The EMBO Journal vol.9 no.11 pp.3631 - 3639, 1990

Thyroid nuclear factor 1 (TTF-1) contains a homeodomain and displays a novel DNA binding specificity

Stefania Guazzi, Melanie Price, Mario De Felice, Giuseppe Damante, Marie-Genevieve Mattei[1] and Roberto Di Lauro

European Molecular Biology Laboratory (EMBL), Meyerhofstrasse 1, 6900 Heidelberg, FRG, and [1]Unité 242 de Physiopathologie Chromosomique de L'Institut National de la Santé et de la Recherche Medicale, Hôpital d'Enfants de la Timone, 13385 Marseille Cedex 5, France

Communicated by I.W.Mattaj

The cDNA for TTF-1, a thyroid nuclear factor that binds to the promoter of thyroid specific genes, has been cloned. The protein encoded by the cDNA shows binding properties indistinguishable from those of TTF-1 present in nuclear extracts of differentiated rat thyroid cells. The DNA binding domain of TTF-1 is a novel mammalian homeodomain that shows considerable sequence homology to the *Drosophila* NK-2 homeodomain. TTF-1 mRNA and corresponding binding activity are detected in thyroid and lung. The chromosomal localization of the TTF-1 gene has been determined in humans and mice and corresponds to chromosomes 14 and 12, respectively, demonstrating that the TTF-1 gene is not located within previously described clusters of homeobox-containing genes.

Key words: DNA binding protein/homeobox/lung/thyroid/TTF-1

Introduction

The thyroid gland originates from an invagination of the primitive pharynx and begins to synthesize thyroid hormones on day 17 of rat development. Thyroid hormone production is preceded, by 2 days, by the appearance of thyroglobulin, the earliest marker of differentiation in the thyroid (Kawaoi and Tsuneda, 1985). In order to determine the primary events leading to thyroid cell differentiation, we have identified several elements involved in thyroid specific activation of the thyroglobulin promoter. Transfection of constructs containing the thyroglobulin gene promoter fused to a reporter gene into the differentiated rat thyroid cell line FRTL-5 (Ambesi-Impiombato *et al.*, 1980) demonstrated that transcriptional control plays a role in the restriction of thyroglobulin mRNA to the thyroid tissue (Musti *et al.*, 1987). The specific expression of the thyroglobulin gene in thyroid tissue *in vivo* and in the FRTL-5 cell line correlates with the presence of a nuclear factor, TTF-1, capable of recognizing four sites in the thyroglobulin promoter (Civitareale *et al.*, 1989). The functional relevance of the interaction between TTF-1 and the thyroglobulin promoter has been demonstrated by analysing the effect of promoter mutations. Changes in the DNA sequence of two of the four binding sites abolish both TTF-1 binding and promoter activity (Civitareale *et al.*, 1989). An additional correlation between the presence of TTF-1 and the expression of the differentiated phenotype comes from studies carried out on FRTL-5 cells transformed by the Kirsten sarcoma virus (Colletta *et al.*, 1983). These cells do not express any differentiated thyroid function and have been shown to contain very little, if any, TTF-1 (Avvedimento *et al.*, 1988), suggesting that the primary event in the process of de-differentiation induced by K-*ras* could be the disappearance of TTF-1 activity. This hypothesis is supported by recent studies on a FRTL-5 cell line transformed with a temperature sensitive K-*ras*, where TTF-1 activity can only be detected at the non-permissive temperature (Avvedimento *et al.*, 1989).

We have recently obtained highly purified TTF-1 from calf thyroids, and determined the sequence of a peptide of 13 amino acids (D.Civitareale, S.Ehrlich-Wiedenmann, R.Di Lauro, in preparation). We report in this paper the cloning of the TTF-1 cDNA. Conceptual translation of the cDNA shows the presence of a homeodomain, which displays a considerable degree of homology to a new family of *Drosophila* homeoboxes, the NK family (Kim and Nirenberg, 1989). The TTF-1 homeodomain alone is able to reproduce the DNA binding properties of the entire protein. The sequence specificity of DNA binding does not conform to rules recently proposed based on the effects of specific amino acid substitutions in some *Drosophila* homeodomains (Hanes and Brent, 1989; Treisman *et al.*, 1989).

TTF-1 mRNA has a restricted tissue distribution, being detectable only in thyroid and lung, suggesting that controls acting at the mRNA level may determine specific expression of the gene in the anterior endoderm.

Results

Isolation and sequencing of a cDNA clone encoding TTF-1

The thyroglobulin promoter contains four binding sites for a protein detected exclusively in extracts of differentiated thyroid cells (Civitareale *et al.*, 1989). We have recently purified the protein from nuclear extracts of calf thyroids (Civitareale *et al.*, 1989). The sequence of a 13 amino acid long peptide was obtained from the purified preparation (Civitareale *et al.*, in preparation), which was then used to deduce the sequence of a degenerate oligonucleotide (TTF-1G in Figure 1A). The TTF-1G oligonucleotide mixture was labelled to a high specific activity and hybridized to 1.3×10^6 plaques of a calf thyroid cDNA library in λgt11. Positive clones were purified and *Sau*3A1 fragments obtained from *Eco*RI inserts were subcloned in Bluescript and sequenced. One clone (λ1) was identified which contained a sequence coding for the TTF-1 peptide (Figure 1A). In order to obtain the rat TTF-1 cDNA, 2.3×10^6 plaques of a rat thyroid cDNA library were screened with the insert of the λ1 clone. Seventeen positive clones were isolated and the one containing the longest insert (λ 12.3)

Fig. 1. Sequence of the TTF-1 cDNA. (**A**) alignment of the 36mer, degenerate oligonucleotide TTF-1-G (derived from the amino acids sequence of peptide 1) with the partial sequence of the calf cDNA clone and with the amino acid sequence of peptide 1 obtained from the purified TTF-1 protein. (**B**) Complete sequence of the rat cDNA insert and deduced amino acid sequence. The homeobox is double underlined and the glutamine/alanine-rich regions are underlined. The putative polyadenylation signal is shown in bold. (**C**) Schematic representation and partial restriction map of TTF-1 cDNA (S: *Sca*I; P: *Pst*I; St: *Sty*I; E: *Eco*RI); the open box indicates the coding region, the black area the homeobox, the thin line the 3′ non-translated region and the thick line the repetitive sequence. The dashed lines (labelled a–d) indicate the probes used in hybridization experiments.

was selected for further analysis. The 2.4 kb insert of λ 12.3 contained an internal *Eco*RI site and was therefore excised by partial *Eco*RI digestion, gel purified and subcloned into Bluescript. The complete sequence of the cDNA insert is presented in Figure 1B. The cDNA insert is 2323 bp long, in agreement with the size observed for the mRNA (see

below). The only long open reading frame begins at the ATG at nucleotide 4 and extends for 1134 nt, encoding a protein of 378 amino acids. The most striking feature of the deduced sequence of the TTF-1 protein is the presence of a homeodomain that cannot be placed in any of the previously described classes, even if it matches the proposed homeodomain consensus sequence (Scott *et al.*, 1989) in 18 positions out of 21. A more detailed discussion of the homeodomain is found below. Other notable features are the presence, downstream of the homeodomain, of an alanine and glutamine-rich (38%) sequence, which is interrupted by a run of nine consecutive glycine residues, and a serine rich region (21%) at the C-terminus of the protein. The 3′ untranslated region (1209 nt) accounts for 52% of the mRNA and contains a long stretch of a repetitive sequence, whose basic motif is the trinucleotide CTT. A putative polyadenylation site is found at nucleotide 2300.

The cloned cDNA encodes a protein that displays the same binding properties as TTF-1 found in thyroid

The cDNA insert was used as a template to obtain the TTF-1 protein by *in vitro* transcription−translation. Rabbit reticulocyte lysates, programmed with RNA transcribed from the T7 promoter in Bluescript (Figure 2A), synthesize a protein of 38 kd (Figure 2B), the size expected for the length of the open reading frame. The two bands observed on SDS−PAGE could derive from alternative initiation of translation from the two closely spaced ATG codons at position 4 and 82 in the nucleotide sequence shown in Figure 1. In band-shift experiments, the same lysates are capable of forming two complexes with oligonucleotide C (5′-CACTGCCCAGTCAAGTGTTCTTGA-3′) which contains the sequence of the proximal TTF-1 binding site of the thyroglobulin promoter (Figure 2C, T7) (Civitareale *et al.*, 1989). These complexes are competed by a 1000-fold excess of unlabelled oligonucleotide C, but not by the unrelated oligonucleotide K (5′-TGACTAGCAGAGAAAACAAA-GTGA-3′) (Figure 2C, T7+C and T7+K). Both complexes comigrate with the ones observed using nuclear extracts from the differentiated thyroid cell line FRTL-5 (Figure 2C, TL5). Complex formation was not observed with lysates programmed with RNA transcribed from the T3 promoter in Bluescript (Figure 2C, T3). In addition, the TTF-1 cDNA was cloned into plasmid pT7-7 (Studier and Moffat, 1986) in order to express TTF-1 protein in *Escherichia coli*. Crude bacterial extracts from the TTF-1 producing strain gave DNase I footprints on the thyroglobulin promoter (Figure 2D, TTF-1 B.E.) which are identical to those obtained with the protein purified from calf thyroids (Figure 2D, TTF-1 PUR). These data show that the isolated cDNA encodes a protein showing the same binding specificity as TTF-1.

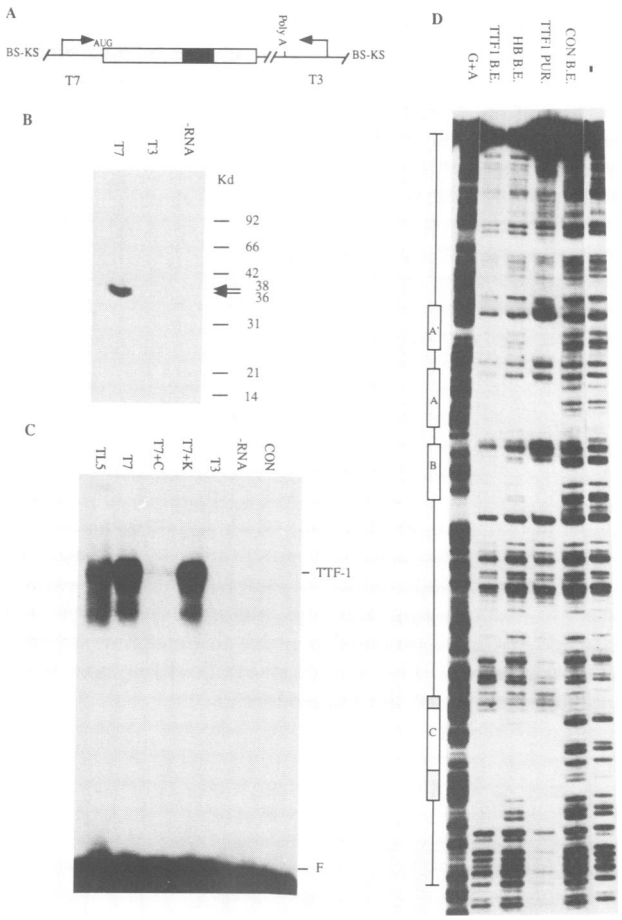

Fig. 2. The cloned cDNA has the same DNA binding properties as TTF-1. RNA was transcribed from the cDNA insert using T3 and T7 RNA polymerase promoters present in the Bluescript vector, and translated using rabbit reticulocyte lysates in presence of [^{35}S]methionine. The products of the translation were analysed on a SDS−PAGE and a gel retardation assay was performed. (**A**) Schematic representation of the cDNA cloned in Bluescript KS$^-$ (BS-KS$^-$). The coding region is an open box, and the homeobox filled. The arrows indicate the direction of the transcription of the T3 and T7 promoters. Poly(A), polyadenylation signal, (**B**) 10% SDS−PAGE of the *in vitro* synthesized TTF-1. The position of molecular weight markers and the apparent molecular weight of the translated protein are shown on the right. (**C**) Gel retardation assay using ^{32}P-labelled oligonucleotide C. TL5, nuclear extracts from the differentiated thyroid cell line, T7, T3, *in vitro* translation products generated from RNA transcribed by the T7 or T3 promoter flanking the cDNA insert. T7+C, T7+K, the same as T7 but competed by a 1000-fold excess of unlabelled oligonucleotide C and K, respectively, −RNA, the lysate used for *in vitro* translation in the absence of added RNA. CON, no protein added, F, free probe; TTF-1, bound probe. (**D**) DNase I protection on a ^{32}P-labelled DNA fragment of the thyroglobulin promoter, spanning from +36 to −284 with respect to the transcription start site (Civitareale *et al.*, 1989). Bacterial extracts were prepared from *E.coli* producing the entire protein (TTF-1 B.E.), the homeobox alone (HB, B.E.), or no TTF-1 (CON B.E.). TTF-1 PUR, purified protein from calf thyroids, −, no protein added to the reaction. G+A, Maxam and Gilbert purine ladder. On the left, the protected regions are boxed, the stippled areas indicate the sequences which are not protected using the homeobox alone.

The TTF-1 homeodomain can reproduce the binding specificity of the entire protein

The DNA binding domain of TTF-1 was defined more precisely using several deletions from the 5′ and 3′ side of the cDNA (see Materials and methods) as shown schematically in Figure 3A. The deletions were used to synthesize the corresponding proteins as described above, and tested for binding to oligonucleotide C. The data, presented in Figure 3B, demonstrate that all deletions extending into the homeodomain abolish DNA binding while no significant decrease in binding activity is observed with deletions which remove sequences either on the NH_2- or the COOH-side of the homeodomain. Also, a construct con-

taining only the homeodomain still displayed efficient DNA binding, suggesting that most, if not all, of the binding properties of TTF-1 are derived from the homeodomain.

The cotranslation of several deleted forms of TTF-1 cDNA with the wild-type cDNA never resulted in complexes of intermediate mobility in band-shift assays, suggesting that TTF-1 binds to DNA as a monomer (unpublished data).

The TTF-1 homeodomain was also expressed in *E.coli*. In footprinting experiments, crude bacterial extracts of the producing strain protected the same regions of the thyroglobulin promoter as the protein purified from calf thyroids (Figure 2D, HB, B.E.). The protected regions observed using the homeodomain containing extract are smaller, as can be expected given the probable reduced steric hindrance to DNase I by a smaller protein. Also, the homeodomain does not protect the A′ and B sites, which have an affinity of TTF-1 lower than the A and C sites (unpublished data), as efficiently as the full-size protein. The significance of this observation is limited by the unknown relative concentration of the two proteins in the crude extracts used for footprinting and will require experiments with purified bacterial proteins in order to determine whether other regions of TTF-1 can cause subtle differences in the binding capacity of the homeodomain.

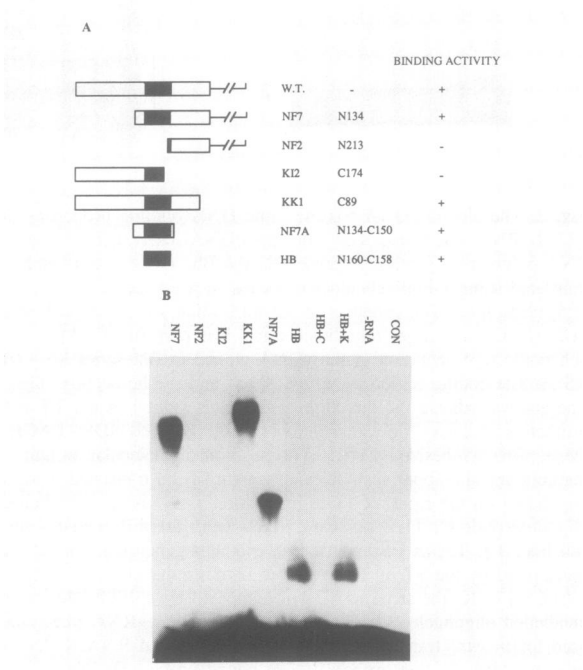

Fig. 3. The homeodomain alone is able to determine the binding specificity of TTF-1. (**A**) Schematic representation of the truncated TTF-1 proteins synthesized *in vitro* from the Bluescript vector using rabbit reticulocyte lysates. The open reading frame of the TTF-1 cDNA is an open box and the homedomain is filled. Next to the deletions is the name (first column), from which side (N for N-terminal and C for C-terminal) the deletion starts and the number of amino acids deleted in the translated protein with respect to the wild type (second column), the presence (+) or absence (−) of DNA binding activity (third column). (**B**) Gel retardation assay of the truncated proteins with ^{32}P-labelled oligonucleotide C. Approximately comparable amounts of ^{35}S-labelled proteins were used in the assay, as judged by SDS−PAGE (not shown). HB+C and HB+K, the assay was performed in the presence of 100 ng of unlabelled oligonucleotides C or K, respectively. −RNA and CON, the assay was performed with the proteins synthesized in absence of mRNA and without any protein, respectively.

The homeodomain of TTF-1 displays a novel binding specificity

It has been proposed that different amino acids sequence in the recognition helix of homeodomains reflect differences in the DNA binding specificity (Desplan *et al.*, 1988). In the homeodomains studied so far, residues at the amino-terminal of the recognition helix have been proposed not to play an important role in the determination of DNA binding specificity (Desplan *et al.*, 1988; Hoey and Levine, 1988). However, the residue at position 9 of the recognition helix does appear to play an important role in DNA binding specificity (Treisman *et al.*, 1989; Hanes and Brent, 1989). The homeodomain of TTF-1 shows considerable divergence along most of its sequence from the *Antennapedia* class homeodomains, while the recognition helix, including the glutamine at position 9, shows a high degree of homology (see Figure 6). Footprints obtained with purified TTF-1 on a total of eight binding sites found in two thyroid specific promoters, thyroglobulin (Civitareale *et al.*, 1989) and thyroperoxidase (Helen Francis-Lang, M.Price and R.Di Lauro, unpublished observations) allowed a consensus recognition sequence to be derived for TTF-1 (Figure 4), which shows no homology to the DNA sequences recognized by the *Antennapedia* class homeodomains. As shown in Figure 5, neither the complete TTF-1 nor the TTF-1 homeodomain bind to the oligonucleotide BS2, containing an *Antennapedia* class binding site (Müller *et al.*, 1988a) (Figure 5, lanes 11−13 and 18−20), which is instead efficiently recognized by the *Antennapedia* homeodomain (Figure 5, lanes 14−165). Conversely, the *Antennapedia* homeodomain is not able to bind the TTF-1 binding site (Figure 5, lanes 4−6). Moreover, the oligonucleotide C does not compete BS2 binding to *Antennapedia* and BS2 does not compete oligonucleotide C binding to TTF-1. TTF-1 is also unable to bind efficiently DNA sequences recognized by other *Antennapedia* class homeodomains, such as the TAA repeat and NP (Desplan *et al.*, 1988) (unpublished data). This finding is surprising since the extensive homology in the putative third helix of the TTF-1 and *Antennapedia* homeodomains, including the glutamine at position 9, might have suggested that TTF-1 would have the *Antennapedia* class DNA recognition. Consequently, it would appear that other residue(s) apart from glutamine at position 9 of the third helix must play an important role in the sequence specificity of DNA site selection by homeodomains.

The homoeodomain of TTF-1 belongs to a new class, with several members in the rat genome

The homeodomain of TTF-1 is very similar to the recently described *Drosophila* NK-2 homeodomain (Kim and

```
TTF-1 BS    G N N C A C T C A A G
            6     6 8 7 7 8 6 6 7

BS-2        G A A A A A G C C A T T A
(TAA) 5     T A A T A A T A A T A A T A A
NP          T C A A T T A A T G A
```

Fig. 4. Consensus binding site of TTF-1. TTF-1 BS shows the result of the alignment of eight TTF-1 binding sites. The numbers below the sequence indicate how many times each nucleotide is found in the eight binding sites aligned. BS-2, (TAA)5 and NP, three binding sites for *Antennapedia* class homeodomains (Müller *et al.*, 1988a; Beachy *et al.*, 1988; Desplan *et al.*, 1988) are shown for comparison.

Nirenberg, 1989). The two homeodomains have 49 identical amino acids residues out of 60 (82% identity), and the homology rises to 93% if conservative substitutions are taken into account (Figure 6). No clear homology between TTF-1 and NK2 is detected outside of the homeodomain (unpublished data), although the limited sequence information available for NK2 does not allow us to conclude whether TTF-1 and NK2 are truly equivalent genes. Comparison of the amino acids sequence of TTF-1 with the sequence of all the other known homeodomains shows an average homology of 27–40%, with the exception of the other members of the NK class (Kim and Nirenberg, 1989) and *Distal-less* (Cohen *et al.*, 1989), which are 50–63% homologous to TTF-1 (Figure 6). The presence of at least four members of the NK class in *Drosophila* prompted us to test whether other TTF-1-like genes are present in the rat genome. A probe derived from the 5′ end of the homeodomain, encompassing residues 1–47 (probe b in Figure 1C), was used since in this region the sequence of TTF-1 is most divergent from the sequences of other homeodomains. Low stringency hybridizaton to rat genomic DNA reveals the presence of four to six bands in several restriction digestions, suggesting that at least four other members of this class exist in the rat genome (Figure 7, lanes 3–6). Identical bands were obtained with a probe derived from the 3′ end of the homeobox encompassing residues 48–60 (probe c, Figure 1C, unpublished data). Only one band for each digestion was detected using a probe derived from the 5′ end of the cDNA (probe a, Figure 1C), which does not include the homeobox, showing that there is a single TTF-1 gene in the rat genome (Figure 7, lanes 1–3).

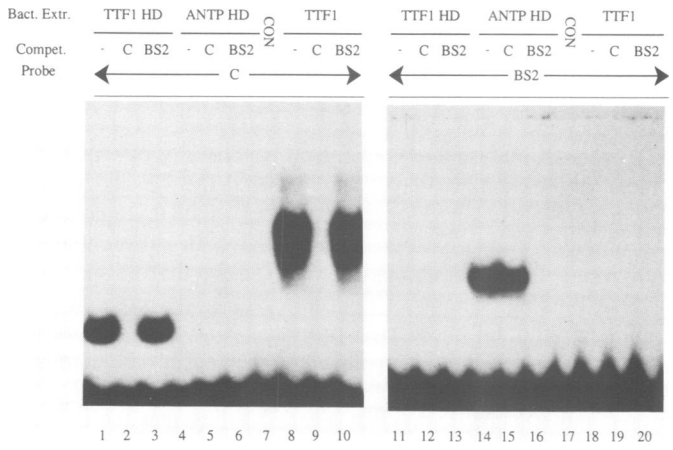

Fig. 5. DNA binding of TTF-1 and *Antennapedia* homeodomains to cognate DNA sequences. TTF-1, TTF-1 homeodomain and *Antennapedia* homeodomain were expressed in bacteria as described in Materials and methods. Crude bacterial extracts containing similar amounts of the expressed proteins were incubated with ^{32}P-labelled double stranded oligonucleotide C (lanes 1–10) and BS2 (Müller *et al.*, 1988a) containing an *Antennapedia* class binding sequence (lanes 11–20). Free and protein-bound probes were resolved on 12% PAGE, run in 0.5× Tris–borate buffer for 3 h at 4°C. Where indicated, cold double-stranded oligonucleotides were used as competitors at a concentration of 50 ng/tube. CON, no protein added.

Chromosomal localization of the TTF-1 gene

To determine whether TTF-1 is associated with one of the described *Hox* clusters (reviewed in Akam, 1989) the chromosomal localization of the gene was determined. A probe (3.9 kb) derived from the 3′ flanking region of a mouse TTF-1 genomic clone (unpublished data) and a 0.6 kb probe from the 5′ end of the rat cDNA clone (probe a, Figure 1C) were used in genomic Southerns and gave rise to unique bands with mouse and with human genomic DNA, respectively (unpublished data and Figure 7). These probes were subsequently used for chromosome mapping by *in situ* hybridization (Figure 8). For mouse gene mapping using the 3′ mouse genomic probe, 200 metaphase spreads were examined. There were 394 silver grains associated with chromosomes, 64 of which (16.2%) were located on chromosome 12. Furthermore, the grain distribution was not random; 68.7% mapped to the C1–C3 region strongly suggesting that the mouse TTF-1 gene lies in the 12C band of the murine genome.

In the human gene mapping, using the 5′ rat cDNA probe, of the 200 metaphase spreads examined, there were 388 silver grains associated with chromosomes, 52 of which (13.4%) were located on chromosome 14 long arm. A maximum was observed in the 14q13 band suggesting that the TTF-1 gene lies in this band in the human genome.

The 14q region of human chromosome 14 and the 12C region of mouse chromosome 12 shows synteny. Neither of these chromosomes have been reported to harbor other homeobox-containing genes. The murine *Hox-4.1* gene, previously located on mouse chromosome 12, has recently been remapped to chromosome 2 (Pravtcheva *et al.*, 1989).

TTF-1 mRNA is present in thyroid and in lung

To examine the tissue distribution of TTF-1 gene expression poly(A)$^+$ RNA was prepared from various tissues of 4

```
                                                                  % IDENTITY
                                                                  TO TTF-1
cons.      R  Y    Q   L   F      Y       R   A  L L    Q KIWFQ NRR K K
TTF1   RRKRRVLFSQ AQVYELERRF KQQKYLSAPE REHLASMIHL TPTQVKIWFQ NHRYKMKRQA
NK2    K-------TK --T------- R--R------ ------L-R- ---------- -----T--AQ      82
NK4    K--P------ ---L---C-- RLK---TGA- --II-QKLN- SA-------- -R---S--GD      63
NK3    KKRS-AA--H ---F------ A--R---G-- -SEM-KSLR- -E-------- -R---T--KQ      62
NK1    P-RA-TA-TY E-LVS--NK- -TTR---VC- -LN--LSLS- -E-------- -R-T-W-K-N      50
Dll    M--P-TIY-S L-LQQ-N--- QRTQ--AL-- -AE--ASLG- -Q-------- -R-S-Y-KMM      50
Antp   -KRG-QTYTR Y-TL---KE- HFNR--TRRR -IEI-HALC- -ER-I----- -R-M-W-KEN      38
Pit1   K----TTI-I -AKDA---H- GEHSKP-SQ- IMRM-EELN- EKEV-RV--C -R-QRE--VK      35
Oct2   ---K-TSIET NVRFA--KS- LANQKPTSE- ILLI-EQL-M EKEV-RV--C -R-Q-E--IN      30
Oct1   ---K-TSIET NIRVA--KS- LENQKPTSE- ITMI-DQLNM EKEVIRV--C -R-Q-E--IN      27
Isl-1  TTRV-TVLNE K-LHT-RTCY AANPRPD-LM K-Q-VE-TG- S-RVIRV--- -K-C-D-KRS      27
               --------------------     ----------     ----------
                        I                   II             III
```

Fig. 6. Comparison of the TTF-1 homeodomain with homeodomains of the NK class, *Antennapedia, Distal-less* and homeodomains found in mammalian promoter/enhancer binding proteins. Dashes indicate identity with the corresponding residue in the TTF-1 homeodomain. The percentage identity of the TTF-1 homeodomain to the other homeodomains is shown in the column on the right. The top line shows the consensus sequences for homeodomains obtained by Scott *et al.* (1989). References for all the listed homeodomains are found in the text.

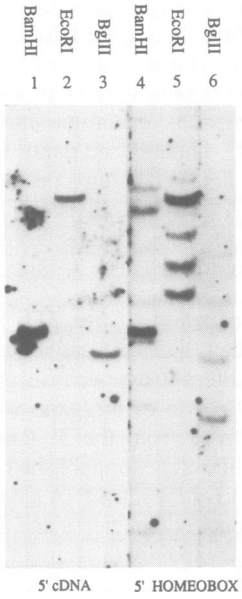

Fig. 7. The rat genome contains sequences related to the TTF-1 homeobox. 8 μg of rat genomic DNA were digested with *Bam*HI (lanes 1 and 4), *Eco*RI (lanes 2 and 5), *Bgl*II (lanes 3 and 6) and electrophoresed in 0.8% agarose gels. After Southern blotting Hybond-N filters were hybridized according to Church and Gilbert (1984) with either a probe from the 5′ end of the TTF-1 cDNA (probe a, Figure 1C) (lanes 1−3) or with a probe containing the 5′ end of the homeobox encompassing amino acids 1−47 of the homeodomain (probe b, Figure 1C) (lanes 4−6). Filters were washed at low stringency in 1× SSC at 65°C revealing several bands with the homeobox probe but single bands with the 5′ cDNA probe.

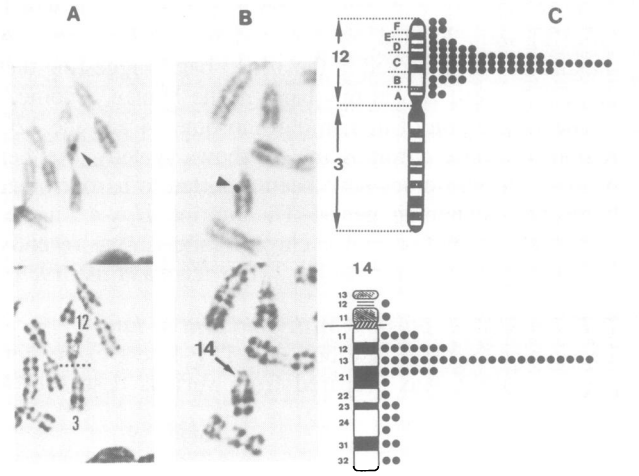

Fig. 8. Localization of the TTf-1 gene in the mouse and human genomes by *in situ* hybridization. **(A)** (upper) indicates the specific site of hybridization to mouse chromosome 12 and **(B)** (upper) to human chromosome 14 in representative metaphase spreads. The arrowheads indicate the silver grains after autoradiography of Geimsa stained chromosomes. In the lower part of panels A and B, the localization of the silver grains were determined by R-banding (fluorochrome-photolysis-geimsa method). The dotted line (lower panel A) refers to the centromeric fusion between mouse chromosomes 12 and 3. The arrowhead (lower panel B) indicates human chromosome 14-long arm. **(C)** **(upper)** G-band diagram of WMP mouse Rb(3;12) chromosome, indicating the distribution of silver grains on 64 labelled chromosomes. Most of the grains localized to the [C1−C3] region of chromosome 12. **(lower)** Idiogram of human G-banded chromosome 14 illustrating the distribution of silver gains on 52 labelled chromosomes. Most of the grains localized to [q12−q21] with a maximum in the 14q13 band.

week old rats and analysed by Northern hybridization using a 0.6 kb probe derived from the 5′ end of the rat cDNA (probe a, Figure 1C). TTF-1 displays tissue specific expression (Figure 9A). A transcript of the expected size (2.4 kb) is observed in thyroid tissue, in the differentiated thyroid cell line FRTL5 and in lung. A higher resolution of the thyroid and lung RNAs (Figure 9B) shows that these tissues contain at least two other transcripts (2.7 and 3.2 kb) related to TTF-1. Furthermore the relative abundance of the transcripts is different, the lung containing a higher proportion of the longer transcripts. The same result is observed using a probe derived from the homeobox (probe c, Figure 1C) and a probe at 0.6 kb derived from the 3′ non-translated region (probe d, Figure 1C) suggesting that all the TTF-1 related mRNAs code for proteins with identical DNA binding specificity (unpublished data). In order to determine whether the TTF-1 mRNA observed in lung gives rise to a protein similar to TTF-1, nuclear extracts were prepared from rat lungs. An activity was observed in these extracts which gives rise to a complex with the oligonucleotide C whose mobility is identical to that of the complex seen in thyroid extracts (Figure 10).

Discussion

In order to analyse the biochemical basis of thyroid specific gene expression we have characterized the *cis* and *trans* acting signals required for the expression of the thyroglobulin promoter. The promoter is only active in differentiated thyroid cells and intact binding sites for two thyroid specific DNA binding proteins, TTF-1 and TTF2, are required for full activity. Of these two proteins, TTF-1 is the most relevant, as mutations in its binding sites have the most dramatic effect on promoter function (Civitareale *et al.*, in preparation). The finding that several binding sites for TTF-1 are also present in the promoter of another thyroid specific gene, thyroperoxidase (Helen Francis-Lang, M.Price and R.Di Lauro, unpublished observations), suggests that the presence of TTF-1 in thyroid is important for the expression of the differentiated thyroid phenotype.

A deeper understanding of the role of the two thyroid specific factors in differentiation requires additional information on their structure and expression. We report in this paper the isolation of a cDNA encoding TTF-1, and an initial characterization of the structure of the TTF-1 protein and of the expression of its mRNA.

The cloned cDNA encodes a protein that binds to the same sites on the thyroglobulin promoter as calf and rat TTF-1. Translation of the longest open reading frame would result in a protein of 38 kd, while the protein purified from calf thyroids was estimated to be 21 kd (Civitareale *et al.*, in preparation), presumably due to proteolysis during the purification procedure. We believe that the 38 kd protein corresponds, or is very close, to the size of the mature protein, based on the length of the cloned cDNA and the size of the TTF-1 mRNA, observed in Northern blot experiments. We have recently isolated the TTF-1 gene and identified the major transcriptional start site. Inspection of the 5′ leader sequence revealed that an additional methionine codon lies adjacent to nucleotide 1 of Figure 1B and no other, in-frame, methionine codons are present (unpublished data). We conclude that the protein sequence shown in Figure 1B

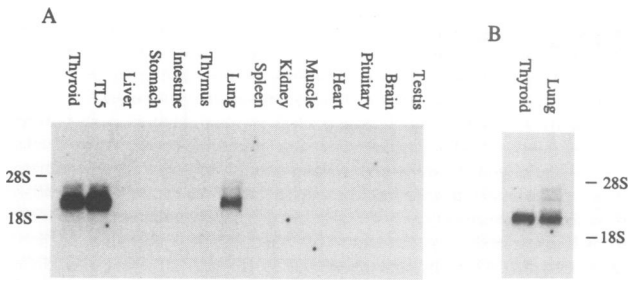

Fig. 9. Specificity of TTF-1 gene expression. (**A**) Various 4 week old rat tissues were analysed for TTF-1 gene expression. Poly(A)$^+$ RNAs (4 μg/lane) were electrophoresed on formaldehyde agarose gels and transferred to nylon membrane. Hybridization was to the 5′ cDNA probe (probe a, Figure 1C). The position of 18S and 28S rRNAs is indicated. (**B**) Poly(A)$^+$ RNAs from thyroid and lung were run in two contiguous lanes at higher resolution on a different gel.

Quality and quantity of the mRNAs was determined by ethidium bromide staining.

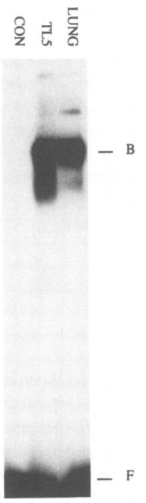

Fig. 10. TTF-1 protein is detectable in lung nuclear extracts. Gel retardation assay with ^{32}P-labelled oligonucleotide C of nucelar extracts from FRTL-5 cells and lung. CON, no extract added. F, free probe; B, bound probe.

represents the full sequence of TTF-1, minus the initiator methionine. The amino acid sequence of the TTF-1 protein shows some features described for other transcription factors: it contains a homeodomain, which we have demonstrated to be necessary and sufficient for DNA binding. A region rich in glutamine and alanine is found just downstream of the homeodomain. Similar regions have been identified in some transcription factors as sequences responsible for activation of transcription (Mitchell and Tjian, 1989) and are also found in other homeodomain-containing proteins (Schneuwly *et al.*, 1986).

Four mammalian transcription factors have been shown to contain homeodomains (the POU homeodomains, Figure 6) that diverge significantly from the consensus sequence obtained aligning all the known homeodomains (Scott *et al.*, 1989), showing a maximun of 13 residues identical out of the 21 proposed for consensus (Bodner *et al.*, 1988; Clerc *et al.*, 1988; Ingraham *et al.*, 1988; Müller *et al.*, 1988b; Sturm *et al.*, 1988; Frain *et al.*, 1989). In addition, these transcription factors also contain another common region, that has been called the POU specific domain. The POU homeodomain and the POU specific domain have subsequently been found in several genes whose expression is developmentally regulated (He *et al.*, 1989). Another mammalian DNA binding protein, Isl-1, which binds specifically to the enhancer of the insulin gene (Karlsson *et al.*, 1990), contains a homeodomain (13 residues identical out of the 21 proposed for consensus, Figure 6), divergent from the POU homeodomain, and another region, rich in histidine and cysteine (the Cys-His domain). This region is conserved, since a similar structure has been found in the *Caenorhabditis elegans* cell lineage gene lin-11 (Freyd *et al.*, 1990). TTF-1 represents a third type of mammalian homeodomain-containing gene, since its homedomain matches well to the consensus (18 residues identical out of the 21 proposed for consensus) while the remaining residues still would assign TTF-1 to a new homeodomain class (Figure 6). In addition TTF-1 does not contain either a POU or a Cys-His domain. Whether TTF-1 also shares additional domains with the other genes containing a related homeobox which have been detected in the rat genome (Figure 7), awaits the isolation and sequence of other members of the TTF-1 related family.

Inspection of the TTF-1 binding site shows no detectable homology to several sequences proposed as targets for the *Antennapedia* class homeodomains (Beachy *et al.*, 1988; Desplan *et al.*, 1988; Müller *et al.*, 1988a). This observation is surprising considering that although TTF-1 and *Antennapedia* class homeoboxes diverge substantially upstream, they show a high degree of homology in the region that has been proposed to determine the binding specificity of the homeodomain, i.e. in the region contributing to the third helix (Quian *et al.*, 1989). Direct tests for the ability of TTF-1 to recognize three binding sites which are bound by *Antennapedia* class homeodomains were negative, suggesting that residues outside the putative third helix contribute to the DNA binding specificity of TTF-1. We are currently exchanging segments of the TTF-1 homeodomain with corresponding parts of the *Antennapedia* homeodomain in an attempt to define which residues determine the TTF-1 binding specificity.

The striking homology between the TTF-1 and the NK2 homeodomains suggests that TTF-1 could be the murine homologue of the *Drosophila* NK2 gene. This possibility may not yet be ruled out since only a small amount of NK2 sequence is available. A close correspondence in developmental function would not be expected, since the tissues in which TTF-1 is expressed in rat (thyroid and lung) do not have counterparts in insects. The relevant structural features of the NK homeoboxes have already been described (Kim and Nirenberg, 1989), but the DNA binding specificity could not be described because the target of the NK2 homeobox is unknown. The *Drosophila* NK family has four members and the strong homology between the homeoboxes of TTF-1 and NK-2 suggested that the mammalian genome may also contain homologues of the other members of the *Drosophila* NK family. Preliminary studies indicate that this is indeed the case as probes derived from the homeobox of TTF-1 detect several bands in mouse and rat genomic DNA blots when hybridized at low stringency, whereas at high stringency only one band is detected with several TTF-1 probes. Analysis of 50 kb of DNA sequence surrounding the TTF-1 murine locus failed to detect any TTF-1 related

homeodomain (M.Price and R.Di Lauro, unpublished observation), suggesting that TTF-1 is not a member of a cluster of homeodomain-containing genes, as is observed for all *Antennapedia* type homeobox genes (Akam, 1989).

The TTF-1 gene is located on syntenic regions of mouse chromosome 12 and human chromosome 14, where no other homeodomain-containing genes have so far been mapped. The location of the TTF-1 gene on chromosome 12 raised considerable interest because of the localization to this chromosome of the *hyt* allele, a recessive mutation that causes severe hypothyroidism (Beamer *et al.*, 1981). Preliminary RFLP mapping experiments show, however, that TTF-1 and *hyt* are only loosely linked (B.Taylor, personal communication), therefore excluding the possibility that the TTF-1 gene is the target of the *hyt* mutation.

Previous binding studies, carried out with nuclear extracts from several tissues and cultured cells, only detected TTF-1 binding activity in thyroid cells or tissue. However, analysis of mRNA revealed TTF-1 mRNA in lung and subsequent binding studies demonstrate that the corresponding binding activity is also present in lung nuclear extracts. The expression of the TTF-1 gene in lung and thyroid is, to our knowledge, the first marker specific for these two tissues. An initial study of the pattern of expression of TTF-1 by *in situ* hybridization to 19 days rat fetus confirm the expression of the TTF-1 gene in thyroid and lung and, in addition, demonstrate that the expression in lung is restricted to endodermally derived epithelial cells (unpublished data). Interestingly, lung and thyroid develop in close proximity from the primitive pharynx, suggesting that the activation of the TTF-1 gene may be a response to positional cues specifying the median anterior endoderm.

The TTF-1 gene is one of the few homeodomain-containing genes in vertebrates, together with *Cdx-1* in mouse (Duprey *et al.*, 1988) (the homologue of the *Drosophila Caudal* gene), XlHbox 8 in *Xenopus laevis* (Wright *et al.*, 1988), which are expressed in tissues of endodermal derivation. Amino acid sequence comparison of TTF-1 with the *Cdx-1* and XlHbox 8 genes do not show significant homology, an observation that would argue against the existence of a class of related homeodomains involved in endodermal differentiation (Wright *et al.*, 1988).

The role of TTF-1 in the lung is at present unclear. mRNAs for two thyroid specific genes, thyroglobulin and thyroid peroxidase, are not detectable in this tissue (unpublished data). It is conceivable that a mechanism counteracting TTF-1 action on the thyroid specific genes is operating in the lung. It is interesting that there are at least three species of TTF-1 mRNA and the longest transcript appears to be only present in lung. RNase mapping experiments with fragments derived from the 5′ end of the TTF-1 gene suggest that the larger transcripts may derive from alternative transcription initiation sites and that the longer mRNAs would encode a protein with an amino-terminal extension (unpublished data). Whether the different ratio of the two proteins plays a role in the biological activity of TTF-1 in thyroid and lung awaits the availability of an assay to test the transcriptional activation of the thyroglobulin promoter by TTF-1.

Materials and methods

Construction of a calf thyroid cDNA library and isolation of TTF-1 cDNA clones

5 μg of poly(A)$^+$ RNA, isolated from calf thyroids as described below, were used to synthesize cDNA using a kit from Pharmacia. 150 ng of double stranded cDNA, having *Eco*RI adapters at both ends, were ligated to bacteriophage λgt11 arms cut with *Eco*RI (2 μg). Phage DNA was packaged *in vitro* using commerical packaging extracts (Stratagene) yielding a library of 1.3 × 10^6 p.f.u. A 36 bases long, degenerate olgionucleotide (TTF-1-G, see Figure 1A), encoding 13 amino acids peptide obtained from purified TTF-1, was designed on the basis of codon usage frequencies in man (Lathe, 1985). The oligonucleotide mixture was labelled by polynucleotide kinase. Phage plating and duplicate filter lifts were carried out according to standard procedures (Maniatis *et al.*, 1982). Prehybridization was carried out at 37°C for 4 h in 6× SSC, 5× Denhardt's, 0.2% SDS, 50 μg/ml salmon sperm DNA, 1 mM ATP; hybridization was performed in the same buffer containing 1 pmol/ml of labelled oligonucleotide TTF-1G for 22 h at 37°C. Filters were washed twice in 6× SSC, 0.05% Na pyrophosphate, 0.2% SDS at 37°C for 20 min and the twice for 10 min at 50°C. Phage DNA was prepared from purified plaques and the cDNA inserts were digested with *Sau*3A1 and subcloned in Bluescript KS$^-$ vector, hybridized again with the oligonucleotide TTF1-G and the positive clones were sequenced by the dideoxy method (Hattori and Sakaki, 1986).

A rat cDNA library was screened with the α-^{32}P-labelled 1.8 kb insert of the TTF-1 calf clone in the conditions described above except that hybridization was at 65°C and stringent washes were at 60°C in 2× SSC, 0.05% SDS.

In vitro transcription/translation

Plasmid DNA was linearized with a suitable restriction enzyme at the 3′ side of the insert. Transcription reactions were carried out in 10 μl containing 1 μg of linearized DNA template, 40 mM Tris−HCl pH 8, 8 mM $MgCl_2$, 2 mM spermidine, 50 mM NaCl, 0.4 mM rNTPs, 0.25 U m7GpppG, 40 U RNAsin, 10 U T3 or T7 RNA polymerase (Stratagene) for 1 h at 37°C. The capped RNA was phenol extracted, precipitated with 3 vol of ethanol in 2.5 M ammonium acetate and reprecipitated in 0.3 M Na acetate. Approximately 1 μg of cDNA was translated in a 25 μl reaction containing 17.5 μl of rabbit reticulocyte lysate (Promega), 20 μM amino acids mixture (minus methionine), 1 mCi/ml [^{35}S]methionine for 90 min at 30°C. ^{35}S-labelled translation products were analysed by SDS−PAGE. Gel mobility retardation assays with ^{32}P-labelled oligodeoxynucleotide were performed using comparable amounts of translated proteins (1−4 μl) as judged from SDS−PAGE. Gels were dried and exposed to X-ray films.

Plasmids construction and DNA sequencing

In order to subclone the cDNA insert into a plasmid vector, the insert of λ 12.3 which contains an internal *Eco*RI site, was excised by partial *Eco*RI digestion, purified on agarose gel and cloned into the *Eco*RI site of Bluescript KS$^-$.

Deletions were constructed by the exonuclease III−Mung bean method (Henikoff, 1984). Since no unique sites leaving a 3′ overhang were available at the 5′ side of the insert, the plasmid was digested with *Not*I, and partial double stranded oligonucleotide, containing a 5′ *Not*I end and a 3′ overhang of non self-complementary sequence was ligated at the *Not*I site. The plasmid was then cleaved with *Bam*HI and subjected to sequential treatment with exonuclease III and mung bean nuclease. Deleted DNAs were ligated to *Nco*I linkers (5′-CCCATGGG), digested with *Nco*I and recircularized with T4 DNA ligase. The *Nco*I linkers were selected to provide an ATG codon for initiation of translation of the deletions which would delete the natural ATG. A similar procedure was carried out with the plasmid cDNA clone cleaved with both *Kpn*I and *Sal*I, to obtain deletions from the 3′ side of the insert. The deleted plasmid DNAs were transformed in XL1-Blue and the resulting colonies screened by restriction analysis. Clones containing inserts of different sizes were sequenced by the dideoxy method and the resulting sequences were assembled using the UWGCG software. Using both the wild-type and the deletion mutants the entire insert was sequenced more than once on both strands.

Chromosomal localization

Chromosome mapping was carried out as described (Mattei *et al.*, 1985). The probes (see Results), inserted in the Bluescript vector, were tritium labelled to a specific activity of 1.3 × 10^8 d.p.m./μg.

RNA extraction

RNA isolation was carried out by the acid guanidium thiocyanate phenol procedure (Chomczynski and Sacchi, 1987) and poly(A)$^+$ RNA was prepared from this by oligo(dT) cellulose chromotography. For RNA gel blots, 4 μg of poly(A)$^+$ RNA were electrophoresed in formaldehyde/agarose gels and transferred to Hybond-N membrane (Amersham). Hybridizations and washings were carried out according to Church and Gilbert (1984) with probes labelled by random oligo priming.

DNA binding assays

Gel mobility shift assays and DNase I footprints were performed as described (Civitareale *et al.*, 1989).

Protein expression in bacteria

To express TTF-1 and its homeodomain alone in bacteria, cDNA inserts were cloned into the expression vector pT7.7 (Studier and Moffatt, 1986). The entire TTF-1 cDNA was cloned using the *Hin*dIII−*Cla*I restriction sites of the pT7.7 polylinker, therefore the ATG present in the polylinker was in frame with the first ATG of the TTF-1 cDNA. The predicted fusion protein contains an additional 22 residues at the amino terminus.

The cDNA corresponding to the homeobox was amplified by PCR using the oligos 5′-CCCGGGCATATGCGCCGGAAGCGTCGGGTGCTC-TTCTCCCAG-3′ and 5′-CGCTGTCCTGCTGCAGTTACTGCTGCG-CCGCC-3′. The amplified DNA was cut with *Nde*I and *Pst*I and cloned in the homologous sites of pT7.7. In this way an ATG and a stop codon are positioned respectively at the 5′ and 3′ sides of the insert to give a translation product of 68 amino acids. Plasmids were used to transform the BL21 bacterial strain containing pLysS (Moffatt and Studier, 1987). Recombinants were grown at 37°C to OD_{600} 0.6 and then induced by 1 mM IPTG for 3 h. Induced cells were harvested, washed, resuspended in 0.1 culture volume of buffer A (20 mM HEPES, pH 7.9, 1 mM ETDA, 5% glycerol, 1 mM DTT, 0.4 M NaCl, 0.5 mM PMSF) and sonicated. Disrupted cells were centrifuged at 100 000 *g* for 1 h at 4°C, the supernatant was collected and dialysed against buffer B (20 mM Tris pH 8.5, 0.2 M KCl, 1 mM DTT, 20% glycerol, 0.1 mM EDTA, 0.5 mM PMSF). The amount of recombinant protein was ~2−5% of total protein as judged by SDS−PAGE.

The *Antennapedia* homeodomain was prepared according to the protocol described above using a previously described construct (Müller *et al.*, 1988a).

Acknowledgements

We thank Iain Mattaj, Denis Duboule and Riccardo Cortese for critical reading of the manuscript, Walter Gehring for providing the plasmid expressing the *Antennapedia* homeodomain, the oligonucleotide service of the EMBL for oligonucleotides and the sequencing service of the EMBL for help in completing the sequence. Stacy Wiedenmann and Maria Polycarpou-Schwartz provided excellent technical assistance. M.D.F. is a recipient of an EEC training contract, M.P. was supported by an EMBL fellowship. S.G. and G.D. were supported by fellowships from the Ministero della Pubblica Istruzione, Italy.

References

Akam,M. (1989) *Cell*, **57**, 347−349.
Ambesi-Impiombato,F.S., Parks,L.A.M. and Coon,H.G. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 3455−3459.
Avvedimento,V.E., Musti,A.M., Bonapace, M., Fusco,A. and Di Lauro,R. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 1744−1748.
Avvedimento,V.E., Obici,S,. Sanchez,M., Gallo,A., Musti,A. and Gottesman,M.E. (1989) *Cell*, **58**, 1135−1142.
Beamer,W.G., Eicher,E.M., Maltais,L.J. and Southard,J.L. (1981) *Science*, **212**, 61−62.
Beachy,P.A., Krasnow,M.A., Gavis,E.R. and Hogness,D.S. (1988) *Cell*, **55**, 1069−1081.
Bodner,M., Castrillo,J.-L., Theill,L.E., Deerinck,T., Ellisman,M. and Karin,M. (1988) *Cell*, **55**, 505−518.
Chomczynski,P. and Sacchi,N. (1987) *Anal. Biochem.*, **162**, 156−159.
Church,M.G. and Gilbert,W. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 1991−1995.
Civitareale,D., Lonigro,R., Sinclair,A. and Di Lauro,R. (1989) *EMBO J.*, **8**, 2537−2542.
Clerc,R.G., Corcoran,L.M., LeBowitz,J.H., Baltimore,D. and Sharp,P.A. (1988) *Genes Dev.*, **2**, 1570−1581.
Cohen,S.M., Bronner,G., Kuttner,F., Jurgens,G. and Jackle,H. (1989) *Nature*, **348**, 432−434.
Colletta,G., Pinto,A., Di Fiore,P.P., Fusco,A., Ferrentino,M., Avvedimento,V.E., Tsuchida,N. and Vecchio,G. (1983) *Mol. Cell Biol.*, **3**, 2099−2109.
Desplan,C., Theis,J. and O'Farrell,P.H. (1988) *Cell*, **54**, 1081−1090.
Duprey,P., Chowdury,K., Dressler,G.R., Balling,R., Simon,D., Guenet, J.-L. and Gruss,P. (1988) *Genes Dev.*, **2**,
Frain,M., Swart,G. Monaci,P., Nicosia,A., Stampfli,S., Frank,R. and Cortese,R. (1989) *Cell*, **59**, 145−157.
Freyd,G., Kim,S.K. and Horvitz,R.H. (1990) *Nature*, **344**, 876−879.
Hanes,S.D. and Brent,R. (1989) *Cell*, **57**, 1275−1283.
Hattori,M. and Sakaki,Y. (1986) *Anal. Biochem.*, **152**, 323−328.
He,X., Treacy,M.N., Simmons,D.M., Ingraham,H.A., Swanson,L.W. and Rosenfeld,M.G. (1989) *Nature*, **340**, 35−42.
Henikoff,S. (1984) *Gene*, **28**, 351−359.
Hoey,T. and Levine,M. (1988) *Nature*, **332**, 858−861.
Ingraham,H.A., Chen,R. Mangalam,H.J., Elsholtz,H.P., Flynn,S.E., Lin,C.R., Simons,D.M., Swanson,L. and Rosenfeld,M.G. (1988) *Cell*, **55**, 519−529.
Karlsson,O., Thor,S., Norberg,T., Ohlsson,H. and Edlund,T. (1990) *Nature*, **344**, 879−882.
Kawaoi,A. and Tsuneda,M. (1985)*Acta Endocrinol.*, **108**, 518−524.
Kim,Y. and Nirenberg,M. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 7716−7720.
Lathe,R. (1985) *J. Mol. Biol.*, **183**, 1−12.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Mattei,M.G., Philip,N., Passage,E., Moisan,J.P., Mandel,J.L. and Mattei,J.F. (1985) *Hum. Genet.*, **69**, 268−271.
Mitchell,P.J. and Tjian,R. (1989) *Science*, **245**, 371−378.
Moffatt,B.A. and Studier,F.W. (1987) *Cell*, **49**, 221−227.
Müller,M., Affolter,M., Leupin,W., Otting,G., Wütrich,K. and Gehring,W.J. (1988a) *EMBO J.*, **7**, 4299−4304.
Müller,M.M., Ruppert,S., Schaffner,W. and Matthias,P. (1988b) *Nature*, **336**, 544−551.
Musti,A.M., Ursini,M.V., Avvedimento,V.E., Zimarino,V. and Di Lauro,R. (1987) *Nucleic Acids Res.*, **15**, 8149−8166.
Pravtcheva,D., Newman,M., Hunihan,L., Lonai,P. and Ruddle,F.H. (1989) *Genomics*, **5**, 541−545.
Quian,Y.Q., Billeter,M., Otting,G., Müller,M., Gehring,W.J. and Wüttrich,K. (1989) *Cell*, **59**, 573−580.
Schneuwly,S., Kuroiwa,A., Baumgartner,P. and Gehring,W.J. (1986) *EMBO J.*, **5**, 733−739.
Scott,M.P., Tamkun,J.W. and Hartzell,G.W. (1989) *Biochim. Biophys. Acta*, **989**, 25−48.
Studier,F.W. and Moffatt,B.A. (1986) *J. Mol. Biol.*, **189**, 113−130.
Sturm,R.A. and Herr,W. (1988) *Nature*, **336**, 601−604.
Sturm,R.A., Das,G. and Herr,W. (1988) *Genes Dev.*, **2**, 1582−1599.
Treisman,J., Gönczy,P., Vashishtha,M., Harris,E. and Desplan,C. (1989) *Cell*, **59**, 553−562.
Wright,C.V.E., Schnegelsberg,P. and De Robertis,E.M. (1988) *Development*, **104**, 787−794.

Received on July 4, 1990; revised on July 26, 1990

Note added in proof

The sequence data reported here have been deposited with the EMBL Data Library with the accession number X53858.