Cytogenet Cell Genet 69:75–80 (1995)

Genetic characterization of general transcription factors TFIIF and TFIIB of *Homo sapiens sapiens*

M. Purrello,[1] C. Di Pietro,[1] A. Rapisarda,[2] E. Mirabile,[1] S. Motta,[2] G. Sichel,[1] and K.-H. Grzeschik[3]

[1] Istituto di Biologia generale and [2] Dipartimento di Biologia animale e di Genetica, Università di Catania, Catania (Italy), and [3] Med. Zentrum für Humangenetik, Philipps Universität, Marburg (Germany)

Abstract. Analysis of loci GTF2F1 and GTF2B, encoding Rap 74 (a subunit of TFIIF) and TFIIB, respectively, showed that they are present in a single copy in the human genome and are localized at 19p13.3 and 1p22, respectively. By using as probe a cDNA for Rap 30 (the other subunit of TFIIF), we localized the GTF2F2 locus to 13q14; the same probe also detected a cross-hybridizing sequence at 4q31 whose functional importance remains to be elucidated. These data and those previously published by our group demonstrate that genes coding for class II general transcription factors with reported sequence similarity to bacterial sigma proteins are scattered in different regions of the human genome, with no evidence of clustering. This dispersion and the identification of homologs of both TBP and TFIIB in *Archaea* suggest an early evolutionary origin of the general transcription apparatus of contemporary eukaryotes.

TFIIF and TFIIB are two general transcription factors necessary for RNA synthesis initiation at all prototype class II promoters analyzed to date (reviewed by Greenblatt, 1992; Conaway and Conaway, 1993; and Zawel and Reinberg, 1993).

TFIIF is a heterodimer (Flores et al., 1989). Rap 74, the large subunit of 517 amino acids, has two globular domains connected by a highly charged segment, which is rich in acidic amino acids (Aso et al., 1992; Finkelstein et al., 1992); its carboxyl terminus has sequence similarity to bacterial sigma proteins (Gong et al., 1992). Rap 30, the small subunit of 234 amino acids, has a high number of basic residues; it is characterized by similarity to region 4 (Garrett et al., 1992) and to regions 1b, 2.1, and 2.2 of the sigma proteins (Sopta et al., 1989; Conaway and Conaway, 1990). These last segments, which mediate interaction with *Escherichia coli* RNA polymerase core, are the most conserved regions of sigma proteins (reviewed by Gross et al., 1992). Both sigma 70 and Rap 30 bind to the heterologous enzyme (McCracken and Greenblatt, 1991); interestingly, Rap 30 binds comparably well to RNA polymerase II and to its bacterial ortholog, whereas sigma 70 interaction with the eukaryotic enzyme is weaker (McKracken and Greenblatt, 1991). The association of TFIIF with RNA polymerase II, mediated by Rap 30, allows the enzyme to recognize a partially assembled preinitiation complex (comprised of TFIID, TFIIA, and TFIIB) and to correctly position itself onto the promoter. Rap 74 stabilizes Rap 30 interaction with RNA polymerase (Garrett et al., 1992) and is necessary for dissociating the enzyme from DNA when it binds nonspecifically to it (Killeen and Greenblatt, 1992).

TFIIB is a 33-kDa polypeptide of 316 amino acids with an overall structure resembling that of TBP (Malik et al., 1991; Ha et al., 1993). The protein, which has sequence similarity to regions 2.1 and 2.2 (Malik et al., 1991) and region 4.2 (Ha et al., 1991) of sigma proteins, acts as a bridge between promoter-bound TFIID (whether associated or not with TFIIA) and the complex of RNA polymerase and TFIIF in solution. Besides determining the spacing between the TATA box and the cap site (Yamashita et al., 1992), TFIIB is also the binding target of upstream activators of transcription (Lin and Green, 1991); accordingly, it may have, together with TFIID, an important role in connecting basal and regulated transcription.

Given their fundamental biological role, we have begun the genetic characterization of the human loci for class II general

Supported by funds from CNR (M.P., G.S.), MURST (M.P., S.M., G.S.), and DFG (K.-H.G.).

Received 16 May 1994; accepted 11 July 1994.

Request reprints from Dr. Michele Purrello, Istituto di Biologia generale, via Androne 81, 95124 Catania (Italy); telephone: 095-316127; fax: 095-310374.

0301–0171/95/0692–0075$8.00/0

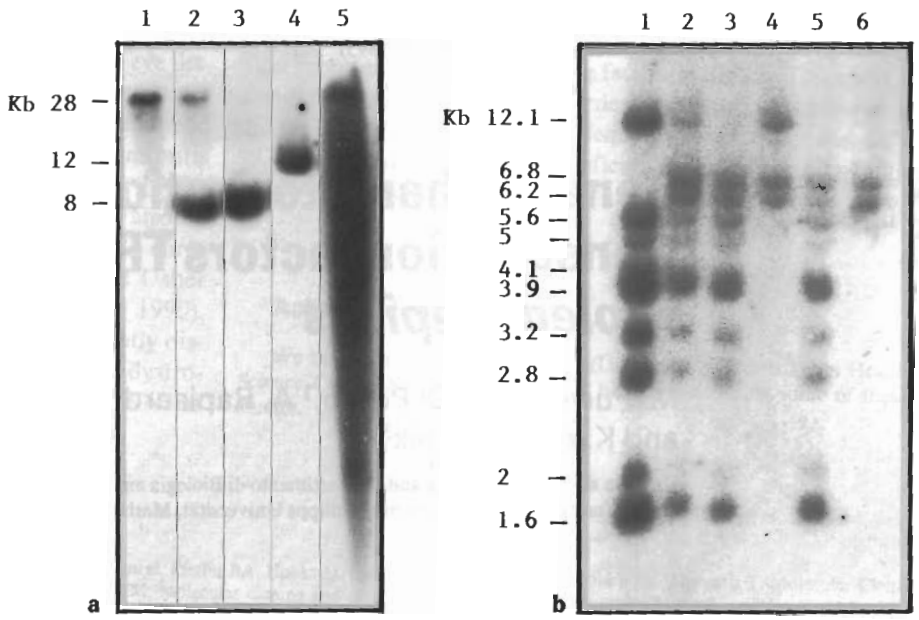

Fig. 1. (**a**) High-stringency hybridization of F74-2 to *Eco*RI-digested DNA from: (1) a human blood donor; (2) RAG × GM194 cl. 7 (a mouse × human hybrid cell line retaining, among others, human chromosome 19); (3) A9 × IT cl. 2-18 (a mouse × human hybrid cell line missing human chromosome 19); and (4) V79 × LY cl. 3-2 (a hamster × human hybrid cell line missing human chromosome 19). The high-stringency hybridization pattern of F74-3 to human DNA is shown in lane 5. Autoradiography was performed at −80° C for 6 d. (**b**) Low-stringency hybridization pattern of purified insert from pF30 to *Hin*dIII-digested DNA from: (1) a human blood donor; (2) RAG × GM 194 cl.5-5 (retaining, among others, both human chromosomes 13 and 4); (3) A9 × SU cl. 1-2 (retaining human chromosome 13); (4) A9 × GM 89 cl. 9c-7 (retaining human chromosome 4); (5) A9 × GM 73 cl. 2b (retaining human chromosome 13); and (6) RAG.

transcription factors (Purrello et al., 1994a, b). In this paper we report the analysis of the genes encoding the two subunits of TFIIF and TFIIB.

Materials and methods

Plasmid pBSII SK− Rap 74 (pF74) contains a cDNA insert of about 2,400 bp with the entire coding sequence for human Rap 74 and part of the 5′ and 3′ untranslated regions, inserted into the *Eco*RI site of pBSII SK− (Finkelstein et al., 1992). Double digestion with *Eco*RI and *Xmn*I cuts the cDNA into three fragments of about 588 (probe F74-1), 910 (probe F74-2), and 853 bp (probe F74-3). F74-1 contains the sequences coding for most of the amino-terminal globular domain, F74-2 contains those coding for the highly charged central segment and for approximately 118 amino acids of the globular carboxyl-terminal domain, and F74-3 has those coding for the remaining part of this last domain. Plasmid p9-3 (pF30) has a full-length cDNA for human Rap 30 of about 2,000 bp inserted into the *Eco* RI site of pEMBL 18 (Sopta et al., 1989). Plasmid pBl3 (pB) has an *Eco*RI insert of about 1,300 bp with the entire open reading frame for human TFIIB, cloned in pBS SK− (Malik et al., 1991).

High- and low-stringency Southern analysis was carried out as described elsewhere (Purrello et al., 1994a), using as probes purified inserts from pF74, pF30, and pB which were labeled by oligolabeling. The chromosomal content of murine × human somatic cell hybrids has been described (Purrello et al., 1991). In situ hybridization was performed following a published protocol (Purrello et al., 1985); in some experiments another method (Pinkel et al., 1988) was utilized. Posthybridization immunological procedures were performed as described by Pinkel et al. (1988). Probes were supercoiled plasmids pF74, pF30, and pB, labeled by nick translation using biotin-16-dUTP as precursor. Prehybridization treatment of pF74 with C_0t-1 human DNA (GIBCO/BRL) was done following standard protocols.

Results

High-stringency hybridization of the cDNA insert from pF74 to human DNA produced a pattern expected from a moderately repetitive sequence, whereas a single *Eco*RI fragment showed homology to the probe both in mouse and hamster genomes. The use of probes F74-1, F74-2, and F74-3 allowed us to demonstrate that the GTF2F1 locus is present in a single copy in the human genome as well, and that the hybridization pattern of F74-3 to human DNA is similar to that of the entire insert (Fig. 1a). GTF2F1 was localized to chromosome 19 based on the segregation of the human *Eco*RI fragment detected by F74-2 in a panel of mouse × human hybrid cell lines (Fig. 1a). F74-1 detected an additional fragment of about 7.5 kb containing the 5′ segment of Rap 74 transcribed region. The hybridization pattern of both F74-1 and F74-2 did not change at low stringency. Chromosomal in situ hybridization of pF74 prehybridized with C_0t-1 human DNA confirmed and refined our localization of GTF2F1. A fluorescent signal was present at 19p13.3 in about 65% (71/109) of metaphases (Fig. 2A), and in about 20% of them both chromatids were labeled at the same site; about 54% (109/202) of the total number of fluorescent grains were found at this position. The mean number of hybridization sites per metaphase was 1.85. No other site besides 19p13.3 was labeled significantly.

The purified insert from pF30 detected in the human genome a series of *Hin*dIII fragments which cosegregated with chromosome 13, with the exception of a fragment of about 12 kb localized to chromosome 4 (Fig. 1b). When washing was performed with 0.5 × SSC at 50° C, all autoradiographic bands had comparable intensity, but at higher stringency (0.5 × SSC at 60° C) the one corresponding to the chromosome 4 sequence became fainter; this suggests that it is only partially homologous to the probe. Both the localization of the GTF2F2 locus and of the related sequence (locus GTF2F2L) were confirmed and refined by in situ hybridization. In about 70% (72/103) of metaphases there was a hybridization site at 13q14, and in about 25% of them the signal was localized at 4q31, with both chromatids labeled at the same sites in about 30%

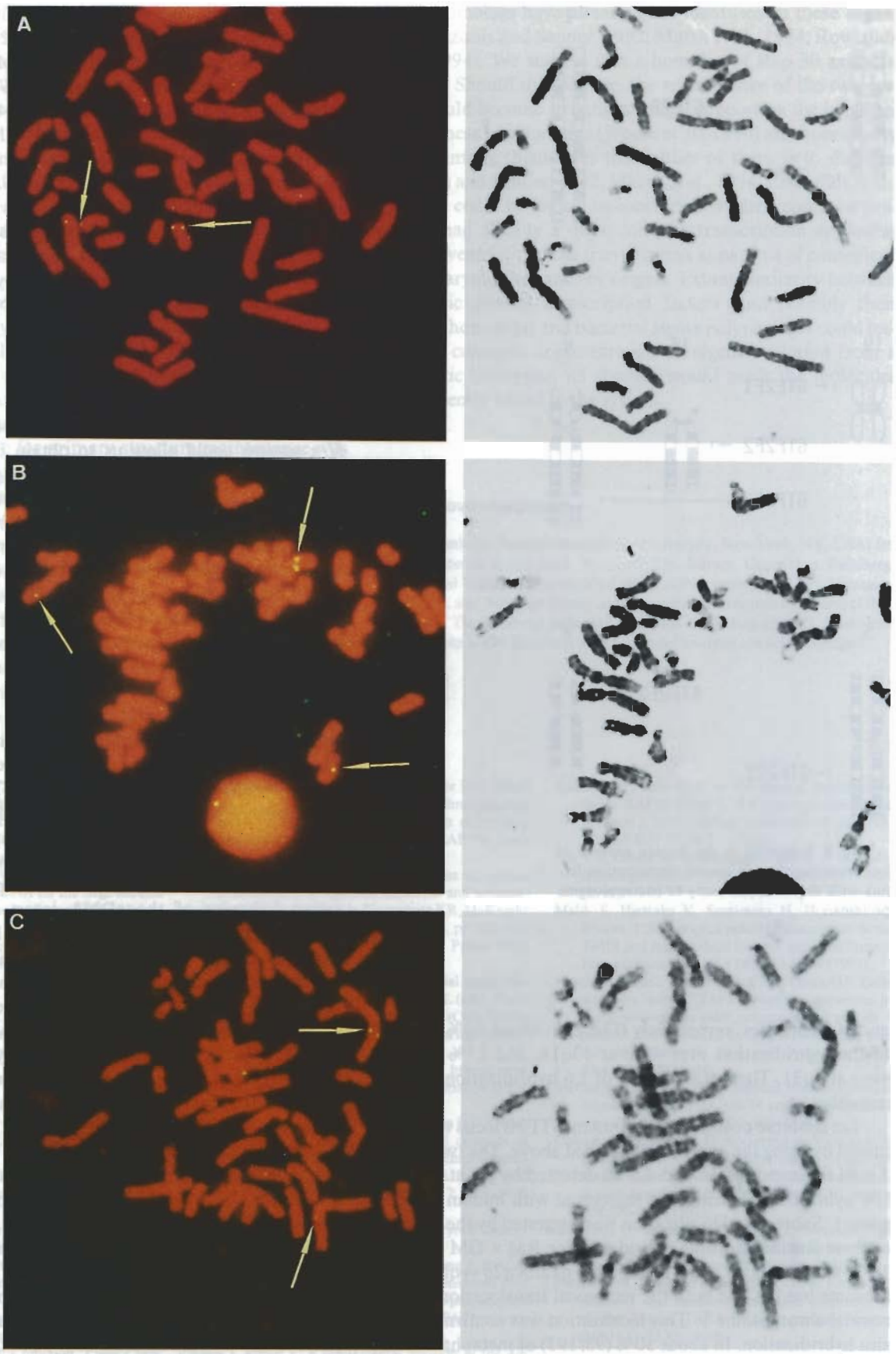

Fig. 2. In situ hybridization of pF74 (**A**), pF30 (**B**), and pB (**C**) to human metaphase chromosomes. Arrows point to the position of the corresponding loci. For details, see text.

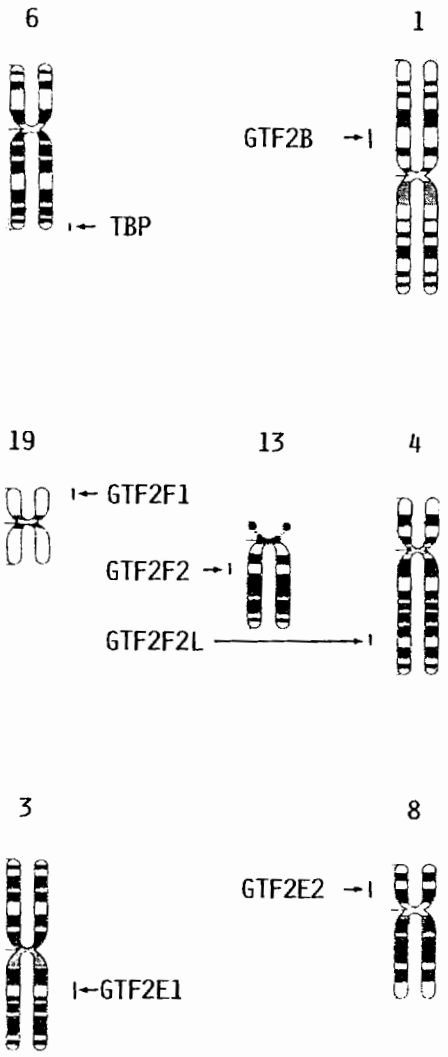

Fig. 3. Scattering in the human genome of loci encoding class II general transcription factors with sequence similarity to bacterial sigma proteins.

and 10% of cases, respectively (Fig. 2B). About 60% (97/165) of the hybridization sites were at 13q14, and 17% (28/165) were at 4q31. There was a mean of 1.6 hybridization sites per metaphase.

The genomic position of the human GTF2B locus was determined by using the approach described above. The two human *Eco*RI fragments of 16 and 4.6 kb detected by pB at high and low hybridization stringency segregated with human chromosome 1. Subregional localization was suggested by the presence of these fragments in the hybrid cell line Rag × GM 97, cl. 8-13-3, which retains a t(1;12)(1pter→q12::Xq26→qter) chromosome but has lost both the reciprocal translocation and the normal chromosome 1. This localization was confirmed by in situ hybridization. In about 50% (95/193) of metaphases there was a signal at 1p22, and about 70% (90/131) of the hybridization sites were localized at band 1p22, with an average of 0.7 hybridization sites per metaphase (Fig. 2C). pB did not detect other cross-hybridizing sites besides 1p22.

Southern analysis of *Eco*RI- and *Taq*I-digested DNA from *Homo sapiens sapiens, Mus musculus, Gallus domesticus, Xenopus laevis,* and *Mugil cephalus* with F74-1 and F74-3 demonstrated that these probes hybridized only to human DNA at both high and low stringency; at high stringency, however, F74-2 detected homologous fragments in DNA from the first three species and at low stringency in *X. laevis* DNA as well. The pF30 probe detected cross-hybridizing sequences only in the genome of *M. musculus* when washing was performed with 0.5 × SSC at 50°C (Fig. 1b). The pB probe hybridized at low stringency to human, murine, and chicken DNA.

Discussion

Genetic analysis of the general transcription apparatus in *H. s. sapiens* could allow us to obtain insights into both its actual structure and its evolutionary origin. When we started this project, our aims were twofold. First, we wished to determine the copy number of genes coding for general transcription factors. Second, we wanted to investigate whether these loci were clustered in specific regions of the human genome.

With the exception of GTF2F2, the loci we have analyzed to date are present as single copies in the human genome (Purrello et al., 1994a, b; this paper), as many of them are in other eukaryotes. Known exceptions are the duplication of TBP in the plant genome, which may be functionally redundant (reviewed by Hernandez, 1993), and the existence of a family of TFIIB-related genes in *Saccharomyces cerevisiae* (Buratowski and Zhou, 1992). The finding of pF30-related sequences at 4q31 was unexpected: since a single form of TBP, TFIIB, and RNA polymerase II has been characterized to date, there could be no need for the synthesis of two different Rap 30 proteins; cloning of the chromosome 4 sequence will show if it is a true gene, a pseudogene, or the result of a retrotransposition. The hybridization behavior of the cDNA insert from pF74 may be explained by the presence of an *Alu*-like (species-specific) sequence in the 3′ transcribed, but untranslated, region of the human gene, possibly the consequence of a retrotransposition that occurred after the separation of the human and murine lineages. Similar events have been previously described (reviewed by Jurka and Smith, 1988). Analysis of DNA from different species with probes for Rap 74 and TFIIB confirmed substantial conservation of the corresponding coding segments (Gong et al., 1992; Yamashita et al., 1992). A notable exception was represented by Rap 30; it is possible that a close structure-function relationship, which would be a strong selective constraint, exists only for the short segment of the protein interacting with RNA polymerase.

The data presented in this paper localize GTF2F2, GTF2F2L, and GTF2B to 13q14, 4q31, and 1p22, respectively, and confirm the previous assignment of GTF2F1 to 19p13.3 (Aso et al., 1993); those previously obtained by our group (Purrello et al., 1994a, b) localized TBP to 6q27, GTF2E1 to 3q13→q21, and GTF2E2 to 8p12 (Fig. 3). Since TBP (a subunit of TFIID), TFIIB, Rap 74 and Rap 30 (subunits of TFIIF),

and E56 and E34 (subunits of TFIIE) have different primary structures, scattering in the genome of the corresponding loci is not surprising and suggests that these proteins do not share a common recent evolutionary history. Intriguingly, all of these polypeptides have been suggested to have localized sequence similarity to bacterial sigma proteins. Partial homology of TBP to sigma region 2.4 and of Rap 30 to regions 1b, 2.1, and 2.2 would correspond to analogy of function (Hoffmann et al., 1990; Peterson et al., 1990; McKracken and Greenblatt, 1991). Sequence similarity of segments of TFIIB, E56, E34, and possibly Rap 74 to sigma regions 2.1 and 2.2 (Ha et al., 1991; Malik et al., 1991; Okhuma et al., 1991; Sumimoto et al., 1991; Gong et al., 1992) could be explained by their involvement in interaction with RNA polymerase; finally, localized similarity of TFIIB to region 4.2 (Ha et al., 1991) and of Rap 30 to region 4 (Garrett et al., 1992) of sigma polypeptides could indicate that they bind to DNA. Assuming that these proteins share a common (ancient) ancestry, these homologies would demonstrate that selective pressure for primary structure maintenance was exerted on those regions performing critical functions, such as binding to DNA or contacting RNA polymerase.

It has been demonstrated that the transcription apparatus of *Archaea* has many features in common with that of eukaryotes. Even though they possess a single type of RNA polymerase, the enzyme has a complex structure (Pühler et al., 1989) and is incapable of correctly initiating RNA synthesis without first interacting with DNA-bound transcription factors; TBP and TFIIB homologs have already been identified in these organisms (Ouzunis and Sander, 1992; Marsh et al., 1994; Rowlands et al., 1994). We suggest that a homolog of Rap 30 exists in *Archaea.* Should this be true, the resemblance of the two systems would become striking, strongly supporting the hypothesis that these two domains (Woese et al., 1990) are more closely related among themselves than either of them is to *Bacteria* (Ouzunis and Sander, 1992; Marsh et al., 1994; Rowlands et al., 1994). It could be logical to conclude that their common progenitor had already a quite complex transcription apparatus and, conversely, that the transcription apparatus of contemporary eukaryotes has ancient origins. Extant similarity between eukaryotic general transcription factors (and possibly their archaeal homologs) and bacterial sigma polypeptides could testify to a common origin through divergent evolution from a progenotic prototype; its absence would mark the molecules more recently added to the system.

Acknowledgements

We thank Dr. Roeder (Rockefeller University, New York, NY, USA) for his kind interest in our work. We thank Drs. Burton, Greenblatt, Reinberg, Roeder, and Weissman for the kind gift of cDNA probes, and Drs. Hisatake, Pan, Sopta, and Wang for kindly communicating restriction maps and cDNA sequences. The skillful collaboration of Mrs. Olberding and Mr. Vasta and the kind help of Dr. Zafarana with computer analysis are acknowledged.

References

Aso T, Tsai P, Kawaguchi T, Menninger JC, Kitajima S, Yasukochi Y, Ward DC, Weissman SM: Assignment of the human GTF2F1 gene to chromosome 19p13.3. Genomics 16:252–253 (1993).

Aso T, Vasavada H A, Kawaguchi T, Germino FJ, Ganguly S, Kitajima S, Weissman SM, Yasukochi Y: Characterization of cDNA for the large subunit of the transcription initiation factor TFIIF. Nature 355:461–464 (1992).

Buratowski S, Zhou H: A suppressor of TBP mutation encodes an RNA polymerase III transcription factor with homology to TFIIB. Cell 71:221–230 (1992).

Conaway JW, Conaway RC: An RNA polymerase II transcription factor shares functional properties with *Escherichia coli* sigma 70. Science 248:1550–1553 (1990).

Conaway RC, Conaway JW: General initiation factors for RNA polymerase II. A Rev Biochem 62:161–190 (1993).

Finkelstein A, Kostrub CF, Li J, Chavez DP, Wang BQ, Fang SM, Greenblatt J, Burton ZF: A cDNA encoding RAP74, a general initiation factor for transcription by RNA polymerase II. Nature 355:464–467 (1992).

Flores O, Maldonado E, Reinberg D: Factors involved in specific transcription by mammalian RNA polymerase II. J biol Chem 264:8913–8921 (1989).

Garrett KT, Serizawa H, Hanley JP, Bradsher JN, Tsuboi A, Arai N, Yokota T, Arai K, Conaway RC, Conaway JW: The carboxyl terminus of RAP30 is similar in sequence to region 4 of bacterial sigma factors and is required for function. J biol Chem 267:23942–23949 (1992).

Gong D-W, Hasegawa S, Wada K, Roeder RG, Nakatani Y, Horikoshi M: Elucidation of three putative structural subdomain by comparison of primary structure of *Xenopus* and human RAP 74. Nucl Acids Res 20:6736 (1992).

Greenblatt J: Protein-protein interactions as critical determinants of regulated initiation and termination of transcription, in Yamamoto KR, McKnight SL (eds): Transcriptional Regulation, pp 203–226 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor 1992).

Gross CA, Lonetto M, Losick R: Bacterial sigma factors, in Yamamoto KR, McKnight SL (eds): Transcriptional Regulation, pp 129–176 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor 1992).

Ha I, Lane S, Reinberg D: Cloning of a human gene encoding the general transcription initiation factor IIB. Nature 352:689–695 (1991).

Ha I, Roberts S, Maldonado E, Sun X, Kim L-U, Green M, Reinberg D: Multiple functional domains of human transcription factor IIB: distinct interactions with two general transcription factors and RNA polymerase II. Genes Dev 7:1021–1032 (1993).

Hernandez N: TBP, a universal eukaryotic factor? Genes Dev 7:1291–1308 (1993).

Hoffmann A, Sinn E, Yamamoto T, Wang J, Roy A, Horikoshi M, Roeder RG: Highly conserved core domain and unique *N*-terminus with presumptive regulatory motifs in a human TATA factor (TFIID). Nature 346:387–390 (1990).

Jurka J, Smith T: A fundamental division in the *Alu* family of repeated sequences. Proc natl Acad Sci, USA 85:4775–4778 (1988).

Killeen MT, Greenblatt J: The general transcription factor RAP30 binds to RNA polymerase II and prevents it from binding nonspecifically to DNA. Mol cell Biol 12:30–37 (1992).

Lin Y-S, Green M: Mechanism of action of an acidic transcriptional activator in vitro. Cell 64:971–981 (1991).

Malik S, Hisatake K, Sumimoto H, Horikoshi M, Roeder R: Sequence of general transcription factor TFIIB and relationships to other initiation factors. Proc natl Acad Sci, USA 88:9553–9557 (1991).

Marsh TL, Reich CI, Whitelock RB, Olsen GJ: Transcription factor IID in the Archaea: sequences in the *Thermococcus celer* genome would encode a product closely related to the TATA-binding protein of eukaryotes. Proc natl Acad Sci, USA 91:4180–4184 (1994).

McCracken S, Greenblatt J: Related RNA polymerase regions in human RAP30/74 and *Escherichia coli* sigma 70. Science 253:900–902 (1991).

Ohkuma Y, Sumimoto H, Hoffmann A, Shimasaki S, Horikoshi M, Roeder R: Structural motif and potential sigma homologies in the large subunit of human general transcription factor TFIIE. Nature 354:398–401 (1991).

Ouzounis C, Sander C: TFIIB, an evolutionary link between the transcription machineries of Archaebacteria and eukaryotes. Cell 71:189–190 (1992).

Peterson MG, Tanese N, Pugh BF, Tjian R: Functional domains and upstream activation properties of cloned human TATA binding protein. Science 248:1625–1630 (1990).

Pinkel D, Landegent J, Collins C, Fuscoe J, Segraves R, Lucas J, Gray JW: Fluorescent in situ hybridization with human chromosome-specific libraries: detection of trisomy 21 and translocations of chromosome 4. Proc natl Acad Sci, USA 85:9138–9142 (1988).

Pühler G, Leffers H, Gropp F, Palm P, Klenk H-P, Lottspeich F, Garrett RA, Zillig W: Archaebacterial DNA-dependent RNA polymerases testify to the evolution of the eukaryotic nuclear genome. Proc natl Acad Sci, USA 86:4569–4573 (1989).

Purrello M, Alhadeff B, Esposito D, Szabo P, Rocchi M, Truett M, Masiarz F, Siniscalco M: The human genes for hemophilia A and hemophilia B flank the X chromosome site at Xq27.3. EMBO J 4:725–729 (1985).

Purrello M, Bettuzzi S, Di Pietro C, Mirabile E, Di Blasi M, Rimini R, Grzeschik K-H, Ingletti C, Corti A, Sichel G: The gene for SP-40,40, human homolog of rat sulfated glycoprotein 2, rat clusterin, and rat testosterone-repressed prostate message 2, maps to chromosome 8. Genomics 10:151–156 (1991).

Purrello M, Di Pietro C, Mirabile E, Rapisarda A, Rimini R, Tinè A, Pavone L, Motta S, Grzeschik K-H, Sichel G: Physical mapping at 6q27 of the locus for the TATA box-binding protein, the DNA-binding subunit of TFIID and a component of SL1 and TFIIIB, strongly suggests that it is single copy in the human genome. Genomics 22:94–100 (1994a).

Purrello M, Di Pietro C, Rapisarda A, Motta S, Pavone L, Grzeschik K-H, Sichel G: Localization of the human genes encoding the two subunits of general transcription factor TFIIE. Genomics 23:253–255 (1994b).

Rowlands T, Baumann P, Jackson SP: The TATA-binding protein: a general transcription factor in eukaryotes and Archaebacteria. Science 264:1326–1329 (1994).

Sopta M, Burton Z, Greenblatt J: Structure and associated DNA-helicase activity of a general transcription initiation factor that binds to RNA polymerase II. Nature 341:410–414 (1989).

Sumimoto H, Ohkuma Y, Sinn E, Kato H, Shimasaki S, Horikoshi M, Roeder R: Conserved sequence motifs in the small subunit of human general transcription factor TFIIE. Nature 354:401–404 (1991).

Woese CR, Kandler O, Wheelis ML: Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc natl Acad Sci, USA 87:4576–4579 (1990).

Yamashita S, Wada K, Horikoshi M, Gong D-W, Kokubo T, Hisatake K, Yokotani N, Malik S, Roeder R, Nakatani Y: Isolation and characterization of a cDNA encoding *Drosophila* transcription factor TFIIB. Proc natl Acad Sci, USA 89:2839–2843 (1992).

Zawel L, Reinberg D: Initiation of transcription by RNA polymerase II: a multi-step process. Progr Nucl Acids Res molec Biol 44:67–108 (1993).