Identification and Nucleotide Sequence of a Developmentally Regulated Gene Encoding a Eukaryotic Histone H1-Like Protein from *Chlamydia trachomatis*

SHA TAO, RAVI KAUL, AND WANDA M. WENMAN*

Division of Infectious Diseases, Department of Pediatrics, University of Alberta, Edmonton, Alberta T6G 2B7, Canada

Received 30 November 1990/Accepted 19 February 1991

A λgt11 recombinant library of *Chlamydia trachomatis* serovar L2 chromosomal DNA was screened with a 29-mer synthetic oligonucleotide specific to the N-terminal amino acids of a predominant 18-kDa chlamydial protein. One recombinant clone, designated λgt11/L2/RKA10, was selected on the basis of its strong hybridization signal. Restriction endonuclease analysis and complete nucleotide sequencing of the recombinant revealed a 2,633-bp insert containing one complete open reading frame (ORF2) and two partial ORFs (ORF1 and ORF3). The deduced amino acid sequence of ORF2 matched perfectly at its N-terminal end with the derived amino acid sequence. The 375-bp ORF is capable of encoding a protein comprising 125 amino acids with a molecular mass of 13,689. A sequence compatible with a Shine-Dalgarno ribosome-binding site was located 9 bp upstream from the initiation codon, while the sequence distal to ORF2 revealed a rho-independent terminator. The protein, designated CTH1, possesses an estimated pI of 10.71 due to its high lysine content. This highly basic protein contains no tryptophan or phenylalanine. A protein data base search identified significant homology between CTH1 and painted sea urchin histone H1. Northern (RNA) blot analysis of *Chlamydia*-infected host cells demonstrated transcripts at 12 h postinfection. The recombinant plasmid encoding ORF2 expressed a gene product of approximately 18 kDa, similar to the native chlamydial protein as analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis. This protein appears to represent one of the few eukaryotic histone-like proteins described to date in prokaryotes.

*Chlamydia trachomatis*, an obligate intracellular parasite of eukaryotic cells, is responsible for a wide spectrum of human diseases, including trachoma and sexually transmitted diseases (16). It possesses a unique and complex developmental cycle involving two distinct forms: extracellular, infectious elementary bodies (EB) and intracellular, noninfectious reticulate bodies. Following attachment and internalization, EB transform into metabolically active reticulate bodies, which divide by binary fission and switch back into the EB form late in the life cycle (22). Marked ultrastructural and metabolic changes occur during the transition between the two developmental forms. Because chlamydiae are not closely related to any other eubacterium, as demonstrated by 16S rRNA sequences (11, 23), they are placed in a distinct order, *Chlamydiales* (18).

The presence of a condensed, discrete, and electron-dense nucleoid in EB is a notable feature, unique among members of the family Eubacteriaceae (2). Wagar and Stephens reported three DNA-binding proteins with molecular masses of 58, 25, and 17 kDa specific to the EB form of chlamydiae (21). Nucleoproteins that are involved in DNA compaction and transcriptional regulation have been identified among prokaryotes (3, 13). These nucleoproteins share biochemical properties, such as DNA packaging, charge, and low molecular weights, with eukaryotic histones. All higher organisms contain small, basic DNA-binding proteins called histones which are conserved in their primary structure (10).

We report here the cloning and sequencing of an 18-kDa developmentally regulated protein-encoding gene from *C. trachomatis* that is homologous to the sea urchin histone H1 protein.

**MATERIALS AND METHODS**

**Reagents.** Restriction endonucleases, T4 polynucleotide kinase, DNA polymerase I, T4 DNA ligase, bacterial alkaline phosphatase, proteinase K, RNase, isopropyl-β-D-thiogalactopyranoside, and 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside were purchased from Boehringer Mannheim, Laval, Quebec, Canada. Nitrocellulose and aminobenzyloxyethyl cellulose paper for Northern (RNA) hybridization were purchased from Schleicher & Schuell, Inc., Keene, N.H. In vitro transcription-translation kits and 32P- and 35S-containing radiotopes were obtained from Amersham Corp., Arlington Heights, Ill. The sequencing reagents and enzymes for dideoxy sequencing were purchased from United States Biochemicals, Cleveland, Ohio. Universal primers were supplied by the Regional DNA Synthesis Facility, University of Calgary. Sequence-specific primers for gene sequencing and primer extension were supplied by the DNA Synthesis Facility, Department of Microbiology, University of Alberta. All other reagents and chemicals were obtained from Sigma Chemical Co., St. Louis, Mo., or Fisher Scientific Ltd., Edmonton, Alberta, Canada.

**Organisms.** *C. trachomatis* serovars L2 (L2/434/Bu), J (J/JUW-36), D (DUW-3), and K (K/UW-31) and *C. psittaci* Mn (meningopneumonitis) were grown in HeLa 229 cells as described by Kuo et al. (7). Host-free infectious EB were purified through Renografin (E. R. Squibb & Sons, Princeton, N.J.) at 48 h postinfection (25). The bacterial strains used were *Escherichia coli* Y1089 and Y1090 for bacteriophage λgt11 transfection (27), DH5α and MV1193 for trans-
fection with recombinant M13mp18/19 (26), and JM83 for transformation with vector pUC (26). The plasmids used were pUC18 and pUC19 (19). E. coli was grown routinely in 2× YT medium (1.6% tryptone, 1% yeast extract, 0.5% sodium chloride), and cells harboring plasmids were grown in 100 µg of ampicillin per ml.

DNA isolation and manipulations. Chromosomal DNAs from C. trachomatis serovars L2, D, J, and K and C. psittaci Mn were prepared as described earlier (6, 24). Standard recombinant DNA techniques were used for nucleic acid preparation and analysis (8). Chromosomal DNA from serovar L2 was digested with EcoRI and then shotgun ligated into EcoRI-digested alkaline phosphatase-treated phage λgt11 DNA. The ligated DNA was packaged by using the Boehringer Mannheim packaging kit. Phage were plated and amplified in E. coli Y1090. Southern blotting of EcoRI-digested C. trachomatis DNA and endonuclease restriction mapping of the λgt11 insert were performed by standard procedures. Plaques showing positive reactions were picked, plated at low density, and reassayed with a synthetic oligonucleotide probe. The process was repeated until all plaques were reactive. Hybridization at either high (polyformamide at 37°C) or low (20% formamide at 37°C) stringency, followed by washing at 70 or 37°C, respectively, was carried out essentially as described earlier (6). Radioactive DNA probes were labeled by nick translation or 5′-end labeling (9).

Amino acid sequence determination. Whole serovar L2 EB cell lysates were resolved by sodium dodecyl sulfate–15% polyacrylamide gel electrophoresis. The resolved polypeptides were then transferred electrophotically onto Immobilon, stained with Coomassie brilliant blue, and destained. Protein strips corresponding to regions of interest were excised with a razor blade and used as a source of protein for amino acid sequencing as described by Matsudaira (9). One group of bands excised was in the 18-kDa region, since one polypeptide in this molecular mass range has been identified as a major eukaryotic cell-binding protein (25). The protein characterized in this study represents the predominant, noneukaryotic cell-binding polypeptide in the 18-kDa range. Protein sequencing was performed by automated Edman degradation at two centers: the protein sequencing facility at the University of Victoria and the protein analysis facility at the University of Wisconsin Biotechnology Center. The two centers reported the same amino acid sequence, which was used as the basis for construction of an oligonucleotide probe.

Synthetic oligonucleotides. Single-stranded oligonucleotide primers were synthesized by using an Applied Biosystems Inc. (Mississauga, Ontario, Canada) 381A DNA synthesizer at the DNA synthesis facility, Department of Microbiology, University of Alberta. The synthetic oligonucleotide used for identification of gene clones was the 29-mer 5′ AAGGA XAC(X)GCA(X)AAAAATGAC(X)GA 3′, in which X indicates bromodeoxyuridine. Its construction was based on the N-terminal amino acid sequence of the 18-kDa protein.

Nucleotide sequence determination. The nucleotide sequences of both strands of the 2.6-kb EcoRI fragment of plasmid pCTH1 were determined by the dideoxynucleotide chain termination method (15) with M13mp18 and M13mp19 recombinant clones. Sequence compilation, open reading frame (ORF) identification, translation, and restriction map construction were all performed with DNA Strider software for the Macintosh computer (C. Mark, Service de Biochimie Centre d’Etudes Nucleaires de Saclay, Gif-sur-Yvette Cedex, France). Codon usage and isoelectric point (pI) determinations were done on NBRF data base software for the Macintosh SE computer.

Plasmid-directed protein synthesis. In vitro transcription-translation studies were carried out with 2 µg of recombinant or vector DNA by the protocol supplied by the manufacturer of the kit used (Amersham Inc., Oakville, Ontario, Canada). Labeled proteins were resolved by sodium dodecyl sulfate–polyacrylamide gel electrophoresis on 12.5% acrylamide, followed by fluorography and examination by autoradiography.

RNA isolation and Northern blotting. Total RNA was isolated by a hot-phenol extraction procedure from infected HeLa cells at various intervals postinfection as described earlier (5). RNA was fractionated on 1.5% agarose gels in the presence of dimethyl sulfoxide and glyoxal, and a ladder of small RNA species (Bethesda Research Laboratories, Burlington, Ontario, Canada) was used to provide markers for accurate size calibration. Northern blotting on diazobenzyloxyethyl-paper was carried out as described earlier (5).

Nucleotide sequence accession number. The nucleotide sequence reported here has been submitted to EMBL and assigned accession no. X57311.

RESULTS

Cloning of an 18-kDa eukaryotic histone H1-like protein-encoding gene from C. trachomatis. A partial N-terminal amino acid sequence of the 18-kDa eukaryotic histone H1-like protein from C. trachomatis was obtained by analyzing approximately 50 pmol of excised protein. The sequence ALKDRTKMTDLESIQQNL was used to design a 29-mer oligonucleotide probe. The degeneracy of the oligonucleotide was minimized by choosing the stretch of amino acids (KDTAKMTD) with minimal degeneracy in their codons and by incorporation of bromodeoxyuridine to optimize the hybridization conditions. Southern blot analysis of total genomic DNA isolated from C. trachomatis was performed by using a 32P-labeled 29-mer oligonucleotide probe. A single radioactive band was identified when hybridization was carried out with 20% formamide at 37°C, followed by washings at 37°C. The size of this band was determined to be approximately 2.6 kb. Higher formamide concentrations resulted in complete loss of any signal (data not shown).

A λgt11 library of C. trachomatis chromosomal DNA fragments was screened by hybridization to the 32P-labeled oligonucleotide under the conditions determined by Southern blot analysis of total genomic DNA. We recovered one strongly hybridizing recombinant clone in λgt11, designated λgt11/L2/RKA10.

Characterization of λgt11/L2/RKA10 insert DNA. Preparations of λgt11/L2/RKA10 insert DNA were obtained from EcoRI digests and separated on agarose gels. The insert was estimated to be a 2.6-kb fragment. A restriction map of this EcoRI insert is shown in Fig. 1A. The relative size and orientation of the CTH1-coding region was determined by hybridization and DNA sequencing. This 2.6-kb EcoRI fragment was subcloned into pUC18 and designated pCTH1. The C. trachomatis 18-kDa CTH1 gene was localized on an internal 670-bp DraI fragment of pCTH1. C. trachomatis DNA samples obtained from serovars L2, D, J, and K and digested with DraI all reacted with a single 670-bp restriction fragment in Southern blot hybridization under stringent conditions. C. psittaci Mn failed to hybridize under similar conditions (data not shown).
FIG. 2. Comparison of the deduced amino acid sequence of C. trachomatis CTH1 with that of painted sea urchin histone H1 protein (SUH1). Dashes indicate gaps introduced to optimize alignment. Colons indicate identities, while dots indicate conserved amino acid substitutions. Numbers to the right refer to amino acid positions.

a 7-bp dyad, followed by eight thymidine residues. The CTH1-encoding gene consists of 375 nucleotide pairs capable of encoding 125 amino acids. The calculated molecular mass of the protein is 13,689 daltons. The ratio of basic (42 residues) to acidic (8 residues) amino acids is 5.25. The protein possesses an estimated pl of 10.71 due to its high lysine content (36 residues). There are three repeat units of Lys-Ala-Ala-Ala and one unit of Lys-Ala-Ala-Pro.

Comparison of the deduced amino acid sequence with the NBRF protein library data base with the FASTP program identified significant homology to H1 subtype histone proteins from different sources. The best alignment of amino acids of CTH1 to sea urchin H1 is shown in Fig. 2. CTH1 shows nearly 74% homology (considering 38 perfectly matched residues and 45 conservative substitutions) over a length of 112 amino acids with painted sea urchin H1 histone protein. Surprisingly, the amino acid identity between the two proteins is mostly limited to alanine and lysine.

Expression of pCTH1 and pCTH2-derived gene products. A 670-bp DraI fragment encoding ORF2 and designated pCTH2 was subcloned into pUC18. Recombinant plasmids pCTH1 and pCTH2 each encode a polypeptide with a molecular mass of approximately 18,000 in an E. coli-derived coupled transcription-translation system (Fig. 3). The vector (pUC18) alone does not express a corresponding gene product.

Transcription pattern during the chlamydial growth cycle. To study the temporal regulation of CTH1-encoding gene transcription, RNA obtained at different times during the growth cycle was evaluated by Northern hybridization. The transcript was detected at 12 h postinfection with chlamydial EB. Increased amounts of the transcript were apparent through 36 h (Fig. 4). However, no transcript was detected at 6 h postinfection (data not shown).

DISCUSSION

In this study, we describe the cloning, identification, and nucleotide sequencing of an 18-kDa protein from C. trachomatis that shows a high degree of similarity to sea urchin histone H1 protein. In eukaryotes, histones play an important role in nucleosome structure, as well as in the formation and maintenance of higher-order chromatin structure (10). However, bacterial DNA-binding proteins that govern DNA folding are structurally unrelated to eukaryotic histones. Identification of eukaryotic histone H1 sequences among prokaryotes is uncommon. Recently, Kato et al. (4) reported homology between the AlgR3 gene product and eukaryotic histone H1; this gene regulates alginate synthesis in Pseu-
domonas aeruginosa. More recently, Perera and Ganem reported a 22-kDa eukaryotic histone H1-like protein from Chlamydia sp. (12).

The nucleotide sequence of CTH1 which starts with the ATG codon matches perfectly with its derived N-terminal amino acid sequence. The methionine codon was preceded by a Shine-Dalgarno-like sequence (17). The sequence distal to the ORF revealed a G+C-rich dyad with potential for forming a stem-and-loop structure, followed by a rho-independent terminator (14). Wagar and Stephens (21) identified three chlamydial DNA-binding proteins by blot analysis. One of them, a basic polypeptide with a molecular mass of 17 kDa, was specific to the EB form only. Whether the appearance of a CTH1-encoding gene transcript as early as 12 h (when reticulate body forms predominate) is due to the coexistence of EB in a nonsynchronous chlamydial culture or simply represents two different DNA-binding proteins (one specific to the EB form and the other present in both forms) remains to be determined. Also, our failure to detect RNA at 6 h may simply represent the low copy number of chlamydiae at that stage. Localization of CTH1 by immune electron microscopy at various time intervals during the developmental cycle, as well as in purified EB and reticulate body forms, will address this question.

Evaluation of RNA by Northern analysis during the chlamydial growth cycle suggests the existence of a monocistronic message. The gradual increase in the amount of transcript correlates with the increased protein synthesis observed with time (5). Analysis of the transcription initiation site by primer extension suggests the presence of a single promoter (6a).

The gene for CTH1 was expressed in an E. coli coupled transcription-translation system as an 18-kDa polypeptide. The mobility of this in vitro gene product is the same as that of the native chlamydial protein. However, a difference of 4.3 kDa between the deduced (13,689 daltons) and apparent (18 kDa) molecular masses appears likely to be due to the presence of highly basic amino acids in the protein. The pI for CTH1 is 10.73 because of its high lysine content. This pI value is in the range determined for a 17-kDa chlamydial DNA-binding protein (pI, >8.0) reported earlier (21). The ratio of basic to acidic amino acid residues is 5.25, compared with 5.4 for eukaryotic histone H1.

The possibility that chlamydial histonelike proteins fulfill a role analogous to that of their eukaryotic counterparts can be envisaged, since EB, unlike other members of the family Eubacteriaceae, contain a condensed, discrete, and electron-dense nucleoid (2). In any case, whether this positively charged protein with sequence homology to eukaryotic histone H1 binds DNA awaits biological evidence. Southern blot analysis of C. trachomatis serovar L2, J, K, and D genomic DNAs implies that the gene for CTH1 is present as a single copy. However, failure to observe hybridization under similar conditions with C. psittaci Mn suggests the existence of interspecies structural diversity. Information about the 18-kDa histone H1-encoding genes from different serovars should provide insight into the structural and functional aspects of this protein. In addition, whether these chlamydial histone H1-like proteins have evolved independently during evolution, by convergent evolution, or have arisen during a horizontal gene transfer event from a eukaryotic host to a prokaryotic parasite must be elucidated. A horizontal gene transfer event has been proposed to explain the existence of glutamate synthetase II in Bradyrhizobium japonicum (1). However, evaluation of CTH1 codon usage reveals a bias towards prokaryotes (20).

Finally, analysis of the 2.6-kb EcoRI fragment in pCTH1 demonstrated the presence of two other partial ORFs. The C-terminal portions of the proposed ORF1 and ORF3 have been deleted because of the presence of an EcoRI restriction site within the gene. One of these partial ORFs lies approximately 400 bp downstream, while the other resides 850 bp upstream of the 18-kDa gene for CTH1. Interestingly, the genes for these potential ORFs are translated from opposite strands, suggesting bidirectional transcription in chlamydiae. This observation attributes another novel feature to this interesting prokaryotic parasite.

---

**FIG. 3.** Identification of the product of the CTH1-encoding gene synthesized in a coupled in vitro transcription-translation system. Lanes 1 to 3 represent autoradiography of a gel electrophoresis pattern of 35S-labeled gene products synthesized by pUC, pCTH1, and pCTH2, respectively. Molecular mass marker sizes are given on the left in kilodaltons. The arrowhead indicates the migration of 18-kDa translated polypeptides.

**FIG. 4.** Temporal pattern of expression of CTH1 mRNA by Northern blot hybridization. Lanes 1 to 3 represent RNA isolated from infected cells at 12, 22, and 36 h postinfection with C. trachomatis serovar L2. Approximately 10 µg of total RNA was loaded on each lane. The arrowhead indicates the mobility of transcripts. 32P-labeled pCTH2 was used as a probe for Northern blots.
ACKNOWLEDGMENTS

This work was supported by a grant from the Canadian Medical Research Council (MA7951). W. M. W. is a scholar of the Alberta Heritage Foundation for Medical Research.

We thank K. L. Roy for helpful discussions, San Vinh for photography, and Fiona Cooper for manuscript preparation.

REFERENCES


