BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS **230,** 164–166 (1997)
ARTICLE NO. RC965877

Sequencing, Analysis and Expression in *Escherichia coli* of a Gene Encoding a 15 kDa *Cryptosporidium parvum* Protein

Nikolai V. Khramtsov,*[,1] Brenda Oppert,*[,2] Beth A. Montelone,* and Steve J. Upton*
**Division of Biology, Kansas State University, Manhattan, Kansas 66506*

Received November 15, 1996

A previous paper presented data on a cDNA sequence encoding a protein associated with the AIDS related pathogen, *Cryptosporidium parvum*. However, the position of the start codon was uncertain, and the 5′ end was continuous, lending doubt about the size and complete sequence of the final protein product. Herein we present the complete gene sequence and conclude the predicted size of the putative protein to be 16.2 kDa.

A previous paper [1] presented data on the cloning, sequencing and expression of a cDNA fragment that encoded a polypeptide with an epitope shared by both 15- and 60 kDa sporozoite proteins of the Auburn isolate of *Cryptosporidium parvum*. This cDNA had a continuous open reading frame at the 5′-end, and the position of the start codon was uncertain. Therefore, the size and complete sequence of the protein product were uncertain. In addition, both Northern and Western blots failed to ascertain whether this cDNA fragment encoded a protein of 15 kDa or 60 kDa. The complete gene sequence and size of the final protein product are presented here.

MATERIALS AND METHODS

Two independent *E. coli* clones, each containing a plasmid with a 2091 base pair (b.p.) insert, were isolated from a *C. parvum* (KSU-1 isolate) EcoRI genomic library in pBluescript II SK(+) (Stratagene) [2]. The clones were detected by hybridization with a ^{32}P-end labeled 21-mer oligodeoxynucleotide (5′-CATGCGTAACTTGAAATCCTG) corresponding to a portion of the coding region of the cDNA (coordinates 30-50 b.p.) [1].

The cloned gene was expressed in *E. coli*. A fragment of DNA, including the coding region, was amplified on the cloned genomic DNA using Taq polymerase and a set of primers (sense 5′-ATTTTTGAATTcTTGATAAAAG 1066-1087 b.p., antisense 5′-CCAGAATTAATGgAATTCCTG 1568-1548 b.p.). Each primer contained one substitution (lowercase letter) to create an EcoRI site (underlined). The amplified products were digested with EcoRI endonuclease and cloned into an EcoRI site of the bacterial expression vector pET-28(a+) (Novagen). The proper orientation of the cloned fragment was verified by restriction analysis and by sequencing. The cloned gene was expressed in *E. coli* strain BL21(DE3) (Novagen) by induction with isopropyl-1-thio-β-D-galactopyranoside. Expressed protein was subjected to 12% SDS-PAGE and transferred to a nitrocellulose filter. Rat polyclonal antiserum to *C. parvum* was incubated with the filter, and monospecific antibodies were eluted from the filter as described [3].

RESULTS AND DISCUSSION

Sequences of both cloned fragments were identical except that they were in opposite orientations in the vector.[3] Comparison of the sequences of genomic DNA (Fig. 1) and cDNA [1] (corrected sequence cDNA, GenBank accession number L08612) revealed identity between sequences except for the following differences: A) The genomic DNA did not have 18 adenosines, which are present on the 3′-end of the cDNA and probably represent the result of processing of *C. parvum* mRNA; B) Upstream from the poly(A) tail in the cDNA and corresponding genomic sequence are two putative poly(A) addition sites (AATAAA [4], coordinates 1592-1597 and 1635-1640 b.p.); and C) At the 5′-end of the cDNA are seven nucleotides, which are absent in the genomic sequence. This last difference may be the result of several possibilities. Two different isolates of *C. parvum* are being compared, and strain differences may be present. Alternatively, some unknown kind of modification may exist in the mRNA. Additional nucleotides may have been synthetically added during *in vitro* synthesis of the cDNA from mRNA, similar to

[1] To whom correspondence should be addressed. Fax: 913-532-6653; email: podolsk@ksu.edu.
[2] Present address: USDA-ARS Grain Marketing Research Laboratory, 1515 College Ave., Manhattan, KS 66502.

[3] Nucleotide sequence data reported in this paper have been submitted to the GenBank data base with the Accession No. U22892.

0006-291X/97 $25.00

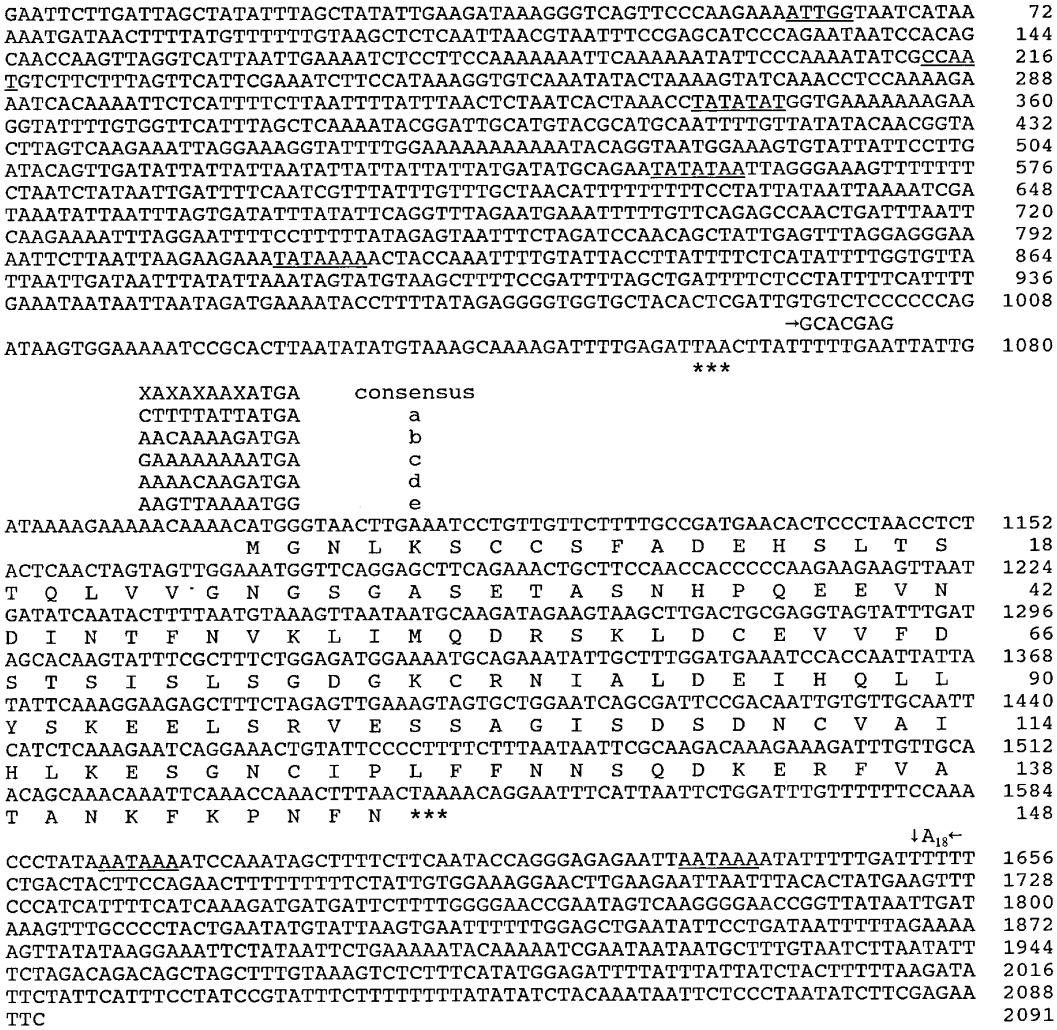

FIG. 1. Nucleotide sequence of a 15 kDa protein gene of *C. parvum* and deduced amino acid sequence. Position in the nucleotide and amino acid sequences are indicated in the right margin of the corresponding lines. Putative TATA-, CAAT-boxes, and poly(A) signal are underlined. Symbols: *** = the first stop codons upstream and downstream in the same reading frame from coding sequence; → = start and ← = end of cDNA sequence [1], respectively. The major differences between genomic and cDNA sequence are shown above the genomic sequence: A_{18} = eighteen adenosines; ↓ = polyA site. The insert between nucleotide sequences represents a local alignment of sequences around the start codon in the gene presented in this paper with sequences of the *C. parvum* genes: a = oocyst wall protein [8]; b = actin [7]; c = acetyl-CoA synthetase [11]; d = hsp 70 [2]; e = elongation factor-2 [10].

that observed during cloning of the cDNA of rod photoreceptor cGMP phosphodiesterase subunits from bovine retina [5].

Analysis of the cloned genomic fragment of DNA (Fig. 1) revealed an open reading frame 480 b.p. long (coordinates 1063-1542 b.p.). It appears that the start codon is at position 1099-1101 b.p. as: A) This is the first AUG codon in the open reading frame downstream from the terminator codon UAA (position 1060-1062 b.p.), which is in accordance with the scanning mechanism for initiation of translation in eukaryotes [6]; B) This AUG codon is found in a consensus sequence proposed for sequences around start codons in other eukaryotic mRNAs [7] (purine, usually A, in position −3 and another purine, usually G, in position +4); C) comparison of the sequences around start codons of the DNA encoding other *C. parvum* proteins, available from GenBank data bases (actin #M69014 [8], oocyst wall protein #Z22537 [9], heat shock 70 #U11761 [2], elongation factor-2 #U21667 [10], and acetyl-CoA synthetase gene #U24082 [11]), reveal homology in this region; and D) Upstream from the predicted first AUG are several putative regulatory elements (TATA- and CAAT-boxes), which are present in eukaryotic genes in the 5′-noncoding region [12,13]. We concluded that the start codon was located at position 1099-1101 b.p., and the resulting open reading frame encoded a putative protein of 148 amino acids with a predicted size of 16.2 kDa.

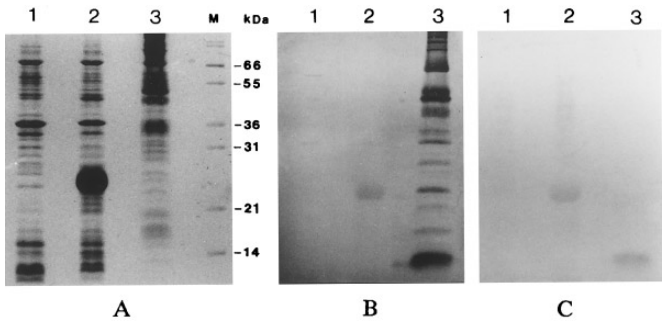

FIG. 2. SDS-PAGE and Western blot analysis of samples: soluble fraction from: (1) uninduced *E. coli* cells; (2) induced *E. coli* cells with recombinant plasmid (amount per lane equals a 30 μl culture with OD_{600}=1.0); (3) total *C. parvum* proteins from 5×10^6 excysted oocysts; and (M) molecular size markers. (A) Samples were fractionated on a 12% SDS-PAGE gel and stained with Coomassie blue. (B, C) Duplicate samples were blotted onto nitrocellulose and probed with rat antiserum raised against an homogenate of purified *C. parvum* oocysts and sporozoites (B) or monospecific eluted antibodies to the expressed recombinant protein (C).

To identify the size of the final protein product, we expressed the cloned gene in *E. coli*. Analysis of bacterial lysates revealed the presence of a protein with a molecular mass of around 22–24 kDa (Fig. 2A, lane 2). This protein was absent in uninduced cells (Fig. 2A, lane 1). The apparent molecular size of this protein was similar to the expected primary structure of the fusion protein (20.9 kDa). The new protein was observed on Western blots (Fig. 2B, lane 2) and occurred in the soluble fraction, but not in the pellet, of the *E. coli* lysate (data not shown). Monospecific antibodies, eluted from recombinant protein, recognized a single 15 kDa band of *C. parvum* (Fig. 2C, lane 3) and protein expressed in *E. coli* (Fig. 2C, lane 2). Rat antiserum and monospecific antibodies did not react with a lysate of uninduced *E. coli* cells (Fig. 2B, lane 1 and Fig. 2C, lane 1, respectively). Preimmune rat serum did not give positive reactions with any of the samples (data not shown).

From the genomic sequence, the complete protein sequence was predicted. *In vitro* translation of the genomic DNA allowed for the isolation of monospecific antibodies. When *C. parvum* proteins were analyzed by Western blot with these antibodies, a 15 kDa protein was recognized. Based on the analysis of the sequence and from Western blot analysis we concluded that this genomic DNA contained the entire gene encoding the 15 kDa *C. parvum* protein of Jenkins et al. [1].

ACKNOWLEDGMENTS

We acknowledge the assistance of M. Tilley in preparation of oocyst samples and rat antiserum. We also thank D.J. Roufa for the SeqAid program for analyses of the nucleotide and amino acid sequences. This work was supported by NIH Grant AI30881 to SJU and BAM. This is Kansas Agricultural Experiment Station contribution No. 95-439-J.

REFERENCES

1. Jenkins, M. C., Fayer, R., Tilley, M., and Upton, S. J. (1993) *Infec. Immun.* **61,** 2377–2382.
2. Khramtsov, N. V., Tilley, M., Blunt, D. S., Montelone, B. A., and Upton, S. J. (1995) *J. Euk. Microbiol.* **42,** 416–422.
3. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed. pp. 18.16–18.18. Cold Spring Harbor Lab. Press, NY.
4. Birnstiel, M. L., Busslinger, M., and Strub, K. (1985) *Cell* **41,** 349–359.
5. Lipkin, V. M., Khramtsov, N. V., Vasilevskaya, I. A., Atabekova, N. V., Muradov, Kh. G., Gubanov, V. V., Li, T., Jonson, J. P., Volpp, K. J., and Applebury, M. L. (1990) *J. Biol. Chem.* **265,** 12955–12959.
6. Kozak, M. (1989) *J. Cell Biol.* **108,** 229–241.
7. Kozak, M. (1987) *Nucleic Acids Res.* **15,** 8125–8148.
8. Kim, K., Gooze, L., Petersen, C., Gut, J., and Nelson, R. G. (1992) *Mol. Biochem. Parasitol.* **50,** 105–114.
9. Ranucci, L., Muller, H. M., La Rosa, G., Reckmann, I., Morales, M. A., Spano, F., Pozio, E., and Crisanti, A. (1993) *Infec. Immun.* **61,** 2347–2356.
10. Jones, D. E., Tu, T. D., Mathur, S., Sweeney, R. W., and Clark, D. P. (1995) *Mol. Biochem. Parasitol.* **71,** 143–147.
11. Khramtsov, N. V., Blunt, D. S., Montelone, B. A., and Upton, S. J. (1996) *J. Parasitol.* **82,** 423–427.
12. Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C., and Chambon, P. (1980) *Science* **209,** 1406–1414.
13. Maniatis, T., Goodbourn, S., and Fisher, J. A. (1987) *Science* **236,** 1237–1244.