4668–4675 *Nucleic Acids Research, 1996, Vol. 24, No. 23* *© 1996 Oxford University Press*

Reverse gyrase gene from *Sulfolobus shibatae* B12: gene structure, transcription unit and comparative sequence analysis of the two domains

Christine Jaxel, Claire Bouthier de la Tour, Michel Duguet and Marc Nadal*

Laboratoire d'Enzymologie des Acides Nucléiques, Institut de Génétique et de Microbiologie, URA 1354, bâtiment 400, Université Paris Sud, 91405 Orsay cedex, France

Received August 22, 1996; Revised and Accepted October 16, 1996

DDBJ/EMBL/GenBank accession no. X98420

ABSTRACT

We cloned and sequenced a DNA fragment from the thermophilic archaeal strain *Sulfolobus shibatae* B12 that includes the gene *topR* encoding the reverse gyrase. The RNA of the reverse gyrase gene was characterized indicating that the *topR* gene is fully functional *in vivo*. We showed by primer extension analysis that transcription of *topR* initiates 28 bp downstream from a consensus A-box promoter. In order to understand how this particular type I DNA topoisomerase introduces positive superturns into the DNA, we compared the amino acid sequence of reverse gyrase from *S.shibatae* with the two other known reverse gyrases. This comparison indicates a common organization of these proteins: the carboxy-terminal domain is related to the type I-5′ topoisomerase family while the amino-terminal domain possesses some motifs of proteins described as RNA or DNA helicases. By using local alignments, we showed that (i) reverse gyrases constitute a new and rather homogenous group within the type I-5′ DNA topoisomerase family; (ii) a careful sequence analysis of the amino-terminal domain allows us to relate the presence of some motifs with an ATP binding and hydrolysis reaction coupled to a DNA binding and unwinding activity.

INTRODUCTION

As noted by Watson and Crick a long time ago, the helical structure of DNA requires enzymes with the ability to eliminate stress in this molecule (1). These enzymes are DNA topoisomerases, ubiquitous enzymes that are required for all DNA metabolism processes such as: replication, transcription, recombination or chromatin assembly (for a review, see 2). DNA topoisomerases act by introducing transient single or double strand breaks in DNA for type I and type II topoisomerases respectively. Whereas the type II DNA topoisomerases appear evolutionarily related (3) in the three domains of living cells (Eucarya, Bacteria and Archaea) (4,5), type I DNA topoisomerases are more puzzling. From a mechanistic point of view, and irrespective of their evolutionary history, we can define two families of type I topoisomerases as proposed by Roca (2). The type I-3′ family groups topoisomerases that are linked to the DNA by a 3′phosphotyrosyl link (essentially eukaryotic topoisomerase I) while the type I-5′ family is constituted by the topoisomerases that are transiently linked to the DNA by a 5′phosphotyrosyl link. Two type I-5′ topoisomerases are present in *Escherichia coli*: protein ω, coded by the gene *topA*, relaxes the DNA in the cell (6) while topoisomerase III, coded by the gene *topB*, has an unknown function (7). This classification of topoisomerases type I-3′ or I-5′ on a biochemical basis is clearly consistent with sequence data (3).

In order to have a better understanding of the different roles of topoisomerases in the cell, we have been studying, for several years, topoisomerases issued from the third domain, the Archaea (previously Archaebacteria) (5). In thermophilic Archaea, a new DNA topoisomerase named reverse gyrase was first described in two different strains of *Sulfolobus* (8,9). This enzyme catalyzes the formation of positively supercoiled DNA in the presence of ATP and magnesium. Surprisingly, although it catalyzes a gyration reaction in an ATP-dependent manner, reverse gyrase is a type I DNA topoisomerase (9–11). Later on, reverse gyrase activity was detected in all thermophilic Archaea tested (12,13) but also in thermophilic Bacteria (14,15) suggesting that this enzyme is a characteristic of life at high temperature rather than an archaeal feature. These results raise questions about the relationships between reverse gyrase and the other topoisomerases and on the phylogenetic origin of the gene encoding it. Since Jaxel *et al.* (16) have shown that reverse gyrase is linked through the 5′ end of the DNA, reverse gyrase is a type I-5′ topoisomerase. The deduced protein sequence of the gene coding for *Sulfolobus acidocaldarius* reverse gyrase supports this view (17). Moreover, reverse gyrase appears to constitute a chimeric protein with two domains: the carboxy-terminal part is clearly related to the type I-5′ DNA topoisomerase family whereas the amino-terminal part exhibits putative helicase motifs.

In order to have a better knowledge of both the phylogeny and the mechanism of reverse gyrase, we decided to clone and sequence the reverse gyrase gene from *Sulfolobus shibatae* B12. Indeed, thermophilic Archaea constitute good models for the understanding of fundamental biological mechanisms. For instance, the

*To whom correspondence should be addressed. Tel: +33 1 69 15 70 03; Fax: +33 1 69 15 72 96; Email: nadal@igmors.u-psud.fr

transcription machinery of these organisms may be related to the eukaryotic one (18). Within the Archaea domain, the strain *S.shibatae* is particularly interesting since homologs of the eukaryotic transcription factors, TFIIB and TBP, were described in this strain (19,20). In addition, *S.shibatae* possesses an inducible virus named SSV1 (21) which is able to infect *Sulfolobus* cells (22). Finally, SSV1 encodes a site specific recombination system (23) and some years ago, we showed that the DNA of SSV1 is positively supercoiled, suggesting a major role of reverse gyrase *in vivo* (24). More recently, we have purified and characterized the reverse gyrase of *S.shibatae*. In particular, we showed that the partial proteolysis of the reverse gyrase gives rise to a topoisomerase, that has only a ATP-independent relaxation activity similar to that of protein ω (25). This fact points out a possible new regulatory level in the cell.

In this paper, we describe the cloning and sequencing of a large DNA fragment of *S.shibatae* encompassing the gene encoding reverse gyrase. A messenger RNA from the reverse gyrase gene is, for the first time, characterized. In particular, from *S.shibatae* cells, we determined the size of this RNA and we localized the transcription initiation site 28 bp downstream from a consensus promoter sequence. Comparison of the *S.shibatae* reverse gyrase amino acid sequence with those of reverse gyrases from *S.acidocaldarius* (17) and *M.kandleri* (26) allows us to define conserved and potentially functional regions of the protein. All the reverse gyrases possess amino acid sequence motifs characteristic of type I-5′ topoisomerase family but exhibit additional common features. We therefore propose that reverse gyrases constitute a new type I-5′ topoisomerases group named topR beside the previously described topA and topB groups. Finally, a more detailed comparative analysis on the amino-terminal part of reverse gyrases suggests that this part possesses a DNA unwinding activity.

MATERIALS AND METHODS

SDS (ultragrade) was from Serva, RnaseA, Sarkosyl and DEPC from Sigma, [α-^{35}S]dATP (1000 Ci/mmol) from ICN, [α-^{32}P]dCTP (3000 Ci/mmol) and [γ-^{32}P]ATP (5000 Ci/mmol), HybondN and Hyperfilm were from Amersham. Acrylamide and bis-acrylamide were purchased from Biorad and phenol was from Appligene. Agarose (indubiose A 37 NA) was from Industrie Biologique Française. Other chemicals were from Carlo Erba.

Escherichia coli alkaline phosphatase and T7 sequencing kit were from Pharmacia Biotech. Random priming kit, Dnase I–RNase free and Rnase inhibitor were from Boehringer (Mannheim, Germany) and proteinase K was from Merck. M-MuLV reverse transcriptase, DNA ligase, restriction enzymes and polynucleotide kinase were purchased from Biolabs.

Purification of genomic DNA

Sulfolobus shibatae B12 frozen cells were thawed and resuspended in 50 mM Tris–HCl pH 7.9, 100 mM EDTA, 100 mM NaCl, and Sarkosyl was dissolved by gentle mixing at room temperature (final concentration 2.5% w/v). After low speed centrifugation, the supernatant was incubated 3 h at 50°C with proteinase K (0.25 mg/ml). The nucleic acid was extracted with phenol (twice) and chloroform/isoamyl alcohol (24/1) followed by an ethanol precipitation. The pellet was dissolved with TE (10 mM Tris–HCl pH 8, 1 mM EDTA) and RNAs were removed by RNaseA treatment (0.1 mg/ml, 4 h at 37°C). The RNaseA was then extracted with chloroform/isoamyl alcohol and the DNA was ethanol-precipitated in the presence of ammonium acetate.

PCR

The two degenerated oligonucleotides P1 and P2 used as primers were defined as in Bouthier de la Tour *et al.* (27). Amplification was done with 400 ng of genomic DNA and 500 pmoles of each primer in a Biomed thermocycler using *Taq* polymerase (Bio-probe). Thirty cycles of amplification were carried out at 72°C.

Sub-bank and screening

The DNA fragment corresponding to the PCR product was purified and labelled with [α-^{32}P]dCTP by the random priming method. This product was used in hybridization experiments. Restriction analysis of *S.shibatae* genomic DNA was performed by using the Southern method. We purified DNA of ~5–8 kb obtained after hydrolysis of genomic DNA by *Eco*RI enzyme. These fragments were ligated with pGEM3Zf(+) DNA previously hydrolyzed by *Eco*RI and dephosphorylated. Transformation of TG1 bacteria cells was achieved by electroporation and plated on LB, 1.5% agar, 100 μg/ml ampicillin, 0.1 mM IPTG, $2^{0}/_{0000}$ X-Gal. The plates were incubated overnight at 37°C. Bacteria were replicated on Hybond N filters and grown for 4 h. Cells were lysed with NaOH as described by the manufacturer. After prehybridization (6 h), filters were hybridized with ^{32}P labelled P 0.85 DNA fragment at 42°C for 16 h in 6× SSC, 5× Denhardts and 50% formamide. Filters were washed at 42°C, three times with 2× SSC and five times with 0.2× SSC, 0.1% SDS. The filters were submitted to an autoradiography by using Hyperfilm (Amersham).

The positive clones were isolated and a second round of selection was done.

DNA sequencing

The double strand DNA fragment was sequenced by using the dideoxy method. Usually, a chase was performed with 200 μM of each dNTP before stopping the reaction (3.5 min at room temperature for the elongation step, 5 min for the termination reaction at 37°C and 2 min at 37°C for the chase). The sequence was performed either with the restriction fragments subcloned in pGEM3Zf(+) or directly on the 6046 bp DNA fragment cloned in pGEM3Zf(+). We used universal or synthetic oligonucleotides as primers. Sequencing reactions were analyzed by polyacrylamide gels (6% acrylamide; 0.2% bis-acrylamide; 7 M urea) in TBE buffer.

RNA purification and analysis

Sulfolobus shibatae B12 cells were grown essentially as described by Zillig *et al.* (28). A mid-log culture (50 ml, OD_{600nm} = 0.68) was quickly cooled by frozen fresh medium at pH 5.5 and centrifuged at 10 000 *g* for 10 min. The cells were resuspended in 50 mM of EDTA and disrupted by addition of SDS (0.25%). Na-acetate (pH 5.2) was then added to a final concentration of 50 mM; RNAs were extracted twice with 1 vol of phenol saturated with water, followed by an extraction with chloroform/isoamyl alcohol. The supernatant was neutralized by 60 mM Tris–HCl pH 8, 150 mM NaCl and then ethanol-precipitated. The pellet was dissolved with DEPC-treated water. The purity and concentration of the RNA preparation were checked spectroscopically.

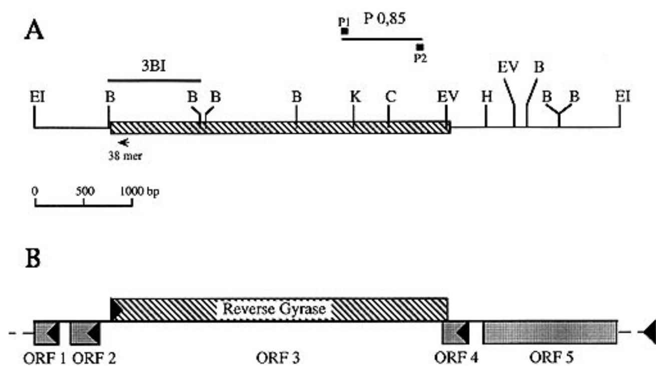

Figure 1. Cloning and sequencing of the *S.shibatae* DNA fragment containing the reverse gyrase gene. (**A**) Restriction analysis. The different enzymes are: *Eco*RI (EI), *Bgl*II (B), *Kpn*I (K), *Cla*I (C), *Eco*RV (EV) and *Hin*dIII (H). P 0.85 is the PCR fragment obtained by using P1 and P2 primers. 3BI is the DNA fragment used as a probe for the Northern blot experiments. 38mer is the oligonucleotide used in the primer extension assay. (**B**) Organization of the ORFs contained in the sequenced DNA fragment. Arrows indicate the amino-terminus of the ORFs. In both schemes, the location of reverse gyrase gene is indicated by a hatched box.

Primer extension was essentially realized as described by Kingston (29). Briefly, 3 and 15 µg of *S.shibatae* RNA (eventually pretreated by RNaseA or RNase-free DNase I) was denatured. A 38mer oligonucleotide described in Figure 1A (position 874–911) was ^{32}P-labelled at the 5′ terminus. The RNA was then hybridized with it during 20 h at 30°C in 80% formamide, 2.6× SSC. After ethanol precipitation, the dried pellets were resuspended in reverse transcriptase buffer. The extension reactions were performed at 37°C for 90 min. The RNA was hydrolyzed by RnaseA treatment and the DNA was ethanol-precipitated and dried. Sequence controls were realized with the same oligonucleotide by using the 6046 bp DNA fragment cloned in pGEM3Zf(+) as a matrix and by the dideoxy method. A part of the reaction products (1/6) were analyzed by sequencing gel electrophoresis.

Computer analyses

Sequence assembly, G+C content, dinucleotide frequency, restriction analyses, ORF translation, codon usage, amino acid composition and deduced molecular mass were performed by using DNA Strider (30) or LGBC (31) softwares.

FASTA (32), BLAST (33), BLITZ (34) and PATTERN (35) softwares were used for retrieving sequences in banks. For this, we used computer facilities of CITI2 (36), NCBI or EBI.

SMARTIES softwares package (unpublished, Atelier de Bio-Informatique, Paris) was used on Macintosh computer for retrieving sequences in banks by FASTA, for primary multialignments and for consensus determination. Primary multialignments were also realized by using the VIZZ program (37).

The data bank used is essentially Swissprot but also GenBank, EMBL library and non-redundant database of the NCBI. The matrices used are essentially PAM (50, 250 or 500) and BLOSUM62.

The sequence of the 6046 bp DNA fragment of *S.shibatae* (Ssh) was deposited with EMBL library with annotations. The accession number is X98420.

The local primary alignments realized by using computer programs were modified manually.

The Swissprot accession number of the proteins used are: for reverse gyrases: *S.acidocaldarius* (Sac) topR: Q08582, *M.kandleri* (Mka) topRa: U41058*, *M.kandleri* topRb: U41059*; for topAs: *E.coli* (Eco): P06612, *H.influenzae*: (Hin): P43012, *Bacillus subtilis* (Bsu): P39814, *Synecoccochus* sp (Ssp): P34185, *T.maritima*: (Tma): P46799; for topBs: *E.coli*: P14294, *H.influenzae*: P43704, *Saccharomyces cerevisiae* (Sce) topIII: P13099, human (Hum) topIII: U43431*. We also used the poorly characterized cellular type I-5′ topoisomerases from *B.anthracis* : P40114, *M.genitalium*: P47368, *B.firmus* (partial): P34184* and Trae from RP4 plasmid: 437697** only for strictly conserved regions.

The unwinding proteins used are: human eIF4A: P04765, *S.cerevisiae* DBP1: P24784 and DBP2: P24783, *Vaccinia virus* (Vacc) NPH-II: M35027**, PriA from *E.coli*: P17888 and *H.influenzae*: P44647, *E.coli* LHR: P30015, RecG from *E.coli*: P24230, *S.cerevisiae* SGS1: P35187, BLM from human: U39817*, RecQ from *E.coli*: P15043 and human: P46063.

*indicated a EMBL library and ** a GenBank accession number instead of a Swissprot accession number.

RESULTS

Cloning and sequencing of the genomic DNA fragment of *S.shibatae* B12 containing reverse gyrase gene

Comparison of *S.acidocaldarius* reverse gyrase amino acid sequence with bacterial topoisomerase I sequence allowed us to define conserved amino acids (17). On this basis, Bouthier de la Tour *et al.* (27) designed a couple of degenerated oligonucleotides (P1, motif 5, and P2, motif 10) (Figs 1A and 4A) used for PCR amplification. By using *S.shibatae* DNA, a DNA fragment with a size of ~0.85 kb was amplified (Fig. 1A, P 0.85). The amino acid sequence deduced from the nucleotide sequence of this fragment exhibited large homology with *S.acidocaldarius* reverse gyrase. We therefore used this DNA fragment as a probe in order to screen a sub-bank of *Eco*RI *S.shibatae* genomic fragments, ranging in size from 5 to 8 kb. We cloned a DNA fragment of 6046 bp (for details, see Materials and Methods). By using the P 0.85 cloned fragment as a probe, restriction analyses show that the cloned DNA fragment and the genomic DNA have the same restriction maps (Fig. 1A). The sequencing was performed by using, either the 6046 bp DNA fragment or *Bgl*II restriction fragments cloned in pGEM 3Z(+) vector as matrices. Sequence analysis of the 6046 bp DNA fragment shows that it contains five ORFs (Fig. 1B). The ORF3 located on the upper strand corresponds to reverse gyrase as seen by homology with *S.acidocaldarius* (see below). ORFs 1, 2, 4 and 5 are located on the bottom strand and the deduced lengths in amino acids are respectively: >102, 97, 89, >453. ORFs 1 and 5 are partial and the ORF 4 overlaps that of reverse gyrase. For these ORFs, we did not find sequences with any significant similarity in data banks by using BLAST or FASTA softwares.

In the case of the ORF encoding reverse gyrase, we found two putative ATG initiation codons separated by three amino acids. Both are included in a putative ribosome binding site (38) as described in Figure 2. By using the first ATG codon, the length of this ORF is 3498 bp corresponding to a protein of 1166 amino acids with a molecular mass of 132 kDa. This predicted molecular mass is slightly higher than that of the purified protein (25), as

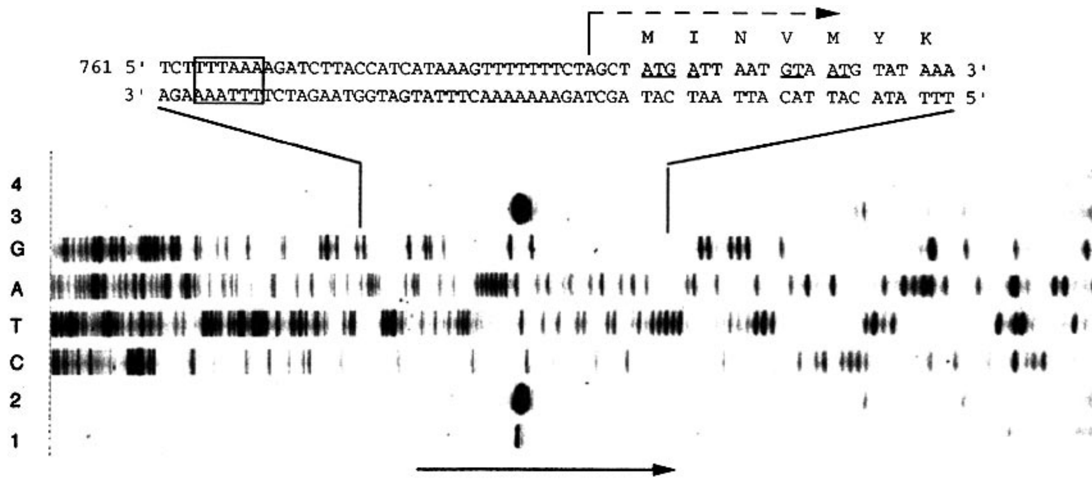

Figure 2. DNA sequence of the 5′ region of *Ssh-topR* and localization of the transcription initiation site. C, T, A and G indicate the dideoxynucleotide used during sequencing assay. For primer extension experiments, we used various amounts of RNA: 3µg of RNA for lane 1 and 15 µg for lanes 2, 3 and 4. Prior to the hybridization and extension, the RNA is incubated with Dnase I (lane 3) and with RnaseA (lane 4). The arrow indicates the direction of the electrophoresis. The double strand DNA sequence of the 5′ region of *Ssh-topR* and the deduced amino acid sequence are shown. The putative promoter box A is framed and the putative ribosome binding sites are underlined. The first nucleotide of the messenger is indicated above the nucleotides by an dashed arrow. We observed an extension arrest one nucleotide below but since this arrest is very weak, we assume that this is an abortive product of the reverse transcriptase reaction with no significance.

described for other thermophilic proteins. The amino acid composition is in good agreement with the biochemical data (25) and in particular, the cysteine content is very low, as for the other Sulfolobales proteins. Upstream from the first initiation codon, we found a putative A-box (Fig. 2) that is a characteristic of the archaeal promoter (39). All these data suggest that this ORF corresponds to the functional gene encoding reverse gyrase.

Transcription of *S.shibatae* reverse gyrase gene

In order to map the transcribed DNA region, we prepared total RNA from mid-log culture of *S.shibatae* cells. By using the DNA fragment 3BI as a probe (Fig. 1A), Northern blot analysis reveals a messenger with a size of ~3700 nt (not shown). This result indicates that this described reverse gyrase gene is functional *in vivo*. In order to map the transcription initiation site of the reverse gyrase gene, a 38 base oligonucleotide was used in a primer extension assay with *S.shibatae* RNA. This yielded a 75 nucleotide run-off product (Fig. 2) mapping to the A (position 798) that is at 28 bp downstream from the 3′ terminus of the putative A-box element (Fig. 2). Finally, the *S.shibatae* reverse gyrase promoter agrees well with the archaeal promoter consensus, emphasing the importance of a TATA-like A-box in transcriptional initiation as proposed by Reiter *et al.* (39). We assume that the first putative initiation codon, located 4 bp downstream from the 5′ end of the messenger, is used in the cell. Indeed, the promoter sequence, the transcription start, the position of the ATG initiation codon and the putative ribosome binding site, exhibit the same features as those described for the transcripts of SSV1, especially for T3 (40). As transcription termination sites in *S.shibatae* genes generally map to thymine-rich regions (41), we looked for this kind of feature in our sequence. We found that downstream from the stop codon, the dinucleotide TT frequency is increased and four putative termination sites are found. They are localized between the positions 4310 and 4531, in agreement with the size of the messenger.

Reverse gyrase sequence characteristics

In order to have a better understanding of the reverse gyrase unusual topoisomerase activity, we compared (Fig. 3) the amino acid sequence obtained in *S.shibatae* with the sequences previously determined in *S.acidocaldarius* and in *M.kandleri*. The two *Sulfolobale* reverse gyrases are constituted by one polypeptide whereas the *M.kandleri* counterpart is composed of two subunits. Nevertheless, except for an additional domain in the *M.kandleri* enzyme (26), the three reverse gyrases are highly homologous, with short insertions or deletions. They are composed of a carboxy-terminal domain (amino acids 582–1167 in *S.shibatae* reverse gyrase) related to the type I-5′ topoisomerases family (motifs 1–10) and an amino-terminal part (amino acids 1–581) exhibiting some motifs (motifs I, Ia, II, III, V and VI) of some DNA or RNA helicases (Fig. 4).

Since reverse gyrases are type I-5′ topoisomerases (16), we compared the amino acids of these proteins with all the available members of this family. A local alignment is realized with all these sequences and in Figure 4A only some of these alignments are presented (reverse gyrases and *T.maritima* topA, *E.coli* topA and topB). The comparative analysis of conserved or non-conserved amino acids is shown in the figure by using different colours; for instance, identical amino acids are in upper case red. All these enzymes are related since the motifs 1–10 are retrieved with no main amino acid change occurring in these motifs.

The topA group, including the enzymes from *E.coli*, *B.subtilis*, *H.influenzae*, *Synecoccochus* sp, is characterized by typical amino acids (magenta coloured in the Fig. 4A) and the presence of at least one tetracysteine motif (putative Zn binding motif) in the carboxy-terminus (Fig. 3).

The topB group, including the enzyme of *E.coli*, *H.influenzae* and *S.cerevisiae* and human TOPIII, is characterized by other amino acids (green coloured in the Fig. 4A) and the presence of a basic carboxy-terminus (an additional putative zinc finger is present both in human and *S.cerevisiae* TopIIIs).

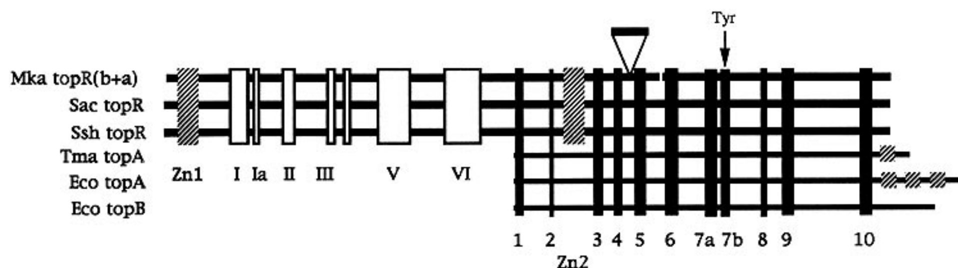

Figure 3. Schematic representation of homology found between: the reverse gyrases of *M.kandleri*, *S.acidocaldarius* and *S.shibatae*; topA of *T.maritima*; topA and topB of *E.coli*. An arrow indicates the position of the active site tyrosine. Motifs I, Ia, II, III, V and VI are characteristic of some helicases. Motifs 1–10 are characteristic of the type I-5′ DNA topoisomerase family. Hatched boxes indicate putative zinc binding motifs. For the abbreviations of organism names, see Materials and Methods.

In order to distinguish between reverse gyrase features and a possible thermophilic selective pressure, we compared reverse gyrases with topA of a thermophilic Bacterium *T.maritima* (27). We do not observe any particular similarity and we therefore assume that the amino acids conserved only in reverse gyrases (amino acids blue coloured in Fig. 4) may constitute a signature of these topoisomerases. Another specific characteristic of reverse gyrases is a putative zinc finger motif (Zn2) between boxes 2 and 3 (Fig. 3). All these features, peculiar to these topoisomerases, allow us to define a third type I-5′ topoisomerase group corresponding to the reverse gyrases and named topR.

At their amino-terminal end, reverse gyrases exhibit a second putative zinc finger (Zn1, Fig. 3). The rest of the amino-terminal part of reverse gyrase from *S.shibatae* (amino acids 76–581) does not match with significant scores with other proteins in the data banks when BLAST, BLITZ and FASTA softwares are used. Comparison of the three reverse gyrases allowed us to define motifs with amino acid similarities (Fig. 3). We searched in data banks for the presence of all these motifs succeeding in the same order by using PATTERN software. No sequence was retrieved.

From a biochemical point of view, reverse gyrase is characterized as an ATPase (42). It possesses the ATPase motifs I (GXGKT) and II (DD) defining the putative helicase superfamily 2 (Fig. 4B) (43–45). Since the *S.acidocaldarius* enzyme appears to possess some motifs found in many putative helicases (17), we searched for helicase motifs, as defined by Linder *et al.* (46) and Schmid and Linder (47), in helicase superfamily 2. We did not find extensive similarities with some putative members of helicase superfamily 2 nor with biochemically defined helicases. By contrast, helicase motifs in common with reverse gyrases were found in other RNA or DNA helicases of superfamily 2 e.g. eIF4a, NPH-II, PriA, RecQ and RecG. In addition, we found these motifs in SGS1, a protein that interacts with the topoisomerase III in yeast, in BLM a human gene involved in Bloom's syndrome and in putative proteins with unknown function like DBP1, DBP2, LHR.

Figure 4B shows the conserved helicase motifs (motifs I, Ia, II, III, V and VI) as defined by Pause and Sonenberg (48) and Pause *et al.* (49). We noted that SGS1, BLM and RecQ are highly homologous for these motifs. The local alignment with all the amino acids sequences indicated that the motifs Ia and V are poorly conserved. Motifs I and III are present in all these proteins. Motifs II and VI are conserved but also possess characteristic amino acids allowing the definition of three groups of proteins (Fig. 4B). Indeed, the sequences DEAD (motif II) with HxxGRxGR (motif VI) are found in RNA helicases such as eIF4a. DEXH (motif II) with QxxGRxGR (motif VI) are found in RNA or DNA helicases like NPH-II or PriA, RecQ and putative DNA helicases like SGS1. Finally, reverse gyrases constitute a new group with DDxD as motif II and QxxGRxSR as motif VI. Moreover, reverse gyrase similarities extend over these motifs. The other helicase motifs proposed by Schmid and Linder (47) were not retrieved in reverse gyrases.

DISCUSSION

In this manuscript, we report the sequence of a 6046 bp DNA fragment from the archaeal strain *S.shibatae* B12 containing the gene encoding reverse gyrase.

Analysis of this DNA fragment shows that it exhibits 33.25% of G+C. This very low G+C content appears as a characteristic of the *Sulfolobus* genus. Interestingly, we found that, in the five coding DNA sequences, the dinucleotide AA frequency is very high (13.3–16.38%) compared to the TT frequency (4.44–8.92%). The reverse gyrase coding region presents the lowest AA frequency (13.3%). We searched AA frequency bias for four other genomic coding sequences reported in *S.shibatae*. We found the same high frequency for AA dinucleotide (11–16.11%) compared with TT (5.88–6.88%). To date, the meaning of this unusual AA content is unknown. Nevertheless, this may be a new criterion for searching putative coding regions in thermophilic organisms.

In order to define the genomic organization, we mapped the transcription initiation site and upstream, we found the consensus sequence of the promoter A-box of *Sulfolobus*. The two putative initiation codons of reverse gyrase are ATG instead of GTG for the *S.acidocaldarius* counterpart (17). The genomic organization is not conserved between *Sulfolobus acidocaldarius* and *shibatae,* since the flanking sequences have no similarity. In *M.kandleri*, the organization is also completely different since reverse gyrase is encoded by two separate genes (26).

Focusing on the coding sequences of the three known reverse gyrases, we observed that apart from the highly conserved motifs, the proteins are different with some insertions or deletions. In particular, the two related *Sulfolobus* genes exhibit slight differences. We conclude that the highly conserved motifs reflect the selective pressure and consequently correspond to the regions involved in the enzymatic activity of the protein: positive supercoiling of the DNA in an ATP-dependent process. Finally, the sequence comparison allows us to define three groups of type I-5′ topoisomerases. The first is represented by topA-related enzymes, the second by topB-related proteins and the third by reverse

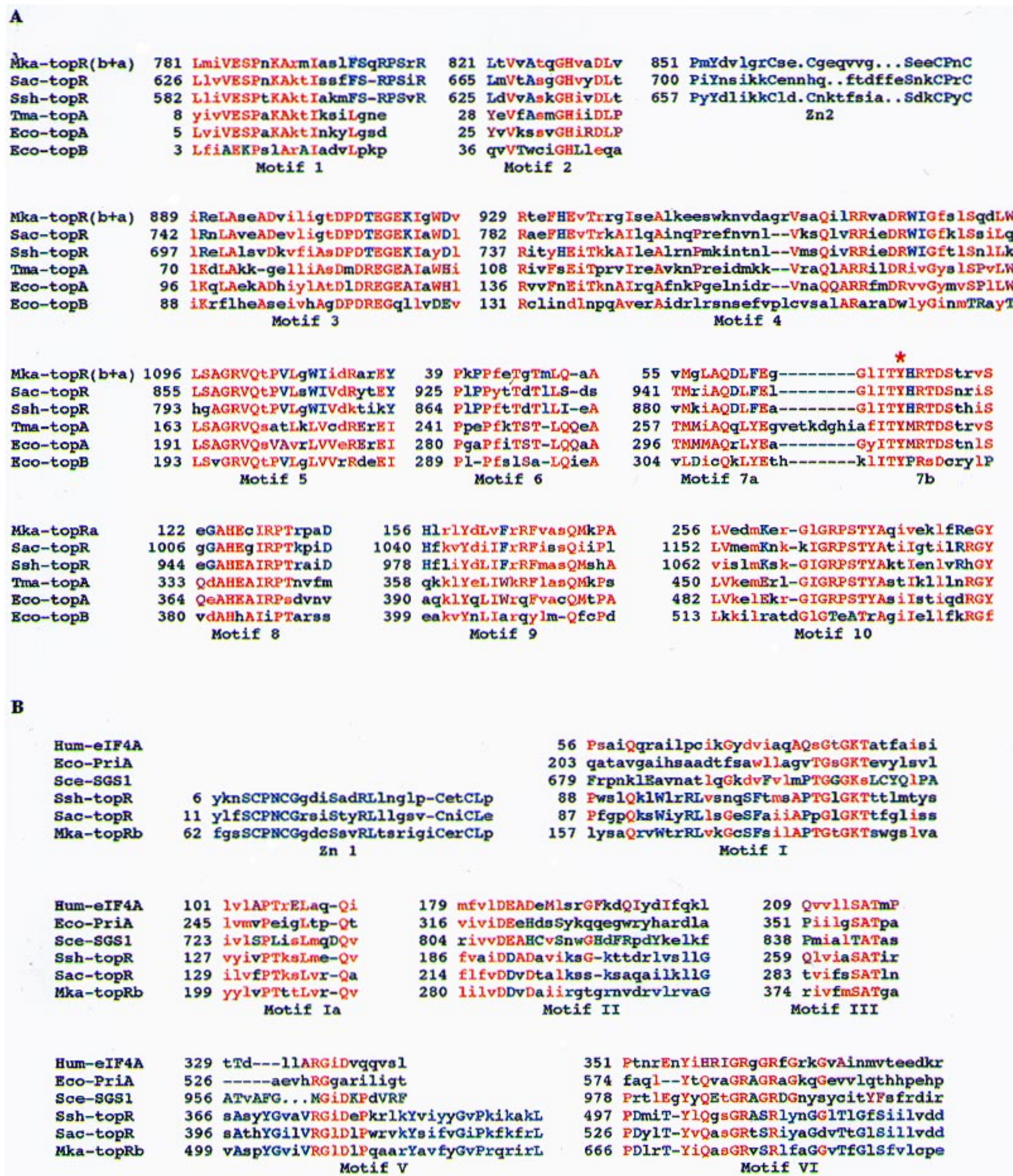

Figure 4. Highly conserved regions found in reverse gyrases. These local alignments were performed as described in Materials and Methods, Computer analyses, by using the three known reverse gyrases. The identical amino acids are in uppercase, the conservative changes are in lower case. The red colour is for amino acids found in 80% of analyzed sequences, and the blue colour for reverse gyrases (score 3/3). (**A**) Comparison with other type I-5′ topoisomerases. For clarity, in addition to the reverse gyrases, only topoisomerases from *E.coli* and *T.maritima* are shown. The magenta colour is for the topA group (score 4/5) and the green colour for the topB group (score 3/4). In the case of *M.kandleri* reverse gyrase, the alignment was performed with the two fused proteins (b+a) but the numbers refer to the amino acid position of each polypeptide. (**B**) Comparison with RNA or DNA helicases. For clarity we only show eIF4A, PriA and SGS1. The pink colour is for the eIF4a family, the green colour for the SGS1 family. Conserved amino acid groups used are: (I, L, V, M); (S, T); (F, W, Y); (D, E); (K, R). For the abbreviations of organisms names, see Materials and Methods.

gyrases. In order to clarify the nomenclature, we propose to name the gene coding for reverse gyrase, *topR*. In the case of *M.kandleri*, *topRb* is the gene coding for the amino-terminal domain and *topRa* the gene coding for the carboxy-terminal domain [topRa is the protein containing the tyrosine of the active site as proposed by Krah *et al.* (26) and by analogy with the gyrase gene].

The carboxy-terminal part of reverse gyrase corresponds approximatively (see Figs 3 and 4A) to the truncated 67 kDa of *E.coli* topA used by Lima *et al.* (50) for the three dimensional structure determination. We have localized the strictly conserved amino acids on this three dimensional structure (not shown). We observed that these amino acids are clustered to a pocket around the tyrosine involved in the transesterification catalytic reaction (51). In terms of enzymatic activity, it is possible to correlate some particular enzymatic reactions with particular regions of the primary structure of the proteins. Indeed, since the 67 kDa of topA is only able to cleave single strand DNA (52), we hypothesize that the 67 kDa part of topA is responsible for the transesterification reaction. The removed carboxy-terminus is presumably responsible for double strand DNA binding (through zinc fingers anchorage) as suggested by Tse-Dinh (53). In the case of topB, the single strand RNA or DNA binding reaction is performed by the

basic amino acids in the carboxy-terminal part (54). Recently, Zhang *et al.* (55) have shown that the carboxy-terminal domain of TopA and TopB is responsible for the substrate binding whereas the rest of the protein is responsible for the transesterification reaction. In the case of reverse gyrase, we showed previously that partial proteolytic products possess an ATP-independent activity which can only relax negatively supercoiled DNA like topA does (25). We assume that these proteolytic products possess all the motifs characteristic of the type I-5′ topoisomerases (motifs 1–10 in Fig. 3) and may bind to DNA by the putative zinc finger Zn2.

Comparison of the amino-terminal conserved regions of the three known reverse gyrases with some DNA or RNA helicases shows that some helicase motifs are present (Fig. 4B). Motif I and motif III are strictly conserved. Moreover, the sequences of motifs II and VI seem characteristic of reverse gyrases with DDxD consensus for motif II and QxxGRxSR for motif VI. Nevertheless, the enzymatic activity of this putative helicase domain is speculative. Indeed, what is a helicase? The first definition is a DNA-dependent ATPase that unwinds DNA and moves along it (56). From a mechanistic point of view, Lohman (57) proposed a 'rolling' mechanism for dimeric DNA helicases like Rep (which possesses a 3′–5′ polarity and a low processivity). The enzymatic cycle can be divided in two parts: the first is the translocation of the unbound monomer to a double strand DNA, the second step is the unwinding of this double strand region coupled to ATP hydrolysis. This ATP hydrolysis permits the recycling of the enzyme. This model may be also proposed for eIF4a (58), NPH-II (59), PriA (60), RecQ (61) and RecG (62) since their biochemical properties are similar to those of Rep. On the other hand, most advanced biochemical studies on these proteins succeeded in relating the different sequence motifs with a precise activity. Thus, the motifs Ia, Ib and II of eIF4a are responsible for ATP binding and hydrolysis, and the motifs III and VI connect the ATP hydrolysis with double strand nucleic acid binding and unwinding (49). Since the conserved motifs extend to ~300–400 amino acids in length, it is possible that the additional translocation activity of helicase is carried out by another part of these proteins or requires oligomerization. In particular, the helicase activity of eIF4A is increased by addition of the eIF4F protein (48). Moreover, by using vaccinia virus RNA helicase NPH-II, Gross and Shuman (59) recently showed that the motif VI is required only for ATP hydrolysis and RNA unwinding. Consequently, we think that the motifs described as helicase motifs in fact define an ATP-dependent unwinding activity rather than a helicase activity *per se*. We therefore propose the amino-terminal domain of reverse gyrase as a DNA unwinding domain.

The biochemical studies on reverse gyrase are consistent with this analysis. Indeed, Shibata *et al.* (42) showed that reverse gyrase possesses a DNA-dependent ATPase activity and we have previously shown that the binding of reverse gyrase induced a DNA unwinding or a left-handed DNA wrapping in an ATP-independent manner (16). Finally, DNA cleavage analysis suggests that reverse gyrase does not translocate along the DNA axis (16,63 and unpublished results of C. J. and M. N.).

Both sequence comparison data and biochemical data are consistent with a very simple hypothetical model. Starting from a relaxed circular DNA, the DNA helix is unwound by the amino-terminal part of the reverse gyrase, defining two topological domains in the DNA molecule: the unwound domain and the rest of the molecule. This unwinding consequently introduces a positive supercoiling in the rest of the molecule. The topoisomerase part of reverse gyrase would relax the unwound DNA domain. After dissociation of reverse gyrase, the result would be an increase in the linking number of the complete DNA molecule and a production of positively supercoiled DNA. In this model, ATP hydrolysis would occur for recycling the amino-terminal part of reverse gyrase.

Finally, reverse gyrase activity, first discovered in thermophilic Archaea, is also present in thermophilic Bacteria, indicating that life at high temperature requires a positive DNA supercoiling activity. It would be interesting to know the relationships between archaeal and bacterial genes, since the genomic organization of *M.kandleri* is very different from that of *Sulfolobus*. In addition, Gangloff *et al.* (64) point out the possible existence of reverse gyrase in yeast, since they demonstrated a direct interaction between SGS1 (a protein which possesses similarities to the reverse gyrase unwinding domain) and TOPIII (a type I-5′ topoisomerase) using the double hybrid method. Genetic studies indicated that *SGS1* and *TOPIII* are involved in the recombination pathway; it is possible that reverse gyrase activity is specialized in the control of this cellular process in mesophilic organisms. This idea is supported by the existence in *E.coli* of a similar pair of proteins, RecQ and topB. Finally, the same kind of interaction may exist in human cells between the*RecQ* or *BLM* gene [involved in Bloom's syndrome (65), a repair and recombination desease] and the *TOPIII* gene product recently reported by Hanai *et al.* (66). Finally, it is interesting to remember that 10 years ago, a not well characterized positive supercoiling activity was described in hypermutating myeloma cell line (67). It is possible that positive DNA supercoiling is not limited to thermophilic organisms but is more crucial for these organisms.

ACKNOWLEDGEMENTS

We thank Valérie Borde, Frédérique Gallison, Joël Pothier and Seamus O'Regan for critical reading of the manuscript, Joël Pothier and Alain Viari (Atelier de BioInformatique, Paris) for the SMARTIES package softwares and for assistance and instruction in the use of MOLSCRIPT, René Perard for modification of the shaker for the *Sulfolobus* cultures, Christophe Cullin for his advices about bank screening and Christiane Portemer for her technical assistance. We acknowledge Christiane Elie for helpful discussions about ATPase sequences comparison. We thank Wolfram Zillig and its team for the gift of *Sulfolobus* strain and for their advices about *Sulfolobus* culture. We thank Alfonso Mondragon for providing coordinates of the *E.coli* topoisomerase I crystal structure prior to their submission to the Brookhaven Protein Data Base. This research was supported by funds from CNRS. M. N. is supported by Université Versailles Saint-Quentin en Yvelines.

REFERENCES

1 Watson, J.D. and Crick, F.H. (1953) *Nature*, **171**, 964–967.
2 Roca, J. (1995) *Trends Biol. Sci.*, **20**, 156–160.
3 Caron, P.R. and Wang, J.C. (1993) In Andoh, T., Ikeda, H. and Oguro, M. (eds) *Molecular biology of DNA topoisomerases and its application to chemotherapy.* CRC Press, Boca Raton, Vol. pp. 1–18.
4 Woese, C.R. and Fox, G.E. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5088–5090.

5 Woese, C.R., Kandler, O. and Wheelis, M.L. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 4576–4579.
6 Drlica, K. (1984) *Microbiol. Rev.*, **48**, 273–289.
7 DiGate, R.J. and Marians, K.J. (1989) *J. Biol. Chem.*, **264**, 17924–17930.
8 Kikuchi, A. and Asai, K. (1984) *Nature*, **309**, 677–681.
9 Forterre, P., Mirambeau, G., Jaxel, C., Nadal, M. and Duguet, M. (1985) *EMBO J.*, **4**, 2123–2128.
10 Nakasu, S. and Kikuchi, A. (1985) *EMBO J.*, **4**, 2705–2710.
11 Nadal, M., Jaxel, C., Portemer, C., Forterre, P., Mirambeau, G. and Duguet, M. (1988) *Biochemistry*, **27**, 9102–9108.
12 Collin, R.G., Morgan, H.W., Musgrave, D.R. and Daniel, R.M. (1988) *FEMS Microbiol. Lett.*, **55**, 235–240.
13 Bouthier de la Tour, C., Portemer, C., Nadal, M., Stetter, K., Forterre, P. and Duguet, M. (1990) *J. Bacteriol.*, **172**, 6803–6808.
14 Andera, L., Mikulik, K. and Savelyeva, N.D. (1993) *FEMS Microbiol. Lett.*, **110**, 107–112.
15 Bouthier de la Tour, C., Portemer, C., Huber, R., Forterre, P. and Duguet, M. (1991) *J. Bacteriol.*, **173**, 3921–3923.
16 Jaxel, C., Nadal, M., Mirambeau, G., Forterre, P., Takahashi, M. and Duguet, M. (1989) *EMBO J.*, **8**, 3135–3139.
17 Confalonieri, F., Elie, C., Nadal, M., Bouthier de la Tour, C., Forterre, P. and Duguet, M. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 4753–4757.
18 Langer, D., Hain, J., Thuriaux, P. and Zillig, W. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 5768–5772.
19 Qureshi, S.A., Baumann, P., Rowlands, T., Khoo, B. and Jackson, S. (1995a) *Nucleic Acids Res.*, **23**, 1775–1781.
20 Qureshi, S.A., Khoo, B., Baumann, P. and Jackson, S. (1995b) *Proc. Natl. Acad. Sci. USA*, **92**, 6077–6081.
21 Yeats, S., McWilliam, P. and Zillig, W. (1982) *EMBO J.*, **1**, 1035–1038.
22 Schleper, C., Kubo, K. and Zillig, W. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 7645–7649.
23 Muskhelishvili, G., Palm, P. and Zillig, W. (1993) *Mol. Gen. Genet.*, **237**, 334–342.
24 Nadal, M., Mirambeau, G., Forterre, P., Reiter, W.-D. and Duguet, M. (1986) *Nature*, **321**, 256–258.
25 Nadal, M., Couderc, E., Duguet, M. and Jaxel, C. (1994) *J. Biol. Chem.*, **269**, 5255–5263.
26 Krah, R., Kozyavkin, S.A., Slesarev, A.I. and Gellert, M. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 106–110.
27 Bouthier de la Tour, C., Kaltoum, H., Portemer, C., Confalonieri, F., Huber, R. and Duguet, M. (1995) *Biochim. Biophys. Acta*, **1264**, 279–283.
28 Zillig, W., Kletzin, A., Schleper, C., Holz, I., Janekovic, D., Hain, J., Lanzendörfer, M. and Kristjansson, J.K. (1994) *System. Appl. Microbiol.*, **16**, 609–628.
29 Kingston, R.E. (1989) In Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (eds) *Current Protocols in Molecular Biology.* Greene Publishing Associates and Wiley-Interscience, New York, Vol. pp. 4–8.
30 Marck, C. (1988) *Nucleic Acids Res.*, **16**, 1829–1836.
31 Bellon, B. (1988) *Nucleic Acids Res.*, **16**, 1837–1846.
32 Lipman, D.J. and Pearson, W.R. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 2444–2448.
33 Altschul, S. F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) *J. Mol. Biol.*, **215**, 403–410.
34 Sturrock, S.S. and Collins, J.F. (1993) *MPsrch version 1.3,* Biocomputing Resarch Unit, University of Edinburgh, UK.
35 Cockwell, K.Y. and Giles, I.G. (1989) *Comput. Appl. Biosci.*, **5**, 227–232.
36 Dessen, P., Fondrat, C., Valencien, C. and Mugnier, C. (1990) *Comput. Appl. Biosci.*, **6**, 355–356.
37 Saurin, W. and Marlière, P. (1986) *C. R. Acad. Sci. Ser. 3*, **13**, 541–546.
38 Reiter, W.-D., Palm, P., Henschen, A., Lottspeich, F., Zillig, W. and Grampp, B. (1987a) *Mol. Gen. Genet.*, **206**, 144–153.
39 Reiter, W.-D., Palm, P., Voos, W., Kaniecki, J., Grampp, B., Schulz, W. and Zillig, W. (1987b) *Nucleic Acids Res.*, **15**, 5581–5595.
40 Reiter, W., Palm, P., Yeats, S. and Zillig, W. (1987c) *Mol. Gen. Genet.*, **209**, 270–275.
41 Reiter, W.-D., Palm, P. and Zillig, W. (1988) *Nucleic Acids Res.*, **16**, 2445–2459.
42 Shibata, T., Nakasu, S., Yasui, K. and Kikuchi, A. (1987) *J. Biol. Chem.*, **262**, 10419–10421.
43 Walker, J.E., Saraste, M., Runswick, M.J. and Gay, N.J. (1982) *EMBO J.*, **1**, 945–951.
44 Gorbalenya, A.E. and Koonin, E.V. (1989) *Nucleic Acids Res.*, **17**, 8413–8440.
45 Reuven, N.B., Koonin, E.V., Rudd, K.E. and Deutscher, M.P. (1995) *J. Bacteriol.*, **117**, 5393–5400.
46 Linder, P., Lasko, P.F., Ashburner, M., Leroy, P., Nielsen, P.J., Nishi, K., Schnier, J. and Slonimsky, P.P. (1989) *Nature*, **337**, 121–122.
47 Schmid, S.R. and Linder, P. (1992) *Mol. Microbiol.*, **6**, 283–292.
48 Pause, A. and Sonenberg, N. (1992) *EMBO J.*, **11**, 2643–2654.
49 Pause, A., Méthot, N. and Sonenberg, N. (1993) *Mol. Cell. Biol.*, **13**, 6789–6798.
50 Lima, C.D., Wang, J.C. and Mondragon, A. (1994) *Nature*, **367**, 138–146.
51 Lynn, R. and Wang, J.C. (1989) *Proteins: Struct. Funct. Genet.*, **6**, 231–239.
52 Lima, C.D., Wang, J.C. and Mondragon, A. (1993) *J. Mol. Biol.*, **232**, 1213–1216.
53 Tse-Dinh, Y.-C. (1991) *J. Biol. Chem.*, **266**, 14317–14320.
54 Zhang, H.L. and DiGate, R.J. (1994) *J. Biol. Chem.*, **269**, 9052–9059.
55 Zhang, H.L., Malpure, S., Li, Z., Hiasa, H. and DiGate, R.J. (1996) *J. Biol. Chem.*, **271**, 9039–9045.
56 Kornberg, A. and Baker, T.A. (1992) In Freeman, New York.
57 Lohman, T.M. (1993) *J. Biol. Chem.*, **268**, 2269–2272.
58 Rozen, F., Edery, I., Meerovitch, K., Dever, T.E., Merrick, W.C. and Sonenberg, N. (1990) *Mol. Cell. Biol.*, **10**, 1134–1144.
59 Gross, C.H. and Shuman, S. (1996) *J. Virol.*, **70**, 1706–1713.
60 Lasken, R.S. and Kornberg, A. (1988) *J. Biol. Chem.*, **263**, 5512–5518.
61 Umezu, K. and Nakayama, H. (1993) *J. Mol. Biol.*, **230**, 1145–1150.
62 Whitby, M.C., Vincent, S.D. and Lloyd, R.G. (1994) *EMBO J.*, **13**, 5220–5228.
63 Kovalsky, O.I., Kozyavkin, S.A. and Slesarev, A.I. (1990) *Nucleic Acids Res.*, **18**, 2801–2805.
64 Gangloff, S., McDonald, J.P., Bendixen, C., Arthur, L. and Rothstein, R. (1994) *Mol. Cell. Biol.*, **14**, 8391–8398.
65 Ellis, N.A., Groden, J., Ye, T.-Z., Straughen, J., Lennon, D.J., Ciocci, S., Proytcheva, M. and German, J. (1995) *Cell*, **83**, 655–666.
66 Hanai, R., Caron, P.R. and Wang, J.C. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 3653–3657.
67 Lazo, P.A. (1985) *Biochem. J.*, **231**, 185–188.