Nucleic Acids Research, Vol. 19, No. 14 3861–3865

Striking homology of the 'variable' N-terminal as well as the 'conserved core' domains of the mouse and human TATA-factors (TFIID)

Taka-aki Tamura[1]*, Kohsuke Sumita[1,2], Ichiroh Fujino[1,2], Atsuo Aoyama[1]+, Masami Horikoshi[3], Alexander Hoffmann[3], Robert G.Roeder[3], Masami Muramatsu[4] and Katsuhiko Mikoshiba[1,3]
[1]Division of Behavior and Neurobiology, National Institute for Basic Biology, Myodaiji-cho, Okazaki 444, [2]Division of Macromolecular Function, Institute for Protein Research, Osaka University, 3-2 Yamada-oka, Suita 565, Japan, [3]Laboratory of Biochemistry and Molecular Biology, The Rockefeller University, New York, NY 10021, USA and [4]Department of Biochemistry, Faculty of Medicine, The University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113, Japan

Received May 12, 1991; Revised and Accepted July 1, 1991 DDBJ accession no. D01034

ABSTRACT

A complementary DNA (cDNA) encoding a mouse TFIID (mIID) was isolated from mouse brain cDNA libraries. The 316 amino acid sequence deduced from cDNA sequences revealed the presence of an amino-terminal region enriched in serine, threonine, and proline (STP-cluster), an uninterrupted stretch of 13 glutamine residues (Q-run), a second STP-cluster, and a conserved carboxy-terminal region. Amino acid sequences of the first STP-cluster and the conserved carboxy-terminal region were identical to those of the human TFIID (hIID). However, the Q-run was considerably shorter than that in hIID and sequences in the second STP-cluster diverged from those of the hIID. The murine TFIID transcript is expressed as a 2 kilobase poly(A)+ RNA in the mouse brain. Southern blot analysis identified a single gene copy per haploid mouse genome.

INTRODUCTION

The general transcription factors for RNA polymerase II identified in human cells have been designated TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIG (1, and references therein). Human TFIID (hIID) binds to the TATA box element (2, 3) and initiates the ordered assembly of RNA polymerase II and the other general factors into a functional preinitiation complex (4, 5). The demonstration of a yeast TFIID (yIID) that was functionally interchangeable with the hIID (6–8) led to the total purification of this factor (9) and the cloning of corresponding cDNA (10–14). This was followed by the cloning of TFIID cDNAs from fission yeast (15, 16), plants (17), *Drosophila* (18, 19), and human (20–22). Sequence comparisons revealed a highly conserved 180 residue C-terminal domain, which earlier mutational studies had shown to be necessary and sufficient for basal level transcription by the yIID (23). In contrast, these analyses failed to reveal significant sequence homologies in the N-terminal domains.

Given previous indications that hIID is a target for the action of at least some regulatory proteins (24–27), as well as functional comparisons of natural and recombinant TFIID species, these results led to the speculation that the variable N-terminal regions might be involved in species-specific interactions with regulatory factors (20, 21, 28). However, the ability of several mammalian activators to enhance promoter activity via yeast TFIID in HeLa cell-derived systems (22, 29) may indicate that at least some interactions with regulatory factors might proceed via the conserved core domain and suggest alternate or accessory roles for the unique N-termini of hIID. To further investigate structure-function and evolutionary relationships in mammalian TFIIDs, we have isolated murine TFIID cDNA and demonstrated striking sequence conservations between human and murine TFIIDs.

MATERIALS AND METHODS

Cloning and Sequencing of the Mouse TFIID cDNA

Two cDNA libraries in *lambda-gt11* were made from cerebellum poly(A)+RNA of ICR mice by priming with oligo(dT) and with a random hexamer (30). Libraries were screened as previously described (31) using a ^{32}P-labeled DNA probe corresponding to nucleotide positions 576–971 (encoding amino acid residues 160–291) in a hIID cDNA (20). Nucleotide sequences were determined by the dideoxy chain termination method in each direction.

* To whom correspondence should be addressed at Department of Biochemistry, Saitama Medical School, Moroyama-machi, Saitama 350-04, Japan

+ Present address: Biotechnology Research Laboratory, Tosoh Co. Ltd, 2743-1 Hayakawa, Ayasa, Kanagawa, Japan

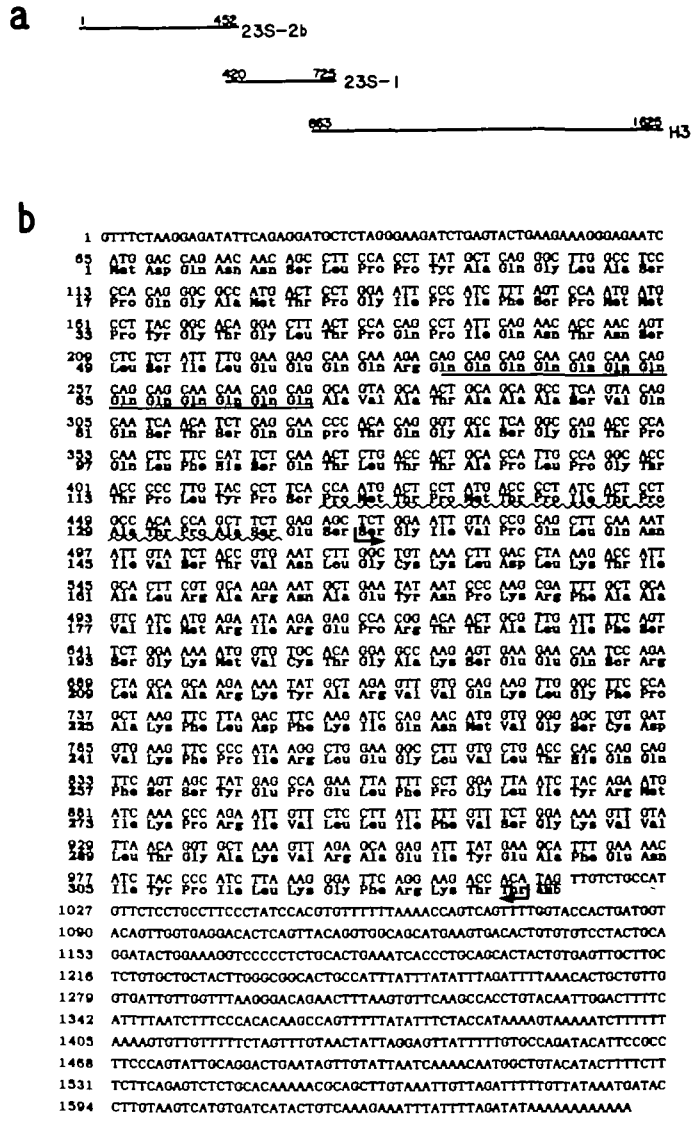

Fig. 1. Structure of the mouse TFIID cDNA. a; Schematic representation of three independent overlapping clones isolated from mouse brain cDNA libraries. b; Nucleotide and predicted amino acid sequences of the mIID Shown is the combined sequence of three overlapping clones, containing a 948 bp ORF encoding the 316 amino acid mIID. Straight and wave lines show the Q-run and the pro-ser-thr or its derivative motifs, respectively. The conserved C-terminal core is specified by the region between the two arrows.

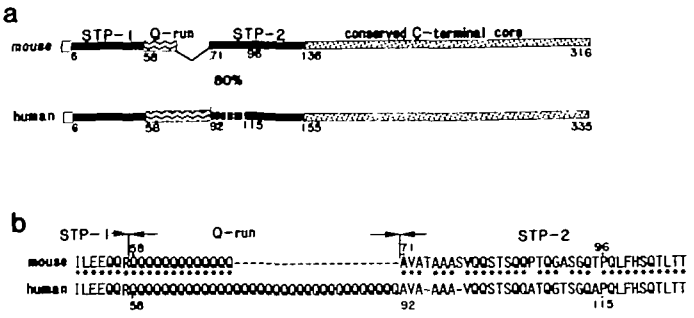

Fig. 2. Comparisons of mouse and human TFIID amino acid sequences. a; Schematic representation of the mouse and human (20) TFIIDs. Amino acid positions of the STP-1, Q-run, STP-2 and conserved C-terminal core are shown Residues 71 to 95 show 80% sequence identity between mIID and hIID. b; Amino acid sequences around the Q-run. Amino acids are indicated by the single letter code. Short bars indicated deleted residues while asterisks indicate identities between the two sequences.

Southern and Northern Blot Analyses

Southern blot analysis under stringent hybridization conditions was carried out by the standard method with minor modifications (32). Briefly, 12 μg amounts of the mouse DNA were digested with *Bam*HI and *Hin*dIII, electrophoresed on a 0.7% agarose gel, blotted onto a nylon membrane, and probed with a random primer-labeled mouse TFIID (mIID) cDNA fragment (8×10^8 cpm/μg) that extended from nucleotide positions 486 to 1006 and covered the conserved C-terminal region. The hybridization solution contained 10^6 cpm/ml probe, 5×Denhardt's solution, 100 μg/ml denatured herring testis DNA, 5×SSC, 50 mM sodium phosphate (pH 6.5), 0.1% SDS and 50% deionized formamide. The blot was incubated with the probe at 42°C for 16 hr, washed in 0.1×SSC, 0.1% SDS at room temperature and exposed to Kodak XAR films. A mouse myelin basic protein (MBP) gene fragment from nucleotides −1318 to +223 relative to the transcription start site (33) was also used as a probe for checking the DNA digestion and for calibration of the copy number. For Northern blot analysis (30), 2 μg of poly(A)$^+$RNA from the mouse brain was analyzed using a mIID probe (8×10^8 cpm/μg) extending from positions 369−675.

RESULTS

The TFIID species encoded by the several cDNAs reported thus far show strong sequence conservations over 180 residues in their respective C-termini (conserved C-terminal core). Assuming sequence conservation at both the protein and DNA level, we employed a hIID cDNA-derived probe encoding 132 residues of the conserved core to screen an oligo(dT)- and random hexamer-primed mouse brain cDNA libraries. The first screens (1×10^7 phage each) yielded 2 positive clones from the oligo(dT)-primed library and 6 from the random hexamer-primed library. These clones (including 23S-1 and H3, Figure 1a) contained sequences related to the conserved C-terminal core of hIID, but lacked the expected N-terminal sequences. The mIID sequence from nucleotide positions 369 to 675 (below) and the hIID sequence from nucleotide positions 1 to 183 were then used as probes to screen the random hexamer-primed library, which yielded one positive clone (23S-2b) carrying a part of the conserved C-terminal core and preceding N-terminal sequences. Overlapping clones 23S-2b, 23S-1 and H3 (Fig. 1a) were then used to construct a 1652 base pair cDNA which contained one long open reading frame (ORF) begining at position 65 (Fig. 1b). The ORF encoded a 316 amino acid polypeptide which was similar in size to the 335 residue hIID (20). The overall amino acid sequence showed 92% identity with the hIID sequence, while the C-terminal 221 amino acid sequence was identical with the corresponding region of the hIID (Fig. 1b, 2a). Nucleotide sequences between mIID and hIID cDNAs revealed a high (90%) sequence identity in the coding regions. Sequence similarities were weak in the short 5'- and proximal 3'-non-coding regions (Fig. 3a), but stronger in the more distal 3'-untranslated region (downstream from 1200) (Fig. 3b).

Analysis of the mIID amino acid sequence revealed motifs observed in other site-specific DNA binding proteins (Figs. 1b

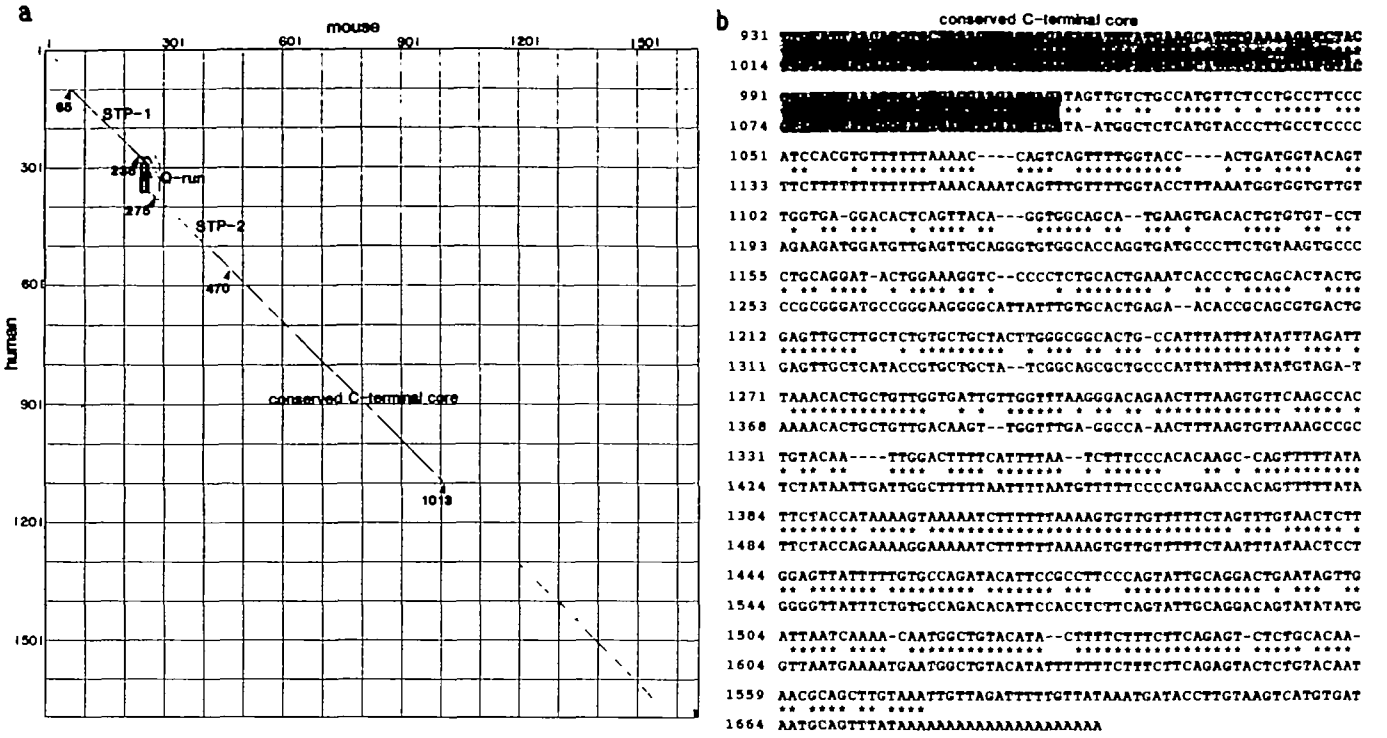

Fig. 3. Comparison of nucleotide sequences between mouse and human TFIIDs. a; Harr plot analysis of the mouse and human (20) TFIID cDNA sequences. One dot represents an identical 9 nucleotides out of 10. Start points of the coding region (65), Q-run (236), STP-2 (275), conserved C-terminal core (470) and the terminus of the coding region (1013) are indicated by arrowheads. b; Nucleotide sequence comparison of the 3'-flanking region between mouse (upper) and human (lower) TFIID cDNAs. Sequences were aligned to have a maximum homology between two cDNAs, including deletion and insertions. The conserved C-terminal core and matching nucleotides are indicated by a shaded area and asterisks, respectively.

and 2b). A stretch of 13 uninterrupted glutamine residues (Q-run) was found between residues 58−70. Regions enriched in ser(S), thr(T), and pro(P) were detected on each side of the Q-run and designated STP-1 and STP-2. The overall contents of ser, thr, and pro residues in STP-1 and STP-2 were 35% and 50%, respectively. The region from residues 119−133 contained an imperfect tripeptide repeat, either pro-ser-thr or a derivative motif (20). In comparing these regions with those of hIID (Fig. 2a, 2b), we found that the N-terminal 57 amino acids (the first 5 amino acids and the STP-1 region) were identical between both TFIIDs. However, the mIID Q-run (13 residues) was considerably shorter than the 34 (20) or 38 (21, 22) residue counterparts in hIID. Sequence deviations between mIID and hIID were most evident in the beginning of STP-2; the mIID sequence of residues 71−95 had 80% identity with the corresponding region of hIID and included two single residue insertions. As described above, the nucleotide sequence of the two cDNAs also diverged in this region.

In a Southern analysis, mouse DNA was digested with *Bam*HI or *Hin*dIII and probed with the conserved region (nucleotide residue 486-to-1006) of the mIID cDNA. This analysis (Fig. 4a) revealed single strong signals of 3.9 (*Bam*HI) or 5.5 (*Hin*dIII) kilobase pairs, respectively. A control with a mouse myelin basic protein (MBP) probe of similar specific activity also identified single strong bands whose intensities were comparable to those observed with the mIID probe. Since a single MBP gene is present per haploid mouse genome (34), these data indicate that the TFIID gene analyzed here exists as a single copy in the haploid mouse genome. It may be noteworthy that the Southern

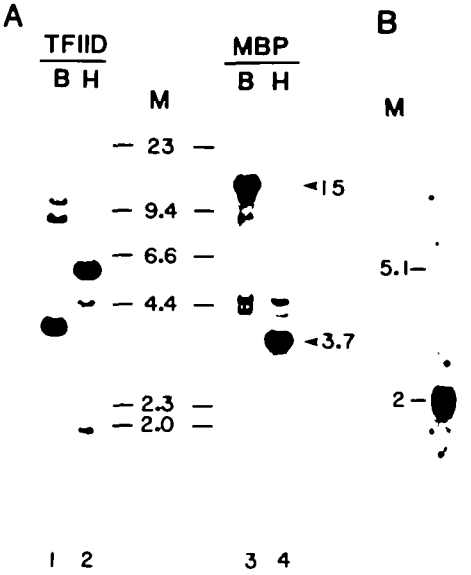

Fig. 4. Analysis of the TFIID DNA in the mouse chromosome and its expression. a; Southern analysis. Mouse DNA digested with *Bam*HI (B) and *Hin*dIII (H) was probed with the mIID (TFIID) or the mouse myelin basic protein (MBP) DNA. Specific activities of the two probes were the same (8×10^8 cpm/μg). The MBP probe detects major signals of 3.7 and 15 kilobase pairs (kb) as demonstrated by Okano et al. (34). Lane M denotes the positions (in kb) of the molecular size markers. b; Northern blot analysis of the mouse poly(A)$^+$RNA. Positions of the 18S (2 kilobases) and 28S (5.1 kilobases) rRNA are shown.

analysis with the mIID probe also revealed a few minor bands at 8.5 and 10 (*Bam*HI) and 1.9 and 4.4 (*Hin*dIII) kilobase pairs (Fig. 4), whose possible significance is discussed below.

Expression of the TFIID sequence in the mouse brain was studied by Northern analysis (Fig. 4b), which revealed a 2 kilobase signal. The size of the mIID RNA was quite similar to that reported for human (HeLa) cells (22). Thus, we estimate that the predominant mIID mRNA contains an additional 200–300 nucleotides (in the upstream non-coding region) beyond that present on the cDNA of Fig. 1.

DISCUSSION

A cDNA from mouse brain libraries was isolated by cross-hybridization with a cDNA encoding hIID. The mouse sequence contained an ORF encoding a 316 amino acid protein with a molecular mass (35 kDa) similar to that of hIID (37–38 kDa) (Fig. 1). The evolutionarily conserved C-terminal core of hIID was totally conserved (below) in the mouse cDNA-encoded protein. Since this region is sufficient for TFIID binding (to the TATA-box) and for function in basal level transcription (22, 23), we conclude that we have obtained a bona fide mouse TFIID cDNA.

Overall amino acid sequence similarities between the mouse and human TFIIDs were striking (Fig. 2a). Each protein had two STP-rich domains separated by Q-run in the N-terminus and a long conserved C-terminal region. We suggest that this structure is a common feature of the mammalian TFIIDs. Most strikingly, the amino acid sequences in the functionally important C-terminal region were identical, indicating that the essential portion of the TFIID remained unchanged during mammalian evolution. Nucleotide sequence comparisons between the mIID and hIID cDNAs (Fig. 3) indicated a high (90%) sequence identity in the coding regions and a lower sequence identity (70%) in the non-coding regions. These data imply that the non-coding flanking sequences of the TFIID varied more rapidly during evolution than did those in the coding region. The amino acid sequence differences observed in two TFIIDs are potentially interesting, notably the short Q-run (13 residues) in mIID versus the longer (34–38 residues) Q-run in hIID. Neither yeast nor plant TFIIDs contain such uniterrupted Q-runs. However, the *Drosophila* TFIID has two short (6 and 8 residue) Q stretches (18, 19). The intermediate length of the mIID Q-run indicates a possible expansion during animal evolution, although present data suggest that the length might vary even between different human cell lines (20–22). We suggest that sequences in the Q-run may mutate more frequently than other portions.

The hIID clones isolated so far were derived from cultured HeLa cells while the present mIID clone was obtained from brain cDNA libraries. Therefore, we cannot eliminate the possibility that the differences found in mIID and hIID may due to the cell or tissue source. Thus far, TFIID heterogeneity has been documented in plants (17) and humans; microheterogeneity in the Q-run (20–22). If there were tissue-specific variants of TFIID, the present study might have selected a TFIID cDNA enriched in the brain. Since several studies (20, 22, 23, 28, 35) have led to speculation of a possible role for the variable N-terminal region of the TFIID molecules in transcriptional activation by sequence-specific DNA-binding proteins, the N-terminal sequence difference between mIID and hIID might possibly be related to tissue-specific transcription.

The size and copy number of mIID transcripts appear to be the same as those reported for hIID (21, 22). As suggested for hIID (22), the present analysis with a conserved C-terminal probe indicated a single TFIID gene in the haploid mouse genome. However, the Southern analysis revealed not only one major signal but several minor signals that most likely represent unique DNA sequences. These minor bands may reflect intron-containing (interrupted) genomic DNA fragments which hybridize only weakly with the probe. Alternatively, the minor bands may reflect TFIID-related sequences in the mouse genome. Consistent with the latter possibility, the variant TFIID species in plants arise from at least two TFIID genes with closely related but distinct nucleotide sequences (17). The clarification of this issue will require isolation of the corresponding genomic DNA fragments.

ACKNOWLEDGEMENTS

We thank T.Furuichi for providing mouse brain cDNA libraries. We also thank Y.Suzuki for encouraging us to continue the work. M.H. is an Alexandrine and Alexander Sinsheimer Scholar. A.H. is an Arnold and Mabel Beckman graduate fellow. This work was supported by Grant-in-Aid for Scientific Research on Priority Areas (T.T. and M.M.) from the Japanese Ministry of Education, Science and Culture, by grants from the National Institute of Health (M.H. and R.G.R.) and by grants from the Pew Trusts to The Rockefeller University.

REFERENCES

1. Sumimoto, H., Ohkuma, Y., Yamamoto, T., Horikoshi, M. and Roeder, R. G. (1990) Proc. Natl. Acad. Sci. USA *87*, 9158–9162.
2. Nakajima, N., Horikoshi, M. and Roeder, R. G. (1988) Mol. Cell. Biol. *8*, 4028–4040.
3. Sawadogo, M. and Roeder, R. G. (1985) Cell *43*, 165–175.
4. Reinberg, D., Horikoshi, M. and Roeder, R. G. (1987) J. Biol. Chem. *262*, 3322–3330.
5. Van Dyke, M. W., Roeder, R. G. and Sawadogo, M. (1988) Science *241*, 1335–1338.
6. Buratowski, S., Hahn, S. Sharp, P. A. and Guarente, L. (1988) Nature *334*, 37–42.
7. Cavallini, B., Huet, J., Plassat, J. L., Sentenac, A., Egly, J. M. and Chambon, P. (1988) Nature *334*, 77–80.
8. Buratowski, S., Hahn, S., Guarente, L. and Sharp, P. A. (1989) Cell *56*, 549–561.
9. Horikoshi, M., Wang, C. K., Fujii, H., Cromlish, J. A., Weil. P. A. and Roeder, R. G. (1989) Proc. Natl. Acad. Sci. USA *86*, 4843–4847.
10. Horikoshi, M., Wang, C. K., Fujii, H., Cromlish, J. A., Weil, P. A. and Roeder, R. G. (1989) Nature *341*, 299–303.
11. Hahn, S., Buratowski, S., Sharp, P. A. and Guarente, L. (1989) Cell *58*, 1173–1181.
12. Eisenmann, D. M., Dollard, C. and Winston, F. (1989) Cell *58*, 1183–1191.
13. Schmidt, M. C., Kao, C. C., Pei, R. and Berk, A. J. (1989) Proc. Natl. Acad. Sci. USA *86*, 7785–7789.
14. Cavallini, B., Faus, I., Matthes, H., Chipoulet, J. M., Winsor, B., Egly, J. M. and Chambon, P. (1989) Proc. Natl. Acad. Sci. USA *86*, 9803–9807.
15. Hoffmann, A., Horikoshi, M., Wang, C. K., Schroeder, S., Weil, P. A. and Roeder, R. G. (1990) Genes Dev. *4*, 1141–1148.
16. Fikes, J. D., Becker, D., Winston, F. and Guarente, L. (1990) Nature, *346*, 291–294.
17. Gasch, A., Hoffmann, A., Horikoshi, M., Roeder, R. G. and Chua, N.-H. (1990) Nature *346*, 390–394.
18. Hoey, T., Dynlacht, B. D., Peterson, M. G., Pugh, B. F. and Tjian, R. (1990) Cell *61*, 1179–1186.
19. Muhich, M. L., Iida, C. T., Horikoshi, M., Roeder, R. G. and Parker, C. S. (1990) Proc. Natl. Acad. Sci. USA *87*, 9148–9152.
20. Hoffmann, A., Sinn, E., Yamamoto, T., Wang, J., Roy, A., Horikoshi, M. and Roeder, R. G. (1990) Nature *346*, 387–390.
21. Peterson, M. G., Tanese, N., Pugh, B. F. and Tjian, R. (1990) Science *248*, 1625–1631.

22. Kao, C. C., Lieberman, P. M., Schmidt, M. C., Zhou, Q., Pei, R. and Berk, A. J. (1990) Science *248*, 1646–1650.
23. Horikoshi, M., Yamamoto, T., Ohkuma, Y., Weil, P. A. and Roeder, R. G. (1990) Cell *61*, 1171–1178.
24. Horikoshi, M., Carey, M. F., Kakidani, H. and Roeder, R. G. (1988) Cell *54*, 665–669.
25. Horikoshi, M., Hai, T., Lin, Y.-S., Green, M. R. and Roeder, R. G. (1988) Cell *54*, 1033–1042.
26. Abmayr, S. M., Workman, J. L. and Roeder, R. G. (1988) Genes Dev. *2*, 542–553.
27. Workman, J. L., Abmayr, S. M., Cromlish, W. A. and Roeder, R. G. (1988) Cell *55*, 211–219.
28 Pugh, B. F. and Tjian, R. (1990) Cell *61*, 1187–1197.
29. Meisterernst, M., Horikoshi, M. and Roeder, R. G. (1990) Proc. Natl. Acad. Sci. USA *87*, 9153–9157.
30. Furuichi, T , Yoshikawa, S., Miyawaki, A., Wada, K., Maeda, N. and Mikoshiba, K. (1989) Nature *432*, 32–38.
31. Inoue, T., Tamura, T., Furuichi, T. and Mikoshiba, K. (1990) J. Biol. Chem. *265*, 19065–19070.
32. Maniatis, T., Fritsch, E. F. and Sambrook, J. (1989) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
33. Popko, B., Puckett, C., Lai, E., Shine, H. D., Readhead, C., Takahashi, N., Hunt, S. W., Sidman, R. L. and Hood, L. (1987) Cell *48*, 713–721
34. Okano, H., Tamura, T., Miura, M., Aoyama, A., Ikenaka, K., Oshimura, M. and Mikoshiba, K. (1988) EMBO J. *7*, 77–83.
35. Smale, S. T., Schmidt, M. C., Berk, A. J. and Baltimore, D. (1990) Proc. Natl. Acad. Sci. USA *87*, 4509–4513.