Human thrombomodulin gene is intron depleted: Nucleic acid sequences of the cDNA and gene predict protein structure and suggest sites of regulatory control

(protein C/coated-pit receptors/endothelial cell)

ROBERT W. JACKMAN*, DAVID L. BEELER*, LINDA FRITZE*†, GERALD SOFF*†, AND ROBERT D. ROSENBERG*†

*Department of Biology and Whitaker College, Massachusetts Institute of Technology, Cambridge, MA 02139; and †the Department of Medicine, Harvard Medical School and the Beth Israel Hospital, Boston, MA 02215

Communicated by Phillips W. Robbins, June 19, 1987

ABSTRACT We have isolated a human thrombomodulin cDNA, and a human genomic clone containing the putative promoter domain, as well as the translated and untranslated regions of the endothelial cell receptor. The nucleotide sequence of the thrombomodulin cDNA allows us to provide a complete picture of the structure of this endothelial cell receptor, and to confirm its homology to the human low density lipoprotein receptor. The nucleotide sequence of the thrombomodulin gene suggests areas within the putative promoter domain that may be critical for regulating expression of the human endothelial cell receptor, indicates a potential signal peptide, and shows that no introns are present within the coding region. The overall organization of the human thrombomodulin gene is surprising because it represents an example of a gene that contains epidermal growth factor type B repeats and a membrane spanning region, which are not isolated within discrete exons.

Thrombomodulin (TM) is a specific endothelial cell receptor that forms a 1:1 molecular complex with thrombin (1). This interaction product is capable of rapidly converting protein C to activated protein C, which proteolytically destroys the activated cofactors of the coagulation mechanism and thereby dramatically suppresses the amount of thrombin generated. The structure of TM was completely unknown until the recent cloning and sequencing of a cDNA that encoded the C-terminal two-thirds of the bovine endothelial cell receptor (2). These studies demonstrated that bovine TM is structurally similar to coated pit receptors and is organized into domains that resemble those of the low density lipoprotein (LDL) receptor. In the present communication, we report the isolation and sequencing of human TM cDNAs, which almost span the endothelial cell receptor transcript, and a human genomic clone, which contains the putative promoter domain as well as the translated and untranslated regions of this membrane component.‡ The nucleotide sequence of the cDNAs allows us to provide a complete picture of the structure of human TM and to confirm the homology of the endothelial cell receptor to the human LDL receptor. The nucleotide sequence of the genomic clone suggests areas within the putative promoter domain that may be critical for regulating expression of the human endothelial cell receptor, indicates a potential signal peptide, and shows that no introns are present within the coding region. The overall organization of the TM gene is surprising because it represents a unique example of a gene that contains epidermal growth factor (EGF) type B repeats and a membrane spanning region, which are not isolated within discrete exons.

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

METHODS

Screening of Human cDNA and Genomic Libraries. The human cDNA and genomic libraries were screened (3) on *Escherichia coli* Y 1088 or LE 392 (cDNA library), and LE 392 (genomic library) with Colony Plaque Screen (New England Nuclear) according to the following modifications of the manufacturer's instructions. The filters were lifted in duplicate, air-dried, autoclaved for 1 min, air-dried again, and then baked at 80°C for 1 hr (4). Prior to hybridization, the filters were washed twice at room temperature for 30 min in $1\times$ SSC/0.1% NaDodSO₄ ($1\times$ SSC = 0.15 M NaCl/0.015 M sodium citrate). The prehybridization of filters was carried out overnight with hybridization subsequently conducted at 42°C for 16–24 hr in standard solution (5) with 30% formamide added for bovine cDNA probes or 50% formamide added for human cDNA probes. The bovine and human cDNA probes were labeled by nick-translation (Bethesda Research Laboratories kit), and then added at $2-6\times 10^6$ cpm per filter. After hybridization, the filters were washed to a final stringency of $2\times$ SSC/0.1% NaDodSO₄, 50°C for the bovine cDNA probe and $0.5\times$ SSC/0.1% NaDodSO₄, 55°C for the human cDNA probes, and then examined by autoradiography. All of the above operations were carried out in parafilm-sealed 150-mm Petri dishes with 10–15 filters per dish.

Southern Blot Analyses of Human Genomic DNA. Human lymphocytes were obtained by Ficoll-Paque density gradient centrifugation of whole blood (Pharmacia), and human genomic DNA was isolated from the lymphocytes (6). The DNA was precipitated at 10 µg/ml with EtOH, rehydrated in TE0.1, and completely digested with various restriction endonucleases. Samples of 7.5 µg of digested DNA were electrophoresed on 1% agarose at 1 V/cm, and then the gel was short UV-irradiated for 2 min prior to capillary blotting onto Zeta-Probe (Bio-Rad) in 0.4 M NaOH for 2 days. After drying at 80°C, the filters were prewashed at 68°C in $0.2\times$ SSC/1% NaDodSO₄ for 2 hr, and then prehybridized and hybridized in standard solution (5) at 42°C with 50% formamide added. Final washes were conducted at 68°C in $0.1\times$ SSC/0.5% NaDodSO₄.

Nucleotide Sequence Determinations and Computer Analyses of Structural Data. Nucleotide sequences were usually determined by the dideoxy chain-termination method of Biggin *et al.* (7). However, G+C-rich regions that exhibit major band compressions were also sequenced by the dideoxy chain-termination method using Klenow DNA polymerase and deoxy-7-deazaguanosine triphosphate as described

Abbreviations: LDL, low density lipoprotein; EGF, epidermal growth factor; TM, thrombomodulin; HUVE, human umbilical vein endothelial cells.

‡This sequence is being deposited in the EMBL/GenBank data base (Bolt, Beranek, and Newman Laboratories, Cambridge, MA, and Eur. Mol. Biol. Lab., Heidelberg) (accession no. J02973).

by Mizusawa *et al.* (8) or Sequenase and deoxyinosine triphosphate as outlined in the manufacturer's instructions (United States Biochemical, Cleveland). Computer analyses of sequence data were performed using the programs of the Protein Identification Resource and University of Wisconsin Biotechnology Center as adapted for The Whitaker College, Massachusetts Institute of Technology, by W. Gilbert.

RESULTS

Isolation of a Human TM cDNA Clone and the Nucleotide Sequence of the Endothelial Cell Receptor Transcript. Two bovine TM cDNA fragments spanning nucleotides 308–1908 (BovA) (coding region), and 2133–2901 (BovB) (3' untranslated region) (sequence numbers from ref. 2) were used to screen 6×10^5 recombinant clones from a λ gt11 cDNA library constructed by Ginsburg *et al.* (9) using oligo(dT) priming of poly(A)⁺ RNA isolated from human umbilical vein endothelial cells (HUVE). Seven positive phage plaques were detected. The largest insert (λ gt11-TM1) exhibited a size of ≈ 2500 base pairs (bp) and was characterized by subcloning into M13mp18 and M13mp19 with the subsequent sequencing of 5' and 3' ends. The data revealed that λ gt11-TM1 exhibited 5' and 3' termini, which were $\approx 70\%$ identical to two regions found within the bovine TM cDNA (2). A new 492-bp probe, from the 5' end of λ gt11-TM1 to an internal *Kpn* I site, was used to screen an additional 6×10^5 recombinant clones from the same oligo(dT)-primed λ gt11 cDNA library without obtaining additional 5' sequence. Therefore, the same probe was then used to screen 2×10^6 recombinant clones from a randomly primed λ gt11 cDNA expression library constructed by Ginsburg *et al.* (9) from HUVE poly(A)⁺ RNA. Five positive phage were detected with one recombinant clone (λ gt11-TM2) possessing 500 bp of nucleotide sequence 5' to λ gt11-TM1. A new 500-bp probe, from the 5' end of λ gt11-TM2 to an internal *Hinc*II site, was used to screen an additional 8×10^6 recombinant clones from the same random primed library described above. Six positive phage plaques were detected that possessed inserts exhibiting two classes of patterns on agarose gel electrophoresis. One of the classes consisted of three identical clones (λ gt11-TM3), which extended the sequence of the endothelial cell receptor by 332 bp to a total of 3470 bp. The nucleotide sequence of TM is depicted in Fig. 1. The extent to which the above cloned cDNA spans the length of the TM transcript was examined by carrying out RNA blotting analyses of HUVE poly(A)⁺ RNA. The results show that the human TM transcript is ≈ 3.8 kilobases (kb), which suggests that the cloned TM cDNA is nearly full length (data not shown).

Isolation and Structure of the Human TM Gene. The genomic library employed to isolate the human TM gene was prepared by E. Frisch (Genetics Institute) from a partial *Sau*3AI digest of peripheral blood lymphocytes cloned into the EMBL3B vector. After screening 2×10^6 plaques with both BovA and BovB probes, five positive clones were identified, which were then subjected to additional rounds of plaque purification. The cloned phage DNAs containing the putative TM gene were isolated, cut with *Sal* I, *Hind*III, or *Bam*HI, and then examined by gel electrophoresis on 0.35% agarose. All genomic clones appeared to be identical, possessing an insert of ≈ 19 kb, and exhibiting similar digestion patterns with *Hind*III (0.9 kb, 1.2 kb, 4.4 kb, 6.8 kb, 10 kb, and ≈ 24 kb) and *Bam*HI (0.5 kb, 2.0 kb, 3.5 kb, 6.0 kb, 10 kb, and ≈ 26 kb). One of the above genomic clones was extensively mapped by partially digesting the phage DNA with nine separate restriction enzymes, annealing a ³²P-labeled synthetic left cos oligomer to the right cos end of the digested DNA, electrophoresing the samples on 0.35% agarose at 3 V/cm, and then examining the resultant products by autoradiography. Southern blot analyses of the digested

products were also carried out with BovA and BovB probes. Based on the above data, we concluded that the 3' end of the human TM cDNA is ≈ 8 kb from the right side of the genomic insert and that the 5' end of the human TM cDNA lay some distance upstream of this site. Therefore, the phage DNA was cut at a unique *Hpa* I site, 4.5 kb upstream of the right side of the insert, as well as the *Sal* I sites, and the resultant 15-kb insert was force cloned into the *Sma* I and *Sal* I sites of a vector purchased from Stratagene Cloning Systems (San Diego, CA) (Bluescript KsM13+).

The insert was further characterized by digesting the above construct with either *Sal* I or *Spe* I, radiolabeling either end with [³²P]dGTP or [³²P]dATP using Klenow DNA polymerase, and then releasing the inserts by cutting with *Spe* I or *Sal* I, respectively. The released labeled inserts were isolated by preparative agarose gel electrophoresis and individually subjected to partial restriction digestion with 20 separate hexanucleotide-specific restriction endonucleases. Those enzymes that cut the inserts at multiple sites were then used to digest the plasmid construct to completion, and, where possible, the resultant products were radiolabeled with appropriate ³²P nucleotide triphosphates using Klenow DNA polymerase. The sizes of the various products were determined by gel electrophoresis on 0.8% agarose using a *Bst*EII digest of λ phage DNA as a standard. The restriction map of the genomic clone from ≈ 3.5 kb to ≈ 6.0 kb upstream of the *Sma* I site showed a remarkable correspondence to the cDNA restriction map within the resolution limits of the above procedure.

The sequence of the genomic clone was defined by isolating five partially overlapping restriction fragments from ≈ 3.3 kb to ≈ 9.3 kb upstream of the *Sma* I site. These fragments were inserted either into the *Sma* I site of M13mp18 after blunting cohesive termini with Klenow DNA polymerase or into the *Hind*III site of the same vector. Both orientations of the replicative form DNA of M13 recombinants containing the restriction fragment spanning the region from 5.8 kb to 8.0 kb were purified, cut with *Sph* I and *Sal* I, sequentially deleted with exonuclease III and S1 nuclease, blunt ended and religated with Klenow DNA polymerase and DNA ligase, and then used to transform competent cells (10). The viral DNAs were isolated from a large number of the deletion subclones, as well as the undeleted M13 recombinants possessing the other restriction fragments, and the nucleotide sequences were determined.

Fig. 1 provides a partial structure of the above genomic clone, which includes the putative promoter region upstream of the translation initiation codon, the entire region that encodes the translated portion of the TM cDNA, partial segments of the region that correspond to the 3' untranslated portion of the TM cDNA, and the region of and immediately after the polyadenylation signal. It is readily apparent that the sequenced portion of the endothelial cell receptor gene contains no introns. A portion of the 3' untranslated region within the TM gene has not yet been sequenced and might possess one or more small introns. This possibility was examined by isolating *Kpn* I to *Bgl* II cDNA and genomic fragments that span the above region, cutting both of these species with either *Dde* I or *Hpa* II, labeling the ends with ³²P nucleotide triphosphates using Klenow DNA polymerase, and examining the products by polyacrylamide gel electrophoresis. The cDNA and genomic fragments exhibited products of identical size, which further reduces the likelihood that one or more small introns are present within this segment of the TM gene (data not shown).

We next sought to determine whether the TM gene, whose structure is described above, constitutes the only locus that encodes the endothelial cell receptor, or whether a second TM gene that possesses multiple exons/introns might also be present within the human genome. To settle this issue, we

FIG. 1. Nucleic acid sequence of the cDNA and the gene of human TM. The cDNA, without polyadenylation, spans the sequence from position 632 to 4050. The gene has been sequenced from position 1 to 2263, 3244 to 3620, and 3901 to 4212. The motifs in the promoter region described in the text are shown with solid underlining, while the direct repeats are underlined with dashes. The probable signal peptidase cut site is marked with an arrow, the hydrophobic areas of the amino terminal one-third are shaded, and the asparagines of the four possible sites of N-linked glycosylation are marked with asterisks. The EGF precursor-like cysteine repeats are numbered, and the probable membrane-spanning domain is marked by wavy underlining. A polyadenylation signal is bracketed and a short stretch of poly(A) sequenced in the cDNA is offset beneath the corresponding gene sequence. Amino acids are designated by the single-letter code.

extensively digested human genomic DNA with several restriction enzymes and carried out Southern blot analyses with a *Kpn* I/*Pst* I TM cDNA probe that spans the last EGF type B repeat, the Ser+Thr-rich region, the transmembrane spanning region, as well as a portion of the C-terminal cytoplasmic tail. Genes of known structure that contain several functional regions appear to always position these domains on separate exons (11, 12). Thus, it seems reasonable to use the above probe in seeking a TM gene that exhibits

multiple exons/introns. Fig. 2 shows that the complete digestion of human genomic DNA with *Pvu* II, *Bam*HI, *Pst* I, or *Kpn* I generates single restriction fragments of 1550, 3460, 912, or 5130 bp, respectively, as detected by hybridization with the *Kpn* I/*Pst* I probe. The experimentally observed sizes of the various fragments are in excellent accord with the known structures of the TM cDNA and gene (average error, $\approx 3\%$). Furthermore, Southern blot analyses of human genomic DNA cut with *Eco*RI, which has no site in

FIG. 2. Autoradiogram of a Southern blot of human genomic DNA restriction digests probed with cDNA fragment as described in the text. Lanes: A, *Pvu* II; B, *Bam* HI; C, *Pst* I; D, *Kpn* I. Positions of *Hind* III-digested phage λ DNA are marked on the left.

either the cDNA or in >21 kb of the cloned gene, reveals only a single fragment of >23 kb as detected by hybridization with the *Kpn* I/*Pst* I cDNA probe (data not shown).

DISCUSSION

The amino acid sequence of human TM, as translated from the consensus of cDNA and cloned gene, is shown in Fig. 1. The endothelial cell receptor consists of 575 amino acid residues, including the initiation methionine as well as a subsequent region that appears to represent a signal peptide (13). The putative signal peptide exhibits a hydrophobic core as well as a termination region enriched in proline residues, but it does not possess any of the basic amino acid moieties, which often mark the beginning and end of such sequences. It is likely that the signal peptidase scissions the TM polypeptide chain between amino acid residues 16 and 17, as judged by the published consensus sequence (14), to generate a mature endothelial cell receptor of 559 amino acid residues.

The C-terminal two-thirds of human TM is quite homologous to the similarly designated segment of bovine TM, whose sequence has been reported by workers in our laboratory (2). The alignment of the two endothelial cell receptor species within this region reveals a 65% identity of amino acid residues, which increases to 75% when conservative changes are accepted. The structural similarity previously noted between bovine TM and human LDL receptor is also apparent when comparing the carboxyl-terminal cytoplasmic tails, membrane spanning domains, the Ser+Thr-rich regions and EGF type B repeats of human TM and human LDL receptor. Detailed examination of the human and bovine endothelial cell receptors shows that the two cytoplasmic tails have only charge-conservative changes from the invariant single cysteine residue at the cytoplasmic boundary to the carboxyl terminus and that the transmembrane regions are >90% conserved. This may indicate an important biologic function for this region in endocytosis (15). The greatest difference between the two TM species is observed to occur in the Ser+Thr-rich regions where the human endothelial cell receptor is 20% shorter than the bovine membrane component. However, we find that a central core of 14 amino acid residues is highly conserved between the two TM species, which may be of functional significance, especially if this region is a site of O-linked glycosylation as is the case for the

LDL receptor (11). The six EGF type B repeat domains present in bovine TM are also found within the human endothelial cell receptor with the same average number of amino acid residues between the 6 cysteine residues within a given repeat. Detailed comparisons of these repeat domains reveal that the spacings between only 2 of the 36 cysteine residues of the human endothelial cell receptor are larger by one or two amino acid residues than the similarly designated region on the bovine membrane component. Each TM species also possesses 2 potential sites for N-linked glycosylation within the EGF type B repeat domains. In the human endothelial cell receptor, these sites are placed at residues 382 and 409. The more proximal site is identically located on both endothelial cell receptor species, whereas the more distal site is found within the second EGF repeat for the human membrane component, and the fourth EGF repeat for the bovine membrane component.

The amino terminal one-third of human TM exhibits no homology to the LDL receptor and has two sites of potential N-linked glycosylation. Preliminary comparison of this region to the current genetic database suggests a weak homology to mullerian inhibiting substance (16). Hydropathy plots of this domain, constructed by the method of Kyte and Doolittle (17), reveal a series of four hydrophobic segments. This observation is intriguing because a region of TM has recently been shown to interact with phospholipid and exert an enhancing effect on the activation of γ -carboxy glutamic acid containing protein C (1). Alternatively, these hydrophobic segments may be critically involved in the correct three-dimensional folding of the amino-terminal domain.

The nucleotide sequence of the human TM gene provided in Fig. 1 is noteworthy with regard to the apparent absence of introns within the coding region and much of the 3' untranslated region. One or more small introns may exist in the yet unsequenced portion of the 3' untranslated region. However, the presence of introns is unlikely because of the excellent correspondence of cDNA and genomic fragment restriction maps that span this segment. Many of the internalized receptor molecules such as the LDL receptor, the transferrin receptor, and the asialoglycoprotein receptor are composed of several discrete functional modules, each of which is encoded by one or more exons of the respective genes. The LDL receptor gene possesses multiple exons, which contain the C9 repeat domains, the EGF type B repeat domains, the Ser+Thr-rich domain, and the membrane-spanning domain (12). In view of the extensive structural homology between the LDL receptor and TM, it is quite surprising that the endothelial cell receptor gene does not possess separate exons for the EGF type B repeat segments, the Ser+Thr-rich segment, and the membrane-spanning segment. We also note that EGF type B repeats contained in viral proteins, blood clotting proteins, and growth factors are isolated in separate exons of the respective genes, although four repeats are found in two exons of the human uromodulin gene (18). Thus, the TM gene appears to represent an example of a gene in which all EGF type B repeats are found within a single exon.

The apparent absence of introns within the TM gene forced us to consider whether this gene might actually represent a functional or nonfunctional processed pseudogene similar to those described for chicken calmodulin, human phosphoglycerate kinase, human metallothionein II, mouse α -hemoglobin, and human β -tubulin (19–21). Our data (see *Results*) argue strongly against this conclusion. First, Southern blot analyses of multiple restriction enzyme digests of human genomic DNA, utilizing a TM cDNA hybridization probe, revealed the presence of a single gene whose overall structure corresponded to that presented in Fig. 1. Second, the isolation and examination of four additional recombinant genomic clones demonstrated that all inserts possessed the

same overall structure observed with the initial phage clone. Third, the TM gene does not possess a poly(A) tail immediately after the polyadenylation signal. However, we note the presence of two short direct repeats that flank the 5' untranslated region and the segment immediately following the polyadenylation signal. Fourth, the 5' region of the TM gene, upstream of the translation initiation codon, possesses several motifs commonly observed in active eukaryotic promoters. These include a probable cap site 164 bp upstream of the translation initiation codon as well as a "TATA" box sequence 30 bp upstream of the probable cap site (22), a GCAAT pentanucleotide "CAT" homology sequence 114 bp upstream of the probable cap site to which the trans-acting factor CAT binding protein BP might bind (23, 24), and three GGGCGG hexanucleotide sequences as well as one complementary sequence 132, 144, 168, 210, and 273 bp upstream of the probable cap site, which are potential regions for SP1 interaction (25).

The occurrence of a functional intron-less TM gene would not be unique. The hamster β -adrenergic receptor gene (26), mammalian α - and β -interferon genes (27), sea urchin histone genes (28), and certain mammalian heat shock genes (29) have been described, which completely lack introns. It has been suggested that this structural feature could permit the encoded protein to be synthesized under conditions that might detrimentally affect the RNA splicing mechanism or that a lack of introns might be critically involved in limiting the biosynthesis of the encoded protein (26). However, the true significance of the absence of introns within a gene remains enigmatic. What is surprising is that TM is structurally similar to coated pit receptors, blood coagulation proteins, viral proteins, and growth factors, which are encoded by genes that possess introns. The presence of putative direct repeats surrounding the sites of initiation and termination of the endothelial cell receptor suggests that this gene could have arisen from an RNA transcript. The progenitor TM gene could be related to one of the many genes that bear EGF type B repeat domains or may have been lost or extensively modified during evolutionary history.

The present availability of the human TM cDNA and gene should allow us to investigate several important issues concerning the structure-function relationships and the tissue-specific regulation of this endothelial cell receptor. Given that the protein C-TM system represents a major natural anticoagulant mechanism of the vessel wall, it is likely that our growing knowledge of the molecular biology of this endothelial cell receptor will permit a better understanding of those abnormalities that lead to thrombotic disease in humans.

Note Added in Proof. We have completed the sequencing of the human TM gene presented in Fig. 1 and continue to find an exact correspondence to the human TM cDNA. The data confirm our supposition that the human TM gene has no introns.

We thank Drs. Peter Davies and Henry Warren for the HUVE poly(A)⁺ RNA, Drs. Stuart Orkin and David Ginsburg for the HUVE cDNA libraries, and Dr. Edward Frisch for the peripheral blood lymphocyte genomic library. We should also like to express our appreciation to Dr. Monty Krieger, Dr. David Kingsley, and Dr. David Housman for many helpful discussions. This work was

supported by Grants P01HL33014 and R01HL38798 from the National Institutes of Health and by funds donated by Merck Sharp & Dohme.

1. Esmon, C. T. (1987) *Science* **235**, 1348-1352.
2. Jackman, R. W., Beeler, D. L., VanDeWater, L., & Rosenberg, R. D. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 8834-8838.
3. Benton, W. D. & Davis, R. W. (1977) *Science* **196**, 180-182.
4. Weeks, D. L., Rebagliati, M. R., Harvey, R. P. & Melton, D. A. (1983) *Cold Spring Harbor Symp. Quant. Biol.* **50**, 21-30.
5. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1972) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
6. Glaser, T., Lewis, W. H., Bruns, G. A. P., Watkins, P. C., Rogler, C. E., Shows, T. B., Powers, V. E., Willard, H. F., Goguen, J. M., Simola, K. O. J. & Housman, D. E. (1986) *Nature (London)* **321**, 882-887.
7. Biggin, M. D., Gibson, T. J. & Hong, G. F. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 3963-3965.
8. Mizusawa, S., Nishimura, S. & Seela, F. (1986) *Nucleic Acids Res.* **14**, 1319-1324.
9. Ginsburg, D., Handin, R. I., Bonthron, D. T., Donlon, T. A., Bruns, G. A. P., Latt, S. A. & Orkin, S. H. (1985) *Science* **228**, 1401-1406.
10. Henikoff, S. (1984) *Gene* **28**, 351-359.
11. Sudhof, T. C., Goldstein, J. L., Brown, M. S. & Russell, D. W. (1985) *Science* **228**, 815-822.
12. McClelland, A., Lukas, C. K. & Ruddle, F. H. (1984) *Cell* **39**, 267-274.
13. Wickner, W. T. & Lodish, H. F. (1985) *Science* **230**, 400-407.
14. Von Heijne, G. (1983) *Eur. J. Biochem.* **133**, 17-21.
15. Maruyama, I. & Majerus, P. W. (1985) *J. Biol. Chem.* **260**, 15432-15438.
16. Cate, R. L., Mattaliano, R. J., Hession, C., Tizard, R., Farber, N. M., Cheung, A., Ninfa, E. G., Frey, A. Z., Gash, D. J., Chow, E. P., Fisher, R. A., Bertonis, J. M., Torres, G., Wallner, B. P., Ramachandran, K. L., Ragin, R. C., Mangano, T. F., MacLaughlin, D. T. & Donahoe, P. K. (1986) *Cell* **45**, 685-698.
17. Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105-132.
18. Pennica, D., Kohs, W. J., Kuang, W.-J., Gleister, D., Aggarwal, B. B., Chen, E. Y. & Goeddel, D. V. (1987) *Science* **236**, 83-88.
19. Fruskin, K. D., Smith, T. F. & Goodman, M. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 1605-1608.
20. McCarrey, J. R. & Thomas, K. (1987) *Nature (London)* **326**, 501-505.
21. Vanin, E. F. (1985) *Annu. Rev. Genet.* **19**, 253-272.
22. Bucher, P. & Trifonov, E. N. (1986) *Nucleic Acids Res.* **14**, 10009-10026.
23. Klarsfeld, A., Daubas, P., Bourachot, B. & Changeux, J. P. (1987) *Mol. Cell. Biol.* **7**, 951-955.
24. Grosschedl, R. & Baltimore, D. (1985) *Cell* **41**, 885-897.
25. McKnight, S. & Tjian, R. (1986) *Cell* **46**, 795-805.
26. Dixon, R. A. F., Kobilka, B. K., Strader, D. J., Benovic, J. L., Dohlman, H. G., Frielle, T., Bolanowski, M. A., Bennett, C. D., Rands, E., Diehl, R. E., Mumford, R. A., Slater, E. E., Sigal, I. S., Caron, M. G., Lefkowitz, R. J. & Strader, C. D. (1986) *Nature (London)* **321**, 75-79.
27. Nagala, S., Mantei, M. & Weissman, C. (1980) *Nature (London)* **287**, 401-408.
28. Schaffner, W., Kunz, G., Daetwyler, H., Telford, J., Smith, H. O. & Birnstiel, M. L. (1978) *Cell* **14**, 655-671.
29. Mues, G. I., Munn, T. Z. & Raese, J. D. (1986) *J. Biol. Chem.* **261**, 874-877.