Molecular cloning of the transcription factor TFIIB homolog from *Sulfolobus shibatae*

(Archaea/archaeobacteria/evolution/transcription)

SOHAIL A. QURESHI, BERNARD KHOO, PETER BAUMANN, AND STEPHEN P. JACKSON*

Wellcome/Cancer Research Campaign Institute, Tennis Court Road, Cambridge, CB2 1QR, United Kingdom; and Department of Zoology, University of Cambridge, Downing Street, Cambridge, CB2 3EJ, United Kingdom

Communicated by Carl R. Woese, University of Illinois, Urbana, IL, March 7, 1995

ABSTRACT The Archaea (archaeobacteria) constitute a group of prokaryotes that are phylogenetically distinct from Eucarya (eukaryotes) and Bacteria (eubacteria). Although Archaea possess only one RNA polymerase, evidence suggests that their transcriptional apparatus is similar to that of Eucarya. For example, Archaea contain a homolog of the TATA-binding protein which interacts with the TATA-box like A-box sequence upstream of many archaeal genes. Here, we report the cloning of a *Sulfolobus shibatae* gene that encodes a protein (transcription factor TFB) with striking homology to the eukaryotic basal transcription factor TFIIB. We show by primer extension analysis that transcription of the *S. shibatae* TFB gene initiates 27 bp downstream from a consensus A-box element. Significantly, *S. shibatae* TFB contains an N-terminal putative metal-binding region and two imperfect direct repeats—structural features that are well conserved in eukaryotic TFIIBs. This suggests that TFB may perform analogous functions in Archaea and Eucarya. Consistent with this, we demonstrate that *S. shibatae* TFB promotes the binding of *S. shibatae* TBP to the A-box element of the *Sulfolobus* 16S/23S rRNA gene. Finally, we show that *S. shibatae* TFB is significantly more related to TFB of the archaeon *Pyrococcus woesei* than it is to eukaryotic TFIIBs. These data suggest that TFB arose in the common archaeal/eukaryotic ancestor and that the lineages leading to *P. woesei* and *S. shibatae* separated after the divergence of the archaeal and eukaryotic lines of descent.

Archaea (formerly archaeobacteria) constitute a heterogeneous group of prokaryotes, many of which thrive under extremes of pH, salinity, and/or temperature (1–3). Although Archaea resemble eubacteria in morphology, 16S rRNA sequence analyses by Woese and colleagues in the late 1970s established that they are at least as diverged evolutionarily from eubacteria as they are from eukaryotes (4). Subsequent sequence comparisons of other molecules have reinforced this conclusion and have led to the proposal that life on earth is divided into three distinct and equivalent primary domains: Eucarya, Bacteria, and Archaea (5).

The molecular organization of Archaea appears to resemble that of eukaryotic cells in many ways. Of the similarities noted so far, perhaps the most striking are at the level of transcription. For example, approximately 27 bp upstream from the transcription initiation site of most archaeal genes is an A-box motif (consensus TTTAWA; where W = A or T) that plays an important role in specifying efficient and accurate transcription initiation (6, 7). Archaeal A boxes therefore closely resemble TATA boxes of eukaryotic RNA polymerase II promoters in both sequence and function. Furthermore, as in eukaryotes, transcription from archaeal promoters depends on factors that are readily separated from the core RNA polymerase (8–10). Most significantly perhaps, the TATA-binding

protein (TBP), which is an essential component of the eukaryotic RNA polymerase I, II, and III transcription systems (11), has been found recently in *Pyrococcus woesei*, *Thermococcus celer*, and *Sulfolobus shibatae* (12–14). In addition, a *P. woesei* protein has been identified that bears strong homology to eukaryotic basal transcription factors TFIIB and BRF (15, 16), which recruit RNA polymerases II and III, respectively, to appropriate transcriptional promoters (17–20).

Evolutionary comparisons indicate that Archaea are separated into two highly divergent kingdoms: the Crenarchaeota, comprising sulfur-metabolizing thermophiles such as *Sulfolobus*, and the Euryarchaeota, which includes the methanogens and hyperthermophiles such as *T. celer* and *P. woesei* (5). Most studies suggest that the Archaea are monophyletic—that is, the Crenarchaeota and Euryarchaeota are more closely related to one another than they are to either Eucarya or Bacteria. However, some analyses suggest that the two archaeal kingdoms are not specifically related, and that Eucarya might be more related to the Crenarchaeota than the Euryarchaeota are (21, 22). In light of this controversy, it would be informative to determine whether TBP and transcription factor TFB exist in Crenarchaeota and, if so, to investigate their relationships to homologs in the other systems. Here, we report the cloning of a gene encoding *S. shibatae* TFB,[†] and we show that this protein contains all the structural features that are conserved in eukaryotic TFIIBs. In addition to studying the functional properties of *S. shibatae* TFB, we use the sequence of this factor to investigate the phylogenetic relationships between *S. shibatae*, Eucarya, and other Archaea.

MATERIALS AND METHODS

Expression of Proteins and Preparation of Antisera. *P. woesei* TFB was synthesized in *Escherichia coli* using expression vector pQE30 (Qiagen) and purified on a Ni-NTA matrix (Qiagen) according to the manufacturer's instructions. Antiserum was generated by injecting 100 μ g into a rabbit at monthly intervals. For expression of *S. shibatae* TFB, the open reading frame (ORF) was inserted into pGEX-2TKP and the resulting protein was purified on glutathione-agarose beads (Pharmacia) as previously (20). *S. shibatae* TBP was expressed and purified as described elsewhere (14).

Bacterial Growth and Preparation of Genomic DNA. *S. shibatae* and *Sulfolobus solfataricus* were grown as described elsewhere (14). To prepare genomic DNA, cells were resuspended in TE (10 mM Tris-HCl, pH 8.0/1 mM EDTA) containing 0.5% SDS, and proteinase K (Sigma) was added to 1 mg/ml. After incubation for 2–4 hr at 37°C and extraction

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

Abbreviations: ORF, open reading frame; TBP, TATA-binding protein; GST, glutathione S-transferase; EMSA, electrophoretic mobility shift assay.

*To whom reprint requests should be addressed.

[†]The sequence reported in this paper has been deposited in the GenBank data base (accession no. U20899).

with phenol, DNA was precipitated with ethanol and spooled onto a glass rod.

Construction and Screening of a λ Expression Library. *S. solfataricus* genomic DNA was partially digested with *Sau*3A and fragments from 0.5 to 2.5 kb were ligated to *Bam*HI-digested and dephosphorylated λ ZAP Express arms (Stratagene); packaging was performed according to the manufacturer's instructions. Plaques were screened with anti-TFB antiserum and alkaline phosphatase-conjugated secondary antibody according to standard procedures (23). Plasmids from the recombinant phage were excised by using the ExAssist system (Stratagene).

DNA Sequencing and Sequence Analysis. DNA sequencing for the *S. shibatae* TFB was by Sequenase version 2.0 (United States Biochemical) according to the manufacturer's instructions. Data base searches were performed by using the FASTA program and the electronic-mail-based BLAST server (24, 25). Sequence alignments were with the GCG package, using the BESTFIT, PILEUP, and LINEUP programs (26), and with CLUSTALW (14) (27). Phylogenetic comparisons were conducted by using CLUSTALW and the PROTDIST and KITSCH programs in the Phylogeny Inference Package (PHYLIP) (28). Trees were displayed by using TREE TOOL 2.0.1 (29).

Electrophoretic Mobility Shift Assay (EMSA). A complementary 19-mer was annealed with a 65-base oligonucleotide containing the A and B boxes of the *S. shibatae* 16/23S rRNA promoter, and the resulting product was made completely double-stranded by using the Klenow fragment of DNA polymerase (Boehringer Mannheim) in the presence of [α - 32 P]dATP. Binding conditions were as described previously (12), except that incubations were at 37°C for 40 min.

Isolation of RNA and Primer Extension Analysis. RNA was isolated from *S. shibatae* by using the Ultraspec-II RNA isolation system (Biotecx Laboratories, Houston). A 20-base oligonucleotide (5'-AAACCCCTATCGTGAATAGT-3'; corresponding to nucleotides 260–241 of the *S. shibatae* TFB gene) was labeled with T4 polynucleotide kinase (Boehringer Mannheim) and [γ - 32 P]ATP and was coprecipitated with 50 μ g of *S. shibatae* RNA. The precipitate was then resuspended in 10 μ l of 100 mM Tris-HCl, pH 8.5/16 mM MgCl₂/60 mM KCl/2 mM dithiothreitol, and annealing was performed at 40°C for 25 min. Subsequently, 10 μ l of the extension mix, containing 0.25 mM each of the dNTPs and 25 units of avian myeloblastosis virus reverse transcriptase (Boehringer Mann-

heim), was added and the reaction mixture was incubated for 1 hr at 40°C. Nucleic acids were then precipitated with ethanol and resolved on a denaturing 6% polyacrylamide gel.

RESULTS

Cloning the Gene Encoding *S. shibatae* TFB. When we used an antiserum against recombinant *P. woesei* TFB in Western blot assays, it was found to recognize a single polypeptide of around 34 kDa in extracts of *Sulfolobus* cells (data not shown). To clone the gene encoding this *Sulfolobus* protein, the antibody was used to screen a *S. solfataricus* expression library. By screening 10⁵ plaques, five positive clones were identified and isolated, and plasmids were obtained by excision from the phage DNA. Restriction enzyme analysis and preliminary DNA sequencing revealed that all clones contained an identical insert of approximately 3.0 kb.

The complete sequencing of one clone identified an ORF able to encode a polypeptide with striking homology to eukaryotic TFIIB. Using primers based on the *S. solfataricus* TFB gene, we also amplified and cloned the homologous gene from *S. shibatae*. The DNA sequence and inferred translation product of *S. shibatae* TFB are shown in Fig. 1. Consistent with the close evolutionary relationship between the two *Sulfolobus* species, the TFB proteins of *S. shibatae* and *S. solfataricus* are identical. Since there is no appropriately positioned ATG codon, we assume that translation initiates at the closely related valine codon GTG, and that *S. shibatae* TFB is composed of 309 amino acid residues, has a molecular mass of 35 kDa, and has a pI of 9.5.

Mapping the Transcriptional Initiation Site of the *S. shibatae* TFB Gene. To map the transcription initiation site of the TFB gene, a 20-base oligonucleotide was used in a primer extension assay with *S. shibatae* RNA. This yielded a 260-nucleotide run-off product mapping to the sequence AAAA (Fig. 2) that is approximately 30 bp upstream of the predicted *S. shibatae* TFB translation initiation codon and 27 bp downstream of a TATA-like A-box element (Fig. 1). Since the *S. shibatae* TFB promoter conforms well to the archaeal promoter consensus, this A-box motif is likely to play an important role in transcriptional initiation. As transcription termination sites in *S. shibatae* genes generally map to thymine-rich regions (30), we speculate that transcription will terminate at the

FIG. 1. DNA sequence of the *S. shibatae* TFB gene and its translation product. The putative A box and transcriptional terminator are highlighted by reverse shading. In the absence of an appropriately positioned AUG codon, translation is assumed to initiate at the underlined Val residue. The four putative metal-coordinating residues in the metal-binding region and the direct repeat elements are indicated.

FIG. 2. Mapping the transcriptional initiation site of the *S. shibatae* TFB gene. Primer extension analysis was carried out with total *S. shibatae* RNA and a 20-mer oligonucleotide. The run-off product (P) was coelectrophoresed with the products of sequencing reactions G, A, T, and C carried out with the same oligonucleotide. The 260-nt band is shown, as is the DNA sequence from where transcription initiates.

sequence TTTTCTTT located approximately 50 bp downstream from the *S. shibatae* TFB translation stop codon (Fig. 1).

Functional Properties of *S. shibatae* TFB. In eukaryotes, TFIIB promotes the binding of TBP to promoter sequences (17–19). To determine whether *S. shibatae* TFB functions analogously, we expressed it in *E. coli* fused to glutathione S-transferase (GST). After purification to homogeneity, this polypeptide was employed in EMSAs together with purified recombinant *S. shibatae* TBP and an oligonucleotide probe corresponding to the *S. shibatae* 16/23S rRNA promoter. Consistent with previous observations (12), TBP alone binds poorly to the A box, and a TBP-DNA complex is not observed unless a large amount of TBP is used (Fig. 3). Addition of the *S. shibatae* GST-TFB fusion protein, however, stimulates binding of TBP to DNA dramatically and results in the appearance of a ternary complex containing TFB, TBP, and DNA. Importantly, addition of unfused GST does not stimulate TBP binding and does not yield a ternary complex (data not shown). Furthermore, competition experiments with wild-type and A-box-mutated oligonucleotides reveal that TBP and TBP/TFB interact specifically with A-box sequences (data not shown). These results demonstrate that *S. shibatae* TFB increases the rate of TBP binding to DNA and/or increases the stability of TBP-DNA complexes.

Protein Sequence Comparisons. The sequence of *S. shibatae* TFB is shown in Fig. 4A aligned with *P. woesei* TFB, TFIIBs of various eukaryotes, and BRF proteins of three yeast species.

FIG. 3. *S. shibatae* TFB stimulates binding of *S. shibatae* TBP to the archaeal A-box element. EMSAs were conducted with a radiolabeled A-box-containing oligonucleotide, either in the absence of TBP (lane 1) or in the presence of 20 ng of TBP (lanes 2–6) or 200 ng of TBP (lanes 7–10). Reactions were supplemented with 2, 20, or 200 ng (lanes 3–5 and 8–10) and 400 ng (lane 6) of purified GST-TFB or no GST-TFB (lanes 2 and 7).

As with other “B factors,” *Sulfolobus* TFB contains a putative metal-binding region towards its N terminus. However, whereas the four putative metal-coordinating residues are always Cys or His in other B factors characterized (asterisks in Fig. 4A), in *Sulfolobus* TFB, two of these are replaced by Ser or Thr. Although unusual, this is not necessarily incompatible with metal binding, because the Ser and Thr residues might contribute to this process by means of the lone oxygen atom electron pairs in their hydroxyl groups.

Another conserved feature of eukaryotic B factors is a region containing two imperfect direct repeats of approximately 75 amino acid residues. This feature is also present in *S. shibatae* TFB. Interestingly, the similarity between the two repeats of *S. shibatae* TFB is greater than the similarity between the two repeats of individual TFIIBs or BRFs. Moreover, inspection of the *S. shibatae* sequence reveals that the repeats extend further than has been noted previously from analyzing the eukaryotic factors, and they continue to the extreme C terminus of the protein (Fig. 4A). For this extended repeat unit, the fraction of identical residues between the repeats of human TFIIB, *Saccharomyces cerevisiae* BRF, *P. woesei* TFB, and *S. shibatae* TFB are 21.5%, 15.2%, 27.6%, and 39.1%, respectively. One interpretation for this is that a duplication event generating *S. shibatae* TFB occurred independently of and subsequent to the event(s) generating the other factors. However, sequence comparisons based on the individual direct repeats suggest strongly that a single duplication event took place in the common archaeal/eukaryal ancestor and imply that the rate of repeat sequence divergence has been less rapid in *S. shibatae* TFBs than in the other factors.

Numerical values for the degree of homology between various B factors are presented in Table 1. *S. shibatae* TFB is clearly more related to *P. woesei* TFB (51.7% identity) than to eukaryotic TFIIBs (around 30% identity) and BRFs (average 22% identity). To assess more precisely the relationship between the various B factors, the aligned sequences were used to construct a phylogenetic tree (Fig. 4B). Importantly, in spite of the great divergence between *S. shibatae* and *P. woesei* TFBs, these factors are clearly more related to one another than either is to the B factors of eukaryotes. An attempt to root the tree was made on the basis of the assumption that the two repeats arose by a duplication event before the Archaea and Eucarya diverged, and this showed that the root, although difficult to place exactly, definitely does not lie

FIG. 4. *S. shibatae* TFB is highly related to other B factors. (A) Alscript output (31) of a CLUSTALW 1.4 alignment of *S. shibatae* TFB and related factors; for BRF proteins, only the N-terminal homologous region is displayed. Highlighted in reverse shading are residues of other proteins that are identical to the *S. shibatae* sequence. Locations of the putative metal-binding region and direct repeats are indicated. Swiss-Prot (S) and European Molecular Biology Laboratory (E) accession numbers are as follows: *Homo sapiens* TFIIB (Hs IIB), Q00403 (S); *Rattus norvegicus* TFIIB (Rn IIB), P29053 (S); *Drosophila melanogaster* TFIIB (Dm IIB), P29052 (S); *Xenopus laevis* TFIIB (Xi IIB), P29054 (S); *Kluyveromyces fragilis* TFIIB (Kf IIB), P29055 (S); *Saccharomyces cerevisiae* TFIIB (Sc IIB), P29056 (S); *S. cerevisiae* BRF (Sc BRF), P29056 (S); *K. lactis* BRF (Kl BRF), Z47203 (E); *Candida albicans* BRF (Ca BRF), Z47202 (E); and *P. woesei* TFB (Pw B), X70668 (E). (B) Phylogenetic tree of TFB and related factors. This is a radial plot of a phylogenetic tree inferred from the alignments in A. The number next to a node refers to the number of bootstrap trials (out of 1000) which support the branching pattern at that node. Scale bar, 10 mutations per 100 positions.

within the archaeal grouping (data not shown). Our results therefore strongly suggest that the Archaea constitute a monophyletic group, and they imply that neither the euryarchaeotic nor the crenarchaeotic line of descent is specifically related to eukaryotes.

DISCUSSION

We have described the cloning of a *S. shibatae* gene encoding a protein with strong homology to eukaryotic basal transcrip-

tion factors TFIIB and BRF. The existence of TFB in *Sulfolobus* suggests that the transcriptional apparatus of this organism is fundamentally homologous to the nuclear transcription systems of eukaryotes and is very similar to that of other Archaea such as *P. woesei*. Consistent with this, we have recently isolated the gene for a TBP homolog from *S. shibatae* (14). Given that TBP and TFIIB homologs have been identified in both archaeal kingdoms, it seems reasonable to speculate that these molecules will be universal in Archaea.

Table 1. Numerical comparisons between *S. shibatae* TFB and related factors for repeats plus the metal-binding region

	Ss B	Pw B	Kl IIB	Sc IIB	Dm IIB	Xl IIB	Rn IIB	Hs IIB	Ca BRF	Kl BRF	Sc BRF
Ss B		0.517	0.299	0.284	0.349	0.330	0.330	0.333	0.199	0.230	0.222
Pw B	0.729		0.307	0.276	0.314	0.326	0.322	0.322	0.234	0.257	0.249
Kl IIB	1.855	1.904		0.705	0.368	0.372	0.356	0.356	0.207	0.203	0.249
Sc IIB	1.947	1.999	0.398		0.352	0.356	0.345	0.345	0.176	0.176	0.203
Dm IIB	1.352	1.437	1.236	1.343		0.808	0.812	0.812	0.238	0.226	0.245
Xl IIB	1.407	1.393	1.198	1.270	0.227		0.950	0.950	0.218	0.230	0.245
Rn IIB	1.422	1.407	1.279	1.323	0.232	0.055		0.996	0.218	0.222	0.238
Hs IIB	1.412	1.401	1.282	1.326	0.233	0.055	0.003		0.218	0.222	0.238
Ca BRF	2.165	1.974	2.511	2.565	2.080	2.122	2.174	2.167		0.640	0.686
Kl BRF	2.021	1.873	2.511	2.673	2.116	2.121	2.201	2.193	0.481		0.785
Sc BRF	1.980	1.904	2.173	2.378	2.051	2.018	2.100	2.092	0.395	0.252	

Figures constituting the upper right triangle are the fraction of identical amino acids between pairs of sequences aligned in Fig. 4A. Figures constituting the lower left triangle are the average number of amino acid replacements, per position, between pairs, calculated by using the Dayhoff PAM (accepted point mutation) option in the PROTDIST program (28). Abbreviations are as in Fig. 4A.

The homology between TFB and eukaryotic B factors can be divided into three domains. The first corresponds to the N-terminal putative metal-binding region that, in eukaryotic TFIIBs, plays an important role in recruiting RNA polymerase II/transcription factor TFIIF to the transcription complex and may also function in transcriptional activation (17–19, 32). The second area of homology is the region of approximately 95 amino acid residues immediately C-terminal to the putative metal finger. Although no specific function for this region has been ascribed, its conservation suggests that it plays an important role in transcription in Archaea and Eucarya. Consistent with this, it has been shown that, in *Saccharomyces cerevisiae* TFIIB, Glu-62 and Arg-78 within this region play important roles in specifying transcription start site recognition (33). It is interesting to note that these amino acid residues are present in all eukaryotic TFIIBs as well as in *P. woesei* and *S. shibatae* TFB. The third region of homology spans the two direct repeat elements, which, in eukaryotic TFIIBs and BRFs, have been implicated in interacting with TBP, RNA polymerase, and transcriptional activators (17–20, 32).

These homologies suggest that *S. shibatae* TFB functions in an analogous manner to eukaryotic TFIIB and BRF. Indeed, we have shown that, like eukaryotic TFIIB, *S. shibatae* TFB enters into a ternary complex with TBP and promoter DNA, and it increases the rate of assembly and/or stability of TBP-DNA complexes. Since TFIIB recruits RNA polymerase II, and BRF contacts the 34-kDa subunit of RNA polymerase III (20), it will be of interest to see whether *S. shibatae* TFB interacts with the archaeal RNA polymerase. Finally, in light of the fact that TFIIB interacts with certain upstream activators, it is tempting to speculate that similar molecules might exist in Archaea. The availability of an *S. shibatae* *in vitro* transcription system should prove very useful in identifying additional components of the archaeal transcriptional apparatus, as well as characterizing TFB, TBP, and RNA polymerase. These studies should provide a framework for understanding gene regulation in Archaea and will also yield important insights into the evolution and molecular mechanisms of transcription in eukaryotes.

We are grateful to Carl Woese for suggestions and helpful comments, B. Cherolambus and H. Connaris for a starter culture of *S. shibatae*, and members of the S.P.J. laboratory for advice and encouragement. B.K. is supported by a Wellcome Trust Prize Studentship. This work was funded by Grant G9420496 MB from the Medical Research Council.

- Kates, M., Kushner, D. J. & Matheson, A. T. (1993) *The Biochemistry of Archaea (Archaeobacteria)*, New Comprehensive Biochemistry (Elsevier, Amsterdam), Vol. 26.
- Barns, S. M., Fundyga, R. E., Jeffries, M. W. & Pace, N. R. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 1609–1613.
- DeLong, E. F., Wu, K. Y., Prezelin, B. B. & Jovine, R. V. M. (1994) *Nature (London)* **371**, 695–697.
- Woese, C. R. & Fox, G. E. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5088–5090.
- Woese, C. R., Kandler, O. & Wheelis, M. L. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 4576–4579.
- Reiter, W.-D., Hüdepohl, W. & Zillig, W. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 9509–9513.
- Hausner, W., Frey, G. & Thomm, M. (1991) *J. Mol. Biol.* **222**, 495–508.
- Frey, G., Thomm, M., Brudigam, B., Gohl, H. P. & Hausner, W. (1990) *Nucleic Acids Res.* **18**, 1361–1367.
- Hüdepohl, U., Reiter, W.-D. & Zillig, W. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 5851–5855.
- Hausner, W. & Thomm, M. (1993) *J. Biol. Chem.* **268**, 24047–24052.
- White, R. J. & Jackson, S. P. (1992) *Trends Genet.* **8**, 284–288.
- Rowlands, T., Baumann, P. & Jackson, S. P. (1994) *Science* **264**, 1326–1329.
- Marsh, T. L., Reich, C. I., Whitelock, R. B. & Olsen, G. J. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 4180–4184.
- Qureshi, S. A., Baumann, P., Rowlands, T., Khoo, B. & Jackson, S. P. (1995) *Nucleic Acids Res.*, in press.
- Creti, R., Londei, P. & Cammarano, P. (1993) *Nucleic Acids Res.* **21**, 2942.
- Ouzounis, C. & Sander, C. (1992) *Cell* **71**, 189–190.
- Buratowski, S. & Zhou, H. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 5633–5637.
- Hisatake, K., Roeder, R. G. & Horikoshi, M. (1993) *Nature (London)* **363**, 744–747.
- Ha, I., Roberts, S., Maldonado, E., Sun, X., Kim, L.-U., Green, M. & Reinberg, D. (1993) *Genes Dev.* **7**, 1021–1032.
- Khoo, B., Brophy, B. & Jackson, S. P. (1994) *Genes Dev.* **8**, 2879–2890.
- Lake, J. A. (1988) *Nature (London)* **331**, 184–186.
- Rivera, M. C. & Lake, J. A. (1992) *Science* **257**, 74–76.
- Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K. (1991) *Current Protocols in Molecular Biology*, ed. K. Janssen (Wiley, New York).
- Pearson, W. R. & Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 2444–2448.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipmann, D. (1990) *J. Mol. Biol.* **215**, 403–410.
- Deveraux, J., Haeberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395.
- Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994) *Nucleic Acids Res.* **22**, 4673–4680.
- Felsenstein, J. (1993) PHYLIP, Phylogeny Inference Package (Dept. of Genetics, Univ. of Washington, Seattle), Version 3.5C.
- Maciukenas, M. (1994) TREE TOOL 2.0.1 (Ribosomal RNA Database Project, Illinois State Univ., Normal).
- Reiter, W.-D., Palm, P. & Zillig, W. (1988) *Nucleic Acids Res.* **16**, 2445–2459.
- Barton, G. J. (1993) *Protein Eng.* **6**, 37–40.
- Roberts, S. G. E. & Green, M. R. (1994) *Nature (London)* **371**, 717–720.
- Pinto, I., Wu, W.-H., Na, J. G. & Hampsey, M. (1994) *J. Biol. Chem.* **269**, 30569–30573.