J. Biochem. 125, 143-150 (1999)

Cloning of the RNA Polymerase α Subunit Gene from *Thermus thermophilus* HB8 and Characterization of the Protein[1]

Takashi Wada,* Toshio Yamazaki,* Seiki Kuramitsu,† and Yoshimasa Kyogoku*,[2]

**Institute for Protein Research, Osaka University, Suita, Osaka 565-0871; and* †*Department of Biology, Graduate School of Science, Osaka University, Toyonaka, Osaka 560-0043*

Received September 4, 1998; accepted October 3, 1998

The region containing the RNA polymerase α subunit (RNAPα) gene (*rpoA*) and the ribosomal protein genes of a thermophilic eubacterial strain, *Thermus thermophilus* (*Tt*) HB8, was cloned from a genomic DNA library by Southern hybridization. The gene order in this region is *rpl36-rps13-rps11-rps4-rpoA-rpl17*, which is identical to that in some other eubacteria. The *rpoA* gene encodes a 315 amino acid residue protein with a molecular weight of 35,013, the amino acid sequence showing 42% identity to that of *Escherichia coli* (*Ec*). From the results of comparison of the amino acid sequence and the predicted secondary structure of the C-terminal domain of *Tt* RNAPα (*Tt* αCTD) with those of *Ec*, the overall folding is expected to be similar. However, amino acid residues Asn268 and Cys269 in *Ec* αCTD, which are essential for its interaction with DNA or regulatory proteins, were replaced by His and Ser, respectively, in *Tt* αCTD. By means of a T7-based expression system in *Ec* cells, *Tt* RNAPα was overexpressed and purified. The high thermostability of *Tt* RNAPα was demonstrated by the CD spectra.

Key words: gene cloning, ribosomal protein, RNA polymerase α subunit, thermostability, *Thermus thermophilus*.

A eubacterial RNA polymerase holoenzyme is composed of five subunits, $\alpha_2\beta\beta'\sigma$. As seen on the investigation of *Escherichia coli* (*Ec*), core enzyme $\alpha_2\beta\beta'$ is fully active in the polymerization of RNA, whereas the binding of different species of the σ subunit is required for the specific initiation from their promoters. The α subunit (RNAPα) is composed of two structural domains. The N-terminal domain of RNAPα (αNTD) plays a key role in RNA polymerase assembly (*1-4*), and the C-terminal domain (αCTD) is necessary for transcription regulation. The latter interacts with transcription activators, for example, cyclic AMP receptor protein (CRP) (*2, 5*), OmpR (*6*), and a transcription repressor, GalR (*7*). Furthermore, αCTD recognizes promoter upstream (UP) elements, which consist of AT-rich sequences, and enhances transcription initiation (*8*). Two αCTDs bind in tandem to the *rrnB* P1 UP element (*9*). After the three-dimensional structure of αCTD was determined by Jeon *et al.* (*10*), it was proposed that the same protein surface on RNAPα was involved in the contact with both CRP and the UP element (*11*).

For structural analysis of RNAPα and αNTD, RNAPα derived from a thermophilic bacterium would be more stable and suitable than that of *Ec*. In addition, comparison of the structures of αCTDs from a thermophilic bacterium and *Ec* may provide information on the thermostability and biological variety of αCTDs. Thus, we carried out gene cloning of an extremely thermophilic bacterium, *Thermus thermophilus* (*Tt*) HB8, and identified the *rpoA* gene. *Tt* RNAPα was overexpressed in *Ec* and purified, and its thermostability was characterized by CD measurement.

EXPERIMENTAL PROCEDURES

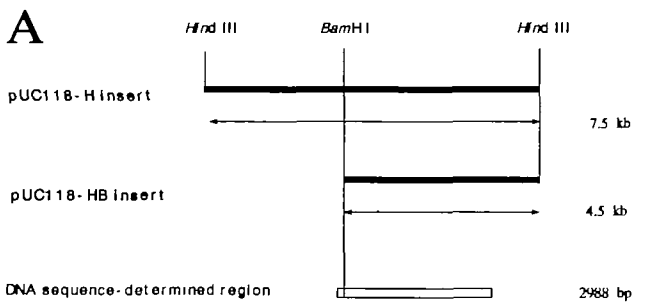
Bacterial Strains and Materials—The *Ec* strains used here were DH5α for plasmid DNA manipulation and BL21 (DE3) for the overexpression of *Tt* RNAPα. They were grown in LB medium. *Tt* genomic DNA was obtained by the procedure described previously (*12*). The enzymes, reagents, and chromatographic materials were purchased from commercial sources.

Design of Oligonucleotide Primers and Polymerase Chain Reaction—Highly conserved regions of ribosomal proteins, S11 and RNAPα, were chosen for the synthesis of oligonucleotide primers for the polymerase chain reaction (PCR). The mixed primers for S11 (KF1) were 5′-C(G/C)-TACATCCACGC(G/C)(A/T)C(G/C)TACAACAA-3′ (25 mer), and the mixed primers for RNAPα (R3) were 5′-(G/

[1] This work was supported by a Grant-in-Aid from the Ministry of Education, Science, Sports and Culture of Japan (09480176). The *rpl36, rps13, rps11, rps4,* and *rpl17* genes are named *rpmJ, rpsM, rpsK, rpsD,* and *rplQ* depending on the species. In this paper, the former naming is used throughout.

[2] To whom correspondence should be addressed. E-mail:kyogoku@protein.osaka-u.ac.jp

Abbreviations: *Aa, Aquifex aeolicus;* αCTD, C-terminal domain of RNA polymerase α subunit; αNTD, N-terminal domain of RNA polymerase α subunit; *Bb, Borrelia burgdorferi; Bp, Bordetella pertussis; Bsp, Bacillus* sp.; *Bsu, Bacillus subtilis; Ct, Chlamydia trachomatis; Hi, Haemophilus influenzae; Hp, Helicobacter pylori; Ec, Escherichia coli;* IPTG, isopropylthio-β-D-galactoside; LB, Luria-Bertani; *Mg, Mycoplasma genitalium; Mp, Mycoplasmapneumoniae; Mt, Mycobacterium tuberculosis;* RNAPα, RNA polymerase α subunit; *Sc, Streptomyces coelicolor; Sh, Shewanella* sp.; *Syo, Synechococcus* sp.; *Syy, Synechocystis* sp.; *Tt, Thermus thermophilus.*



B

```
CGACCCCACGCGGGGCCGGATCGTTTACCGCAAGTAGGAGGCAGGCATGAAGGTGCGCGC   60
                                          M  K  V  R  A
                                          L36 start
                  BamHI
GTCGGTCAAGAGGATCTGCGACAAGTGCAAGGTGATCCGCCGGCACGGGCGGGTCTACGT  120
S  V  K  R  I  C  D  K  C  K  V  I  R  R  H  G  R  V  Y  V

CATCTGCGAGAACCCCAAGCATAAGCAGCGGCAGGGTTAGGGAGGAGGACTGTGGCGAGG  180
I  C  E  N  P  K  H  K  Q  R  Q  G  *              M  A  R
                                                    S13 start

ATCGCAGGGGTAGAGATTCCCAGGAACAAGCGGGTGGACGTCGCCCTCACCTACATCTAC  240
I  A  G  V  E  I  P  R  N  K  R  V  D  V  A  L  T  Y  I  Y

GGCATCGGCAAGGCCCGGGCCAAGGAGGCCCTGGAGAAGACGGGGATCAACCCCGCCACC  300
G  I  G  K  A  R  A  K  E  A  L  E  K  T  G  I  N  P  A  T

CGGGTGAAGGACCTCACCGAGGCGGAGGTGGTCCGCCTCCGGGAGTACGTGGAGAACACC  360
R  V  K  D  L  T  E  A  E  V  V  R  L  R  E  Y  V  E  N  T

TGGAAGCTGGAGGGGGAGCTCAGGGCCGAGGTGGCGGCCAACATCAAGCGCCTCATGGAC  420
W  K  L  E  G  E  L  R  A  E  V  A  A  N  I  K  R  L  M  D

ATCGGCTGCTACCGGGGCCTCCGGCATCGGCGGGGGCTGCCGGTGCGGGGCCAGCGGACC  480
I  G  C  Y  R  G  L  R  H  R  R  G  L  P  V  R  G  Q  R  T

CGCACCAACGCCCGCACCCGCAAGGGGCCCCGCAAGACGGTGGCGGGCAAGAAGAAGGCC  540
R  T  N  A  R  T  R  K  G  P  R  K  T  V  A  G  K  K  K  A

CCGAGGAAGTGAGCCTGAGGGCCAGAGGATACTCCGGAGAGGGAGAGTATGGCCAAGAAG  600
P  R  K  *                          M  A  K  K
                                    S11 start
                        KF1
CCGAGCAAGAAAAAGGTCAAGCGGCAGGTGGCGAGCGGTCGCGCCTACATCCACGCCTCC  660
P  S  K  K  K  V  K  R  Q  V  A  S  G  R  A  Y  I  H  A  S

TACAACAACACCATCGTCACCATCACCGACCCGGACGGCAACCCCATCACCTGGTCTTCG  720
Y  N  N  T  I  V  T  I  T  D  P  D  G  N  P  I  T  W  S  S

GGCGGCGTCATCGGCTACAAGGGAAGCCGTAAGGGCACCCCTTACGCCGCCCAGCTCGCG  780
G  G  V  I  G  Y  K  G  S  R  K  G  T  P  Y  A  A  Q  L  A

GCCCTGGACGCCGCCAAGAAGGCCATGGCCTACGGCATGCAGAGCGTGGACGTGATCGTG  840
A  L  D  A  A  K  K  A  M  A  Y  G  M  Q  S  V  D  V  I  V

CGGGGCACCGGGGCGGGCCGGGAGCAGGCCATAAGGGCCCTCCAGGCCTCCGGCCTCCAG  900
R  G  T  G  A  G  R  E  Q  A  I  R  A  L  Q  A  S  G  L  Q

GTGAAGTCCATCGTGGACGACACCCCCGTCCCCCACAACGGCTGCAGCCCCAAGAAGAAG  960
V  K  S  I  V  D  D  T  P  V  P  H  N  G  C  R  P  K  K  K

TTCCGTAAGGCTTCCTAGAAGGAGTGAGCGATGGGTCGTTACATTGGTCCAGTTTGCCGT 1020
F  R  K  A  S  *                    M  G  R  Y  I  G  P  V  C  R
                                    S4 start

CTTTGCCGCCGGGAAGGCGTGAAGCTCTACCTCAAGGGGGAGCGGTGCTACAGCCCCAAG 1080
L  C  R  R  E  G  V  K  L  Y  L  K  G  E  R  C  Y  S  P  K
   S?                                       D

TGCGCCATGGAGCGCCGCCCCTACCCCCCTGGGCAGCACGGGCAGAAGCGGGCGCGCCGC 1140
C  A  M  E  R  R  P  Y  P  P  G  Q  H  G  Q  K  R  A  R  R

CCCTCCGACTACGCGGTGCGGCTTAGGGAGAAGCAGAAGCTCCGCCGGATCTACGGCATC 1200
P  S  D  Y  A  V  R  L  R  E  K  Q  K  L  R  R  I  Y  G  I

TCCGAGCGCCAGTTCCGCAACCTGTTTGAGGAGGCGAGCAAGAAGAAGGGCGTCACGGGC 1260
S  E  R  Q  F  R  N  L  F  E  E  A  S  K  K  K  G  V  T  G
```

```
TCCGTCTTCCTGGGGCTTTTGGAGTCCCGTCTGGACAACGTGGTCTACCGGCTGGGCTTC 1320
S  V  F  L  G  L  L  E  S  R  L  D  N  V  V  Y  R  L  G  F

GCCGTAAGCCGCCCCCAGGCCCGCCAGCTGGTGCGCCACGGGCACATCACCGTGAACGGG 1380
A  V  S  R  R  Q  A  R  Q  L  V  R  H  G  H  I  T  V  N  G

CGCCGGGTGGACCTCCCCTCCTACCGGGTCCGGCCCGGGGACGAGATCGCCGTGGCCGAG 1440
R  R  V  D  L  P  S  Y  R  V  R  P  G  D  E  I  A  V  A  E

AAGAGCCGCAACCTCGAGCTCATCCGCCAGAACCTCGAGGCCATGAAGGCCGGAAGGTG 1500
K  S  R  N  L  E  L  I  R  Q  N  L  E  A  M  K  G  R  K  V

GGCCCCTGGCTCTCCCTGGACGTGGAGGGCATGAAGGGCAAGTTCCTCCGCCTGCCCGAC 1560
G  P  W  L  S  L  D  V  E  G  M  K  G  K  F  L  R  L  P  D

CGGGAGGACCTGGCCCTGCCCGTGAACGAGCAGCTGGTGATCGAGTTCTACTCCAGGTAA 1620
R  E  D  L  A  L  P  V  N  E  Q  L  V  I  E  F  Y  S  R  *

GCCCCTGGAAGGAGGCGTATGTTGGATTCCAAGCTCAAGGCCCCGGTCTTCACGGTGCGC 1680
                M  L  D  S  K  L  K  A  P  V  F  T  V  R
                RNAP$_{II}$ start

ACCCAGGGGCGGGAGTACGGGGAGTTCGTCCTGGAGCCCCTGGAGCGGGGCTTCGGCGTC 1740
T  Q  G  R  E  Y  G  E  F  V  L  E  P  L  E  R  G  F  G  V

ACCCTGGGAAACCCCCTGCGGCGCATCCTCCTTTCCTCCATCCCTGGGACCGCGGTCACC 1800
T  L  G  N  P  L  R  R  I  L  L  S  S  I  P  G  T  A  V  T

AGCGTCTACATTGAGGACGTCCTCCACGAGTTCTCCACCATCCCCGGGGTCAAGGAGGAC 1860
S  V  Y  I  E  D  V  L  H  E  F  S  T  I  P  G  V  K  E  D

GTGGTGGAGATCATCCTCAACCTCAAGGAGCTGGTGGTCCGGTTCCTGAACCCCAGCCTC 1920
V  V  E  I  I  L  N  L  K  E  L  V  V  R  F  L  N  P  S  L

CAGACCGTGACCCTCCTCCTCAAGGCCGAGGGCCCCAAGGAGGTGAAGCGCGGGACTTC 1980
Q  T  V  T  L  L  L  K  A  E  G  P  K  E  V  K  A  R  D  F

CTTCCCGTGGCCGACGTGGAGATCATGAACCCCGACCTCCACATCGCCACCCTGGAGGAG 2040
L  P  V  A  D  V  E  I  M  N  P  D  L  H  I  A  T  L  E  E

GGGGGTCGGCTCAACATGGAGGTCCGGGTGGACCGGGGGGTGGGGTACGTCCCGGCGGAG 2100
G  G  R  L  N  M  E  V  R  V  D  R  G  V  G  Y  V  P  A  E

AAGCACGGCATCAAGGACCGGATCAACGCCATCCCGGTGGACGCCGTCTTCTCCCCGGTG 2160
K  H  G  I  K  D  R  I  N  A  I  P  V  D  A  V  F  S  P  V

CGCCGCGTGGCCTTCCAGGTGGAGGACACCCGCCTGGGGCAGCGCACCGACCTGGACAAG 2220
R  R  V  A  F  Q  V  E  D  T  R  L  G  Q  R  T  D  L  D  K

CTCACCCTCAGGATCTGGACCGACGGCTCCGTCACCCCCTCGAGGCCCTGAACCAGGCG 2280
L  T  L  R  I  W  T  D  G  S  V  T  P  L  E  A  L  N  Q  A

GTGGAGATCCTCCGGGAGCACCTCACCTACTTCTCCAACCCCAGGCGGCCGCCGTGGCG 2340
V  E  I  L  R  E  H  L  T  Y  F  S  N  P  Q  A  A  A  V  A

GCCCCCGAGGAGGCCAAGGAGCCCGAGGCGCCGCCAGAGCAGGAGGAGGAGCTGGACCTG 2400
A  P  E  E  A  K  E  P  E  A  P  P  E  Q  E  E  E  L  D  L
                                                    R3

CCCTTGGAGGAGCTCGGCCTCTCCACCCGGGTGCTCCACAGCCTCAAGGAGGAGGGCATT 2460
P  L  E  E  L  G  L  S  T  R  V  L  H  S  L  K  E  E  G  I

GAGTCCGTCCGGGCCCTTCTCGCCCTCAACCTCAAGGACCTGAAGAACATCCCCGGGATT 2520
E  S  V  R  A  L  L  A  L  N  L  K  D  L  K  N  I  P  G  I

GGGGAGCGGAGCCTGGAGGAGATCAAGGAAGCCCTGGAGAAGAAGGGCTTCACCCTGAAG 2580
G  E  R  S  L  E  E  I  K  E  A  L  E  K  K  G  F  T  L  K

GAGTGAGACCATGCGCCACCTGAAGTCCGGACGTAAGCTGAACCGCCACTCTTCCCACCG 2640
E  *     M  R  H  L  K  S  G  R  K  L  N  R  H  S  S  H  R
         L17 start

CTTGGCCCTTTACCGCAACCAGGCGAAAAGCCTCCTCACCCACGGCCGCATCACCACCAC 2700
L  A  L  Y  R  N  Q  A  K  S  L  L  T  H  G  R  I  T  T  T

CGTGCCCAAGGCCAAGGAGCTCAGGGGGTTTGTGGACCACCTCATCCACCTGGCCAAGCG 2760
V  P  K  A  K  E  L  R  G  F  V  D  H  L  I  H  L  A  K  R

GGGCGACCTCCACGCCCGCAGGTTGGTCCTTAGGGACCTCCAGGACGTGAAGCTGGTGCG 2820
G  D  L  H  A  R  R  L  V  L  R  D  L  Q  D  V  K  L  V  R

GAAGCTCTTTGACGAGATCGCCCCCCGCTACCGGGACCGCCAGGGGGGGTACACCCGGGT 2880
K  L  F  D  E  I  A  P  R  Y  R  D  R  Q  G  G  Y  T  R  V

CCTCAAGCTGGCGGAGCGCCGGCGGGGGGACGGGGCGCCGCTCGCCCTGGTGGAGCTGGT 2940
L  K  L  A  E  R  R  R  G  D  G  A  P  L  A  L  V  E  L  V

GGAGTAGCCCCGCCGGGACCGGCCCCCCCCGGAGCCTCGGCCCGGGGGG 2988
E  *
```

Fig. 1

C)GCCTC(G/C)GCCTT(G/C)AGGCAGTT-3′ (21 mer) (Fig. 1B). PCR was carried out under the conditions described previously (*12*). An amplified DNA fragment was used as a probe for Southern and colony hybridization. The nucleotide sequence of the amplified DNA fragment was determined. It was confirmed that the translated amino acid sequence included the conserved regions of RNAPα.

Cloning and Sequencing of the Tt rpoA Gene—*Tt* genomic DNA was digested with a restriction endonuclease, *Hin*dIII. Southern hybridization was carried out using a DIG-labeled PCR DNA fragment and a DIG detection system (Boehringer Mannheim Biochemica). An 8 kb fragment was inserted into the pUC 118 vector. From this mini DNA library, a plasmid containing the target insert (pUC 118-H) was obtained by colony hybridization (Fig. 1A). Furthermore, a *Hin*dIII-*Bam*HI digested fragment (4.5 kb, pUC 118-HB insert) was subcloned (Fig. 1A). A nucleotide sequence of about 3 kb of pUC 118-HB and pUC 118-H was determined on both strands by the dideoxy termination method using a *Taq* cycle sequencing system and DNA auto-sequencers (Applied Biosystems and LI-COR) (Fig. 1B). pET-*rpoA* for overexpression was obtained by inserting a PCR fragment of the *Tt rpoA* gene into the pET11 vector (Novagen), and then sequenced for identification.

Overproduction and Purification of Recombinant Tt RNAPα—*Tt* RNAPα was overexpressed in *Ec* BL21 (DE3) cells transformed with pET-*rpoA*, as described previously (*13*). To a liquid culture, IPTG was added to 1 mM at $A_{620}=0.8$. After 3 h induction the cells were harvested and stored at −80°C until use.

The following procedure, except for the heat treatment, was carried out at 4°C. Wet cells from 600 ml of culture were suspended in 50 ml of deionized water and then sonicated on ice. The crude extract was then incubated at 65°C for 30 min and centrifuged. The supernatant was recovered, and ammonium sulfate was added to it to 50% saturation. After centrifugation, the precipitate was resuspended in 10 ml of the gel-filtration buffer [50 mM K_2HPO_4/KH_2PO_4 (pH 6.8), 500 mM KCl]. The suspension was applied to a Superdex75 gel-filtration column (2.6×60 cm) (Pharmacia), and proteins were eluted with the gel-filtration buffer. The peak fractions of *Tt* RNAPα, as judged on SDS-PAGE, were pooled. The pooled fractions were dialyzed against deionized water and stored at 4°C after lyophilization.

Spectroscopy—The molar extinction coefficient, ε_M, of *Tt* RNAPα determined by the reported method (*14, 15*) was 1.1×10^4 $M^{-1}\cdot cm^{-1}$ at 280 nm and pH 6.0. The concentration of *Tt* RNAPα was determined spectrometrically using the ε_M value.

Fig. 1. **(A) The inserted DNA fragments of pUC118-H and pUC118-HB, and the position of the determined sequence. (B) The determined nucleotide sequence and the translated amino acid sequences.** The nucleotide sequence numbers are given on the right. The two arrows indicate the hybridized positions in the *Tt* genomic DNA, and the elongation direction with the KF1 and R3 DNA primers on PCR. The underlined amino acids denote that the translated amino acid sequence of each protein was identical with that determined on N-terminal amino acid analysis (*17*). The bold letters in S4 indicate amino acid analysis results incompatible with the translated sequence. The underlined nucleotide sequence indicates the *Bam*HI site, whose sequence is different from the usual recognition sequence of *Bam*HI (GGATCC).

The circular dichroic (CD) spectra of *Tt* RNAPα, at 18 μM protein concentration in a 0.1 cm cell, were measured with a Jasco J-720 spectropolarimeter. The buffer solution comprised 20 mM K_2HPO_4/KH_2PO_4 (pH 6.0), 30 mM KCl, and 1 mM DTT. Thermal stability was monitored by measuring the temperature dependence of the molar ellipticity at 222 nm with a heating rate of 0.5°C/min.

RESULTS

Identification of the Tt rpoA Gene and Tt Ribosomal Protein Genes—Six open reading frames were found in the sequence-determined region (Fig. 1, A and B). The order of these genes was *rpl36-rps13-rps11-rps4-rpoA-rpl17*, which was identical with those of *Ec, Hi, Hp*, and *Mt* among eubacterial strains. The G+C content of the determined sequence was 67.6%, which coincides well with that of *Tt* genomic DNA, 69% (*16*). Incidentally, the nucleotide sequence-determined region has a *Bam*HI site, whose nucleotide sequence is TGATCC, which is different from the usual recognition sequence of *Bam*HI (GGATCC) (Fig. 1B). This may be caused by the star activity of *Bam*HI. Alignment of the amino acid sequences of RNAPα is shown in Fig. 2. The identities of the amino acid sequences of the encoded proteins to those of the corresponding proteins of *Ec* were 63% (L36), 58% (S13), 54% (S11), 53% (S4), 55% (L17), and 42% (RNAPα). The N-terminal amino acid sequences of *Tt* L36, S13, S11, and S4 had already been determined (Fig. 1B, underlined sequences) (*17* and GenBank accession No. P80256), and the results were identical with the translated amino acid sequences except for two amino acids of S4 (shown in bold letters in Fig. 1B).

Overexpression and Purification of Tt RNAPα—By means of a T7-based expression system in *Ec*, *Tt* RNAPα was overexpressed and purified (Fig. 3). Although the calculated molecular weight of *Tt* RNAPα is 35,013, there was a trace at 42 k on SDS-PAGE. This value is close to those obtained by Date *et al.* (42 k), and Wnendt *et al.* (40 k), who purified the *Tt* RNAP holoenzyme from *Tt* (*18, 19*). Through the expression and purification procedure, 25 mg of *Tt* RNAPα was obtained from 600 ml of LB medium culture.

Thermostability of Tt RNAPα—The UV CD pattern around 220 nm of *Tt* RNAPα at 30°C (Fig. 4A) shows that *Tt* RNAPα mainly consists of α-helical structures. The thermostability of *Tt* RNAPα was characterized by a change in ellipticity at 222 nm from 20 to 95°C (Fig. 4B). Although *Ec* RNAPα was denatured at about 60°C, *Tt* RNAPα retained the native structure up to 85°C.

DISCUSSION

The α operon of *Ec* consists of five genes, *rps13* to *rpl17*, and the DNA sequence of *Tt* determined here contains all of them. However, an apparent promoter sequence was not found upstream of the *Tt rps13* gene, and the *rpl36* gene directly precedes it. Thus the *Tt* operon containing the *rpoA* gene might have a different structure from the α operon of *Ec*. There is a 35 bp sequence between the *rps13* and *rps11* genes, and this region might be related to the gene regulation of the *Tt* operon.

The *rps4* gene is located in front of the *rpoA* gene in the genomes of *Aa* (*20*), *Ec* (*21*), *Hi* (*22*), *Hp* (*23*), *Mt* (*24*),

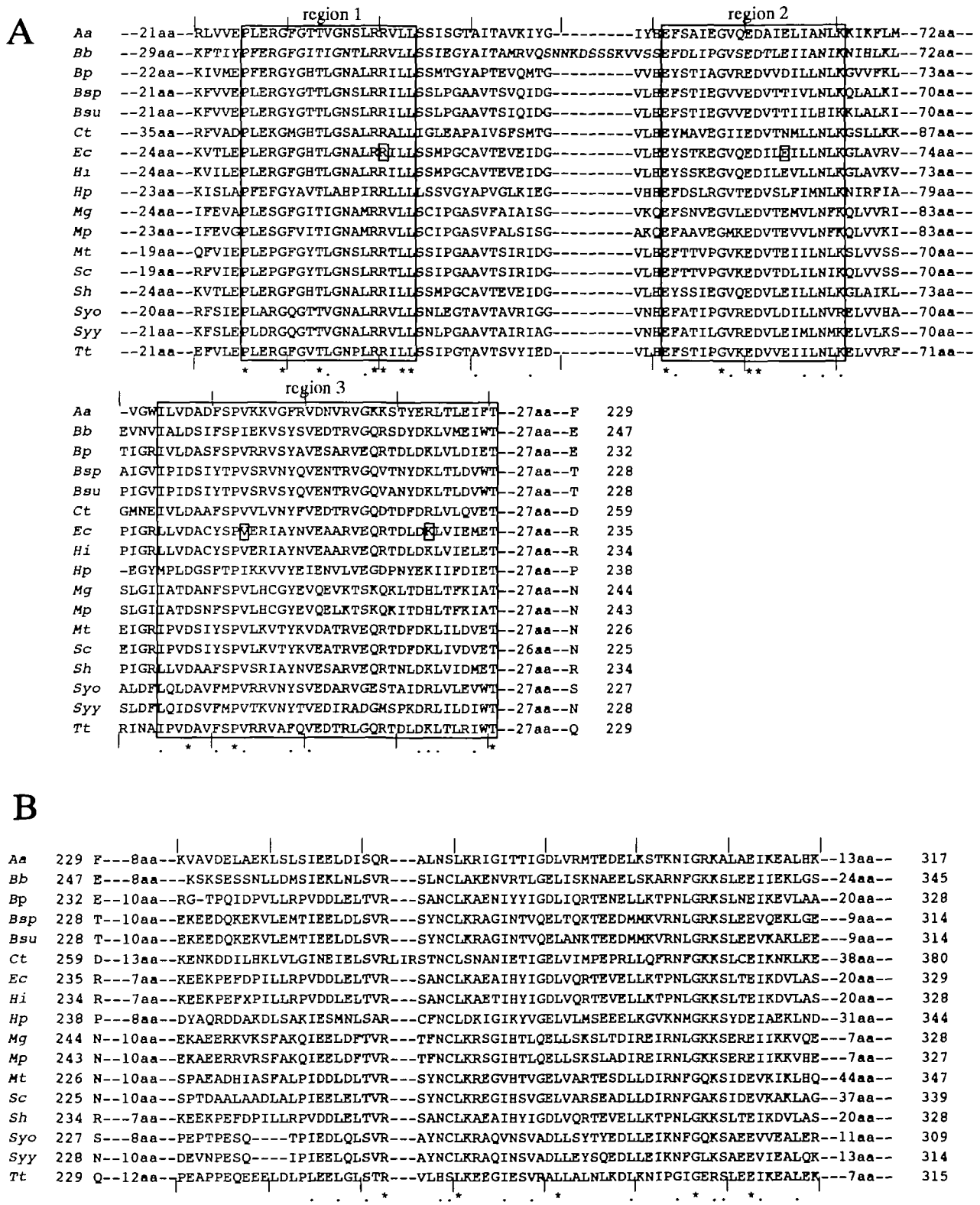


Fig. 2. **Multiple amino acid sequence alignment (CLUSTALW) (*40*) of (A) αNTDs, and (B) αCTDs of *Aa* (*20*), *Bb* (*26*), *Bp* (*41*), *Bsp* (GenBank accession No. AB010082), *Bsu* (*25*), *Ct* (*27*), *Ec* (*21*), *Hi* (*22*), *Hp* (*23*), *Mg* (*28*), *Mp* (*29*), *Mt* (*24*), *Sc* (*42*), *Sh* (*43*), *Syo* (*30*), *Syy* (*31*), and *Tt* (this paper).** The αNTDs and αCTDs are defined by the position of Arg235 of *Ec* RNAPα. The numbers on the right are the total amino acids of individual proteins. The numbers in the sequences indicate the numbers of nonidentical amino acid residues. Vertical lines indicate every ten amino acids. Dashes indicate gaps inserted to optimize the alignment. Asterisks and dots indicate identical and homologous amino acid residues, respectively. The boxed areas in (A) are conserved regions of the amino acid sequence. The boxed letters in regions 1, 2, and 3 in (A) show Arg45, Glu80, Val180, and Lys200, respectively (see the text).

and *Tt*. On the other hand, the gene order in *Bsp* (GenBank accession No. AB010082), *Bsu* (*25*), *Bb* (*26*), *Ct* (*27*), *Mg* (*28*), *Mp* (*29*), *Syo* (*30*), and *Syy* (*31*) is *rps13-rps11-rpoA-rpl17*, the *rps4* gene being lacking in the region. The complete genome sequences of *Bsu, Bb, Mg, Mp*, and *Syy* were determined, and the *rps4* gene was found to exist at another site in their genomes. In *Ec*, ribosomal protein S4

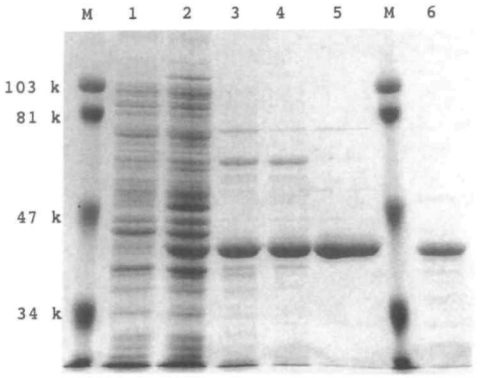


Fig. 3. **SDS-PAGE pattern of *Tt* RNAPα at various steps of purification.** Before induction (lane 1), and after induction by the addition of IPTG (lane 2). After cell sonication, heat treatment and removal of debris by centrifugation, *Tt* RNAPα was rather purified (lane 3). After 50% ammonium sulfate precipitation of *Tt* RNAPα (lane 4), it was loaded onto a Superdex75 gel-filtration column (Pharmacia) and eluted with the gel-filtration buffer (lane 5). Although the calculated molecular weight of *Tt* RNAPα is 35,013, it showed a molecular weight of about 42 k on 10% SDS-PAGE, which is almost the same as that of *Ec* RNAPα (lane 6). Lane M contains molecular weight markers (Bio Rad).

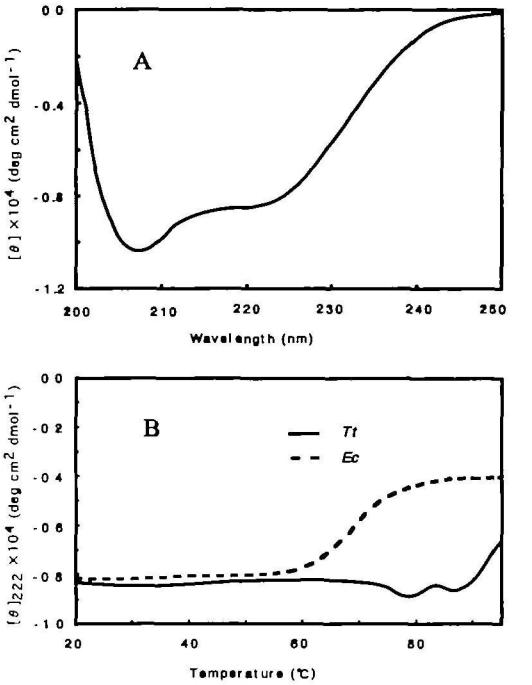


Fig. 4. **UV CD spectra of *Tt* RNAPα.** (A) CD spectrum of *Tt* RNAPα at 30°C. (B) Thermal dependence of the ellipticity at 222 nm for *Tt* RNAPα and *Ec* RNAPα. *Ec* RNAPα was prepared as described previously (*2*), and its buffer content on CD measurement was the same as that of *Tt*.

```
predicted   HHHHHHHHHHH       HHHHHHHHHHHH    HHHHHHH     HHHHHHHHH
            231       240       250       260       270       280
Tt          AAVAAPEEAKEPEAPPEQEEELDLPLEELGLSTRVLHSLKEEGIESVRALLA
              *  **   *    ** .  *  *...* *. *  ..** * *  .  *.
Ec          RDVRQPEVKEEK---PEFDPILLRPVDDLELTVRSANCLKAEAIHYIGDLVQ
            235  240          250       260  ▲  ▲▲270       280
observed                      hhhh  hhh    HHHHHHHHHH    HHHHHH
predicted   HHHH   HHHH                 HHHHHH         EHEH  HHH

predicted   HHHHHHH          HHHHHHHHHH
                 290       300       310
Tt          LNLKDLKHIPGIGERSLEEIKEALEKKGFTLKE----------
             .  .*   *..* .** ***. *  .* .*
Ec          RTEVELLKTPNLGKKSLTEIKDVLASRGLSLGMRLEWPPASIADE
                 290       300       310       320
observed      HHHHHHH     HHHHHHHHHHHH
predicted   HHHHHHHH         HHHHHHHHHH       HHHH
```

Fig. 5. **Comparison of the primary sequences and the predicted secondary structures between *Tt* αCTD and *Ec* αCTD.** The prediction of the secondary structure was based on the 3D-1D (threading) method (*44*). The observed secondary structure of *Ec* αCTD was that determined by NMR analysis (*10*). Asterisks and dots indicate identical and homologous amino acid residues, respectively. H and E indicate the predicted alpha-helix and beta-strand, respectively. Bold H and h denote the observed alpha-helix and helical-turn, respectively. The numbers are the total amino acids of individual proteins. Arg265, Asn268, and Cys269, which are important for the interaction with CRP or the UP element, in *Ec* αCTD are indicated by arrows.

TABLE I. **Comparison of the contents of Pro residues of RNAPαs of eubacteria.**

Species of RNAPα	*Aa*	*Bb*	*Bp*	*Bsp*	*Bsu*	*Ct*	*Ec*	*Hi*	
Total number of amino acids	317	345	328	314	314	380	329	328	
Number of Pro	11	5	18	8	9	12	16	16	
Pro content (%)	3.5	1.4	5.5	2.5	2.9	3.2	4.9	4.9	
Species of RNAPα	*Hp*	*Mg*	*Mp*	*Mt*	*Sc*	*Sh*	*Syo*	*Syy*	*Tt*
Total number of amino acids	344	328	327	347	339	328	309	314	315
Number of Pro	17	10	10	20	19	16	14	13	20
Pro content (%)	4.9	3.0	3.1	5.8	5.6	4.9	4.5	4.1	6.3

binds to not only the 16S ribosomal RNA but also the mRNA of the α operon, and causes repression of the expression of the S13, S11, S4, and L17 proteins, but not that of RNAPα (*32*). The variation in the operon structure is relevant to the function of S4.

The RNAP core enzyme of *Ec* is assembled in the sequence of α_2-$\alpha_2\beta$-$\alpha_2\beta\beta'$ (*33*). The amino-terminal domain down to residue 235 of *Ec* RNAPα (αNTD), includes three important regions for this process. They are one near residue 45, one around residue 80, and one between residues 180 and 200 of *Ec* RNAPα. The first and second regions are involved in the association with the β subunit, and the second and third ones in the β' association (*4, 34–36*). These regions are highly conserved in eubacterial αNTDs including the present determined sequence of *Tt* αNTD (Fig. 2A). Thus, the process of assembly of the RNAP core enzyme should be the same as not only in *Tt* but also in all eubacteria.

Ec αCTD, which consists of residues 235 to 329, plays a regulatory role by providing the contact surface for *trans*-acting protein factors and *cis*-acting DNA elements (*2, 6, 11*). The three-dimensional structure of *Ec* αCTD and the flexibility of the linker were determined by NMR (*10, 37*). Structurally it is composed of three regions, *i.e.*, the flexible linker (residues 235 to 248), the hydrophobic core (residues 249 to 315), and the C-terminal loop (residues 316 to 329). *Tt* αCTD does not have the C-terminal loop region (Fig. 5), which extends from the hydrophobic core in *Ec* αCTD and wraps around it (*10*). Not only *Tt* αCTD but also those of *Bsp, Bsu, Mg*, and *Mp* do not contain such a region (Fig. 2B). On the other hand, the length of the C-terminal loop of *Mt* αCTD is longer than that of *Ec* (Fig. 2B, not shown in detail). This variety in the C-terminal loop region might reflect some biological functions of the region, such as interaction with activators, depending on the kind of eubacterium.

Prediction of the secondary structure of *Tt* αCTD indicated that the α helices in the core domain are located at almost the same positions as in *Ec* αCTD (Fig. 5), and their folding topology may not be so different from that of *Ec*. Arg265 in helix 1 (residues 264 to 273) of *Ec* αCTD, which is conserved in the αCTDs of all eubacteria including *Tt* (Fig. 2B), plays an important role in the response to both a transcription activator, CRP, and the DNA UP element. However, Asn268 and Cys269 of *Ec* αCTD are substituted by His and Ser, respectively, in *Tt*. In *Ec*, Asn268 is necessary for interaction with both CRP and the UP element, and Cys269 with CRP. Therefore, *Tt* αCTD might have a different mode of interaction with protein and DNA transcription factors from those of other eubacteria.

Although the N-terminal region of *Tt* αCTD was predicted to be a helical domain, it may be flexible because the corresponding region of *Ec* αCTD is flexible in spite of the similar prediction for *Ec* αCTD (Fig. 5).

It has been shown that the thermostability of *Tt* RNAPα is higher than that of *Ec*. *Tt* RNAPα has a high content of Pro residues (6.3%) compared with those of other eubacteria (Table I). A high Pro content is a general characteristic of *Tt* proteins, and the importance of Pro residues for thermostability has been pointed out (*38*). However, the Pro content of RNAPα of a hyperthermophilic bacterium, *Aa*, is 3.5%, *i.e.*, less than that of *Ec* (4.9%). The Pro content, therefore, would not be directly related with thermostability.

Although the *Tt* RNAP holoenzyme has been purified from *Tt* and the core enzyme has already been crystallized (*39*), its three-dimensional structure remains unknown. Since, the amino acid sequence of *Tt* RNAPα has now been determined, and the protein has been overexpressed and purified, structural analyses of *Tt* RNAPα by means of NMR and X-ray are feasible.

We wish to thank Dr. S. Tsunasawa (Takara Shuzo, Kusatsu) for the N-terminal amino acid analysis.

REFERENCES

1. Hayward, R.S., Igarashi, K., and Ishihama, A. (1991) Functional specialization within the α-subunit of *Escherichia coli* RNA polymerase. *J. Mol. Biol.* **221**, 23-29
2. Igarashi, K. and Ishihama, A. (1991) Bipartite functional map of the *E. coli* RNA polymerase α subunit: involvement of the C-terminal region in transcription activation by cAMP-CRP. *Cell* **65**, 1015-1022
3. Igarashi, K., Fujita, N., and Ishihama, A. (1991) Identification of a subunit assembly domain in the α subunit of *Escherichia coli* RNA polymerase. *J. Mol. Biol.* **218**, 1-6
4. Kimura, M., Fujita, N., and Ishihama, A. (1994) Functional map of the alphα subunit of *Escherichia coli* RNA polymerase. Deletion analysis of the amino-terminal assembly domain. *J. Mol. Biol.* **242**, 107-115
5. Murakami, K., Owens, J.T., Belyaeva, T.A., Meares, C.F., Busby, S.J., and Ishihama, A. (1997) Positioning of two alpha subunit carboxy-terminal domains of RNA polymerase at promoters by two transcription factors. *Proc. Natl. Acad. Sci. USA* **94**, 11274-11278
6. Igarashi, K., Hanamura, A., Makino, K., Aiba, H., Aiba, H., Mizuno, T., Nakata, A., and Ishihama, A. (1991) Functional map of the α subunit of *Escherichia coli* RNA polymerase: two modes of transcription activation by positive factors. *Proc. Natl. Acad. Sci. USA* **88**, 8958-8962
7. Choy, H.E., Park, S.W., Aki, T., Parrack, P., Fujita, N., Ishihama, A., and Adhya, S. (1995) Repression and activation of transcription by Gal and Lac repressors: involvement of α subunit of RNA polymerase. *EMBO J.* **14**, 4523-4529
8. Ross, W., Gosink, K.K., Salomon, J., Igarashi, K., Zou, C., Ishihama, A., Severinov, K., and Gourse, R.L. (1993) A third recognition element in bacterial promoters: DNA binding by the α subunit of RNA polymerase. *Science* **262**, 1407-1413
9. Murakami, K., Kimura, M., Owens, J.T., Meares, C.F., and Ishihama, A. (1997) The two α subunits of *Escherichia coli* RNA polymerase are asymmetrically arranged and contact different halves of the DNA upstream element. *Proc. Natl. Acad. Sci. USA* **94**, 1709-1714
10. Jeon, Y.H., Negishi, T., Shirakawa, M., Yamazaki, T., Fujita, N., Ishihama, A., and Kyogoku, Y. (1995) Solution structure of the activator contact domain of the RNA polymerase α subunit. *Science* **270**, 1495-1497
11. Murakami, K., Fujita, N., and Ishihama, A. (1996) Transcription factor recognition surface on the RNA polymerase α subunit is involved in contact with the DNA enhancer element. *EMBO J.* **15**, 4358-4367
12. Kato, R. and Kuramitsu, S. (1993) RecA protein from an extremely thermophilic bacterium, *Thermus thermophilus* HB8. *J. Biochem.* **114**, 926-929
13. Moffatt, B.A. and Studier, F.W. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* **189**, 113-130
14. Gill, S.C. and von Hippel, P.H. (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* **182**, 319-326
15. Gill, S.C. and von Hippel, P.H. (1990) Erratum. *Anal. Biochem.* **189**, 283
16. Oshima, T. and Imahori, K. (1974) Description of *Thermus*

thermophilus (Yoshida and Oshima) comb. nov., a nonsporulating thermophilic bacterium from a Japanese thermal spa. *Int. J. Syst. Bacteriol.* **24**, 102-112
17. Tsiboli, P., Herfurth, E., and Choli, T. (1994) Purification and characterization of the 30S ribosomal proteins from the bacterium *Thermus thermophilus*. *Eur. J. Biochem.* **226**, 169-177
18. Date, T., Suzuki, K., and Imahori, K. (1975) Purification and some properties of DNA-dependent RNA polymerase from an extreme thermophile, *Thermus thermophilus* HB8. *J. Biochem.* **78**, 845-858
19. Wnendt, S., Hartmann, R.K., Ulbrich, N., and Erdmann, V.A. (1990) Isolation and physical properties of the DNA-directed RNA polymerase from *Thermus thermophilus* HB8. *Eur. J. Biochem.* **191**, 467-472
20. Deckert, G., Warren, P.V., Gaasterland, T., Young, W.G., Lenox, A.L., Graham, D.E., Overbeek, R., Snead, M.A., Keller, M., Aujay, M., Huber, R., Feldman, R.A., Short, J.M., Olson, G.J., and Swanson, R.V. (1998) The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*. *Nature* **392**, 353-358
21. Bedwell, D., Davis, G., Gosink, M., Post, L., Nomura, M., Kestler, H., Zengel, J.M., and Lindahl, L. (1985) Nucleotide sequence of the alpha ribosomal protein operon of *Escherichia coli*. *Nucleic Acids Res.* **13**, 3891-3903
22. Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.F., Dougherty, B.A., Merrick, J.M., McKenney, K., Sutton, G., Fitzhugh, W., Fields, C., Gocayne, J.D., Scott, J., Shirley, R., Liu, L.I., Glodek, A., Kelley, J.M., Weidman, J.F., Phillips, C.A., Spriggs, T., Hedblom, E., Cotton, M.D., Utterback, T.R., Hanna, M.C., Nguyen, D.T., Saudek, D.M., Brandon, R.C., Fine, L.D., Fritchman, J.L., Fuhrmann, J.L., Geoghagen, N.S.M., Gnehm, C.L., McDonald, L.A., Small, K.V., Fraser, C.M., Smith, H.O., and Venter, J.C. (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**, 496-512
23. Tomb, J.F., White, O., Kerlavage, A.R., Clayton, R.A., Sutton, G.G., Fleischmann, R.D., Ketchum, K.A., Klenk, H.P., Gill, S., Dougherty, B.A., Nelson, K., Quackenbush, J., Zhou, L., Kirkness, E.F., Peterson, S., Loftus, B., Richardson, D., Dodson, R., Khalak, H.G., Glodek, A., McKenney, K., Fitzegerald, L.M., Lee, N., Adams, M.D., Hickey, E.K., Berg, D.E., Gocayne, J.D., Utterback, T.R., Peterson, J.D., Kelley, J.M., Cotton, M.D., Weidman, J.M., Fujii, C., Bowman, C., Watthey, L., Wallin, E., Hayes, W.S., Borodovsky, M., Karp, P.D., Smith, H.O., Frase, C.M., and Venter, J.C. (1997) The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* **388**, 539-547
24. Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry III, C.E., Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A., Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, S., Squares, S., Squares, R., Sulston, J.E., Taylor, K., Whitehead, S., and Barrell, B.G. (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**, 537-544
25. Boylan, S.A., Suh, J.W., Thomas, S.M., and Price, C.W. (1989) Gene encoding the alpha core subunit of *Bacillus subtilis* RNA polymerase is cotranscribed with the genes for initiation factor 1 and ribosomal proteins B, S13, S11, and L17. *J. Bacteriol.* **171**, 2553-2562
26. Fraser, C.M., Casjens, S., Huang, W.M., Sutton, G.G., Clayton, R.A., Lathigra, R., White, O., Ketchum, K.A., Dodson, R., Hickey, E.K., Gwinn, M., Dougherty, B., Tomb, J.-F., Fleischmann, R.D., Richardson, D., Peterson, J., Kerlavage, A.R., Quackenbush, J., Salzberg, S., Hanson, M., van-Vugt, R., Palmer, N., Adams, M.D., Gocayne, J.D., Weidman, J., Utterback, T., Watthey, L., McDonald, L., Artiach, P., Bowman, C., Garland, S., Fujii, C., Cotton, M.D., Horst, K., Roberts, K., Hatch, B., Smith, H.O., and Venter, J.C. (1997) Genomic sequence of a Lyme disease spirochete, *Borrelia burgdorferi*. *Nature* **390**, 580-586
27. Gu, L., Wenman, W.M., Remacha, M., Meuser, R., Coffin, J., and Kaul, R. (1995) *Chlamydia trachomatis* RNA polymerase alpha subunit: sequence and structural analysis. *J. Bacteriol.* **177**, 2594-2601
28. Fraser, C.M., Gocayne, J.D., White, O., Adams, M.D., Clayton, R.A., Fleischmann, R.D., Bult, C.J., Kerlavage, A.R., Sutton, G., Kelley, J.M., Fritchman, J.L., Weidman, J.F., Small, K.V., Sandusky, M., Fuhrmann, J.L., Nguyen, D.T., Utterback, T.R., Saudek, D.M., Phillips, C.A., Merrick, J.M., Tomb, J.F., Dougherty, B.A., Bott, K.F., Hu, P.C., Lucier, T.S., Peterson, S.N., Smith, H.O., Hutchison, C.A. III, and Venter, J.C. (1995) The minimal gene complement of *Mycoplasma genitalium*. *Science* **270**, 397-403
29. Himmelreich, R., Hilbert, H., Plagens, H., Pirkl, E., Li, B.C., and Herrmann, R. (1996) Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*. *Nucleic Acids Res.* **24**, 4420-4449
30. Sugita, M., Sugishita, H., Fujishiro, T., Tsuboi, M., Sugita, C., Endo, T., and Sugiura, M. (1997) Organization of a large gene cluster encoding ribosomal proteins in the cyanobacterium *Synechococcus* sp. strain PCC 6301: comparison of gene clusters among cyanobacteria, eubacteria and chloroplast genomes. *Gene* **195**, 73-79
31. Kaneko, T., Sato, S., Kotani, H., Tanaka, A., Asamizu, E., Nakamura, Y., Miyajima, N., Hirosawa, M., Sugiura, M., Sasamoto, S., Kimura, T., Hosouchi, T., Matsuno, A., Muraki, A., Nakazaki, N., Naruo, K., Okumura, S., Shimpo, S., Takeuchi, C., Wada, T., Watanabe, A., Yamada, M., Yasuda, M., and Tabata, S. (1996) Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res.* **3**, 109-136
32. Thomas, M.S., Bedwell, D.M., and Nomura, M. (1987) Regulation of α operon gene expression in *Escherichia coli*. *J. Mol. Biol.* **196**, 333-345
33. Ishihama, A. (1981) Subunit of assembly of *Escherichia coli* RNA polymerase. *Adv. Biophys.* **14**, 1-35
34. Kimura, M. and Ishihama, A. (1995) Functional map of the alpha subunit of *Escherichia coli* RNA polymerase. Insertion analysis of the amino-terminal assembly domain. *J. Mol. Biol.* **248**, 756-767
35. Kimura, M. and Ishihama, A. (1995) Functional map of the alpha subunit of *Escherichia coli* RNA polymerase. Amino acid substitution within the amino-terminal assembly domain. *J. Mol. Biol.* **254**, 342-349
36. Kimura, M. and Ishihama, A. (1996) Subunit assembly *in vivo* of *Escherichia coli* RNA polymerase: role of the amino-terminal assembly domain of alpha subunit. *Genes Cells* **1**, 517-528
37. Jeon, Y.H., Yamazaki, T., Otomo, T., Ishihama, A., and Kyogoku, Y. (1997) Flexible linker in the RNA polymerase alpha subunit facilitates the independent motion of the C-terminal activator contact domain. *J. Mol. Biol.* **267**, 953-962
38. Okamoto, A., Kato, R., Masui, R., Yamagishi, A., Oshima, T., and Kuramitsu, S. (1996) An aspartate aminotransferase from an extremely thermophilic bacterium, *Thermus thermophilus* HB8. *J. Biochem.* **119**, 135-144
39. Tsuji, S., Suzuki, K., and Imahori, K. (1976) Crystallisation of DNA-dependent RNA polymerase from *Thermus thermophilus* HB8. *Nature* **261**, 725-726
40. Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673-4680
41. Carbonetti, N.H., Fuchs, T.M., Patamawenu, A.A., Irish, T.J., Deppisch, H., and Gross, R. (1994) Effect of mutations causing overexpression of RNA polymerase alpha subunit on regulation of virulence factors in *Bordetella pertussis*. *J. Bacteriol.* **176**, 7267-7273
42. Cho, E.J., Bae, J.B., Kang, J.G., and Roe, J.H. (1996) Molecular analysis of RNA polymerase alpha subunit gene from *Strep-*

tomyces coelicolor A3(2). *Nucleic Acids Res.* **24**, 4565-4571
43. Nakasone, K., Kato, C., and Horikoshi, K. (1996) Molecular cloning of the gene encoding RNA polymerase alpha subunit from deep-sea barophilic bacterium. *Biochim. Biophys. Acta* **1308**, 107-110
44. Ito, M., Matsuo, Y., and Nishikawa, K. (1997) Prediction of protein secondary structure using the 3D-1D compatibility algorithm. *Comput. Appl. Biosci.* **13**, 415-423