyl terminus of the enterotoxins nor any other NAD-binding enzymes and the analogous region in the S1 subunit. This suggests that the NAD-binding function of the ADP-ribosylating enzymes is dependent more on the secondary or tertiary structures than on the primary structures. We propose that the two enzy-

Since all pertussis toxin subunits are closely linked and probably expressed in a very precise ratio, it is possible that they are arranged in a polycistronic operon. A polycistronic arrangement for the subunit cistrons also has been described for other bacterial toxins bearing similar enzymatic functions, such as diphtheria, cholera, and *E. coli* heat labile toxins (14, 21, 32). We therefore analyzed the flanking regions for the presence of transcriptional signals. In the 5'-flanking region, starting at position 469, the sequence TAAAATA was found, which matches six of the seven nucleotides found in the ideal TATAATA Pribnow or -10 box. An identical sequence can be found in several other bacterial promoters, including the phage λ L57 promoter (33). Given the fact that most transcripts start at a purine

residue about five to seven nucleotides downstream from the Pribnow box, we tentatively located the transcriptional start site at the adenine residue at position 482. This residue is located in the sequence CAT, often found at transcriptional start sites (33). Upstream from the proposed -10 box, the sequence CTGACC starts at position 442. This sequence matches four of the six nucleotides found in the ideal *E. coli* -35 box TTGACA. The mismatching nucleotides in the proposed pertussis toxin -35 box are the two end nucleotides, of which the 3'-residue is the less important nucleotide in the *E. coli* -35 consensus box. A replacement of the T by a C in the first position of the consensus sequence can also be found in several *E. coli* promoters (33). The distance between the two proposed promoter boxes is 21 nucleotides, a distance too long to be efficiently functional in *E. coli*, although a distance of the same length has been found in the galP1 promoter (34) and in several plasmid promoters (33). The proposed -35 box is immediately preceded by

two overlapping, short inverted repeats with calculated free energies (35) of -15.6 kcal and -8.6 kcal, respectively. Inverted repeats can also be found at the 5'-end of the cholera toxin promoter. In both cases they may be involved in positive regulation of the toxin promoters. Evidence for a positive regulation of toxin biosynthesis in *B. pertussis* came through the discovery of the *vir* gene, the product of which is essential for the expression of many virulence factors, including pertussis toxin (36). None of the ORF's assigned to the other subunit is closely preceded by a similar promoter-like structure.

The 3'-flanking region has been examined for the presence of possible transcriptional termination sites. Several inverted repeats could be found: the most significant is located in the region extending from position 4031 to 4089 and has a calculated free energy (35) of -41.4 kcal. None of the inverted repeats are immediately followed by an oligo(dT) stretch, which may suggest that they function in a rho-dependent fash-

| Pertussis toxin | | | | | | | E. coli | | Pertussis toxin | | | | | | | E. coli | | | | |
|-----------------|-----|----|----|----|----|----|---------|-----|-----------------|------|-----|-----|-----|----|----|---------|-----|----|-----|----|
| | | S1 | S2 | S3 | S4 | S5 | PTX | S | W | | | S1 | S2 | S3 | S4 | S5 | PTX | S | W | |
| Ala | GCU | 3 | 0 | 1 | 0 | 1 | 5 | 33 | 17 | Lys | AAA | 0 | 2 | 0 | 1 | 1 | 4 | 49 | 31 | |
| | GCC | 17 | 7 | 14 | 9 | 4 | 52 | 9 | 34 | | AAG | 0 | 5 | 7 | 7 | 4 | 24 | 20 | 8 | |
| | GCA | 5 | 3 | 2 | 1 | 1 | 12 | 23 | 20 | | Met | AUG | 4 | 3 | 4 | 9 | 2 | 22 | 27 | 25 |
| | GCG | 9 | 5 | 8 | 5 | 5 | 33 | 25 | 28 | | | Phe | UUU | 0 | 1 | 0 | 1 | 1 | 3 | 7 |
| Arg | CGU | 3 | 2 | 0 | 1 | 0 | 6 | 42 | 19 | UUC | 7 | | 4 | 5 | 4 | 4 | 25 | 22 | 19 | |
| | CGC | 12 | 7 | 9 | 4 | 0 | 33 | 19 | 25 | Pro | CCU | | 1 | 1 | 0 | 1 | 0 | 3 | 4 | 6 |
| | CGA | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 5 | | CCC | | 5 | 3 | 2 | 6 | 1 | 17 | 0.4 | 9 |
| | CGG | 5 | 3 | 1 | 2 | 2 | 13 | 0.2 | 8 | | CCA | 0 | 1 | 2 | 0 | 0 | 3 | 5 | 9 | |
| | AGA | 1 | 1 | 1 | 0 | 1 | 4 | 1 | 5 | | CCG | 4 | 6 | 7 | 5 | 5 | 28 | 31 | 19 | |
| | AGG | 3 | 1 | 3 | 0 | 0 | 7 | 0.2 | 3 | Ser | UCU | 0 | 1 | 0 | 0 | 0 | 1 | 18 | 7 | |
| Asn | AAU | 4 | 2 | 0 | 1 | 1 | 8 | 2 | 19 | | UCC | 7 | 6 | 3 | 2 | 4 | 23 | 17 | 9 | |
| | AAC | 9 | 3 | 6 | 0 | 2 | 20 | 30 | 19 | | UCA | 0 | 2 | 0 | 0 | 0 | 2 | 1 | 7 | |
| Asp | GAU | 2 | 3 | 1 | 2 | 1 | 9 | 22 | 35 | | UCG | 5 | 0 | 2 | 0 | 2 | 9 | 2 | 12 | |
| | GAC | 8 | 6 | 7 | 2 | 5 | 29 | 39 | 20 | AGU | 0 | 0 | 0 | 1 | 0 | 1 | 2 | 11 | | |
| Cys | UGU | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 6 | AGC | 12 | 10 | 5 | 5 | 3 | 36 | 9 | 12 | | |
| | UGC | 3 | 7 | 6 | 4 | 4 | 25 | 4 | 7 | Thr | ACU | 4 | 2 | 1 | 1 | 2 | 10 | 20 | 9 | |
| Gln | CAA | 1 | 2 | 3 | 3 | 0 | 9 | 7 | 17 | | ACC | 10 | 9 | 8 | 3 | 4 | 35 | 26 | 23 | |
| | CAG | 7 | 5 | 7 | 1 | 3 | 24 | 32 | 32 | | ACA | 3 | 1 | 1 | 0 | 0 | 5 | 3 | 6 | |
| Glu | GAA | 10 | 5 | 5 | 5 | 3 | 29 | 63 | 40 | | ACG | 6 | 9 | 7 | 2 | 2 | 27 | 5 | 15 | |
| | GAG | 7 | 3 | 2 | 0 | 3 | 15 | 20 | 19 | Trp | UGG | 5 | 2 | 1 | 1 | 1 | 10 | 5 | 13 | |
| Gly | GGU | 1 | 1 | 2 | 1 | 0 | 5 | 43 | 24 | | Tyr | UAU | 8 | 6 | 8 | 2 | 3 | 28 | 6 | 18 |
| | GGC | 15 | 16 | 13 | 7 | 7 | 59 | 33 | 27 | UAC | | 11 | 10 | 11 | 0 | 2 | 35 | 19 | 12 | |
| | GGA | 3 | 4 | 3 | 0 | 2 | 12 | 1 | 8 | Val | GUU | 2 | 1 | 1 | 1 | 0 | 5 | 37 | 21 | |
| | GGG | 0 | 1 | 3 | 0 | 0 | 4 | 3 | 13 | | GUC | 10 | 7 | 6 | 6 | 3 | 33 | 8 | 13 | |
| His | CAU | 3 | 4 | 1 | 1 | 2 | 11 | 4 | 18 | GUA | 3 | 1 | 2 | 1 | 0 | 7 | 23 | 9 | | |
| | CAC | 3 | 2 | 3 | 1 | 2 | 11 | 14 | 11 | GUG | 4 | 5 | 2 | 4 | 2 | 17 | 16 | 24 | | |
| Ile | AUU | 3 | 3 | 3 | 0 | 0 | 9 | 13 | 30 | End | UAA | - | - | - | - | - | 0 | ND | ND | |
| | AUC | 7 | 8 | 9 | 2 | 4 | 31 | 15 | 23 | | UAG | 1 | - | - | - | - | 1 | ND | ND | |
| | AUA | 0 | 1 | 4 | 0 | 2 | 7 | 0.4 | 5 | | UGA | - | 1 | 1 | 1 | 1 | 4 | ND | ND | |
| Leu | UUA | 0 | 1 | 0 | 0 | 0 | 1 | 2 | 14 | fMet | AUG | 1 | 1 | 1 | - | 1 | 4 | ND | ND | |
| | UUG | 1 | 2 | 3 | 2 | 3 | 11 | 3 | 12 | | GUG | - | - | - | 1 | - | 1 | ND | ND | |
| | CUU | 1 | 2 | 2 | 1 | 1 | 7 | 5 | 14 | | | | | | | | | | | |
| | CUC | 4 | 7 | 5 | 3 | 4 | 24 | 6 | 13 | | | | | | | | | | | |
| | CUA | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 4 | | | | | | | | | | | |
| | CUG | 5 | 9 | 14 | 9 | 10 | 48 | 66 | 56 | | | | | | | | | | | |

Fig. 4. Comparison of codon usage between pertussis toxin and strongly and weakly expressed *E. coli* genes. Absolute codon usage for the subunit cistrons of pertussis toxin include the signal peptides. The number of codons in the five individual subunits are 269(S1), 227(S2), 228(S3), 132(S4), and 121(S5). The codon frequencies for *E. coli* are deduced from Grosjean and Fiers (42). S, strongly expressed genes; W, moderately to weakly expressed

genes. All *E. coli* and the PTX data are expressed in relative codon usage per 1000 codons. The pertussis usage is based on 977 codons for the pertussis toxin gene (PTX). *Escherichia coli* usage is based on 5253 codons for highly expressed genes (S) and 5231 codons for moderately to weakly expressed genes (W). ND, not determined.

ion (37). Whether any of these inverted repeats is functional in *B. pertussis* remains to be established and can be investigated by a small deletion or by site-directed mutagenesis experiments, which are feasible now that the DNA sequence is known. However, the gene may contain other, not yet identified, cistrons related to expression and possibly involved in regulation, assembly, or transport of functionally active pertussis toxin or other virulence factors. Nonstructural proteins involved in the posttranslational processing of *E. coli* heat labile toxin have been proposed (38). However, no significantly long ORF was found at the 3'-end of the nucleotide sequence shown in Fig. 1b. If other proteins are encoded by the same polycistronic operon, their coding regions must be located further downstream.

Finally, we examined the 5'-flanking region of each cistron for the presence of ribosomal binding sites. Neither the ribosomal binding sequences for *B. pertussis* genes, nor the 3'-end sequence of the 16S ribosomal RNA (rRNA) are known. Therefore, we could only compare the flanking regions with the ribosomal binding sequences of heterologous prokaryotic organisms represented by the Shine-Dalgarno sequence (39). Preceding the S1 initiation codon we found the sequence GGGGAAG, starting at position 495. This sequence shares four out of seven nucleotides with the ideal Shine-Dalgarno sequence AAG-GAGG. The two first mismatching nucleotides in the pertussis toxin gene would not destabilize the hybridization to the 3'-end of the *E. coli* 16S rRNA. This putative ribosomal binding site is close enough to the initiation codon for S1 to be functional in *E. coli*. Another possible Shine-Dalgarno sequence overlaps the first one and also matches four out of seven nucleotides to the consensus sequence. The mismatching nucleotides, however, have a more destabilizing effect than the ones found in the first sequence. The S2 subunit ORF is not closely preceded by a ribosomal binding sequence, which may suggest that S2 is translated through a mechanism not involving the detachment and reattachment of the ribosome between the coding regions for S1 and S2. The short distance between the S1 and S2 cistrons and the absence of a ribosomal binding site are characteristic of this mechanism. A possible ribosomal binding site for S4 is the sequence CAGGGCGGC, starting at position 2066. The ORF for S5 is preceded by the sequence AAGGCG, starting at position 2485, which matches five out of six nucleotides in the consensus sequence AAGGAG. Finally, S3 is preceded by the sequence GGGAACAC, which is very similar to the proposed ribosomal

binding site for S1, that is, GGGAAGAC.

Two different initiation codons (40) are used for the translation of all subunits in *B. pertussis*, that is, the most frequently used ATG for S1, S2, S3, and S5, and the less frequently used GTG for S4. The initiation codon GTG is also found in the diphtheria toxin gene (14).

Attempts to efficiently express the pertussis toxin gene in *E. coli* have been unsuccessful thus far (12), although very sensitive monoclonal and polyclonal antibodies are available (24). This lack of expression in *E. coli* may be because *B. pertussis* promoters are not efficiently recognized by the *E. coli* RNA polymerase. Analysis of the promoter-like structure of the pertussis toxin gene and comparison of this structure to strong *E. coli* promoters reveal very significant differences indeed, of which the most striking one is the unusually long distance between the proposed -35 and -10 boxes in the pertussis toxin promoter. The distance between those two boxes in strong *E. coli* promoters is around 17 nucleotides (41), whereas the distance in the putative pertussis toxin promoter is 21 nucleotides. By using expression vectors designed to detect heterologous expression signals that are able to function in *E. coli*, we have found further evidence that *B. pertussis* promoters may not be recognized by the *E. coli* expression machinery.

Even in the presence of *E. coli* promoters, such as the inducible strong *lac/trp* fusion promoters, recombinant expression may be hampered by translational difficulties. Indeed, the codon usage (Fig. 4) is unsuitable for efficient translation of the pertussis toxin gene in *E. coli* (42). This is reflected by the codon choice for frequently used amino acids, such as alanine, arginine, glycine, histidine, lysine, proline, serine, and valine. Efficient expression in *E. coli* may require resynthesis of the pertussis toxin operon, respecting the optimal codon usage for *E. coli*. It is not known whether the codon usage for pertussis toxin reflects the optimal codon usage for expression in *B. pertussis*, since no other *B. pertussis* gene has been sequenced. Whether pertussis toxin is a strongly or weakly expressed protein in *B. pertussis* and whether this expression is regulated by the presence of a precise relative amount of the different transfer RNA isoacceptors, possibly different from *E. coli*, can be evaluated by in vitro translation with the use of *E. coli* and *B. pertussis* cell-free extracts.

The cloned and sequenced pertussis toxin gene will facilitate the development of an efficient and safer vaccine against whooping cough. By comparison to other toxin genes with similar biochemical functions, and by physical identification of the active sites either for the ADP-ribosylation in the S1

subunit or the target cell binding in subunits S2 through S4, it is now possible to modify those sites by site-directed mutagenesis of the *B. pertussis* genome. These modifications could abolish the pathobiological activities of pertussis toxin without hampering its immunogenicity and protectivity. Alternatively, by knowing the DNA sequence it will be possible to map protective epitopes. Synthetic oligopeptides that include those epitopes also will be useful in the development of a new generation of vaccines.

REFERENCES AND NOTES

- Pertussis toxin is also known as lymphocytosis-promoting factor, leukocytosis-promoting factor, histamine-sensitizing factor, islets-activating protein, and pertussigen [for a review of pertussis vaccines, see C. R. Manclark and J. L. Cowell, in *Bacterial Vaccines*, R. Germanier, Ed. (Academic Press, New York, 1984), pp. 69-106].
- J. J. Munoz, in *Pertussis Toxin*, R. D. Sekura, J. Moss, M. Vaughan, Eds. (Academic Press, New York, 1985), pp. 1-18.
- J. J. Munoz, H. Arai, R. L. Cole, *Infect. Immun.* **32**, 243 (1981).
- H. Sato and Y. Sato, *ibid.* **46**, 415 (1984).
- Y. Sato, M. Kimura, H. Fukumi, *Lancet* **1984-I**, 122 (1984).
- L. Steinman *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 8733 (1985).
- M. Tamura *et al.*, *Biochemistry* **21**, 5516 (1982).
- T. Katada and M. Ui, *J. Biol. Chem.* **257**, 7210 (1982); T. Katada, M. Tamura, M. Ui, *Arch. Biochem. Biophys.* **224**, 290 (1983).
- M. Tamura, K. Nogimori, M. Yajima, K. Ase, M. Ui, *J. Biol. Chem.* **258**, 6756 (1983).
- M. Ui, K. Nogimori, M. Tamura, in *Pertussis Toxin*, R. D. Sekura, J. Moss, M. Vaughan, Eds. (Academic Press, New York, 1985), pp. 19-43.
- D. M. Gill, in *Bacterial Toxins and Cell Membranes*, J. J. Eljaszewicz and T. Wadstrom, Eds. (Academic Press, New York, 1978), pp. 291-332.
- C. Loch *et al.*, *Nucleic Acids Res.* **14**, 3251 (1986).
- T. Yamamoto, A. Suyama, N. Mori, T. Yokota, A. Wada, *FEBS Lett.* **181**, 377 (1985); G. D. N. Pearson and J. J. Mekalanos, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 2976 (1982).
- M. Kaczorek *et al.*, *Science* **221**, 855 (1983).
- A. M. Maxam and W. Gilbert, *Methods Enzymol.* **65**, 499 (1980).
- F. Sanger, S. Nicklen, A. R. Coulson, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463 (1977).
- J. Messing, *Methods Enzymol.* **101**, 20 (1983).
- K. Kersters *et al.*, *Int. J. Syst. Bacteriol.* **34**, 56 (1984).
- J. W. Fickett, *Nucleic Acids Res.* **10**, 5303 (1982).
- In Tamura *et al.* (7), no error limits for the amino acid composition were given. However, since amino acids must be present in whole number molar ratios, it is easy to see that the percentage of His in S4, for example, is underestimated. In this mature subunit there is one His residue in 110 amino acids. Thus, the experimental value should be about 0.9%, not 0.5%. Operational factors contribute to errors in estimation of amino acid compositions. Such errors can be evaluated by including a known control protein in the analysis.
- T. Yamamoto, T. Nakazawa, T. Miyata, A. Kaji, T. Yokota, *FEBS Lett.* **169**, 241 (1984).
- K. Marchitto, J. E. Coligan, J. M. Keith, unpublished results.
- A. A. Weiss, E. L. Hewlett, G. A. Myers, S. Falkow, *Infect. Immun.* **42**, 33 (1983).
- K. Marchitto, J. J. Munoz, J. M. Keith, in preparation.
- J. Kyte and R. F. Doolittle, *J. Mol. Biol.* **157**, 105 (1982).
- J. Garnier, D. J. Osguthorpe, B. Robson, *ibid.* **120**, 97 (1978).
- G. von Heijne, *Eur. J. Biochem.* **133**, 17 (1983).
- , *J. Mol. Biol.* **173**, 243 (1984).
- , *ibid.* **184**, 99 (1985).
- M. S. Pepper, R. C. Judd, J. J. Munoz, *Dev. Biol. Stand.*, in press.
- C. Y. Lai, Q. C. Xia, P. T. Salotra, *Biochem. Biophys. Res. Commun.* **116**, 341 (1983); L. K. Duffy, A.

- Kurosky, C. Y. Lai, *Arch. Biochem. Biophys.* **239**, 549 (1985).
32. J. J. Mckalanos *et al.*, *Nature (London)* **306**, 551 (1983).
33. D. K. Hawley and W. R. McClure, *Nucleic Acids Res.* **11**, 2237 (1983).
34. R. Musso, R. DiLauro, M. Rosenberg, B. deCrombrughe, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 106 (1977).
35. I. Tinico *et al.*, *Nature (London) New Biol.* **246**, 40 (1973).
36. A. A. Weiss and S. Falkow, *Infect. Immun.* **43**, 263 (1984).
37. S. Adhya and M. Gottesman, *Annu. Rev. Biochem.* **47**, 967 (1978).
38. H. Hofstra and B. Witholt, *J. Biol. Chem.* **260**, 16037 (1985).
39. J. Shine and L. Dalgarno, *Nature (London)* **254**, 34 (1975).
40. M. Kozak, *Microbiol. Rev.* **47**, 1 (1983).
41. J. Brosius, M. Erfle, J. Storella, *J. Biol. Chem.* **260**, 3539 (1985).

42. H. Grosjean and W. Fiers, *Gene* **18**, 199 (1982).
43. T. Maniatis, E. F. Fritsch, J. Sambrook, *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982).
44. We thank Dr. K. Marchitto for reading the manuscript, S. Perryman for technical assistance, G. Hettrick for graphics, S. Smith for discussions, S. Smaus and H. Blahnik for help in preparing the manuscript, and A. Maquet for her patience.

18 February 1986; accepted 8 April 1986

Caffeine-Induced Uncoupling of Mitosis from the Completion of DNA Replication in Mammalian Cells

ROBERT SCHLEGEL* AND ARTHUR B. PARDEE

Caffeine was shown to induce mitotic events in mammalian cells before DNA replication (S phase) was completed. Synchronized BHK cells that were arrested in early S phase underwent premature chromosome condensation, nuclear envelope breakdown, morphological "rounding up," and mitosis-specific phosphoprotein synthesis when they were exposed to caffeine. These mitotic responses occurred only after the cells had entered S phase and only while DNA synthesis was inhibited by more than 70 percent. Inhibitors of protein synthesis blocked these caffeine-induced events, while inhibitors of RNA synthesis had little effect. These results suggest that caffeine induces the translation or stabilizes the protein product (or products) of mitosis-related RNA that accumulates in S-phase cells when DNA replication is suppressed. The ability to chemically manipulate the onset of mitosis should be useful for studying the regulation of this event in mammalian cells.

IN YEAST AND MAMMALIAN CELLS, MITOSIS does not occur when S phase is interrupted by drug treatments or by restrictive conditions for mutants that are

temperature-sensitive for DNA synthesis (1). Although premature chromosome condensation (PCC) and breakdown of the nuclear envelope can be induced in S-phase

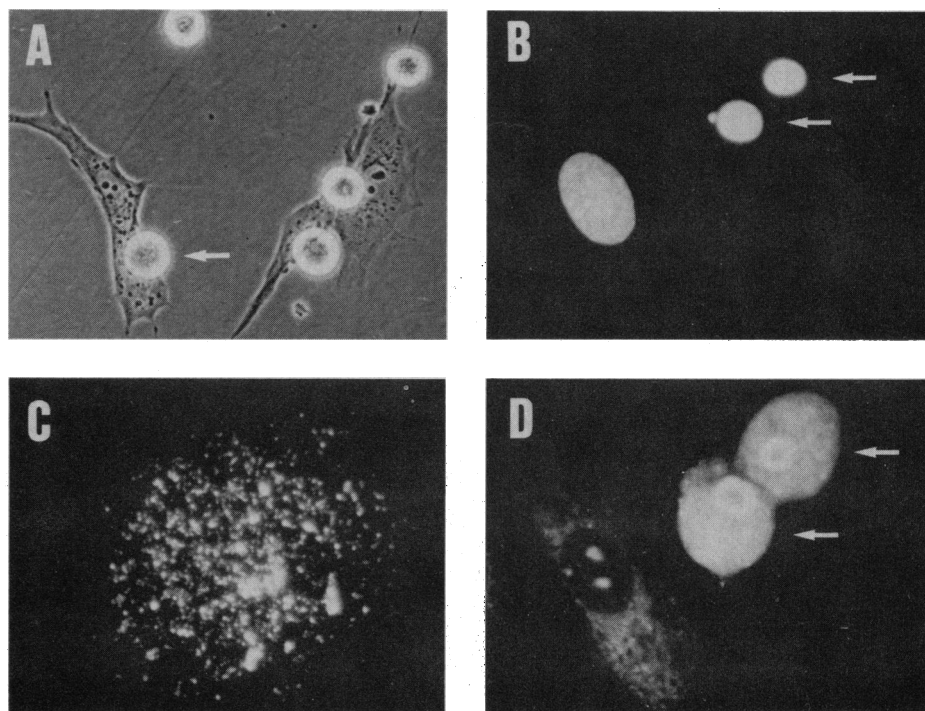
cells by fusing them with mitotic cells (2, 3), these events result from the transfer of preformed mitotic factors, not the uncoupling of endogenously induced mitotic events from the completion of DNA replication. The BHK (Syrian hamster fibroblast) temperature-sensitive mutant tsBN2 (4, 5), which undergoes PCC and other early mitotic events at the restrictive temperature, provides the only example of a cell line in which mitotic events can occur before S phase is completed. The biochemical processes that initiate mitosis are still unknown.

We have found that normal BHK cells that were arrested in S phase (6) and treated with caffeine underwent, within a few hours, the same early mitotic events described by Nishimoto (4, 5) for the tsBN2 mutant. Caffeine could induce PCC at con-

Department of Pharmacology, Harvard Medical School and Division of Cell Growth and Regulation, Dana-Farber Cancer Institute, Boston, MA 02115.

*To whom correspondence should be addressed at the Dana-Farber Cancer Institute.

Fig. 1. Cytoplasmic, chromosomal, and phosphoprotein synthesis changes in PCC-containing cells (indicated by arrows). Cells were plated at 1×10^5 per 60-mm dish in Dulbecco's modified Eagle's medium (DME) + 10 percent fetal calf serum (FCS) and incubated for 24 hours. Synchrony was achieved by incubation in isoleucine-deficient DME + 5 percent dialyzed FCS for 34 hours (16, 17). Cells were then treated with DME + 10 percent FCS + 2.5 mM hydroxyurea for 14 hours, and finally were given the specified drugs ($T = 0$) along with Colcemid (0.3 $\mu\text{g/ml}$). (A) Phase contrast photograph ($\times 840$) of "rounded up" PCC-containing cells 4 hours after the addition of 5 mM caffeine. No "rounded up" cells were seen in untreated controls. (B) Condensed chromatin of cells 8 hours after the addition of 5 mM caffeine. Cells were fixed in absolute methanol and stained for 10 minutes with Hoechst 33242 (1 $\mu\text{g/ml}$) ($\times 2100$). (C) Chromosome preparation showing S-phase PCC stained with Hoechst 33242 ($\times 2100$). Four hours after the addition of 5 mM caffeine and Colcemid (0.3 $\mu\text{g/ml}$) the cells were trypsinized, treated for 20 minutes at 37°C in 0.55 percent KCl, fixed for 30 minutes in methanol:acetic acid (3:1), dropped on wetted slides, and heated at 70°C overnight before staining. (D) Indirect immunofluorescence of mitosis-specific phosphoproteins present in both the cytoplasm and chromatin of the same PCC-containing cells shown in (B). Monoclonal antibody MPM-2 (9) and rhodamine-conjugated, goat-antimouse secondary antibody were diluted 1:500 and 1:1000, respectively, in phosphate-buffered saline (PBS) con-



taining 1 percent bovine serum albumin. Methanol-fixed cells were preincubated at 37°C with DME + 10 percent FCS for 30 minutes; incubated with MPM-2 for 4 hours; rinsed successively

with DME + 10 percent FCS, PBS, and distilled water; air-dried; incubated with the fluorescent secondary antibody for 1 hour; and rinsed as before.