Sequence Analysis of the Gene Encoding the Chlamydia pneumoniae DnaK Protein Homolog

JODI M. KORNAK, CHO-CHOU KUO, AND LEE ANN CAMPBELL*

Department of Pathobiology, University of Washington, Seattle, Washington 98195

Received 13 August 1990/Accepted 5 November 1990

The antigen-coding region of a 4.2-kb *PstI* fragment of *Chlamydia pneumoniae* (pLC3), which encodes a 75-kDa immunoreactive protein recognized during human *C. pneumoniae* infection, was localized to a 2.0-kb *EcoRI* fragment. This subclone expressed an immunoreactive fusion protein of ca. 82 kDa. Nucleotide sequence analysis of the *C. pneumoniae* gene revealed that it consisted of a 1,980-base open reading frame with an inferred 71,550-Da protein of 660 amino acids. Putative *Escherichia coli*-like promoters and a ribosomal binding site were located in the 5' upstream region, and an 11-base dyad forming a stable stem-loop structure following two in-frame stop codons was identified. The *C. pneumoniae* 75- kDa protein is a member of the hsp70 family of heat shock proteins and has 87% amino acid similarity with the *Chlamydia trachomatis* protein.

Chlamydia pneumoniae, formerly known as Chlamydia sp. strain TWAR, is a newly established species of the genus Chlamydia (9). The pathogen has been shown to be an important cause of acute respiratory disease in humans (8). It has been associated with bronchitis and pharyngitis in addition to endemic and epidemic pneumonia (8). Because of the difficulty associated with isolating and growing large amounts of the organism for antigenic and biochemical characterization, we have used recombinant DNA techniques to identify genes encoding immunoreactive proteins. Previously, a 4.2-kb PstI fragment of C. pneumoniae was found to encode a 75-kDa protein. This gene was shown to contain a genus-reactive determinant that was recognized during C. pneumoniae infection (6). Separately, Maclean et al. (16) described a 75-kDa protein in Chlamydia trachomatis that was genus reactive. Monospecific polyclonal rabbit antisera to this protein neutralized infectivity. Sequence analysis of the gene encoding the C. trachomatis 75-kDa protein revealed sequence homology to the hsp70 family of heat shock proteins (2, 7).

The purposes of our studies with the gene encoding a 75-kDa protein of *C. pneumoniae* were to identify the region encoding the genus-reactive determinant, to precisely define the coding sequence of the gene, to investigate the relationship between the *C. pneumoniae* gene and the *C. trachomatis* gene, and to perform DNA and amino acid sequence analyses.

The C. pneumoniae isolate AR-39, which had been adapted to grow in HeLa 229 cells, was harvested and purified in a linear gradient of meglumine diatrizoate (Hypaque-76; Winthrop-Breon Laboratories, New York, N.Y.) (13). The final products usually contained 1.0×10^8 to 5.0×10^8 inclusion-forming units per ml of organisms.

The 4.2-kb *PstI* fragment of *C. pneumoniae* AR-39 (pLC3) (6) was digested with *Eco*RI, and the resulting fragments were ligated into similarly digested pATH1, pATH3, and pATH11 vectors (25) by standard protocol (17). These vectors permit cloning in all possible reading frames and result in the overexpression of a fusion polypeptide linked to the amino terminus of the *Escherichia coli trpE* operon. The *trpE* portion is a 37-kDa polypeptide. After transformation

into *E. coli* HB101 by the method of Hanahan (10), transformants were plated onto Luria-Bertani agar containing 20 μ g of tryptophan per ml and 50 μ g of ampicillin per ml. Transformants were screened for inserts by DNA hybridization, and plasmid inserts were analyzed by restriction endonuclease digestion. The plates were overlaid with nitrocellulose disks (Schleicher & Schuell, Inc., Keene, N.H.), colonies were lysed, and adsorbed DNA was probed with the gel-purified ³²P-labeled 4.2-kb *PstI* fragment from pLC3 at 42°C as previously described (5). Plasmid DNA was isolated from the recombinant clones by standard techniques (17) and then was digested with *Eco*RI.

HB101 strains containing recombinant plasmids were induced for the expression of recombinant fusion protein by using previously described methods to prepare whole cell lysates and insoluble fractions (12, 25). Proteins were separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) on 10% polyacrylamide gels with a 5% stacking gel by the method of Laemmli (14). After protein separation by SDS-PAGE, proteins were transferred electrophoretically to nitrocellulose by the method of Towbin et al. (27). Immunoblotting was performed as described previously by using rabbit immune sera prepared against C. *pneumoniae* AR-39 (6).

The Sanger dideoxy-chain termination method of DNA sequencing (22) was carried out on single-stranded fragments cloned into M13mp18 and M13mp19 (28) by using the Sequenase kit (United States Biochemical Corp., Cleveland, Ohio). Exonuclease III digestions were done to generate nested deletions by using the Erase-a-Base kit (Promega, Madison, Wis.). Sequence analyses were performed by the Pustell sequence analysis program (IBI) and the University of Wisconsin Genetics Computer Group programs.

In order to localize the region encoding the genus-reactive determinant in the *C. pneumoniae* clone pLC3, *Eco*RI restriction fragments making up the 4.2-kb *Pst*I fragment (Fig. 1) were subcloned into the pATH expression vectors. Protein profiles demonstrated the induction of a ca. 82-kDa fusion protein in a pATH11 recombinant clone containing the 2.0-kb *Eco*RI fragment (Fig. 2A). Immunoblots revealed that this protein was recognized by anti-*C. pneumoniae* rabbit immune sera (Fig. 2B). Recombinant clones containing the 1.1-kb *Eco*RI fragment did not express any novel or immunoreactive fusion proteins (data not shown).

^{*} Corresponding author.



500 bp

FIG. 1. Restriction map of and sequencing strategy for the 4.2-kb *PstI C. pneumoniae* fragment. The 2.0-kb *Eco*RI and the adjacent 1.1-kb *Eco*RI fragments were sequenced in addition to the overlapping 1.2-kb *Hind*III fragment. The resulting ATG start and TAA stop codons for the gene encoding the *C. pneumoniae* 75-kDa protein are indicated. Restriction sites: A, *AvaI*; B, *Bam*HI; E, *Eco*RI; H, *Hind*III; P, *PstI*; X, XhoI.

Southern blot analyses comparing the C. pneumoniae gene (pLC3) to the C. trachomatis dnaK homolog (7) showed that the 2.0- and 1.1-kb EcoRI fragments and an overlapping 1.2-kb HindIII fragment of C. pneumoniae had sequence homology with the C. trachomatis gene (this clone was kindly provided by Robert C. Brunham, Department of Medical Microbiology, University of Manitoba) (data not shown). Subsequently, these three fragments were cloned into M13mp18 and M13mp19 and sequenced by generating nested deletions. The complete nucleotide sequence encoding the 75-kDa C. pneumoniae protein and the inferred amino acid sequence are shown in Fig. 3. The open reading frame, beginning with a codon for methionine, consists of 1,980 nucleotides and ends with two in-frame stop codons. Comparison of this open reading frame with the C. trachomatis serovar D sequence reported by Danilition et al. (7) revealed that the two genes are 73% homologous at the DNA level. Thirty bases downstream from the second stop codon, an 11-base dyad, beginning at base 2011 and ending at base 2036, was identified.

Upstream from the translational start site, the Shine-Dalgarno sequence (24) is indicated by the boxed 5-bp sequence in Fig. 3 centered at -11. This sequence was identified by searching for homology with the 3' end of the *E. coli* 16S rRNA. Further upstream, potential transcription promoters were identified by their homologies to typical *E. coli* promoters (11). The -35 and -10 consensus regions are centered around -129 and -108, respectively.

The C. pneumoniae open reading frame encodes a protein of 660 amino acids with a calculated molecular weight of 71,550. The protein is 40% hydrophobic with areas of intervening hydrophilicity. Comparison of the inferred amino acid sequence of the C. pneumoniae 75-kDa protein with that of the C. trachomatis serovar D protein indicated 87% similarity. In addition, regions that are heterogeneous among the hsp70 proteins but conserved between the 75-kDa proteins of C. trachomatis and C. pneumoniae have been identified (amino acids 23 to 36, 69 to 140, 281 to 326, 529 to 559, and 603 to 647) (Fig. 4). There is also a region which is relatively conserved among the hsp70 proteins but unique to the 75-kDa protein of C. trachomatis (amino acids 198 to 215).

A comparison of the nucleotide sequence of the 75-kDa gene of *C. pneumoniae* with sequences of the GenBank nucleic acid sequence data base (release 63) revealed a large



FIG. 2. SDS-10% polyacrylamide gel stained with Coomassie blue (A) and corresponding immunoblot (B). Shown are whole-cell lysates and the insoluble fraction, respectively, of *E. coli* HB101 containing induced pATH11 (lanes 1 and 3), uninduced pATH11 (lanes 2 and 4), induced pATH11 with 2.0-kb *Eco*RI insert (lanes 5 and 7), and uninduced pATH11 with 2.0-kb *Eco*RI insert (lanes 6 and 8).

degree of similarity between the 75-kDa gene and genes encoding members of the hsp70 family (Fig. 4). As previously reported with the *C. trachomatis* protein (2, 7), the hsp70 amino acid sequences of *Bacillus megaterium* DnaK and *E. coli* DnaK showed the greatest degrees of similarity (74 and 72%, respectively) to that of the *C. pneumoniae* 75-kDa protein.

In general, the molecular genetic characterization of Chlamydia spp. has been difficult because of the lack of good mutagenesis techniques and the absence of known mechanisms of genetic exchange. Although recombinant DNA techniques have proved invaluable in chlamydial genetic studies, very few chlamydial genes are expressed well in E. coli (20). The problems encountered with C. trachomatis and Chlamydia psittaci gene expression in E. coli have also been found with C. pneumoniae (4). However, the genes encoding the 75-kDa proteins of both C. trachomatis and C. pneumoniae are readily expressed in the E. coli host (2, 6, 7), and the C. pneumoniae gene is expressed in an E. coli in vitro transcription-translation system (6). These observations suggest that this chlamydial promoter is recognized by the E. coli transcription and translation machinery. Nucleotide sequence analysis of the C. trachomatis gene revealed E. coli-like promoter sequences (7, 21). Similarly, putative promoters homologous to typical E. coli promoters were identified in the C. pneumoniae gene. In contrast, Birkelund et al. (2) described a mixed promoter for the C. trachomatis L2 dnaK gene in which the -35 region is similar to a heat shock promoter and the -10 region is a classical TATA box. However, unlike most heat shock genes, which have highly conserved promoters, no classical heat shock promoters were identified upstream of the C. pneumoniae gene. Pollack et al. (21) also reported the absence of typical heat shock promoter sequences in the C. trachomatis mouse biovar gene encoding a DnaK homolog. Whether or not the C. pneumoniae gene is being transcribed from the E. coli-like promoters will remain unclear until S1 nuclease experiments

-181	CEMECOSETECHMCAGAMEAGACTICTEMETICCTGAGESEACGAT TITTAGA AGAGTTTECHAGEGA TATAMA ATAGGAG	
-97		
1 1	C A A G G T A T C C G G C T C TG AGT GAA CAA AAA TCA AGC AMA ATT ATA GGT ATA GAC TTA GGC ACA ACA AAC TCC TGC GTA TCT GTT ATG et Serr Giu His Lys Lys Ser Ser Lys Iie Iie Giy Iie Asp Leu Giy Thr Thr Asn Ser Cys Val Ser Val Met Lys Arg Asn	
76 26	T C C T G C T T T T T T T T MA GGA GGA CAA GCT AAA GTA ATT ACA TCA TCC GAA GGA ACA AGA ACC AGG CCA TCG ATC GTT GCC TTC AAA GGT Ju Giy Giy Gia Ala Lys Val lie Thr Ser Ser Giu Giy Thr Ang Thr Thr Pro Ser lie Val Ala Phe Lys Giy Pro Ala	
151 51	GC ACTCTTA TGCTTG ATGAG AMA TTAGTG GGG ATT CCAGCA AMA OGT CAAGCA GTG ACA AMT CCAGAA AMA ACT CTC GGC TCT ACA AMA an Gìu Lya Leu Val Gìy 11e Pro A1a Lys Ang Gìn A1a Val Thr Asn Pro Gìu Lys Thr Leu Gìy Ser Thr Lys Iy Thr Ala	
226 76	A C C TAA A T C AA T TA A C C C A TG TC TAAC G GC TTT ATT GGC CGT AAG TAC TCT GAA GTA GCT TCG GAA ACC GTT CCT TAT ACA GTC ACC TCC GGA TCT rg Phe I'le Giy Ang Lys Tyr Ser Giu Val Ala Ser Giu I edin Thr Val Pro Tyr Thr Val Thr Ser Giy Ser Phe Giu Lys Lys Ala Pro Asn	
301 101	A G C T T G A CMA TG C T G C C G M GGT GAT GCT GAT GTT GAT GGC AMA CAA TAC ACT CCA GAA GAA ATT GGC GCA CAA ATC TTA ATG AMA ys Giy Asp Ala Val Phe Giu Val Asp Giy Lys Gin Tyr Thr Pho Giu Giu Ile Giy Ala Gin Ile Leu Met Lys Asp Giy Gin Leu	
376 126	GATTG CAAAGC TAATTC TG AAAGAG ACA GCA GAAGCTTATCTA GGC GAAACTGTC ACA GAAGCAGTG ATC ACC GTC CCC GCA TAC TTC AAT et Lys Glu Thr Ala Glu Ala Tyr Leu Gly Glu Thr Val Thr Glu Ala Val Ile Thr Val Pro Ala Tyr Phe Asn	
451 151	A T T C T T A AT TCT CAA CGA GCA TCC ACA AAA GAT GCT GGA CGC ATT GCA GGT CTA GAT GTA AAA GGT ATC ATT CCA GAA CCT sp Ser Gin Ang Ala Ser Thr Lys Asp Ala Gly Ang Ile Ala Gly Leu Asp Val Lys Ang Ile Ile Pro Glu Pro	
526 176	A G C T T T T G AA A CAGTCTOGA A G A G A G DC GCA GCA GCTCTTGC TAC GCA ATC GAT AMA GTC GGT GAT AMA AMA ATC GCT GT TTC GAC CTT GGT GGA GGA hr A1a A1a A1a Leu A1a Tyr G1y I1e Asp Lys Va1 G1y Asp Lys Lys I1e A1a Va1 Phe Asp Leu G1y G1y G1u Ser Leu Arg Arg Arg Arg	
601 201	AC COSA TAT T TAT T G AT C G TOG A T T C A C C G A T G CT TTT GAT ATC TCC ATC CTA GAA ATC GAT GAT GAC GT TTC GAA GTT CTA TCT ACA AAT GAA GAT ACT CTC CTC TTP Preakspillesen Tieleue Guille Gity Asp Gity Val Phe Giu Val Leu Ser Thr Asn Gity Asp Thr Leu sn Ang Tyr Phe Tyr Gity Asn Ang Trp	
676 226	A C A A A CA CALL AND A CA A CAA GAA GAA GAA GAA GAA GAA GAA	
751 251	C T G T G T G A T G A T G T T A G 3C AMA GAT AAT ATG GOC TTA CAA AGA CTT AAA GAT GCT GCT GAG AAA GCA AMA ATA GAA CTT TCA GGA GTC TCT ar Lys Asp Asn Met A Ta Lew Gin Ang Lew Lys Asp A Ta A Ta Giu Lys A Ta Lys I Te Giu Lew Ser Giy Val Ser	
826 276	T T T C C T A T T T G T A T A T C CC ACA GAA ATC AAT CAG CCA TTC ATC ACA ATG GAT CCA CCAA GGA CCT AAA CAC CTT GCA TTG ACA CTC ACA CST mr Thr Glu lie Asn Gin Pro Phe lie Thr Het Asp Ala Gin Giy Pro Lys His Leu Ala Leu Thr Leu Thr Ang lie Asn	
901 301	T A C C A T T C T G C C CAA T T GCT CAG T T A AA T T 36 CAA TTC GAG AMA CTC GCA GCC TCT TCTA ATC GAA AGA ACA AAA TCT CCA TGC ATC AAA GCA CTC AGT GAC GCA 1a Gìn Phe Gìu Lys Leu A 1a A1a Ser Leu 11e Gìu Ang Thr Lys Ser Pro Cys 11e Lys A1a Leu Ser Asp A1a Hris Ser Hris Ser	
976 326	TG TCT CT TC CA CT GA CGT VACTT TO CGCT AAG GAT ATC GAT GAT GAT CTC TA GAT ATC TO CGCA GTA GAT ATC GOC GCA GTA ACT VS Leu Ser Ala Lys Asp Tie Asp Asp Val Leu Leu Val Gly Gly Het Ser Arg Het Pro: Ala Val Gin Glu Thr Ser	
1051 351	GAG TCT G TA GAGC TA TA G CC T A T A G T T TA AAA GAL CTC TCT GEC AAA GAG CCT AAT AAA GGA GTC AAC COC GAC GAA GTT GTT GCT ATT GGA GCC GCA ATT al Lys Gil De He GTy Lys Gil PPO Asn Lys GIy Val Asn Pro Asp Giu Val Val Ala Ila Giy Ala Ala Ila Arg Ser Leu Val Ser Leu Ila Ala	
1126 176	G C C G A C G T G T G T C T A A T G AA GET GET GIT CTT GEC GEA GAA GIT AAG GAT GIT CTA CAC GIT ATC COC CTA TCT CTG GET ATC GAA In Gly Gly Val Leu Gly Gly Glu Val Lys Asp Val Leu Leu Leu Asp Val I le Pro Leu Ser Leu Gly I le Glu	
1201 401	T G T C T C G T CT CTA GGA GGC GTC ATG AGG ACT CTG GTA GAG AGA AAT ACA ACT ACA ACT CCT ACA CAG AAA AAA CAA ATC TTC TCC hr Lew Gly Gly Val Met Thr Thr Lew Val Glu Arg Asn Thr Thr Ile Pro Thr Gln Lys Lys Gln Ile Phe Ser Pro	
1276 426	C C T A A G T C T T A G T G C T CA GCT GAT AAC CAG GCT GGG GTT ACC ATC GTA GTT CTC CAA GGA GAG CGT COC ATG GCC AAA GAT AAC AAG hr A1a A1a Asp Asn G1n Pro A1a Va1 Thr I1e Va1 Va1 Leu G1n G1y G1u Arg Pro Met A1a Lys Asp Asn Lys	
1351 451	T T A C T T C A T G A A AA ATC GGA AGA HTC GAT CTT ACA GAT ATC COT COG GCT CCT CGA GGC CAT CCT CAA ATC GAA GTC TCC TTC GAT Iu 11e G1y Ang Phe Asp Leu Thr Asp 11e Pro Pro A1a Pro Ang G1y His Pro G1n 11e G1u Va1 Ser Phe Asp Thr	

	т		с					С	т	T				с	T	т	A	CGC		A		с		т	
1426	ATC	GAT	GCA	AAC	GGA	ATT	TTC	CAT	GTC	TCA	GCT	***	GAT	GTT	GCC	AGC	GGT	AAA	GAA	CAG	***	ATT	CGT	ATC	GAA
476	IJе	Asp	Ala	Asn	Gly	Пe	Phe	His	Va1	Ser	Ala	L.ys	Asp	Val	Ala	Ser	Gly	Lys	Glu	Gln	Lys	I۱e	Arg	I۱e	Glu
							Leu							Ala				Arg							
			т		TA					т		CA		A C	с		A	G	с	с	A	G		С	
1501	GCA	AGC	TCA	GGA	стт	CAA	GAA	GAT	GAA	ATC	CAA	AGA	ATG	GTT	CGA	GAT	GCC	GAA	ATT	AAT	AAG	GAA	GAA	GAT	AAA
501	Ala	Ser	Ser	Gly	Leu	Gln	Glu	Asp	Glu	Ile	Gln	Arg	Met	Val	Arg	Asp	Ala	Glu	Ile	Asn	Lys	Glu	Glu	Asp	Lys
						Lys						Gln		I١e					Leu	His					
	с		***			т		TG						G A			т						GG		
1576	***	CGT	OGT	GAA	GCT	TCA	GAT	GCT	***	AAT	GAA	ecc	GAT	AGC	ATG	ATC	TTC	AGA	GCC	GAA	A AA	GCT	ATT	AAA	GAT
526	Lys	Arg	Ang	Glu	Ala	Ser	Asp	Ala	Lys	Asn	Glu	Ala	Asp	Ser	Met	I le	Phe	Arg	A1a	Glu	Lys	Ala	Ile	Lys	Asp
	Gln		Lys					Val						Gly									Val		
	с	сс	с	•			с	GAA	СТ				т			AT	т	G	A	A		C A		A	
1651	TAT	AAG	GAG	CAA	ATT	сст	GAA	ACT	TTA	GTT	AAA	GAA	ATC	GAA	GAG	CGA	ATC	GAA	AAC	GTG	CGC	AAC	GCA	стс	AAA
551	Tyr	Lys	Glu	Gln	IJе	Pro	Glu	Thr	Leu	Val	Lys	Glu	I۱e	Glu	Glu	Arg	Ile	Glu	Asn	Val	Arg	Asn	Ala	Leu	Lys
		His	Asp	Lys			Ala	Glu								His			Lys			Gln		I le	
		т		тс	CA	AC	GCT	с		CA	с	т	т	G	ΤG	т	СТ					с		A	G
1726	GAT	GAC	CCT	αст	ATT	GAA	***	ATT	***	GAG	GTT	ACT	GAA	GAC	СТА	AGC	AAG	CAT	ATG	CAA	AAA	ATT	GGA	GAG	TCT
576	Asp	Asp	Ala	Pro	IJe	Glu	Lys	Ile	Lys	Glu	Val	Thr	Glu	Asp	Leu	Ser	Lys	His	Met	Gln	Lys	I le	Gly	Glu	Ser
	Gly			Ser	Thr	Thr	Ala			Ala	Ala	Ser	Asp	Glu			Thr								Ala
		G	GΤ		c		с				т	Т		G			с	A	G	A		т	с	тс	
1801	ATG	CAA	TCG	CAG	тст	GCA	TCA	GCA	GCA	GCA	TCA	106	GCA	GCC	AAT	GCT	AAA	GGT	GGA	сст	AAC	ATC	AAT	ACA	GAA
601	Met	Gln	Ser	Gln	Ser	Ala	Ser	Ala	Ala	Ala	Ser	\$er	Ala	Ala	Asn	Ala	Lys	Gly	Gly	Pro	Asn	Ile	Asn	Thr	Glu
			Ala														Gln							Ser	
																		_							
1076		с —						c	A	CGA			G	GGA	GGA	A C	GC	Т	TCT	ACA	GC	A	ATT		A
18/6	GAI				CAI	AGI	TIC	AGI	ACG	AAG	ai	-	ICA	AAT	AAC	GGT	TCT	TCA	GAA	GAC	CAT	ATC	GAA	GAA	GCT
020	ASP	Leu	Lys	Lys	HIS	Ser	Phe	Ser	Ihr	Lys	Pro	Pro	Ser	Asn	Asn	Gly	Ser	Ser	Glu	Asp	His	I le	Glu	Glu	Ala
										Arg			Ala	Gly	Gly	Ser	Ala		Ser	Thr	Asp	Asn	Ile		Asp
	~		-				с т		~	~															
1051	GAT	AI CTA	C44	417		CAT	440	A A	~~~		T.4.4	TC-									>	•			
1921	40	UIA Val	000	A11	11-	GAI An:	ANG	GAC A.	GAI A-	AAG	IAA	ICA.	AMAI	nic	AAII	TAAG	IIIC	ICTA	1100	LATO	LICA	AAG	AGGA	IGGG	AAAC
031	ASP	-	910	110	116	ASP	ASN	ASP	ASP	LYS															
	~18	лsр	val	ulu		vai	ASP	LYS	rro	619															
2039	πο	CTTA	TAAA	CAGA		AGTT	CCAT	тстс	TAT	тстс	TGAT	CAAG	GAGT	TGCA	ATAA	CAGA	сстт	стп	AGTG	CAAT	TGGC	птG	AATT	TGAG	ACTG

2138 CTOCTTTCATAATCACAAAACCCACTTAAAAGGGAAAATTTTGTTGAGCCACTCA

FIG. 3. Nucleotide sequence and inferred amino acid sequence of the gene encoding a 75-kDa protein of *C. pneumoniae*. The nucleotide and amino acid sequences of the corresponding *C. trachomatis dnaK* gene are also shown (above and below, respectively) where they differ from those of the *C. pneumoniae* gene. A dot (\cdot) indicates a gap in the sequence. The gene contains an ATG start codon followed by a 1,980-nucleotide open reading frame and ends with two in-frame TAA stop codons. The putative promoter is shown as a boxed region upstream centered at -108 and -129. The ribosomal binding site is shown as a boxed region centered at -11. The 11-base transcription terminator dyad is indicated by the inverted arrows from 2011 to 2036.

are done. In C. trachomatis, the promoter recognized by the chlamydial RNA polymerase during heat shock-induced transcription of the chlamydial dnaK gene appears to be different from the E. coli-like promoter sequence upstream from the gene (21).

While putative promoter and terminator regulatory regions in other chlamydial genes have been identified (7, 23, 26), no chlamydial consensus promoter sequences have been identified. The terminator sequences reported for other chlamydial genes demonstrate typical rho-independent stemloop structures (7, 26). In contrast, the region downstream of the translational stop signal in the C. pneumoniae gene revealed an 11-base dyad which forms a stem-loop structure, but there are no thymidine residues characteristic of the rho-independent terminator. The terminator sequence for the C. pneumoniae major outer membrane protein gene also contains a stem-loop structure which lacks a poly(T) tail (20a). However, there is another predicted stem-loop structure formed from bases 2036 to 2090. Transcription of the 11-base dyad would result in the folding of the mRNA into a stem-and-loop structure with a calculated least free energy of -15.0 kcal. If the transcription of the second stem-loop structure is included in this calculation, a lower calculated least free energy of -22.7 kcal results. Perhaps the second stem-loop structure acts in conjunction with the 11-base dyad to terminate transcription.

C. pneumoniae C. trachomatis MSEHKKSSKIIGIDLGTTNSCVSVMEGGQAKVITSSEGTRTTPSIVAF.K E. coli megaterium 120 51 GNEKLVGIPAKRQAVTNPEKTLGSTKRFIGRKY...SEVASEIQTVPYTVTSGSKGDAVFEVDGKQYTPEE -G-T-------Q-----QN---PAI--L-----F.-----K----K-PN-----D-EQ-L-----DG-T---Q------QN---PAI--L-----RQDE--QRDVSIMF-KIIAADN---WV--K-QKMA-PQ 190 121 IGAQILMKMKETAEAYLGETVTEAVITVPAYFNDSQRASTKDAGRIAGLDVKRIIPEPTAAALAYGIDK. MS-I--QHL-GY-E----P--K-----AE-QA-----K----E-E---N------LL-T 191 260 VGDKKIAVFDLGGGTFDISILEI....GDGVFEVLSTNGDTLLGGDDFDEVIIKWMIEEFKKQEGIDLSK DE-QTVL-Y-----V----L....-RA-A--NR-----Q---DYLVA----EN-K----261 330 DNMALQRLKDAAEKAKIELSGVSSTEINQPFITMDAQGPKHLALTLTRAQFEKLAASLIERTKSPCIKAL -----H--S-----0--AO---PL-M----E------SAQQ-DV-L-Y-A-T---MNIKV---KL-S-VED-VN-SIE-LKV---K-----KD----T-Q-SL----AGEA--L--EVS-S--K-DE-S-G-V---MA-VRQ--331 SDAKLSAKDIDDVLLVGGMSRMPAVQETVKELFGKEPNKGVNPDEVVAIGAAIQGGVLGGEVKDVLLLDV Q--G--VS----I---QT---KK-A-F----R-D---A---V--T-D------K--G---SEL-K-I----ST-I----DAL-KET-QD-H-----L-----T-D----V----470 IPLSLGIETLGGVMTTLVERNTTIPTQKKQIFSTAADNQPAVTIVVLQGERPMAKDNKEIGRFDLTDIPP 471 APRGHPOIEVSFDIDANGIFHVSAKDVASGKEOKIRIEASSGLOEDEIORMVRDAEINKEEDKKRREASD ----V------K----VN-R---LGTN---A-T-KS-T--SD---D---KE--E-ADA--Q-K-EVE 541 AKNEADSMIFRAEKAIKDYKEOIPETLVKEIEERIENVRNALK..DDAPIEKIKEVTEDLSKHMOKIGES V----G-----V---HDK--AE----H-K--Q-I-..E--STTA--AASDE--T-----A TR-QG-HLLHSTR-QVEEAGDKL-ADDKTA--SALTALET---GE-K-A--AKMQELAQV-QKLME-AQQ --QLV-TT--TL--LEGKVE-AE-TKAN-AKDALKA-IEKN-...L-E--AKKDE-...-E-VQA 671 MQSQSASAAASSAANAKGGPNINTEDLKKHSFSTKPPSNNGSSEDHIEEADVEIIDNDDK-LTVK-YEQ-Q.....QAQQAGEQ-AQN-DVVD-EF-EVND-K-

FIG. 4. Comparison of the amino acid sequence of the 75-kDa protein of *C. pneumoniae* with those of hsp70 proteins of *C. trachomatis* (75-kDa protein), *E. coli* (DnaK), and *B. megaterium* (DnaK). A dash indicates the same amino acid as in the *C. pneumoniae* 75-kDa protein. A dot represents a gap in the sequence.

Antigenic properties of the 75-kDa chlamydial protein suggest that it is important in the elicitation of antichlamydial responses and that it may provide a broad-based target for intervention in chlamydial infection (1, 3, 16). The 75-kDa proteins of C. pneumoniae and C. trachomatis were compared to identify shared regions that might be potential targets of such a response, and they were found to have a degree of amino acid similarity higher than that of other hsp70 proteins (15). In comparing amino acid sequences of the C. trachomatis 75-kDa protein to other bacterial hsp70 proteins, Danilition et al. (7) identified conserved regions and variable regions. Comparison of the inferred amino acid sequences of the C. pneumoniae and C. trachomatis 75-kDa proteins identified conserved sequences shared by the two Chlamydia species that are found in generally divergent regions of the hsp70 proteins. Interestingly, these regions are also identified as regions of high antigenic index when a computer algorithm that compiles data which predicts secondary structure, flexibility, mobility, hydrophilicity, and surface probability is used. For other intracellular parasites, heat shock proteins have been shown to be important in the elicitation of the immune response. T-lymphocyte responses and B-cell responses against heat shock proteins have been reported for Mycobacterium tuberculosis (30) and Plasmodium falciparum (29), respectively. In the C. trachomatis immune response, the hsp60 homolog has been associated

with the delayed-hypersensitivity immunopathology of chlamydial infection (18, 19).

The heat shock response is universal, and the stress proteins are among the most highly conserved genetic elements known (15). In addition to their immunogenic roles, a multitude of biologic functions has been ascribed to the proteins of the hsp70 family. They have been associated with a role in the developmental cycles of various organisms, including leishmanias, trypanosomes, plasmodia, histoplasmas, and chlamydiae (15, 29, 30). The hsp70 protein has also been implicated in a variety of other processes, including DNA replication, protein transport, protein binding, and the uncoating of coated vesicles (15). Perhaps the *Chlamydia* 75-kDa protein utilizes one of these mechanisms in order to perpetuate or maintain its pathogenicity.

This study was supported by grant 16889 from the Edna McConnell Clark Foundation and by U.S. Public Health Service grants EY 00219 and AI 21885 from the National Institutes of Health.

We thank Lynne C. Kikuta for technical assistance and David T. Grayston for assistance in computer analysis of the data.

REFERENCES

- Birkelund, S., A. G. Lundemose, and G. Christiansen. 1989. Characterization of native and recombinant 75-kilodalton immunogens from *Chlamydia trachomatis* serovar L2. Infect. Immun. 57:2683-2690.
- Birkelund, S., A. G. Lundemose, and G. Christiansen. 1990. The 75-kilodalton cytoplasmic *Chlamydia trachomatis* L2 polypeptide is a DnaK-like protein. Infect. Immun. 58:2098–2104.
- 3. Brunham, R. C., I. Maclean, J. McDowell, K. Persson, and S. Osser. 1987. Postabortal *Chlamydia trachomatis* salpingitis: correlating risk with antigen specific serologic responses and with neutralization. J. Infect. Dis. 155:749–755.
- 4. Campbell, L. A., L. C. Kikuta, C.-C. Kuo, M. Perez Melgosa, and J. T. Grayston. 1990. Characterization of *Chlamydia pneu*moniae genes, p. 414–417. *In* W. R. Bowie et al. (ed.), Chlamydial infections: proceedings of the 7th International Symposium on Human Chlamydial Infections. Cambridge University Press, Cambridge.
- Campbell, L. A., C.-C. Kuo, and J. T. Grayston. 1987. Characterization of the new *Chlamydia* agent, TWAR, as a unique organism by restriction endonuclease analysis and DNA-DNA hybridization. J. Clin. Microbiol. 25:1911–1916.
- Campbell, L. A., C.-C. Kuo, R. W. Thissen, and J. T. Grayston. 1989. Isolation of a gene encoding a *Chlamydia* sp. strain TWAR protein that is recognized during infection of humans. Infect. Immun. 57:71–75.
- Danilition, S., I. W. Maclean, R. Peeling, S. Winston, and R. C. Brunham. 1990. The 75-kilodalton protein of *Chlamydia trachomatis*: a member of the heat shock protein 70 family? Infect. Immun. 58:189–196.
- Grayston, J. T., L. A. Campbell, C.-C. Kuo, C. H. Mordhorst, P. Saikku, D. H. Thom, and S.-P. Wang. 1990. A new respiratory tract pathogen: *Chlamydia pneumoniae* strain TWAR. J. Infect. Dis. 161:618–625.
- Grayston, J. T., C.-C. Kuo, L. A. Campbell, and S.-P. Wang. 1989. Chlamydia pneumoniae sp. nov. for Chlamydia sp. strain TWAR. Int. J. Syst. Bacteriol. 39:88–90.
- Hanahan, D. 1983. Studies on transformation of *Escherichia coli* with plasmids J. Mol. Biol. 166:557-580.
- 11. Hawley, D. K., and W. R. McClure. 1983. Compilation and analysis of *Escherichia coli* promoter DNA sequences. Nucleic Acids Res. 11:2237-2249.
- Kleid, D. G., D. Yansura, B. Small, D. Dowbenko, D. M. Moore, M. J. Grubman, P. D. McKercher, D. O. Morgan, B. H. Robertson, and H. L. Bachrach. 1981. Cloned viral protein vaccine for foot-and-mouth disease: responses in cattle and swine. Science 214:1125–1129.
- 13. Kuo, C.-C., H.-H. Chen, S.-P. Wang, and J. T. Grayston. 1986. Identification of a new group of *Chlamydia psittaci* strains

called TWAR. J. Clin. Microbiol. 24:1034-1037.

- 14. Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) 227:680-685.
- 15. Lindquist, S., and E. A. Craig. 1988. The heat shock proteins. Annu. Rev. Genet. 22:631-677.
- Maclean, I. W., R. W. Peeling, and R. C. Brunham. 1987. Characterization of *Chlamydia trachomatis* antigens with monoclonal and polyclonal antibodies. Can. J. Microbiol. 34: 141–147.
- 17. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- Morrison, R. P., R. J. Belland, K. Lyng, and H. D. Caldwell. 1989. Chlamydial disease pathogenesis. The 57 kDa chlamydial hypersensitivity antigen is a stress response protein. J. Exp. Med. 170:1271-1283.
- Morrison, R. P., K. Lyng, and H. D. Caldwell. 1989. Chlamydial disease pathogenesis. Ocular hypersensitivity elicited by a genus-specific 57 kDa protein. J. Exp. Med. 169:663–675.
- Palmer, L., and S. Falkow. 1986. Characterization of cloned genes from *Chlamydia trachomatis*, p. 91–95. In L. Leive (ed.), Microbiology—1986. American Society for Microbiology, Washington, D.C.
- 20a. Perez Melgosa, M., et al. Unpublished data.
- Pollack, J., J. N. Engel, E. Perara, and D. Ganem. 1990. The heat shock response of *Chlamydia trachomatis*, p. 137-140. *In* W. R. Bowie et al. (ed.), Chlamydial infections: proceedings of the 7th International Symposium on Human Chlamydial Infections. Cambridge University Press, Cambridge.
- 22. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequenc-

ing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463-5467.

- Sardinia, L. M., J. N. Engel, and D. Ganem. 1989. Chlamydial gene encoding a 70-kilodalton antigen in *Escherichia coli*: analysis of expression signals and identification of the gene product. J. Bacteriol. 171:335-341.
- 24. Shine, J., and L. Dalgarno. 1974. The 3' terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementary to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. USA 71:1342-1346.
- 25. Spindler, K. R., D. S. E. Rosser, and A. J. Berk. 1984. Analysis of adenovirus transforming proteins from early regions 1A and 1B with antisera to inducible fusion antigens produced in *Escherichia coli*. J. Virol. 49:132-141.
- Stephens, R. S., G. Mullenbach, R. Sanchez-Pescador, and N. Agabian. 1986. Sequence analysis of the major outer membrane protein gene from *Chlamydia trachomatis* serovar L2. J. Bacteriol. 168:1277-1282.
- 27. Towbin, H., T. Staehelin, and J. Gordon. 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedures and some applications. Proc. Natl. Acad. Sci. USA 76:4350-4354.
- Yanisch-Perron, C., J. Vieira, and J. Messing. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33:103–119.
- Young, D., R. Lathigra, and A. Mehlert. 1989. Stress-induced proteins as antigens in infectious diseases, p. 275–285. In M. L. Pardue, J. R. Feramisco, and S. Lindquist (ed.), Stress-induced proteins. Alan R. Liss, Inc., New York.
- Young, R. A., and T. J. Elliott. 1989. Stress proteins, infection, and immune surveillance. Cell. 59:5-8.