Identification and sequence determination of the host factor gene for bacteriophage Q_β

Masayuki Kajitani[1,2] and Akira Ishihama[1]*
[1]Department of Molecular Genetics, National Institute of Genetics, Mishima, Shizuoka 411 and
[2]Basic Research Laboratories, Toray Industries, Inc., Kamakura, Kanagawa 248, Japan

Received December 7, 1990; Revised and Accepted February 5, 1991 DDBJ accession no. D00743

ABSTRACT

The host factor (HF-I) required for phage Q_β RNA-directed synthesis of complementary minus-strand RNA was purified to homogeneity from phage-infected *Escherichia coli* cells. The *hfq* gene encoding HF-I was cloned using synthetic probes designed based on the partial amino acid sequence of HF-I, and mapped at 94.8 min on the *E. coli* chromosome downstream of the *miaA* gene involved in 2-methylthio-N^6-(isopentyl)-adenosine (ms^2i^6A) tRNA modification. Sequence determination of the cloned *hfq* gene indicated that HF-I is a small protein of Mr 11,166 consisting of 102 amino acid residues.

INTRODUCTION

Replication of the bacteriophage RNA genome is catalyzed by the RNA-dependent RNA polymerase or RNA replicase. RNA replicases purified from *Escherichia coli* cells infected with various RNA phages recognize the genome RNA of only the same species (or in some cases RNA from closely related phages) as template for *in vitro* RNA synthesis (1), indicating that the purified enzymes retain the specificity of template selection. This strict template selectivity is needed for the selective replication of only phage RNA in infected *E. coli* cells.

The basic architecture of phage RNA replicase with such template selectivity is composed of four distinct subunit polypeptides, α, β, γ and δ (2,3). Subunit β is encoded by the phage genome, while other subunits are provided by the host bacteria. Subunit α is ribosomal protein S1 (4,5), and subunits γ and δ are translation elongation factors Tu and Ts, respectively (6). For replication of phage Q_β, prototype of group III RNA phages (7), an additional yet poorly characterized host factor, hitherto designated HF-I, is required.

HF-I has been purified to apparent homogeneity from both phage-infected and uninfected *E. coli* cells (8). Purified HF-I is a hexamer of a heat-stable small protein with the molecular weight of 12−13 kDa (kilo dalton) (9), and plays a role in plus-strand (genome polarity) directed synthesis of minus-strand RNA (10). The function of HF-I in uninfected host cells remains unknown.

For a better understanding of the molecular basis of host range for replication of the RNA genome in bacteria, animals and plants, we have been concerned with the identification and characterization of host factors for various RNA viruses. In this report, we describe cloning of the *E. coli* gene (*hfq*) coding for HF-I for phage Q_β, mapping of the *hfq* gene on the *E. coli* chromosome, and determination of its DNA sequence.

MATERIALS AND METHODS

Bacteria, phages and plasmids

Escherichia coli P4X8 (Hfr metB), strain number ME7784 of the Bacterial Stock Center of this institute (NIG), was used for preparation of wild-type Q_β phage. Phage particles were purified by CsCl equilibrium centrifugation. For purification of RNA replicase, *E. coli* Q13 (Hfr met rna pnp tyr) was infected with Q_β am12 phage (a gift from Dr. Y. Fukami) carrying a mutation in the lysis gene (11). For cloning the HF-I gene, plasmid pUC19 was used as a vector, and *E. coli* HB101 was used as a recipient. The ordered λ phage clone library of the *E. coli* W3110 chromosome constructed by Kohara *et al.* (12) was obtained from a stock maintained at the NIG Gene Library Laboratory.

Purification of RNA replicase and HF-I

Q_β RNA replicase was purified essentially according to the method of Eoyang *et al.* (13). HF-I was copurified along with RNA replicase up to the step of liquid polymer phase separation by the purification method of Franze de Fernandez *et al.* (9), but they were separated by chromatography on a DEAE-Toyopearl column; RNA replicase was eluted at 0.15 M NaCl, while HF-I was at 0.10 M. HF-I fractions were pooled, heated at 95°C for 5 min, and centrifuged to remove insoluble materials. The supernatant was subjected to chromatography on a phosphocellulose column previously equilibrated with 50 mM Tris-HCl buffer (pH 7.6 at 4°C) containing 5 mM 2-mercaptoethanol, 1 mM EDTA and 20% glycerol. HF-I was eluted at 0.3 M NaCl. For HPLC on a C18 reverse-phase column, HF-I was solubilized in 30% acetonitrile containing 0.1% trifluoroacetic acid, and eluted from the column by a linear gradient of acetonitrile from 35 to 55%.

* To whom correspondence should be addressed

Assay of RNA replicase activities

Poly(G) polymerase activity was determined by measuring poly(C)-dependent incorporation of [α-^{32}]GMP into acid-insoluble materials. Poly(G) synthesis was carried out for 20 min at 30°C in a reaction mixture, which contained, in a final volume of 50 μl, 100 mM Tris-HCl (pH 7.6 at 30°C), 12 mM $MgCl_2$, 10 mM $MnSO_4$, 4 mM 2-mercaptoethanol, 0.4 mM [α-^{32}P]GTP (1×10^5 cpm/nmol; Amersham, UK), 1.0 μg poly(C) and enzyme. One unit activity was defined as the amount of enzyme which catalyzed the incorporation of 1 nmol GMP under the standard reaction conditions.

Q_β RNA-dependent RNA polymerase activity was determined by measuring incorporation of [α-^{32}P]GMP into acid-insoluble materials. The reaction was carried out for 20 min at 35°C in a reaction mixture, which contained, in a final volume of 50 μl, 100 mM Tris-HCl (pH 7.6 at 35°C), 12 mM $MgCl_2$, 4 mM 2-mercaptoethanol, 0.8 mM each of ATP, UTP and CTP (Pharmacia PL-Biochemicals, USA), 0.4 mM [α-^{32}P]GTP (1×10^5 cpm/nmol; Amersham, UK), 1.0 μg Q_β RNA and 1 unit of Q_β RNA replicase, and in the presence or absence of HF-I. Products were analyzed by agarose gel electrophoresis and autoradiography.

Cloning of HF-I gene

Amino acid sequence of HF-I was determined using an automatic amino acid sequencer (Applied Biosystems, model 470A). Oligonucleotide probes were designed based on the partial amino acid sequence of HF-I, and synthesized using a DNA synthesizer (Applied Biosystems, model 381A). After 5'-terminal labeling with ^{32}P, the probes were hybridized with a DNA filter carrying the mini-set of Kohara's ordered lambda clone library (12), for screening a lambda phage clone which carries the HF-I gene.

For subcloning, the phage DNA which hybridized with the synthetic HF-I probes was digested with various restriction enzymes (Takara Shuzo, Japan) and directly subjected to Southern blot analysis. Hybridized bands were eluted from the gel and cloned into pUC19 vector.

Determination of DNA sequences

DNA sequence was determined by the dideoxy chain termination method (14) using universal primers, RV and M4 (products of Takara Shuzo, Japan). To determine the internal regions, specific internal primers were synthesized on the basis of nucleotide sequences determined in this study.

RESULTS AND DISCUSSION

Purification of HF-I

Replication *in vitro* of Q_β phage RNA requires an *E. coli* protein designated as host factor I (HF-I) besides RNA replicase (8). Purification of HF-I was carried out by measuring stimulation activity of Q_β RNA-dependent RNA synthesis in the presence of excess RNA replicase. HF-I was copurified along RNA replicase up to the step of liquid polymer phase separation according to the procedure of Franze de Fernandez *et al.* (9), but was separated from RNA replicase by chromatography on a DEAE-Toyopearl column. After heat treatment, the soluble fraction of DEAE eluate gave a doublet band on SDS-gel electrophoresis with the apparent molecular weight of 13 kDa (data not shown). Analysis by reverse-phase (RP) HPLC indicated the presence of two components, A and B in this order of elution. Separation of the two components was then achieved by chromatography on a phosphocellulose column. As shown in Fig. 1A, a single peak of HF-I activity was found, but RP-HPLC analysis of PC fractions, shown in Fig. 1(B and C), indicated that the distribution pattern of component B agreed well with that of HF-I activity, whereas component A was eluted at a higher salt concentration, although there was an overlapped region (Fig. 1C).

To confirm our tentative conclusion that component B was HF-I, N-terminal sequences of 26 and 44 amino acid residues were determined for component A and B, respectively (Fig. 2). Previously, Carmichael *et al.* (15) published the N-terminal sequence of 5 amino acid residues for HF-I. This sequence was found in the N-terminal region of component B. On the other hand, the sequence of component A agreed with that of *E. coli*

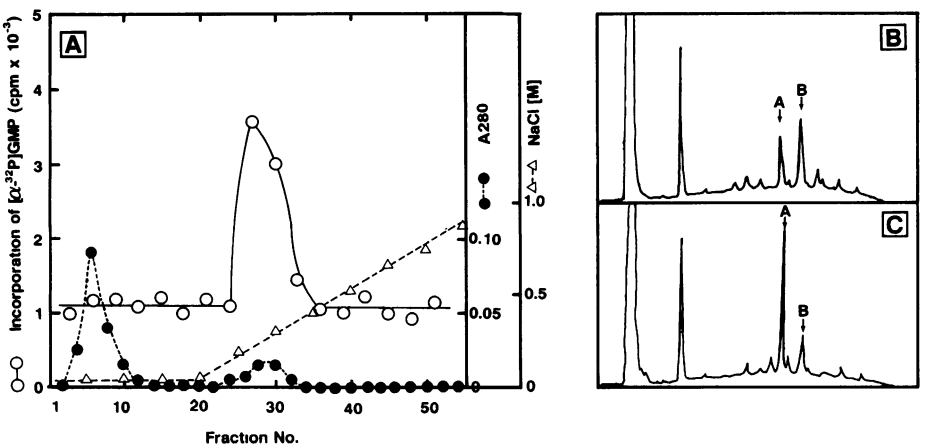

Figure 1. Column chromatograms of HF-I. [A] Phosphocellulose column chromatography. HF-I was purified as described in Materials and Methods. DEAE-Toyopearl fraction was fractionated by phosphocellulose column chromatography. [B] C18 reverse phase-HPLC. Phosphocellulose fraction number 27 with HF-I activity was analyzed by C18 RP-HPLC using 35–55% linear gradient of acetonitrile in the presence of 0.1% trifluoroacetic acid. Peak A, component A (H-NS); peak B, component B (HF-I). [C] C18 reverse phase-HPLC. Phosphocellulose fraction number 30 was analyzed as in [B].

H-NS, a heat-stable histone-like DNA-binding protein (16,17). Isolated H-NS alone has no host-factor activity. Co-purification of HF-I and H-NS might imply their molecular interaction.

Identification of the HF-I Gene

To identify the gene encoding HF-I, we synthesized three kinds of DNA probes with the nucleotide sequences which were capable of encoding peptides indicated in the HF-I sequence (see Fig. 2). All three probes hybridized with clone 652 of the mini-set library of the ordered lambda clones (or clone 3A1 of the original full set library). This clone contains 94.6−95.0 min region of the *E. coli* chromosome (12).

	1 5 10 15 20 25 30 35 40 45
Component A	SEALKILNNIRTLRAQARE*TL*TLE
Component B	AKGQSLQDPFLNALRRERVPVSIYLVNGIKLQGQI*SFDQFVIL
	#1 #2 #3
Published	AKGZS
Probe #1	CARGAYCCITTYYTIAA
#2	AARYTICARGGICARAT
#3	TTYGAYCARTTYGTIAT

Figure 2. Amino-terminal sequence of HF-I. Amino (N)-terminal sequences were determined for both component A (H-NS) and B (HF-I), and compared with that of published N-terminal 5 amino acid residues of HF-I (15). Probes for cloning the *hfq* gene were designed based on the sequences #1, #2 and #3. *, unidentified amino acid residues. I, inosine; R, purine (A plus G); Y, pyrimidine (C plus T).

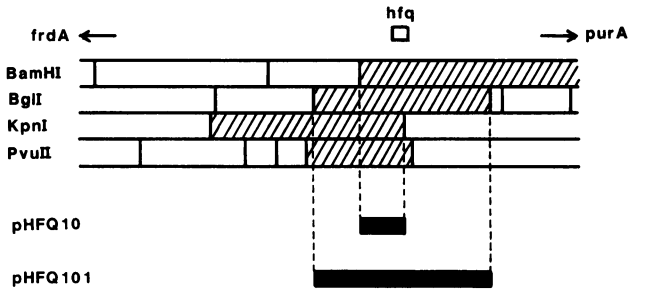

Figure 3. Physical map and cloning of the *hfq* gene. The *hfq* gene encoding HF-I was identified in clone 652 (3A1) of the Kohara's gene library. Cleavage sites for restriction endonuclease, *Bam*HI, *Bgl*I, *Kpn*I and *Pvu*II, agreed with those in clone 652. Shaded fragments hybridized with the HF-I probes. Two DNA fragments, 1.1 kbp *Bam*HI-*Kpn*I and 5,5 kbp *Bgl*I-*Bgl*I fragments, were isolated from the clone, and sub-cloned into pUC19, yielding pHFQ10 and pHFQ101, respectively.

Table 1. Amino acid composition of HF-I

Amino acid	Analyzed	Published*	Amino acid	Analyzed	Published*
Hydrophobic residues			Asn	7	(10)**
Gly	7	8	Gln	9	(13)***
Ala	6	6	Cys	0	–
Val	9	8	*Hydrophilic residues*		
Leu	7	7	Asp	3	(10)**
Ile	5	4	Glu	5	(13)***
Met	2	1	Lys	4	4
Phe	3	3	His	5	4
Trp	0	–	Arg	4	4
Pro	4	4	Tyr	3	3
Neutral residues					
Ser	14	14	Total	102	96
Thr	5	6			

* Carmichael, G. G. *et al.* (15)
** Combined values of Asp and Asn
*** Combined values of Glu and Gln

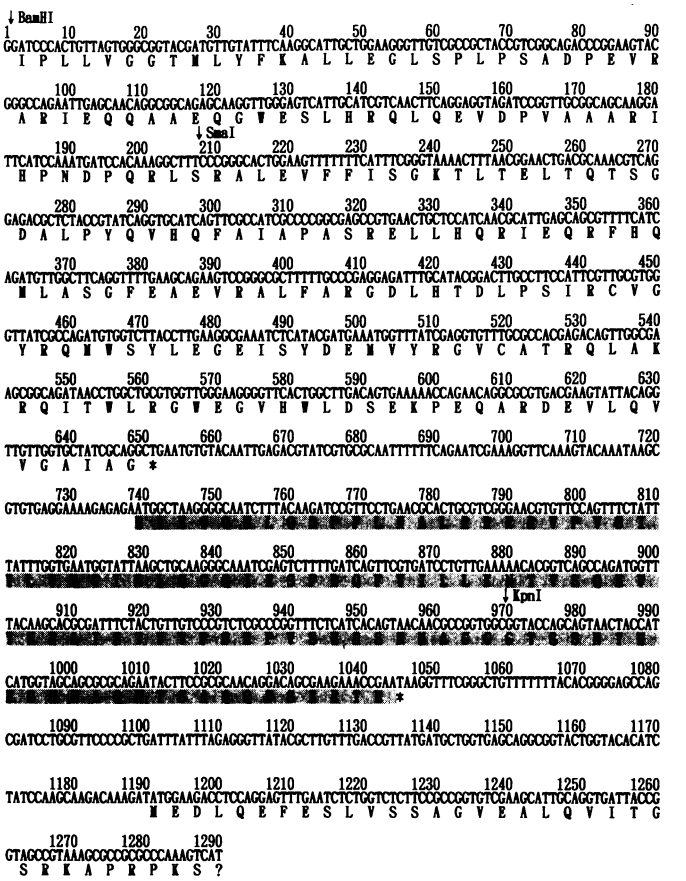

Figure 4. Nucleotide sequence of the *hfq* gene. Nucleotide sequence of the *hfq* gene and its 5′ and 3′ flanking regions was determined. The reading frame of HF-I protein starts at nucleotide 739 and ends at nucleotide 1,044. The sequence from nucleotide 1 to 213 is completely identical with that published by Connolly and Wickler (16) and corresponds to a part of the *miaA* gene.

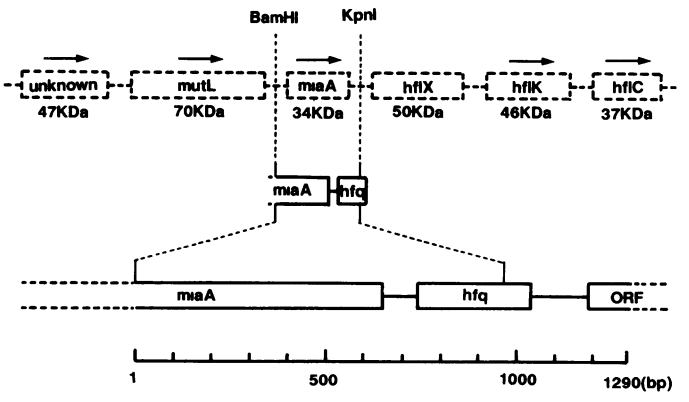

Figure 5. Organization of the *hfq* region of the *E. coli* chromosome. The organization of the *hfq* and its neighboring genes on the *E. coli* chromosome was constructed taking the present and past observations together (5′ region from ref. 18 and 3′ region from ref. 20). For details see text.

For fine mapping, the clone 652 phage DNA was digested with various restriction endonucleases and subjected to Southern hybridization analysis using the same DNA probes used for cloning (data not shown). Results indicated that the *hfq* gene encoding HF-I for Q_b phage is located on the 1.1 kbp *Bam*H1-*Kpn*I fragment, which is derived from 94.8 min region of the *E. coli* chromosome (Fig. 3).

Sequencing of the HF-I gene

Based on the result of fine mapping, we constructed two subclones: pHFQ10 carrying 1.1 kbp *Bam*H1-*Kpn*I fragment in pUC19 previously digested with *Bam*H1 and *Kpn*I; and pHFQ101 carrying 5.5 kbp *Bgl*I fragment in pUC19 at *Sma*I site, respectively (Fig. 3). Using the universal primers, we first sequenced the whole insert of 1.1 kbp in length in pHFQ10. The sequence of N-ternimal two-thirds of HF-I was found in this fragment (Fig. 4). To determine the remaining C-terminal portion of HF-I, we prepared additional primers, which corresponded to the internal regions of HF-I, and sequenced the downstream region from the *Kpn*I site using pHFQ101 as template. Altogether 1,290 bp fragment was sequenced (Fig. 4), including an open reading frame of 102 amino acid residues (nucleotide number 739 to 1,044), which contains the same N-terminal sequence of 44 amino acid residues with that of HF-I protein (component B described above). The amino acid composition of this open reading frame was in good agreement with the published amino acid composition of HF-I (8; see also Table 1). We then concluded that the open reading frame thus determined represents the structural gene (*hfq*) for HF-I.

The *hfq* gene has never been registered in the DNA data bases of EMBL, Gen Bank and DDBJ. However, in upstream region of the *hfq* gene, we found a sequence similar to that of the *miaA* gene (Fig. 5)(18), which codes for an enzyme involved in 2-methylthio-N^6-(isopentyl)-adenosine (ms^2i^6A) tRNA modification. Connolly and Wickler (18) indicated that translation of the *miaA* gene starts immediate downstream of *BamH*I site. However, the translation initiation site might be located further upstream, because the yeast *mod*5 gene product, yeast homologue of the *E. coli miaA* gene, contains additional amino acid sequence at the N-terminus (19). In downstream of the *hfq* gene, there exists the *hflA* locus (20), which controls high frequency lysogenization of phage λ and contains at least three genes, *hflX*, *hflK* and *hflC* in this order (Fig. 5). Since the 5′ proximal *hflX* gene is considered to code for a protein of about 50 KDa (20), the *hfq* gene might be different from the *hflX* gene. It remains to be determined whether the open reading frame starting from nucleotide number 1,191 at downstream of the *hfq* gene represents the *hflX* gene or not. At present, it is also not clear whether there is any functional relationship between the *hfq* gene and the genes present in the upstream or downstream regions.

Analysis of the amino acid sequence of HF-I indicated neither similar proteins hitherto sequenced nor known motifs, but a unique long stretch (about 40% the length of HF-I) rich in neutral hydrophilic amino acid residues was found at the C-terminus. Studies on the physiological function(s) of HF-I in uninfected *E. coli* is in progress.

ACKNOWLEDGEMENTS

We thank Dr. H.Nakamura for advice, Drs. T.Kunitomo and Y.Inokuchi for discussion, and the Toray Research Center for technical assistance. This work was supported in part by Grants-in-Aid from the Ministry of Education, Science and Culture of Japan, the NIG Cooperative Research Programm, the Joint Studies Program of The Graduate University of Advanced Studies, and the Toray Industries, Inc.

REFERENCES

1. Yonesaki, T. and Aoyama, A. (1981) *J. Biochem.* **89**, 751–757.
2. Kamen, R. (1970). *Nature* **228**, 527–533.
3. Kondo, M., Gallerani, R. and Weissmann, C. (1970). *Nature* **228**, 525–527.
4. Groner, Y., Pollack, Y., Berissi, H. and Revel, M. (1972). *Nature New Biol.* **239**, 16–19.
5. Groner, Y., Scheps, R., Kamen, R., Kolakofsky, D. and Revel, M. (1972). *Nature New Biol.* **239**, 19–20.
6. Blumenthal, T., Landers, T. A. and Weber, K. (1982). *Proc. Natl. Acad. Sci. USA* **69**, 1312–1317.
7. Miyake, T., Haruna, I., Shiba, T., Itoh, Y.H., Yamane, K. and Watanabe, I. (1971) *Proc. Natl. Acad. Sci. USA*, **68**, 2022–2024.
8. Shapiro, L., Franze de Fernandez, M. T. and August, T. (1972). *Nature* **220**, 478–480.
9. Franze de Fernandez, M. T., Hayward, W. S. and August, J. T. (1972). *J. Biol. Chem.* **247**, 824–831.
10. Franze de Fernandez, M. T., Eoyang, L. and August, J. T. (1968). *Nature* **219**, 585–590.
11. Horiuchi, K. and Matsuhashi, S. (1970) *Virology*, **42**, 49–60.
12. Kohara, Y., Akiyama, K. and Isono, K. (1987). *Cell* **50**, 495–508.
13. Eoyang, L. and August, J. T. (1968). *Methods in Enzymol.* **12B**, 530–540.
14. Sanger, S., Nicklen, S. and Coulson, A. (1977) *Proc. Natl. Acad. Sci. USA*, 74, 5463–5467.
15. Carmichael, G. G., Weber, K., Niveleau, A. and Wahba, A. J. (1975). *J. Biol. Chem.* **250**, 3607–3612.
16. Falconi, M., Gualtieri, M. T., LaTeana, A., Losso, M. A. and Pon, C. L. (1988). *Molec. Microbiol.* **2**, 323–329.
17. Pon, C. L., Calogero, R. A. and Gualerzi, C. O. (1988). *Mol. Gen. Genet.* **212**, 197–202.
18. Connolly, D. M. and Wickler, M. E. (1989). *J. Bacteriol.* **171**, 3233–3246.
19. Najarian, D., Dihanich, M. E., Martin, N. C. and Hopper, A. K. (1987) *Mol. Cell Biol.* **7**, 185–191.
20. Banuett, F. and Herskowitz, I. (1987). *J. Bacteriol.* **169**, 4076–4085.