Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 10783–10787, December 1991
Plant Biology

The plastid genome of *Cryptomonas* Φ encodes an hsp70-like protein, a histone-like protein, and an acyl carrier protein

SHENGLONG WANG AND XIANG-QIN LIU*

Canadian Institute for Advanced Research, Department of Biochemistry, Dalhousie University, Halifax, NS B3H 4H7, Canada

Communicated by André T. Jagendorf, September 5, 1991

ABSTRACT **The plastid genome of *Cryptomonas* Φ, a cryptomonad alga, contains three genes that have not previously been found in any organellar genome. Each of these genes encodes a functional class of organellar gene product not previously reported. The first gene, *dnaK*, encodes a polypeptide of the hsp70 heat shock protein family. The predicted amino acid sequence of the DnaK protein is 54% identical to that of the *Escherichia coli* hsp70 protein (DnaK), 50–53% identical to that of two nucleus-encoded mitochondrial hsp70 proteins, and 43–46% identical to that of several eukaryotic cytoplasmic members of the hsp70 protein family. The second gene, *hlpA*, encodes a polypeptide resembling bacterial histone-like proteins. The predicted amino acid sequence of the HlpA protein is 25–53% identical to that of several bacterial histone-like proteins, and the identity increases to 39–76% over a conserved region corresponding to the long arm that binds DNA. The third gene, *acpA*, encodes an acyl carrier protein, which is a key cofactor in the synthesis and metabolism of fatty acids. Its predicted amino acid sequence is 36–59% identical to that of eubacterial and plant chloroplast (nucleus-encoded) acyl carrier proteins.**

The endosymbiont hypothesis suggests that contemporary plastids and mitochondria evolved from free-living eubacteria that entered early eukaryote cells as endosymbionts (1, 2). A central theme of this hypothesis is the loss or transfer to the nuclear genome of a majority of the endosymbiont genes and the retention of some genes in the organellar genomes. Nucleotide sequences of three plant chloroplast genomes have been completely determined, revealing their gene contents (3–5). In addition to tRNA and rRNA genes, each chloroplast genome contains over 60 identified protein-encoding genes and about 30 unidentified open reading frames (6). The identified genes encode proteins involved in transcription (RNA polymerase subunits), translation (ribosomal proteins, translation factors), photosynthesis, and probably chlororespiration (*ndh* genes). Various chloroplast genomes studied so far show similar gene contents, while several exceptions have been found in which a gene is present in the chloroplast genome of one organism but absent from that of another (e.g., see refs. 7–9). Such exceptions have proven useful in studying the process of gene transfer from chloroplasts to the nucleus (10).

Cryptomonas Φ is a chlorophyll c- and phycobiliprotein-containing alga, belonging to the Kingdom Chromista. Other members of this kindom include chromophytes, dinoflagellates, and diatoms (11). Plastids of this group of organisms are suggested to have originated from a secondary endosymbiosis, in which the endosymbiont was a photosynthetic eukaryote rather than a eubacterium (11–13). In contrast to chloroplasts of metaphytes (land plants) and chlorophytes (green algae), which have two membranes surrounding the thylakoids, plastids of Chromista possess two additional membranes (termed chloroplast endoplasmic reticulum, or CER), the outermost of which is continuous with the nuclear membrane. In addition, a nucleomorph (likely a degenerate nucleus) is still present inside the CER of cryptomonads. The plastid genome of *Cryptomonas* Φ has been shown to be a circular DNA of 118 kilobase pairs (kbp), containing an inverted repeat of less than 6 kbp within which the plastid rRNA genes are located (14). Several genes of this genome have been described previously, including the *rbcS* gene, whose counterpart in plants is encoded in the nuclear genome (7). In this paper, we present the structures of three genes found in the plastid genome of *Cryptomonas* Φ that encode an hsp70-like heat shock protein, a histone-like protein, and an acyl carrier protein.† To date, so far as we know, these genes have not been reported for any other organellar genome, and each of them represents a functional class of organellar gene product not described previously.

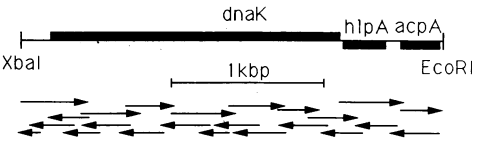

FIG. 1. Gene map of the 2.8-kbp *Eco*RI–*Xba* I DNA fragment of *Cryptomonas* Φ plastid DNA. The solid bars represent coding sequences of the genes. The *dnaK* gene is transcribed from left to right, while *hlpA* and *acpA* genes are transcribed from right to left. The arrows illustrate the length and direction of the sequenced fragments.

MATERIALS AND METHODS

A plasmid clone (BS-7) containing a 12-kbp *Bam*HI–*Sal* I plastid DNA fragment was provided by Susan E. Douglas (Institute of Marine Biosciences, Halifax, NS, Canada). The *Cryptomonas* Φ strain, from which the plastid DNA fragment was originally isolated (14), was from the Culture Collection of Marine Phytoplankton (Bigelow Laboratory for Ocean Sciences, West Boothbay Harbor, ME). A 2.8-kbp *Xba* I–*Eco*RI DNA fragment was isolated from the BS-7 DNA, cloned in plasmid vector pUC118, and subjected to DNA sequencing.

DNA sequencing was performed with the dideoxy chain-termination method (15), using the Sequenase enzyme system (United States Biochemical) and deoxyadenosine 5′-[α-[^{35}S]thio]triphosphate (DuPont/NEN) according to the instructions of the suppliers. Sequencing clones were generated by performing systematic deletions using exonuclease III and S1 nuclease (16). In some cases, oligonucleotides were synthesized and used as primers to sequence small gaps or the second strand. The deduced amino acid se-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: CER, chloroplast endoplasmic reticulum.

*To whom reprint requests should be addressed.

†The sequence reported in this paper has been deposited in the GenBank data base (accession no. M76547).

quences were aligned by using the sequence alignment program of Corpet (17), with a gap penalty of 8.

RESULTS

Gene Structure and Sequences. Complete sequence determination of the 2.8-kbp *Xba* I–*Eco*RI DNA fragment revealed three open reading frames, designated *dnaK*, *hlpA*, and *acpA* (Fig. 1). The *dnaK* gene is encoded on the 5′-to-3′ DNA strand in a direction from *Xba* I to *Eco*RI, while *hlpA* and *acpA* are encoded on the 3′-to-5′ DNA strand. There is a 105-bp noncoding sequence between *hlpA* and *acpA*, and a 13-bp sequence between the stop codon of *dnaK* and that of *hlpA*.†

DnaK Protein. The deduced protein sequence of the plastid DnaK is 627 amino acid residues long, containing 98 acidic, 78 basic, and 184 hydrophobic amino acid residues. The calculated molecular mass is 68.5 kDa, with a calculated pI

```
                                                  10        20        30        40        50        60        70        80
P. DnaK                                    MGKVVGIDLGTTNSVVAVMEGGKPAVIQNAEGFRTTPSVVAYTKTGDRLVGQIAKRQAVINPDNTFYSVKRFIGRR--SEEV
E. DnaK                                    ...II.........C..I.D.TT.R.LE....D.....II...QD.ET....P......T..Q..LFAI..L....FQD...
Sscl     MLAAKNILNRSSLSSSFRIATRLQSTKVQ.S.I.........A..I...KV.KI.E....S........F..E.E....IP......V..E..LFAT..L....FEDA..
Mtp70       MFARRLRGAGSLAAASLARWQSSKVT.D.I.......Y.C......D..R.LE.T....A......F-.GQEK...LA......T..QS..FA...L....FEDSNI
Tryp.                                   MTYE.-AI.......Y.C.G.WQNERVEI.A.DQ.N.....Y..F.-DSE..I.DA..N.VAM..T..VFDA..L...KFSDSV.
Yeast                                     .S.A........Y.C..HFANDRVDI.A.DQ.N.....F..F.-DTE..I.DA..N..AM..S..VFDA..L...NFNDP..
Maize                                  MAKSE.PAI.......Y.C.GLWQHDRVEI.A.DQ.N.....Y.GF.-DTE..I.DA..N.VAM..T..VFDA..L....FS.PA.
Human                                    MAKAAA........Y.C.G.FQH..VEI.A.DQ.N.....Y..F.-DTE..I.DA..N.VAL..Q..VFDA..L...KFGDPV.

                90        100       110       120       130       140       150       160       170       180
P. DnaK SEELKQVSYIVKTDSNGNIKL--DCPSLKKEFASEEISAEVLRKLVDDASKYLGESVKQAVITVPAYFNDSQRQATKDAGRIAGLEVLRIINEPTAASLAYGLDKK-N--
E. DnaK QRDVSIMPFKIIAAD..DAWV--EVKG--QKM.PPQ.....K.MKKT.ED....P.TE..........A..............K.........A.......GTG--
Sscl     QRDI...P.KIVKH...DAWV--EARG--QTYSPAQ.GGF..N.MKET.EA...KP..N..V..................Q.V..N...VV......A.....E.S-D--
Mtp70    QHDI.N.P.KIGRS...DAWVQ-.ANG--.QYSPSQVG.F..E.MKET.ENF..RK.SN..V.C.....GP.........T....N.I.VV.G....A.......T-K--
Tryp.    QSDM.HWPFK.V.KGDDKPVIQVQFRGET.T.NP....SM..L.MKEV.ES...KQ.AK..V..................T................AI......ADEGK
Yeast    QADM.HFPFKLI-.VD.KPQIQVEFKGET.N.TP.Q..SM..G.MKET.ES...AK.ND..V..................T....N...........AI......--KGK
Maize    QSSM.LWPSRHL-GLGDKPMIVFNYKGEE.Q..A....SM..I.MKEI.EA...STI.N..V..................V....N.M.........AI.......ATSS
Human    QSDM.HWPFQ.IN-DGDKP.VQVSYKGET.A.YP....SM..T.MKEI.EA...YP.TN...................V....N...........AI......RTG--K

          190       200       210       220       230       240       250       260       270       280
P. DnaK -NETILVFDLGGGTFDVSVLEV----GDGVFEVLSTSGDTHLGGDDFDDKIVQWLLKEFETEHS-INLKSDRQALQRLTEASEKAKIELSNLSQTEINLPFLTATETGPK
E. DnaK -.R..A.Y........I.II.IDEVD.EKT....A.N.......E...SRLINY.VE..KKDQG-.D.RN.PL.M...K..A.........SAQ..DV...YI..DA....
Sscl     -SKVVA..........I.I.DI----DN.....K..N.......E...IYLLREIVSR.K..TG-.D.EN..M.I..IR..A.........STVS.......I..DAS...
Mtp70    -DSM.A.Y........I....I----AG.....KA.N.......E...LCLSDYI.T..KKSTG-.D.SNE.M....IR..A....C...TTME..V....I..NQD.AQ
Tryp.    -ERNV.I..........TL.TI----DG.I...KA.N.......E...NRL.AHFTE..KRKNKGKD.S.NLR..R..RT.C.R..RT..SAA.AT.EIDA.FENI----
Yeast    -E.HV.I..........L.FI----E..I...KA.A.......E...NRL.NHFIQ..KRKNK-KD.STNQR..R..RT.C.R..RT..SSA..SVEIDS.FEGI----
Maize    GEKNV.I..........L.TI----EE.I...KA.A.......E...NRM.NHFVQ..KRKNK-KDISGNPR..R..RT.C.R..RT..STA..T.EIDS.FEGI----
Human    GERNV.I..........I.TI----D..I...KA.A.......E...NRL.NHFVE..KRK.K-KDISQNKR.VR..RT.C.RFEGIDFYT.I.RARFEE.AKRT----

      290       300       310       320       330       340       350       360       370       380           390
P. DnaK HLERSITRAKFEELCSDLINRVKIPVENALKDAKLDSSKIDEVVLVGGSTRIPAIQELVKRIL-NKTPNQTVNPDEVVAIGAAVQAGVLAGEVK-----DILLLDVTPLS
E. DnaK .MNIKV....L.S.VE..V..SIE.LKV..Q..G.SV.D..D.I....Q..M.MV.KK.AEFF-G.E.RKD.....A........G...T.D..-----.V.........
Sscl     .INMKFS..Q..T.TAP.VK.TVD..KK.....G.ST.D.S..L....MS.M.KVV.T..SLF-G.D.SKA.....A........GA..S...T-----.V.........
Mtp70    .VQMTVS.S...S.AEK.VQ.SLG.CKQCI...AV.LKE.S.......M..M.KVI.A..QFF-GRD.FRG.....A..L.G.TLG...RRD..-----GLV.........
Tryp.    DFQAT....R.....G..FRGTLQ...RV.Q...M.KRAVHD..........KVMQ..SDFFGG.EL.KSI....A.XY......FI.T.GKSKQT-EGL.....A..T
Yeast    DFYT.....R.....A..FRSTLD...KV.R.....K.QV..I..........KV.K..TDYFNG.E..RSI....A..Y......AI.T.DESSKT-Q.L.....A...
Maize    DFTPRSS..R....NM..FRKCME...KC.R...M.K.SVHD..........KV.Q.-QDFFNG.ELCKSI....A..Y......AI.S..GNERS--.L.........
Human    LSSSTQASLEIDS.....FRSTLE...K..R.....KAQ.HDL.........KV.K.LQDFFNGRDL.KSI....A.GY......AI.M.DKSENV-Q.L.....A...

           400       410       420       430       440       450       460       470       480       490       500
P. DnaK LGVETLGGVTTRIIPRNTTIPTKKSE-VFSTAVDNQPNVEIHVLQGEREFAKDNKSLGTFRLDGILPAPRGVPQIEVTFDIDANGILSVTAKDKGTGKEQSITITGAST-
E. DnaK ..I..M...M.TL.AK.......H.Q-.....E...SA.T.........KR.A......Q.N....N.....M...........D...H.S....NS....K...KAS.G-
Sscl     ..I......F..L.............Q-I....AAG.TS...R.F.....LVR...LI.N.T.A..P...K.............D..IN.S.R..A.N.DS...VA.S.G-
Mtp70    .........F..M..K.........QTF....AF..TQ.G.K.F.....M.A..QMM.Q.D.V..P..............EP...CH.......A...T.N....ASGG-
Ttyp.    ..I..A...M.AL.K..........Q-I...YS....G.H.Q.FE...TMT..CHL....D.S..P..............L.........S.EE.....RNQ.V..NDKGR
Yeast    ..I..A...M.KL....S..S...F.-I...YA....G.L.Q.FE...AKT...NL..K.E.S..P..............V.S....N.S.VE.....SNK....NDKGR
Maize    ..L..A...M.VL...........EQ-....YS....G.L.Q.YE...ART...NL..K.E.S..P.........T......V.N..N.S.E..T..QKNK....NDKGR
Human    ..L..A...M.AL.K..S.....QTQ-I.T.YS....G.L.Q.YE...AMT...NL..R.E.S..P...-..............N...T..S...ANK....NDKGR

          510       520       530       540       550       560       570       580       590       600
P. DnaK LPSDEVERMVNEAQNSAKEDKEKRDKIDLKNQSDSLCYQSEKQLKE--LEGLIDDTNKNKISSMISELRNAINNENY---DEMRDLNSKVQTALMDLGKSVYEKTSKQQT
E. DnaK .NE..IQK..RD.EAN.EA.RKFEELVQTR..G.H.LHSTR..VE.--AGDKLPADD.TA.E.ALTA.ET.LKG.DKA---AIEAKMQELAQVSQK.MEIAQQQHAQ...
Sscl     .SEN.I.Q...D.EKFKSQ.EARKQA.ETA.KA.Q.ANDT.NS...--F..KV.KAEAQ.VRDQ.TS.KELVARVQGG--E.VNA--EELK.KTEE.QT.SMKLFEQLYK
Mtp70    .SKEQI...IRDSESH.ES.RL..ELVEVR.NAETQANTA.R..T.--WK-YVS.AE.ENVRTLLRAC.KSME.P.V.TK..LSAATD.L.K.V.EC.RTE.QQAAAGNS
Tryp.    .SKADI....SD.AKYEA...AHVXX..A..GLENYAFSMKNTINDPNVA.KL..AD..AVTTAVE.ALRWL.DNQEASLE.YNHRQKELEGVCAPILSKM.QGMGG--G
Yeast    .SKEDI.K..A..EKFKE..EKESQR.AS...LE.IA.SLKNTIS.--AGDKLEQAD.DTVTKKAE.TISWLDSNTTASKE.FD.KLKEL.DIANPIMSKL.QA-----G
Maize    .SKE.I.K..Q..EKYKA..E.VKK.V.A..ALENYA.NMRNTI.DDKIASKLPAED.K..EDAVDGAISWLDSNQLAEVE.FE.KMKELEGICNPIIAKM.X-------
Human    .SKE.I....Q..EKYKA..EVQ.ERVSA..ALE.YAFNMKSAVEDEG.K.K.SEAD.K.VLDKCQ.VISWLDANTLAEK..FEHKRKELEQVCNPIISGL.QGA----G

                610       620
P. DnaK STS----------SPTNSNDSVIDADFSETK
E. DnaK AGAD-------ASANNAKD.D.V..E.E.V.DKK
Sscl     ND.----------NN.N.NNGNN.ESG...Q
Mtp70    .S.---------SGN.D.SQGEQQQQGDQQ.Q
Tryp.    DGPGGMPEGMPGGM.GGMPGGMGGGMGGAAASSGPKVEEVD
Yeast    GAPGGAAGGAPGGF.GGAPPAPEAEGPTVEEVD
Maize    GEGAG--MGAAAGMDEDAPSGGSG.GPKIEEVD
Human    GPGPG-----GFGAQGPKGG.GSGPTIE.VD
```

FIG. 2. Comparison of plastid DnaK protein with hsp70 family proteins. The amino acid sequence of *Cryptomonas* plastid DnaK protein (P. DnaK), deduced from the DNA sequence, is aligned with those of seven biologically distinct members of the hsp70 family: a eubacterial member [E. DnaK for *Escherichia coli* DnaK (18)], two mitochondrial members [Sscl of yeast (19) and Mtp70 of *Trypanosoma cruzi* (20), both nucleus encoded], and four members that function in the cytoplasm of eukaryotic cells [*Trypanosoma brucei* hsp70 (21), yeast Ssa1 (22), maize hsp70 (23), and human hsp70 (24)]. ., Residues in the other sequences that are identical to those in P. DnaK; -, computer-generated gaps that improve the alignment.

Table 1. Sequence similarity between plastid DnaK and other hsp70 proteins

	P. DnaK	E. DnaK	Ssc1	Mtp70	Tryp.	Yeast	Maize
E. DnaK	54						
Ssc1	53	56					
Mtp70	50	53	57				
Tryp.	43	43	44	42			
Yeast	46	46	46	44	69		
Maize	44	45	45	41	68	70	
Human	44	45	44	42	65	67	67

Numbers shown are percentages of amino acid residues identical between individual pairs of sequences. Calculations are based on the sequence alignment in Fig. 2, and the abbreviations are the same.

at pH 4.85. Computer-assisted sequence alignment was performed to compare the amino acid sequence of the plastid DnaK with that of selected hsp70 family proteins (Fig. 2). Extensive similarity among these proteins is seen over about 80% of the total sequence (N-terminal part), while the C-terminal 20% of the entire sequence shows little conservation. The extra amino acid residues at the N termini of the Ssc1 and Mtp70 sequences are putative leader sequences for targeting these proteins into mitochondria and may not be present on the mature proteins (19, 20). Table 1 summarizes sequence similarities among the proteins compared. Presented in the form of percent identical amino acid residues, the plastid DnaK protein is more similar to *E. coli* DnaK (54%), Ssc1 (53%), and Mtp70 (50%) than to the other members (43–46%).

HlpA Protein. The deduced protein sequence of HlpA is 93 amino acid residues long, containing 9 acidic, 17 basic, and 31 hydrophobic amino acid residues. The calculated molecular mass is 10.6 kDa, with a calculated pI at pH 10.48. The plastid HlpA protein sequence was aligned with 10 bacterial histone-like protein sequences (Fig. 3). Significant sequence similarity exists between HlpA protein and the other proteins, especially in the region of residues 38–79, which corresponds to the DNA-binding long arm of histone-like proteins (26, 27). At the level of the complete sequence, HlpA protein shares 25–53% identical amino acid residues with the other histone-like proteins (Table 2). If the comparison is limited to the region between residues 38 and 79, the protein sequence of HlpA is 76% identical to that of cyanobacterial histone-like protein HAn and 39–68% identical to those of other bacterial histone-like proteins (Table 2).

AcpA Protein. The deduced protein sequence of AcpA is 81 amino acid residues long, containing 17 acidic, 4 basic, and 26 hydrophobic amino acid residues. The calculated molecular mass is 8.9 kDa, with a calculated pI at pH 3.58. The amino acid sequence of AcpA is aligned with that of eubacterial and plant acyl carrier proteins in Fig. 4. Amino acid sequence identities between AcpA and the others are 56% for *E. coli* (36), 59% for *Anabaena variabilis* (37), 44% for spinach ACP-I (38), 36% for spinach ACP-II (39), 43% for *Arabidopsis thaliana* (40), and 43% for *Brassica campestris* (41). The serine residue at position 38, which is the attachment site for the prosthetic group phosphopantetheine (42), is conserved among all the acyl carrier proteins compared.

DISCUSSION

The three genes (*dnaK, hlpA, acpA*) described in this paper are, to our knowledge, the first such genes to be reported for any organellar genome, and each of them represents a functional class of gene product not previously found in organelles. They most likely originated from an ancient proto-organellar prokaryote, in accordance with the endosymbiont hypothesis of plastid origins (1, 2). The absence of these genes in the chloroplasts of plants and green algae suggests substantial differences between these plastids and that of *Cryptomonas* Φ in the number and/or identity of genes they possess. This suggestion is supported also by the finding of the *rbcS* gene and several ribosomal protein genes (all absent from plant chloroplasts) in plastids of *Cryptomonas* Φ (ref. 7; S.W., S. E. Douglas, and X.-Q.L., unpublished data). Such differences between the plastid of chlorophytes (plants, green algae) and that of *Cryptomonas* Φ may have interesting implications in term of plastid evolution. For example, *Cryptomonas* Φ plastids may have originated from a primary endosymbiosis (between a eubacterium and a eukaryote) separate from that leading to chlorophytic chloroplasts. Another possibility is that plastid CER (i.e., the two extra membranes surrounding the plastid) in *Cryptomonas* Φ may have slowed the transfer of plastid genes to the nucleus.

The *Cryptomonas* plastid DnaK protein is clearly a member of the hsp70 protein family (Fig. 2), which is a group of highly conserved proteins found in eukaryotic as well as

```
                  10        20        30        40        50        60        70        80        90
HlpA    MNKSQLISKIA-YYTKYSKTDIEKIITSMLEIIVDTVATGEKVTLVGFGSFEARERKAREGRNPRTGEKLFLPASRIPTFSVGNFFRNKVNKTF
HAn     ...GE.VDAV..EKASVT.KQADAVL.AA..T.IEA.SR.D..........S..........K.N..MEI..T.V.A..A.KL..E..APPKA
HCp     ...AE..TSM..EKS.LT.K.A.LALKALI.SVEEALEK....Q.....T..T...A.........K.VINI..TTV.V.KA.KE.KD....
HBs     ...TE..NAV..ETSGL..K.AT.AVDAVFDS.TEALRK.D..Q.I...N..V...A..K....Q...EMEI...KV.A.KP.KALKDA.K
HUb     .......D....AGADI..AAAGRALDAIIASVTESLKE.DD.A.....T.AVK..A..T....Q..KEITIA.AKV.S.RA.KALKDA..
HUa     ...T...DV...EKAEL...QAKAALE.T.AA.TESLKE.DA.Q.....T.KVNH.AE.T....Q..KEIKIA.ANV.A.VS.KALKDA.K
IHFb    .T..E..ERL.TQQSHIPAKTV.DAVKE...HMAS.L.Q..RIEIR.....SLHY.AP.T....K..D.VE.EGKYV.H.KP.KEL.DRNIYG
IHFa  MALT.AEMSEYLF.DKLGL..R.AKELVELFF.E.RRALEN..Q.K.S...N.DL.DKNQ.P....K...DIPIT.R.VV..RP.QKLKSR.ENASPKDE
Tf1     ...TE..KA...QD.ELTQVSVS.MLA.FEK.TTE...K.D..Q.T..LNIKPVA.Q..K.F..Q.Q.A.EIAP.VGVSVKP.ESLKKAAEGLKYEDFAK
HRm     ...NE.VAAV..DKAGL..A.ASSAVDAVF.T.QGELKN.GDIR.....N.SVSREASKGR...S..AEVDI..RNV.K.TA.KGLKDA..
HTa     MVGI.E.SKEV..KKANTTQKVARTV.K.F.DE..SEANG.Q.IN.A...I..R.TQFP.KA...Q.KKVIEV.SKKKFV.RASSKIKYQQ
```

FIG. 3. Comparison of plastid HlpA protein with bacterial histone-like proteins. The amino acid sequence of HlpA, deduced from the DNA sequence, is aligned with sequences of bacterial histone-like proteins: HAn of *Anabaena* 7120 (25), HCp of *Clostridium pasteurianum* (26), HBs of *Bacillus stearothermophilus* (26), *E. coli* HU-β (28, 29), *E. coli* HU-α (28, 29), *E. coli* IHF-β (30), *E. coli* IHF-α (31, 32), Tf1 of *Bacillus subtilis* bacteriophage SP01 (33), HRm of *Rhizobium meliloti* (34), and HTa of *Thermoplasma acidophilum* (35). Symbols are as in Fig. 2.

Table 2. Sequence similarity between HlpA and other histone-like proteins

	HAn	HCp	HBs	HU-β	HU-α	IHF-β	IHF-α	Tf1	HRm	HTa
HlpA	53	47	41	37	37	35	29	31	30	25
HlpA-(38–79)	(76)	(68)	(63)	(51)	(46)	(46)	(51)	(39)	(39)	(39)

Numbers shown are percentages of amino acid residues identical between individual pairs of sequences. Calculations are based on the sequence alignment in Fig. 3. Numbers in parentheses are calculated the same way, except that the comparison is limited to sequences between residue 38 and residue 79.

```
                  10        20        30        40        50        60        70        80
AcpA       NEQEIFEKVQTIISEQLGVDKSQ-VTKDANFANDLGADSLDTVELVMAIEEAFNIEIPDDAAEQISNLQQAVDFISQKVAA
E. coli      ST.E.R.KK..G.....KQEE-..DN.S.VE..............L..E.DT....EE..K.TTV.A.I.Y.NGHQ.
A. v.      SQS.T....KK.VI...S.ENPDT..PE.S.....Q..........L..E.D.....E...K.TTV.A...?.N
SP Acp-I   AKK.TID..CD.VK.K.ALGADVV..A.SE.S-K.........I..NL..E.G.NVDE.K.QD..TI...A.V.ESLLEKK
SP Acp-II AAKP.MVT..SD.VKS..ALAEDAK..GETK.S-EI.........I..KL..E.GVTVEEEN.QT.TTI.E.A.M.EALQQNK
A. t.      AAK..TI...SA.VKK..SLTPDKK.VAETK..-..........I..GL..E...QMAEEK.QK.ATVE..AEL.EELINEKK
B. c.      AAKP.TV...SK.VKK..SLKDD.K.VAETK..-..........I..GL..E.D..MAEEK.QK.ATVEE.AEL.EEL.QLKK
```

FIG. 4. Comparison of plastid AcpA protein with other acyl carrier proteins. The amino acid sequence of AcpA, deduced from the DNA sequence, is aligned with sequences of acyl carrier proteins from *E. coli* (36), *Anabaena variabilis* (A. v.) (37), spinach ACP-I (SP Acp-I) (38), spinach ACP-II (SP Acp-II) (39), *Arabidopsis thaliana* (A. t.) (40), and *Brassica campestris* (B. c.) (41). Symbols are as in Fig. 2.

bacterial cells (43). Proteins immunologically related to the *E. coli* DnaK protein have also been detected previously in chloroplasts of plants and *Euglena*, although their sequences and genes (most likely nucleus-encoded) were not studied (44, 45). The *Cryptomonas* plastid DnaK is more similar to *E. coli* DnaK and the mitochondrial hsp70 proteins (yeast Ssc1, *T. cruzi* Mtp70), relative to the eukaryotic cytoplasmic hsp70 proteins, which form a natural group of their own (Table 1). This is consistent with the notion that genes for both the plastid (DnaK) and the mitochondrial (Ssc1 and Mtp70) hsp70 proteins originated from ancient eubacteria through endosymbiosis, although the mitochondrial *ssc1* and *mtp70* genes are now in the nucleus (19, 20). Nothing is known at present about the role of the plastid DnaK protein. Its structural resemblance to other hsp70 proteins, however, suggests similar cellular functions. Both the *E. coli* DnaK protein and the yeast Ssc1 protein are involved in protein translocation across cellular membranes (46–49), most likely by modulating the folding and unfolding of other proteins through protein–protein interactions (50–52). *E. coli* DnaK and *T. cruzi* Mtp70 have been implicated also in DNA replication processes (20, 53–55). The plastid DnaK protein is likely, therefore, to function in plastid protein import as well as other plastid activities such as DNA replication.

The sequence and other structural features (small size, basic charge) of the plastid HlpA protein clearly resemble those of bacterial histone-like proteins (Fig. 3). The HlpA protein is most similar to the histone-like protein of *Anabaena* 7120 (Table 2), consistent with a cyanobacterial origin for the plastid of this organism (2). Most of the sequence similarity is concentrated in a stretch of sequence (residues 38–79) that forms a long arm that binds DNA (26, 27), suggesting a common DNA-binding function of these proteins. Histone-like proteins have previously been extracted from chloroplasts (e.g., see ref. 56) and mitochondria (e.g., see ref. 57), although their sequences and genes (most likely nucleus-encoded) were not studied. Structural resemblance of the plastid HlpA protein to bacterial histone-like proteins suggests similar cellular functions. The *E. coli* histone-like proteins (HU, IHF) have been shown *in vitro* to bind both double- and single-stranded DNAs (58), to mediate very tight DNA curvatures (59), and to introduce negative supercoils into a relaxed closed-circular DNA, resulting in condensed structures resembling nucleosomes (60). *In vivo*, *E. coli* histone-like proteins have been implicated in the coiling of specific DNA sequences as well as in stimulation of transcription, site-specific recombination, and initiation of DNA replication (61–63). In *E. coli* mutant cells lacking the histone-like protein HU, multiple defects are seen, including poor growth, irregular cell cycle, and formation of anucleate cells (64). It is therefore likely that the plastid HlpA protein is involved in organizing the plastid genome as well as in other processes such as DNA replication, DNA recombination, and transcription. It is interesting that chloroplast genomes have previously been observed as nucleoid-like structures attached to the thylakoid membrane (65) and that DNA supercoiling affects *in vitro* transcription of two maize chloroplast genes (66).

Sequence alignment (Fig. 4) clearly identifies the *acpA* gene product as an acyl carrier protein, a key cofactor in the synthesis and metabolism of fatty acids (39). Our finding of the *acpA* gene in the plastid genome of *Cryptomonas* Φ thus identifies lipid biosynthesis as a metabolic pathway involving plastid-encoded enzymes/proteins.

Note Added in Proof. After the submission of this paper, it has come to our attention that an acyl carrier protein gene was found recently in the plastid genome of a marine diatom (67).

We thank Susan E. Douglas for providing the plasmid clone BS-7 used in the study reported here. This work was supported by Operating Grant OGP0041823 from the Natural Sciences and Engineering Research Council of Canada to X.-Q.L.; X.-Q.L. is a Scholar in the Evolutionary Biology Program of the Canadian Institute for Advanced Research.

1. Margulis, L. (1970) *Origin of Eukaryotic Cells* (Yale Univ. Press, New Haven, CT).
2. Gray, M. W. & Doolittle, W. F. (1982) *Microbiol. Rev.* **46**, 1–42.
3. Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., Matsubayashi, T., Zaita, N., Chunwongse, J., Obokata, J., Yamaguchi-Shinozaki, K., Ohto, C., Torazawa, K., Meng, B. Y., Sugita, M., Deno, H., Kamogashira, T., Yamada, K., Kusuda, J., Takaiwa, F., Kato, A., Tohdoh, N., Shimada, H. & Sugiura, M. (1986) *EMBO J.* **5**, 2043–2049.
4. Ohyama, K., Fukuzawa, H., Kohchi, T., Shirai, H., Sano, T., Sano, S., Umesono, K., Shiki, Y., Takeuchi, M., Chang, Z., Aota, S., Inokuchi, H. & Ozeki, H. (1986) *Nature (London)* **322**, 572–574.
5. Hiratsuka, J., Shimada, H., Whittier, R., Shibashi, T., Sakamoto, M., Mori, M., Kondo, C., Honji, Y., Sun, C. R., Meng, B. Y., Li, Y.-Q., Kanno, A., Nishizawa, Y., Hirai, A., Shinozaki, K. & Sugiura, M. (1989) *Mol. Gen. Genet.* **217**, 185–194.
6. Shimada, H. & Sugiura, M. (1991) *Nucleic Acids Res.* **19**, 983–995.
7. Douglas, S. E. & Durnford, D. G. (1989) *Plant Mol. Biol.* **13**, 13–20.
8. Martin, W., Lagrange, T., Li, Y. F., Bisanz, S. C. & Mache, R. (1990) *Curr. Genet.* **18**, 553–556.
9. Smooker, P. M., Kruft, V. & Subramanian, A. R. (1990) *J. Biol. Chem.* **265**, 16699–16703.
10. Baldauf, S. L., Manhart, J. R. & Palmer, J. D. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 5317–5321.
11. Cavalier-Smith, T. S. (1986) *Prog. Phycol. Res.* **4**, 309–347.
12. McFadden, G. I. (1990) *J. Cell Sci.* **95**, 303–308.
13. Douglas, S. E., Murphy, C. A., Spencer, D. F. & Gray, M. W. (1991) *Nature (London)* **350**, 148–151.
14. Douglas, S. E. (1988) *Curr. Genet.* **14**, 591–598.
15. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
16. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY), 2nd Ed., pp. 13.39–13.41.
17. Corpet, F. (1988) *Nucleic Acids Res.* **16**, 10881–10890.
18. Bardwell, J. C. A. & Craig, E. A. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 848–852.
19. Craig, E. A., Kramer, J., Shilling, J., Werner-Washburne, M., Holmes, S., Kosic-Smithers, J. & Nicolet, C. M. (1989) *Mol. Cell. Biol.* **9**, 3000–3008.
20. Engman, D. M., Kirchhoff, L. V. & Donelson, J. E. (1989) *Mol. Cell. Biol.* **9**, 5163–5168.

21. Glass, D. J., Polvere, R. I. & Van der Ploeg, L. H. T. (1986) *Mol. Cell. Biol.* **6,** 4657–4666.
22. Slater, M. J. & Craig, E. A. (1989) *Nucleic Acids Res.* **17,** 805–806.
23. Rochester, D. E., Winer, J. A. & Shah, D. M. (1986) *EMBO J.* **5,** 451–458.
24. Hunt, C. & Morimoto, R. I. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 6455–6459.
25. Nagaraja, R. (1986) Ph.D. thesis (Univ. of Chicago, Chicago).
26. Tanaka, I., Appelt, K., Dijk, J., White, S. & Wilson, K. (1984) *Nature (London)* **310,** 376–381.
27. Drlica, K. & Rouviere, Y. J. (1987) *Microbiol. Rev.* **51,** 301–319.
28. Laine, B., Kmiecik, D., Sautiere, P., Biserte, G. & Cohen-Solal, M. (1980) *Eur. J. Biochem.* **103,** 447–461.
29. Mende, L., Timm, B. & Subramanian, A. R. (1978) *FEBS Lett.* **96,** 395–398.
30. Flamm, E. L. & Weisberg, R. A. (1985) *J. Mol. Biol.* **183,** 117–128.
31. Miller, H. I. (1984) *Cold Spring Harbor Symp. Quant. Biol.* **49,** 691–698.
32. Mechulam, Y., Fayat, G. & Blanquet, S. (1985) *J. Bacteriol.* **163,** 787–791.
33. Greene, J., Brennan, S., Andrew, D., Thompson, C., Richards, S., Heinrikson, R. & Geiduschek, E. P. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 7031–7035.
34. Laine, B., Belaiche, D., Khanaka, H. & Sautiere, P. (1983) *Eur. J. Biochem.* **131,** 325–331.
35. DeLange, R. J., Williams, L. C. & Searcy, D. G. (1981) *J. Biol. Chem.* **256,** 905–911.
36. Vanaman, T. C., Wakil, S. J. & Hill, R. L. (1968) *J. Biol. Chem.* **243,** 6420–6431.
37. Froehlich, J. E., Poorman, R., Reardon, E., Barnum, S. R. & Jaworski, J. G. (1990) *Eur. J. Biochem.* **193,** 817–825.
38. Kuo, T. M. & Ohlrogge, J. B. (1984) *Arch. Biochem. Biophys.* **234,** 290–296.
39. Schmid, K. M. & Ohlrogge, J. B. (1990) *Plant Mol. Biol.* **15,** 765–778.
40. Post-Beittenmiller, M. A., Hlousek-Radojcic, A. & Ohlrogge, J. B. (1989) *Nucleic Acids Res.* **17,** 1777.
41. Rose, R. E., DeJesus, C. E., Moylan, S. L., Ridge, N. P., Scherer, D. E. & Knauf, V. C. (1987) *Nucleic Acids Res.* **15,** 7197.
42. Jaworski, J. G., Post, B. M. & Ohlrogge, J. B. (1989) *Eur. J. Biochem.* **184,** 603–609.
43. Craig, E. A. & Gross, C. A. (1991) *Trends Biochem. Sci.* **16,** 135–140.
44. Marshall, J. S., DeRocher, A. E., Keegstra, K. & Vierling, E. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 374–378.
45. Amir, S. D., Leustek, T., Dalie, B., Weissbach, H. & Brot, N. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 1749–1752.
46. Phillips, G. J. & Silhavy, T. J. (1990) *Nature (London)* **344,** 882–884.
47. Altman, E., Kumamoto, C. A. & Emr, S. D. (1991) *EMBO J.* **10,** 239–245.
48. Kang, P. J., Ostermann, J., Shilling, J., Neupert, W., Craig, E. A. & Pfanner, N. (1990) *Nature (London)* **348,** 137–143.
49. Smith, B. J. & Yaffe, M. P. (1991) *Mol. Cell. Biol.* **11,** 2647–2655.
50. Ostermann, J., Voos, W., Kang, P. J., Craig, E. A., Neupert, W. & Pfanner, N. (1990) *FEBS Lett.* **277,** 281–284.
51. Scherer, P. E., Krieg, U. C., Hwang, S. T., Vestweber, D. & Schatz, G. (1990) *EMBO J.* **9,** 4315–4322.
52. Sheffield, W. P., Shore, G. C. & Randall, S. K. (1990) *J. Biol. Chem.* **265,** 11069–11076.
53. Kawasaki, Y., Wada, C. & Yura, T. (1990) *Mol. Gen. Genet.* **220,** 277–282.
54. Wickner, S. H. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 2690–2694.
55. Wickner, S., Hoskins, J. & Mckenney, K. (1991) *Nature (London)* **350,** 165–167.
56. Briat, J., Letoffe, S., Mache, R. & Rouviere-Yaniv, J. (1984) *FEBS Lett.* **172,** 75–79.
57. Yamada, E. W., Dotzlaw, H. & Huzel, N. J. (1991) *Prep. Biochem.* **21,** 11–23.
58. Bonnefoy, E. & Rouviere-Yaniv, J. (1991) *EMBO J.* **10,** 687–696.
59. Hodges-Garcia, Y., Hagerman, P. J. & Pettijohn, D. E. (1989) *J. Biol. Chem.* **264,** 14621–14623.
60. Broyles, S. & Petitjohn, D. E. (1986) *J. Mol. Biol.* **187,** 47–60.
61. Flashner, Y. & Gralla, J. (1988) *Cell* **54,** 713–721.
62. Pettijohn, D. (1988) *J. Biol. Chem.* **263,** 12793–12796.
63. Schmid, M. B. (1990) *Cell* **63,** 451–453.
64. Huisman, O., Faelen, M., Girard, D., Jaffe, A., Toussaint, A. & Rouviere-Yaniv, J. (1989) *J. Bacteriol.* **171,** 3704–3712.
65. Nemoto, Y., Kawano, S., Nagata, T. & Kuroiwa, T. (1991) *Plant Cell. Physiol.* **32,** 131–141.
66. Stirdivant, S. M., Crossland, L. D. & Bogorad, L. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 4886–4890.
67. Hwang, S.-R. & Tabita, F. R. (1991) *J. Biol. Chem.* **266,** 13492–13494.