Structure of the Mouse Dipeptidyl Peptidase IV (CD26) Gene[†,‡]

Anne-Marie Bernard,[§] Marie-Geneviève Mattei,[||] Michel Pierres,[§] and Didier Marguet*[,§]

Centre d'Immunologie INSERM−CNRS de Marseille-Luminy, Case 906, 13288 Marseille Cedex 9, France, and INSERM, Unité 406, Faculté de Médecine, Boulevard Jean Moulin, 13385 Marseille Cedex 5, France

Received May 12, 1994; Revised Manuscript Received October 5, 1994[⊗]

ABSTRACT: Dipeptidyl peptidase IV (DPP IV, EC 3.4.14.5) is an ectopeptidase whose expression is modulated during thymocyte differentiation and T cell activation. We describe here the organization of the mouse DPP IV gene. This gene, which encompasses more than 90 kb, is composed of 26 exons separated by introns, the lengths of which vary from 100 bp to more than 20 kb. Reverse PCR performed on RNA from different tissues indicated that DPP IV transcripts do not contain alternatively spliced CDS sequences and, therefore, are supposed to yield a single polypeptide. However, two types of specific mRNA have been detected that differ in their 3′UTR sequences. They derive from alternative polyadenylation of the DPP IV primary transcript, since the different 3′UTR sequences are contiguous in the mouse DPP IV gene. Sequence analysis of the gene 5′-flanking region revealed several structural features found in the TATAA-box-less promoters, including a G+C-rich segment, a high frequency of dinucleotide CpG, and an imperfect symmetrical dyad. The DPP IV gene was assigned by *in situ* hybridization to the mouse [2C2−2D] region, which is syntenic with human chromosome 2. These data indicate that the human *Dpp4* locus is located within this synteny region (*i.e.*, 2q14−q37). The genomic organization of the mouse DPP IV gene is compared to that of classical serine proteases and serine hydrolases. As structural and mechanistic conservation in the absence of sequence similarity is the most remarkable feature among α/β hydrolases [Ollis, D. L., et al. (1992) *Protein Eng. 5*, 197-211], we report the possible evolutionary link between the DPP IV related family and α/β hydrolases.

Cell surface enzymes capable of inactivating peptides have recently become a matter of major interest since they appear to regulate the activity of a variety of bioactive transmitters [for a review, see Erdös and Skidgel (1989)]. Accumulating evidence indicates that these enzymes participate in the control of peptide degradation *in vivo*. Special interest has focused on the peptidases localized on the plasma membrane of cells of the immune system, *e.g.*, neutral endopeptidase (CD10, EC 3.4.24.11) (Shipp et al., 1988), aminopeptidase N (CD13, EC 3.4.11.2) (Look et al., 1989), aminopeptidase A (BP-1/6C3, EC 3.4.11.7) (Wu et al., 1991), or dipeptidyl peptidase IV (DPP IV,[1] CD26, EC 3.4.14.5) [for a review, see Marguet et al. (1993)]. These ectoenzymes play a major role in both cellular homeostasis and metabolism. One of these molecules, DPP IV, has been actively studied especially in the context of its additional functions in T cell activation.

DPP IV is a serine protease cleaving X-proline or X-alanine dipeptides from the NH_2-terminus of several biological peptides such as the growth hormone-releasing hormone (Frohman et al., 1989) or substance P (Nausch & Heymann, 1985; Ahmad et al., 1992). This enzyme is expressed and regulated at all developmental stages in a variety of epithelial tissues including kidney, liver, and intestine (Hartel et al., 1988; Hildebrandt et al., 1991). Furthermore, DPP IV has been identified as CD26, a cell surface differentiation marker involved in signal transduction in the T cell lineage (Fox et al., 1984; Fleischer, 1987; Naquet et al., 1988; Vivier et al., 1991), the expression of which is also regulated during T cell ontogeny. Recent experimental evidence has shed some light on the receptor functions of this molecule. A direct association between adenosine deaminase (ADA) and DPP IV has been demonstrated in two biological models: First, DPP IV was found to be identical to the ADA-binding protein (ADAbp) known as a marker for malignant transformation of melanocytes and melanoma progression (Morrison et al., 1993); in addition, it has recently been shown that ADA can associate itself with DPP IV at the T cell surface (Kameoka et al., 1993). Finally, a recent study has suggested that DPP IV (CD26) participates, along with CD4, the human immunodeficiency virus (HIV) primary receptor molecule, in HIV entry into permissive cells (Callebaut et al., 1993).

We previously reported the purification and characterization of the mouse thymocyte-activating molecule (THAM), a cell surface peptidase with DPP IV activity, which was found to be the counterpart of the human T cell-activating

[†] This work was supported by institutional funding from the Institut National de la Santé et de la Recherche Médicale (INSERM) and the Centre National de la Recherche Scientifique (CNRS) and grants from the Association pour la Recherche contre le Cancer (ARC) and the Ligue Nationale Française Contre le Cancer (LNFCC).

[‡] The nucleic acid sequences reported in this paper have been deposited in the GenBank/EMBL Data Bank under Accession Numbers U12599, U12600, U12601, U12602, U12603, U12604, U12605, U12606, U12607, U12608, U12609, U12610, U12611, U12612, U12613, U12614, U12615, U12616, U12617, U12618, U12619, and U12620.

* To whom all correspondence should be addressed: Centre d'Immunologie INSERM−CNRS de Marseille-Luminy, Case 906, 13288 Marseille Cedex 9, France. Tel: (33) 91-26-94-00. Fax: (33) 91-26-94-30. e-mail: marguet@ciml-cd.univ-mrs.fr.

[§] Centre d'Immunologie INSERM−CNRS de Marseille-Luminy.

[||] INSERM.

[⊗] Abstract published in *Advance ACS Abstracts,* November 15, 1994.

[1] Abbreviations: DPP IV, dipeptidyl peptidase IV; ADA, adenosine deaminase; HIV, human immunodeficiency virus; RT−PCR, reverse transcriptase−polymerase chain reaction; 5′UTR, 5′ untranslated region; 3′UTR, 3′ untranslated region; CDS, coding sequence; bp, base pair(s); kb, kilobase(s).

0006-2960/94/0433-15204$04.50/0

molecule CD26 (Vivier et al., 1991; Marguet et al., 1992). Mouse DPP IV shares with several nonclassical serine hydrolases (including acylaminoacyl peptidase and prolyl oligopeptidase) a conserved stretch of ~200 amino acids in which the catalytic triad residues (*i.e.*, Ser^{624}, Asp^{702}, and His^{734}) were found to be organized in a novel sequential order (Ogata et al., 1992; Marguet et al., 1992; David et al., 1993). Interestingly, structural similarities emerged from the comparison of the topological arrangement of the catalytic site residues of nonclassical serine proteases and unrelated serine hydrolases, including members of the lipase and esterase families (Polgár, 1992; David et al., 1993).

In the present study, which represents the first structural analysis of a gene encoding a member of this nonclassical serine protease family, we have analyzed the structural features of the mouse DPP IV gene and its chromosomal location. We report also data including the origin of DPP IV transcripts and the promoter and regulatory *cis* elements controlling its expression.

EXPERIMENTAL PROCEDURES

Materials and Reagents. Modifying enzymes and other molecular biology reagents were purchased from Promega Biotec (Madison, WI), Boehringer Mannheim (Mannheim, Germany), Bethesda Research Laboratories (Bethesda, MD), or Stratagene (La Jolla, CA) and used according to the manufacturers' instructions. All synthetic oligonucleotides were synthesized in our local facilities using an Applied Biosystems synthesizer. Mouse DPP IV cDNA in pTHAM-3.3 (Marguet et al., 1992), pGEM-5Zf(+), and pGEM-7Zf-(+) from Promega (Madison, WI) and pBluescript-KS from Stratagene (La Jolla, CA) were used in XL1-Blue cells for probe screening material and recombinant construction. ^{32}P-Labeled nucleotides and [α-^{35}S]dATP were purchased from Amersham Corp. (Buckinghamshire, U.K.). Standard recombinant DNA techniques (Sambrook et al., 1989) were utilized, unless otherwise indicated.

Isolation and Characterization of Genomic Sequences. A library of *Sau*3A partially digested mouse B10.A liver DNA cloned into the λ bacteriophage vector EMBL-3 (Frischauf et al., 1983) was obtained from M. Malissen (Hue et al., 1990). Mouse DPP IV gene was isolated by plaque hybridization performed with ^{32}P-labeled mouse DPP IV cDNA clone pTHAM-3.3 (Marguet et al., 1992). Hybridization was performed overnight at 65 °C in 5× SSC, 5× Denhart's, 10% dextran sulfate, 1% sodium dodecyl sulfate, and 150 μg/mL salmon sperm DNA. The filters were then washed twice at room temperature with 0.2× SSC and 0.1% SDS and four times at 65 °C with the same buffer. Insert DNA fragments were initially characterized by restriction digestion and Southern blot analysis. Mouse genomic DNA fragments that hybridized to DPP IV cDNA probe were subcloned into classical plasmid vectors.

To isolate the mouse DPP IV locus as an uninterrupted fragment, we screened yeast artificial chromosome (YAC) library no. 902 from the Reference Library Data Base of the Imperial Cancer Research Fund (London) (Lehrach, 1990). This YAC library was constructed from mouse spleen DNA digests, ligated into pYAC4, and expressed in host *Saccharomyces cerevisiae* AB1380 (Larin et al., 1991). Recombinant clones were spotted on high-density filters. Approximately 160 000 clones were screened with the mouse DPP IV cDNA. The positive ones were analyzed by pulsed field gel electrophoresis using a contour clamped homogenous electric field (CHEF) (2015 Pulsaphor; LKB, Bromma, Sweden) following standard procedures (Chimini et al., 1990).

DNA Sequence Determinations. Complete sequence determination was performed for all genomic exon regions (except exon 23) to determine intron interruptions and to ascertain that the genomic sequences were in complete agreement with available DPP IV cDNA information. Double-stranded sequencing was performed using Sequenase (U.S. Biochemicals, Cleveland, OH) by the dideoxy chain termination method of Sanger (Sanger et al., 1977). Sequences were analyzed using the BISANCE system (Dessen et al., 1990) of programs and data bank implemented on a VAX-8530 of the CITI 2 (Centre Inter-Universitaire de Traitement de l'Information), Paris.

RNA Analysis. Total RNA from mouse tissues was isolated by the guanidium thiocyanate method (Sambrook et al., 1989), phenol extracted, and precipitated. For reverse transcriptase−polymerase chain reaction (RT−PCR), 2.5 μg of total RNA from different tissues was reverse transcribed to cDNA using the Moloney murine leukemia virus RNase H^- reverse transcriptase (BRL, Eragny, France) and random hexamer primers (Pharmacia, Uppsala, Sweden) in a 20-μL reaction volume for 1 h at 42 °C. The reverse transcriptase was inactivated at 80 °C for 5 min. Twenty picomoles of specific oligonucleotide primers was added to the reaction mixture, and the final volume was adjusted to 100 μL before the addition of Taq DNA polymerase for PCR as described below.

Polymerase Chain Reaction (PCR) Analysis. PCR was performed using Taq DNA polymerase. To analyze for the presence of a specific fragment in the DNA matrix, an initial step of denaturation (4 min at 94 °C) was followed by 30 cycles of 30 s at 94 °C, 2 min at the annealing temperature, and 2 min at 72 °C. For each oligonucleotide pair, the annealing temperature was established in order to avoid background on a positive control. The PCR mixtures were separated on appropriate agarose gels [*e.g.*, 4% Metaphor agarose (FMC BioProducts, Rockland, ME) or NuSieve GTG agarose (FMC BioProducts, Rockland, ME), 2%, plus Ultrapure agarose, 2% (BRL, Eragny, France)].

Genomic Southern Analysis. Genomic Southern analysis was conducted as described (Sambrook et al., 1989). Ten micrograms of high molecular weight mouse DNA was digested with restriction endonucleases. All samples were subjected to electrophoresis on a 0.7% agarose gel and transferred to GeneScreen (DuPont-New England Nuclear, Boston, MA) using the alkaline transfer protocol. DNA was cross-linked to the membrane, prehybridized, and probed using a random-primed labeled DNA fragment.

DPP IV Gene Chromosomal Assignment. In situ hybridization experiments were carried out using metaphase spreads from a WMP male mouse, in which all the autosomes except 19 were in the form of metacentric Robertsonian translocations. Concanavalin A-stimulated lymphocytes were cultured at 37 °C for 72 h with 5-bromodeoxyuridine added during the final 6 h of culture (60 μg/mL of medium) to ensure good quality chromosomal R-banding.

The partial cDNA clone containing the *Pvu*II−*Pvu*II insert of 980 bp in pGEM5Z was tritium-labeled by nick translation to yield a specific activity of 2.1×10^8 dpm/μg. The

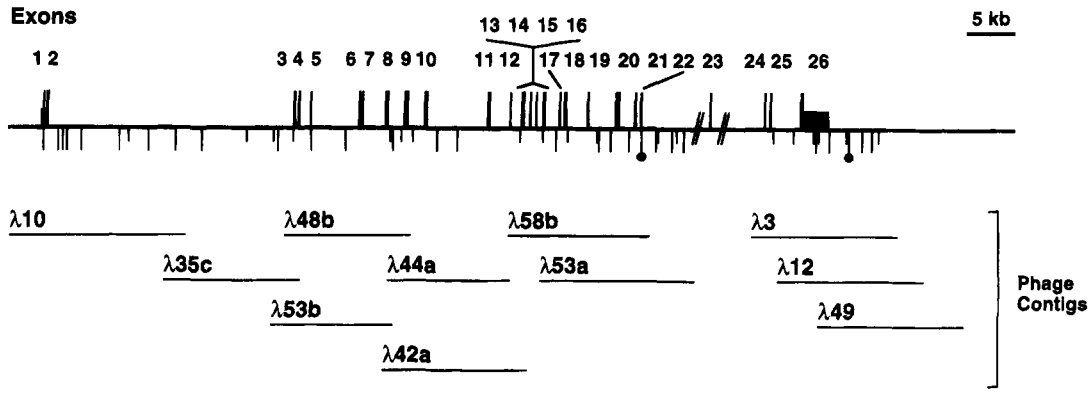

FIGURE 1: General organization of the mouse DPP IV gene. Recombinant phage inserts used to characterize the gene (λ10, λ35c, λ53b, λ48b, λ44a, λ42, λ58b, λ53a, λ3, λ12, and λ49) are indicated below the *Hin*dIII (large bar), *Eco*RI (small bar), and *Sal*I (black circle) genomic restriction map. Exons (black boxes) are represented on a solid line. Exons 1−25 are too small (45−194 bp) to be precisely drawn to scale.

radiolabeled probe was hybridized to metaphase spreads at a final concentration of 25 ng/mL in hybridization solution, as previously described (Mattei et al., 1985).

After being coated with nuclear track emulsion (Kodak NTB_2), the slides were exposed for 14 days at 4 °C and then developed. To avoid any slipping of silver grains during the banding procedure, chromosome spreads were first stained with buffered Giemsa solution and metaphases were photographed. R-banding was then performed by the fluorochrome−photolysis−Giemsa (FPG) method, and metaphases were rephotographed before analysis.

RESULTS

Structural Analysis of the Mouse DPP IV Gene. Approximately 3×10^6 recombinant phages from a mouse liver genomic library were screened with the 3.3-kb DPP IV-specific cDNA. Twenty-six positive clones were isolated, purified, and clustered by means of four probes covering distinct regions of the DPP IV cDNA. Further analyses focused on 11 phages (λ10, λ35c, λ53b, λ48b, λ44a, λ42a, λ58b, λ53a, λ3, λ12, and λ49) that contained various coding segments from the known cDNA sequence. Alignment of these genomic clones with respect to mouse DPP IV cDNA was achieved by means of a series of oligonucleotide probes. Restriction mapping and sequence analysis indicated that the λ10, λ35c, λ53b, λ48b, λ44a, λ42a, λ58b, and λ53a clones contained overlapping DPP IV genomic fragments beginning at the 5′UTR and that λ3, λ12, and λ49 contained another contig covering the end of the CDS and the 3′UTR (Figure 1). However, a small segment of 65 bp corresponding to exon 23 could not be detected in either the 5′ or the 3′ phage contigs. We screened therefore a mouse yeast artificial chromosome library. Three independent recombinant clones (ICRFy902E1151, ICRFy902B1278, and ICRFy902F0323) were isolated and characterized by hybridization after chromosome separation by pulsed-field gel electrophoresis. Inserts were estimated to have a length of ~260, ~470, and ~600 kb, respectively. Moreover, it was found that YAC ICRFy902B1278 and ICRFy902F0323 contained the complete DPP IV gene (data not shown). Mouse genomic DNA, YAC ICRFy902F0323 DNA, and contig λ phage DNA resulted in an identical set of signals indicating the correct representation of the mouse DPP IV genomic sequence on the cloned DNA (Figure 1). Altogether, the above results indicate that the mouse genome contains a single DPP IV gene.

Restriction map and sequence analysis defined the DPP IV intron−exon structures (Figure 1 and Table 1). We have identified all splice junctions except those of exon 23. To demonstrate the lack of introns in the 65 bp composing exon 23, PCR was performed using different oligonucleotide pairs and three types of matrix [DPP IV cDNA as positive control, mouse genomic DNA, and total DNA from YAC positive (ICRFy902F0323) and negative clones] (Figure 2). As expected, the primer pair a + b delimiting the putative exon 23 (Figure 2A) yielded an amplification fragment of exactly 65 bp with DPP IV cDNA as well as with mouse genomic DNA and YAC ICRFy902F0323 DNA matrix (Figure 2B). However, use of a primer pair on exons 22/23 (primers d + b) or exons 23/24 (primers a + c) showed that the fragments obtained with DPP IV cDNA matrix were absent when mouse genomic DNA or YAC ICRFy902F0323 DNA matrix was used (Figure 2B). From a YAC ICRFy902F0323 map with *Eco*RI, *Hin*dIII, and *Sal*I restriction enzymes we were able to estimate the size of the introns flanking exon 23 (Table 1).

Hence, the DPP IV gene is composed of 26 exons separated by introns that vary in length from 110 bp to more than 20 kb (Table 1). Except for exon 23 for which data are not available, all splice junctions followed the GT-AG rule of Breathnach *et al.* (1978) and were found to be similar to the consensus sequences described by Mount (1982). In conclusion, the *Dpp4* locus spans ~90 kb from its 5′UTR to its 3′UTR extremity.

Correlation of Functional Domains with Specific DPP IV Exons. Interestingly, the CDS is encoded by multiple small exons (exons 1−25) ranging in size from 45 to 196 bp, whereas the end of the coding region and the 3′UTR are encoded by a single large exon (exon 26) (Figure 1). This structure divides the protein-coding sequence into segments ranging from 15 to 64 amino acids (exons 12 and 20, respectively). Exon 1 contains all of the 5′UTR and encodes the translation start site and the two first amino acids of the cytoplasmic domain. Exon 2 contains the last four amino acids of this domain and the single hydrophobic sequence that constitutes both the signal peptide and the transmembrane segment of this type II integral membrane protein. Exons 3−26 encode the major part (732 of 760 amino acids) of the DPP IV extracellular domain. Exon 3 contains the stalk segment found in most ectopeptidases.

Table 1: Intron–Exon Splice Junction Sequences in the Mouse DPP IV gene[a]

Exon						Intron
Acceptor		Size & cDNA positions		Donor	Phase	Size
		E1	123 bp (- 117; + 6)	Met Lys ACC ATG AAG **gtgagt**	0	0.48 kb
Thr Pro Trp ttgcag ACA CCG TGG		**E2**	85 bp (+ 7; + 91)	Ser Lys Asp AGC AAA G **gtagaa**	1	~ 26.5 kb
Glu Ala caacag AT GAA GCG		**E3**	96 bp (+ 92; + 187)	Val Ser Asp GTT TCA G **gtaagt**	1	0.54 kb
Phe Glu ctttag AC TTT GAA		**E4**	92 bp (+ 188; + 279)	Ser Thr Phe AGT ACC TTT **gtaggt**	0	~ 1.10 kb
Glu Ser Phe caacag GAA AGC TTT		**E5**	69 bp (+ 280; + 348)	Tyr Val Lys TAC GTG AAG **gtaaag**	0	~ 5.4 kb
Gln Trp Arg ttgcag CAA TGG AGA		**E6**	53 bp (+ 349; + 401)	Asn Lys Arg AAT AAA AG **gtcagt**	2	0.11 kb
Gln Leu ctatag A CAG CTG		**E7**	73 bp (+ 402; + 474)	His Lys Leu CAT AAG TTG **gttagt**	0	~ 3.0 kb
Ala Tyr Val tttaag GCA TAT GTC		**E8**	121 bp (+ 475; + 595)	Tyr Glu Glu TAT GAA G **gtaaag**	1	~ 3.0 kb
Glu Val ttccag AG GAA GTC		**E9**	161 bp (+ 596; + 756)	Tyr Pro Lys TAC CCA AAG **gtctct**	0	~ 1.8 kb
Ala Gly Ala ccctag GCA GGA GCT		**E10**	113 bp (+ 757; + 869)	Ala Arg Gly GCA AGA GG **gtaaga**	2	~ 7.2 kb
Asp His ctccag G GAT CAC		**E11**	136 bp (+ 870; + 1005)	Cys Pro Ser TGT CCA TCC **gtaaga**	0	~ 2.4 kb
Glu Gln Gln ccccag GAG CAG CAG		**E12**	45 bp (+ 1006; + 1050)	VAl Gly Arg GTC GGA AGA **gtaagt**	0	~ 1.0 kb
Phe Arg Pro gtctag TTT AGG CCC		**E13**	108 bp (+ 1051; + 1158)	Asp Lys Lys GAT AAG AAA **gtgagt**	0	~ 0.6 kb
Asp Cys Thr aaccag GAC TGT ACA		**E14**	68 bp (+ 1159; + 1226)	Asp Tyr Leu GAT TAT CT **gtgagt**	2	~ 0.4 kb
Tyr Tyr tctcag A TAC TAC		**E15**	54 bp (+ 1227; + 1280)	Leu Tyr Lys CTC TAT AA **gtaaga**	2	~ 0.8 kb
Ile Gln ttacag A ATT CAA		**E16**	122 bp (+ 1281; + 1402)	Cys Trp Gly TGT TGG G **gtaagt**	1	~ 1.9 kb
Pro Gly tcacag GC CCC GGT		**E17**	48 bp (+ 1403; + 1450)	His Lys Glu CAT AAA G **gtattt**	1	0.25 kb
Leu Arg gtttag AG CTG CGA		**E18**	99 bp (+ 1451; + 1549)	Glu Thr Arg GAA ACA A **gtaagt**	1	~ 2.1 kb
Phe Trp gtttag GA TTT TGG		**E19**	70 bp (+ 1550; + 1619)	Leu Asp Val TTA GAT GT **gtaagt**	2	~ 3.5 kb
Tyr Ala tcacag A TAT GCA		**E20**	195 bp (+ 1620; + 1814)	Ala Ala Arg GCA GCC AG **gtaagc**	2	~ 1.5 kb
Gln Phe ttttag G CAA TTT		**E21**	55 bp (+ 1815; + 1869)	Trp Gly Trp TGG GGC TGG **gtaagt**	0	~ 0.7 kb
Ser Tyr Gly ccccag TCA TAT GGA		**E22**	100 bp (+ 1870; + 1969)	Tyr Tyr Asp TAC TAT G **gtacgt**	1	7 kb < < 12 kb
Ser Val ------ AC TCA GTG		**E23**	65 bp (+ 1970; + 2034)	His Tyr Arg CAT TAC AGG ------	0	0.5 kb < < 5.5 kb
Asn Ser Thr tttcag AAC TCA ACA		**E24**	73 bp (+ 2035; + 2107)	Ala Asp Asp GCA GAT G **gtgagc**	1	~ 0.7 kb
Asn Val ccacag AT AAT GTT		**E25**	74 bp(+ 2108; + 2181)	Gln Ala Met CAA GCA ATG **gtatac**	0	~ 3.2 kb
Trp Tyr Thr ctgcag TGG TAC ACG		**E26**	1008 bp (+ 2182; + 3189) 2648 bp (+ 2182; + 4829)	(First polyadenylation site) (Second polyadenylation site)		

[a]The nucleotide sequences of exon-intron junctions were determined from genomic subclones by using oligonucleotides corresponding to exon sequences as primers. Exon and intron sequence are shown in capital and lower-case letters, respectively. The complete nucleotide sequences of the mouse CD26 gene have been submitted to the GenBank™/EMBL Data Bank with accession numbers U12599, U12600, U12601, U12602, U12603, U12604, U12605, U12606, U12607, U12608, U12609, U12610, U12611, U12612, U12613, U12614, U12615, U12616, U12617, U12618, U12619, U12620.

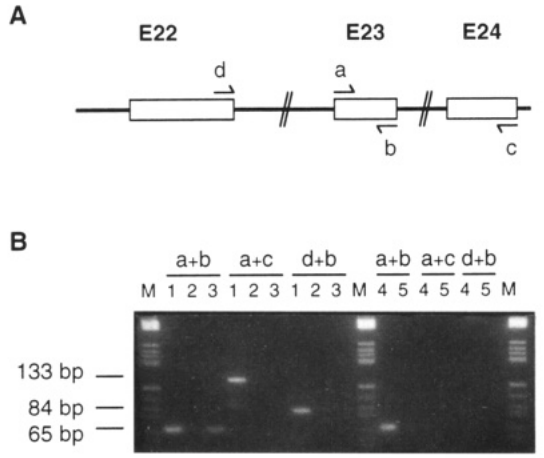

FIGURE 2: Identification of exon 23 by PCR analysis. (A) Primers correspond to the respective position on the published cDNA sequence (Marguet et al., 1992): a, 2087−2109; b, 2151−2134; c, 2223−2200; d, 2068−2085. (B) PCR product analysis on agarose gel. DNA matrices: lane 1, DPP IV cDNA; lane 2, negative control; lane 3, mouse genomic DNA; lane 4, YAC ICRFy902F0323 DNA; lane 5, YAC ICRFy902 DNA; lane M, DNA molecular weight markers (pBR322 digested by *Hae*III).

Small Exons of the Coding Region and Alternative Splicing Hypothesis. The multiple small exon structure of the DPP IV gene led us to explore the presence of alternative splicing events during DPP IV gene expression. To address this question, an exhaustive reverse transcriptase−polymerase chain reaction (RT−PCR) was performed with three different oligonucleotide pairs flanking the exons susceptible to differential splicing (*i.e.*, primers a + a′, b + b′, and c + c′, Figure 3A). Results obtained with total RNA from thymus are shown in Figure 3B. As expected, a single band was obtained with each of the three pairs of primers used. Experiments were performed with RNA from M14T thymoma cells and a variety of tissues (*e.g.*, kidney, thymus, liver, intestine, and lung) that expressed DPP IV. No differentially spliced products could be detected in any of these RNA samples, as shown in Figure 3C, which illustrates an example with primers a + a′.

Alternative Polyadenylation of Mouse DPP IV Gene. Two DPP IV mRNAs from mouse thymus RNA were detected by Northern blot analysis with the 3.3-kb DPP IV cDNA probe (data not shown), a major band and a minor one that corresponded to transcripts of 3300 and 4900 nucleotides, respectively. This observation and the fact that both rat and human DPP IV transcripts have a larger 3′UTR (Ogata et al., 1989; Darmoul et al., 1992) prompted us to evaluate whether the longer mRNA species observed in mouse thymus could result from an alternative polyadenylation site. To explore this issue, the 3′-end of DPP IV transcript was amplified by RT−PCR using a primer in the 3′UTR of the published cDNA [oligonucleotide 3154−3173 in the DPP IV cDNA sequence (Marguet et al., 1992)] and a specific oligonucleotide corresponding to the oligo(dT)-extended 3′-end of the long mRNA from rat. A single band of ~1800 bp was observed, indicating the presence of another polyadenylation site in the gene. In accordance with the genomic map, *Eco*RI digestion of the amplified product yielded a small fragment of ~640 bp (Figure 1). The amplified fragment was cloned in pGEM-T (Promega, Madison, WI) and sequenced. This additional 3′UTR was found to derive from alternative polyadenylation of the DPP IV primary transcript localized in the extension of exon 26 (Figure 4). This conclusion is based on the fact that the different 3′UTR sequences are contiguous in the mouse DPP IV gene.

Long mRNA was identified by Northern blot with a specific probe, and its estimated size of 4.9 kb corresponded to that deduced from the nucleotide sequence of the DPP IV gene. Expression of this longer transcript was also estimated by RT−PCR analysis using oligonucleotide primers d + d′ (Figure 3A). This longer transcript form was found to be present in all tissues expressing DPP IV proteins (Figure 3D). It is worth noting that the 3′UTR of rat (Ogata et al., 1989) and mouse DPP IV mRNA exhibit a high degree of homology (>90% in several 3′UTR segments) (Figure 4).

5′-Flanking and Promoter Sequences. To search for possible promoter and regulatory *cis* elements upstream of the DPP IV structural gene, the 5′-DNA sequence of the transcriptional initiation site was determined from a subclone of mouse DPP IV genomic λ10 bacteriophage by sequencing a 1.8-kb genomic region located immediately upstream of the published cDNA sequence (Marguet et al., 1992) (Figure 5A). No obvious TATAA and CCAAT boxes were detected in the vicinity of the putative transcription initiation site [at the 5′-end of the 5′UTR of DPP IV cDNA (Marguet et al., 1992)]. However, some features of the TATAA-box-less promoters were found in this region (Ackerman et al., 1993). A search for consensus binding sites of transcriptional factors indicated the presence of multiple Sp1 and AP-2 consensus binding sites in the DPP IV gene 5′-upstream region (Figure 5A). Other putative binding sites that have been identified include those for AP-1, GATA, LF-A1 and XREbf (Figure 5A) [for a review of consensus sequences, see Faisst and Meyer (1992)]. G+C nucleotide composition analysis of the DPP IV gene 5′-flanking region is shown in Figure 5C. In the proximity of the first exon, a 300-bp region with a high G+C content was found. To address the question of whether this region represents a putative CpG island, the occurrence of CpG, which is usually present at lower frequency in the genome, was compared to that of GpC (Figure 5C). In the G+C-rich clusters observed in the vicinity of exon 1 of the DPP IV gene, the CpG:GpC ratio was found to be 0.75, whereas outside these clusters this ratio falls to about one-third, in accordance with the CpG deficiency of the bulk genome. Moreover, examination of the sequence composition of the region revealed an imperfect 180° rotational symmetry with respect to the guanosine residue in position +42 (Figure 5B).

Mouse DPP IV Gene Chromosomal Location. The chromosomal location of the mouse DPP IV structural gene was determined by *in situ* hybridization to metaphase chromosome preparations. In the 100 metaphase cells examined, there were 209 silver grains associated with chromosomes, 64 of which (30.6%) were located on chromosome 2. The nonrandom grain distribution (73.4%, 47/64) assigned the DPP IV gene to the [2C2−2D] chromosomal region (Figure 6).

Note that a sequence of dinucleotide repeats was observed in the 5′-flanking region of the DPP IV gene, composed of an AC/GT stretch, extended along ~100 bp, in position −791 to −698 (Figure 5A). These repetitive sequences, referred to as microsatellites (Litt & Luty, 1989; Weber & May, 1989; Tautz, 1989), occur frequently and randomly in eukaryotic species and exhibit high polymorphism between human and

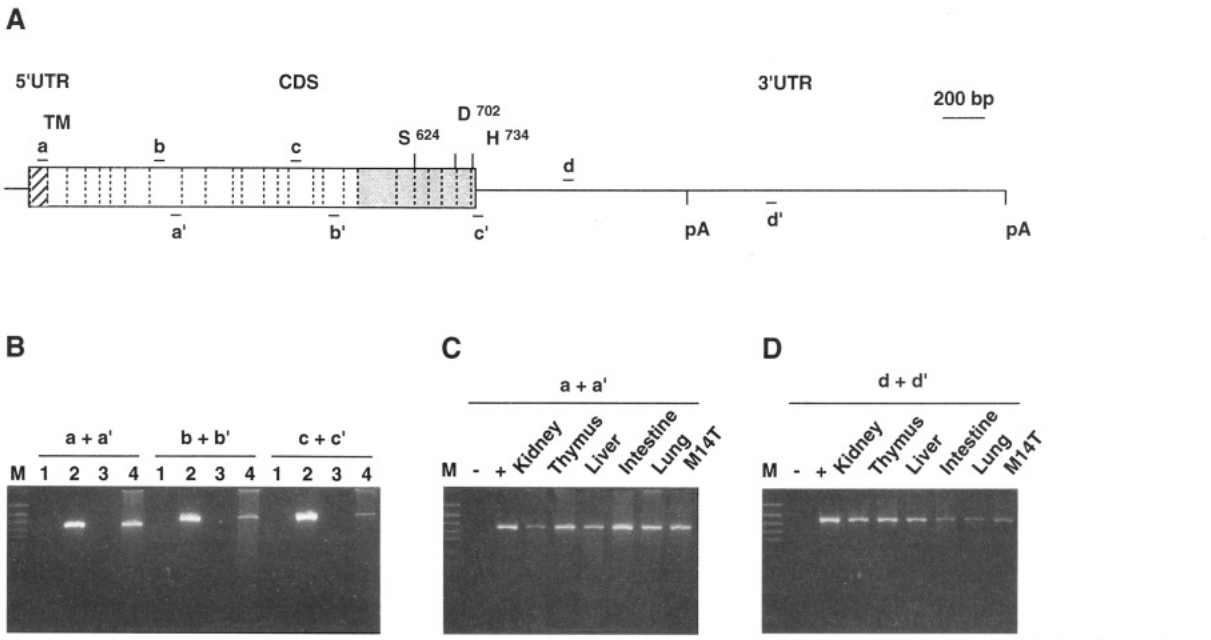

FIGURE 3: Different transcript origins in mouse DPP IV gene. (A) Schematic representation of the mouse DPP IV cDNA. The coding region (CDS) is represented as an open bar in which the transmembrane domain is shown as a hatched bar, and the enzymatic domain is shown as a gray-shaded box. Exon segmentations are represented by vertical dotted lines. Polyadenylation sites (pA) were mapped on 3'UTR. Primers correspond to the respective positions on the published cDNA sequence (Marguet et al., 1992): a, 182−201; a′, 863−844; b, 790−807; b′, 1629−1609; c, 1411−1428; c′, 2325−2308; d, 2773−2792; d′, 3627−3600. (B and C) Small exons of the mouse DPP IV CDS and alternative splicing analysis. Different primer pairs were assayed on total RNA from adult mouse thymus. For each pair, RT−PCR was performed without DNA matrix (lane 1), with cDNA matrix (lanes 2), or with total RNA (lanes 3 and 4), in the absence (lane 3) or presence of reverse transcriptase (lane 4). Results were obtained with the a+a′ oligonucleotide pair with total RNA from different tissues. (D) Alternative polyadenylation of the mouse DPP IV transcript. Total RNA from different tissues was analyzed by RT−PCR with a primer pair (d+d′) included in the 3′ longer UTR. Lane M, DNA molecular weight markers (mix of pUC digested by *Sau*3AI and pUC digested by Taq I).

mouse. This last feature has provided a useful means for constructing detailed human and mouse genetic maps. The microsatellite isolated here could be used therefore for the linkage analysis of the *Dpp4* locus.

Comparison of the Mouse DPP IV Gene with Those of Other Serine Hydrolases. DPP IV-related enzymes present the same order in the primary sequence of their catalytic triad (Ser^{624}, Asp^{702}, and His^{734}) (Marguet et al., 1992; David et al., 1993). A similar order was observed in α/β hydrolases (Ollis et al., 1992), but this order differed from that observed among classical serine proteases like trypsin (His^{57}, Asp^{102}, and Ser^{195}) and subtilisin (Asp^{32}, His^{64}, and Ser^{221}). The detailed analysis of DPP IV genomic organization allowed structural comparison with several serine hydrolase genes for tracing evolutionary events.

Upon initial comparison with archetypal serine proteases such as trypsin (Rogers, 1985; Brenner, 1988; Irwin et al., 1988; Vanderslice et al., 1990), we observed that (1) the catalytic triad residues of mouse DPP IV are encoded by separate exons (Ser^{624} on exon 22, Asp^{702} on exon 24, and His^{734} on exon 26), like in the trypsin gene; (2) the serine consensus motif is split between two exons (Gly-Trp^{623} lies at the end of exon 21, and Ser^{624}-Tyr-Gly, at the beginning of exon 22), a feature differing from typical serine proteases where the complete consensus motif is encoded by a single exon; and (3) the location and phases of the introns within the coding region of the enzymatic domain of the DPP IV gene are distinct from those of the trypsin gene (Figure 7). In the same way, no similarity in genomic organization was observed between the DPP IV gene and those of an updated compilation of serine proteases established by S. Brenner (1988) (data not shown).

Considering the previous results, we performed further comparative analysis between the members of the DPP IV-related family and some α/β hydrolases (lipases and cholinesterases) that presented the same sequential order of the catalytic triad residues as well as structural similarity surrounding these residues (Polgár, 1992; Marguet et al., 1992). A multiple exon organization of the enzymatic domain was observed in both mammalian lipases (Deeb & Peng, 1989; Cai et al., 1989; Ameis et al., 1990) and mouse DPP IV genes. However, the size of the exons coding for the catalytic triad and the position of the corresponding codons in the exons of the lipase gene differ from those in the mouse DPP IV gene (Figure 7). In the case of butyrylcholinesterase and acetylcholinesterase genes, the catalytic triad residues are encoded by a single exon (Arpagauss et al., 1990; Li et al., 1993).

This comparative analysis of the genomic organization of the enzymatic domains of serine proteases and serine hydrolases did not provide clear evidence of a divergent evolutionary process.

DISCUSSION

DPP IV is a member of a serine protease subfamily that shares a similar catalytic triad with other serine hydrolases (Polgár, 1992; Marguet et al., 1992; David et al., 1993). Delineation of the structure of the DPP IV gene gives valuable information as to the evolution of these enzymes. Our data indicate that the mouse DPP IV gene is approximately 90 kb long, comprises 26 exons, and has an intron:exon length ratio of 18. The length of the first 25 exons (ranging from 45 to 194 bp) falls within the range

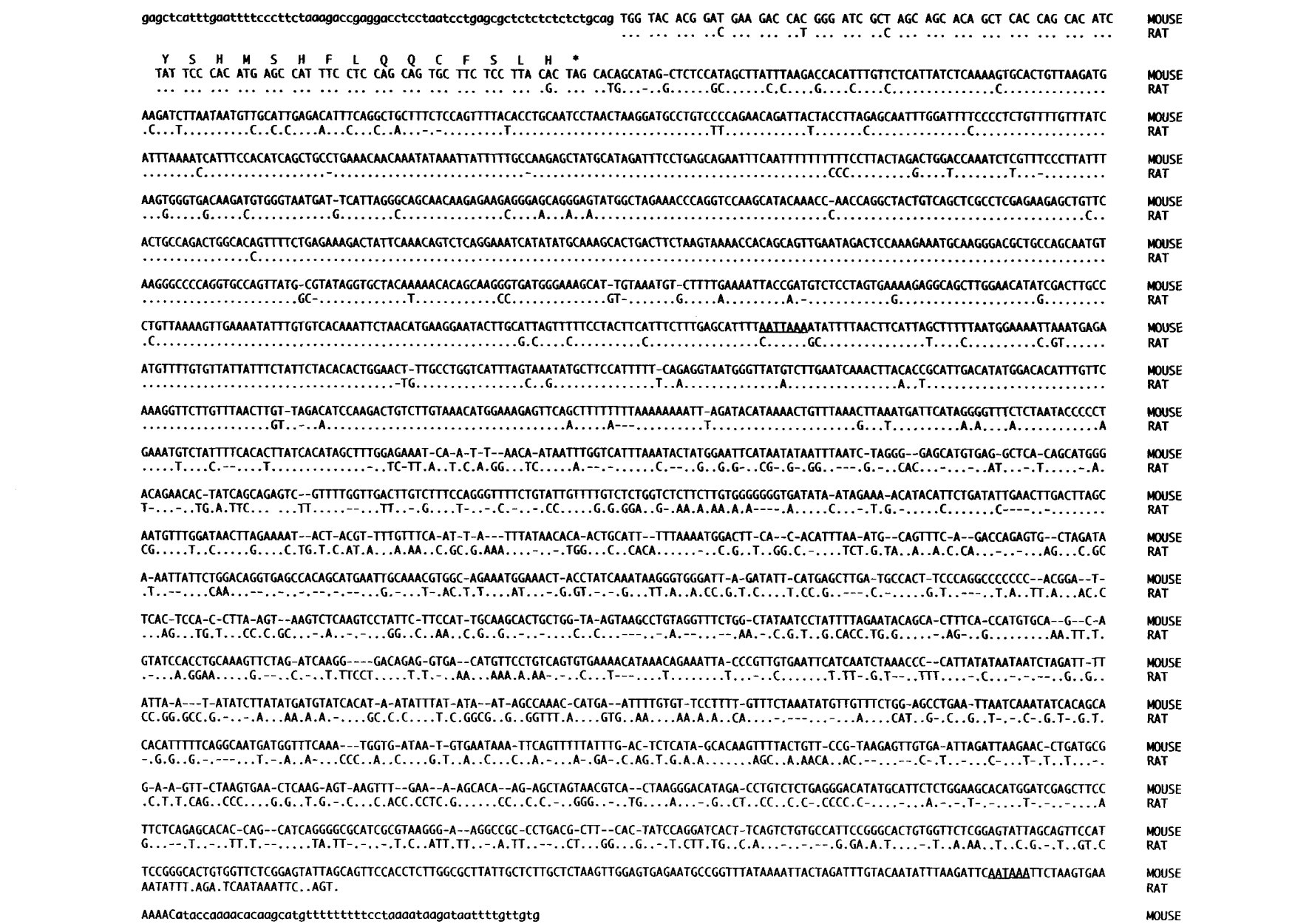

FIGURE 4: Exon 26 of DPP IV and 3′UTR homology. Exon 26 is depicted in capital letters, while introns are shown in lowercase letters. The amino acid sequence derived from CDS of exon 26 is placed above the corresponding DNA sequences. Poly(A) addition sites are marked in boldface, while TGA stop codon is marked with an asterisk. The polyadenylation signals (ATTAAA or AATAAA) are underlined. The 3′UTR sequence from rat DPP IV (Ogata et al., 1989) cDNA is aligned with the 3′UTR of mouse DPP IV. The gaps are left as spaces, and the nucleotides that were identical to those of the DPP IV gene are presented as dots.

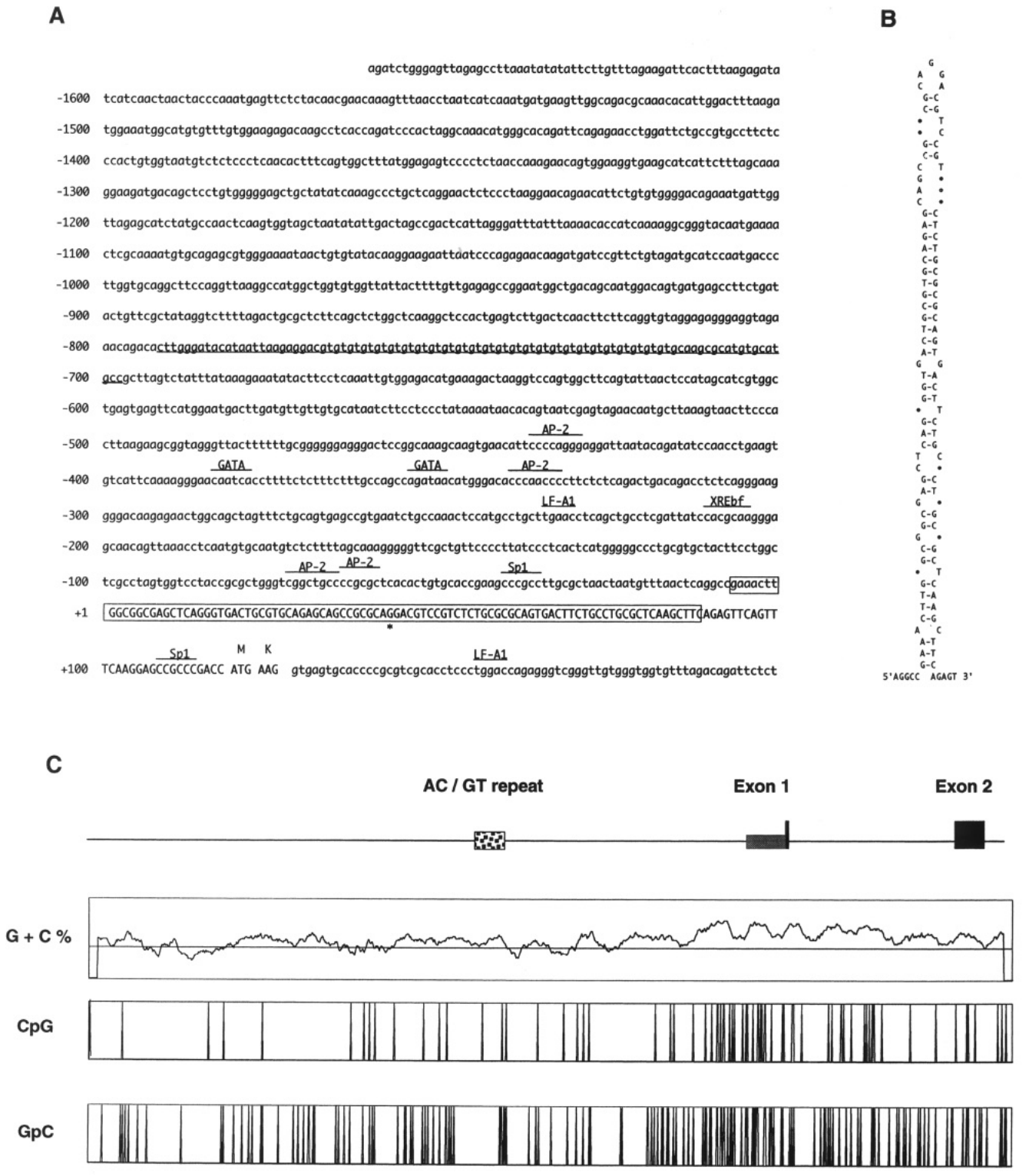

FIGURE 5: Mouse DPP IV gene promoter. (A) Nucleotide sequence of the 5′-flanking region of the DPP IV gene. Capital letters represent exon 1. The sequence organized in an imperfect dyad symmetry is boxed. Putative transcription factor binding sites with perfect homology to published consensus sequences are overlined. (B) Graphic representation of the potential stem-and-loop structure deduced from the primary structure of G+C-rich putative DPP IV promoter. (C) Schematic representation of the CpG island in the mouse DPP IV gene. G+C content across the 2.0-kb genomic region containing the mouse DPP IV gene. A window of 50 bp shifted in steps of 7 bp was used in the computer analysis. Map plot compares the densities of CpG and GpC across the same region. Microsatellite AC/GT repeat region and exons 1 and 2 are boxed.

generally observed in mammalian genes (Breathnach & Chambon, 1981; Gilbert, 1985). The 26th exon encodes the last 33 COOH-terminal amino acids and the 3′UTR (2.6 kb for the longest mRNA).

A Single Gene Encodes Mouse DPP IV. Although alternative processing events involving internal exons have been documented in several cell surface molecule genes, our results indicate that DPP IV transcripts do not contain alternatively spliced CDS exons. This conclusion is reinforced by the fact that there is no missing nucleotide sequence corresponding to a specific mouse exon in DPP IV cDNA from human jejunum (Darmoul et al., 1992) or

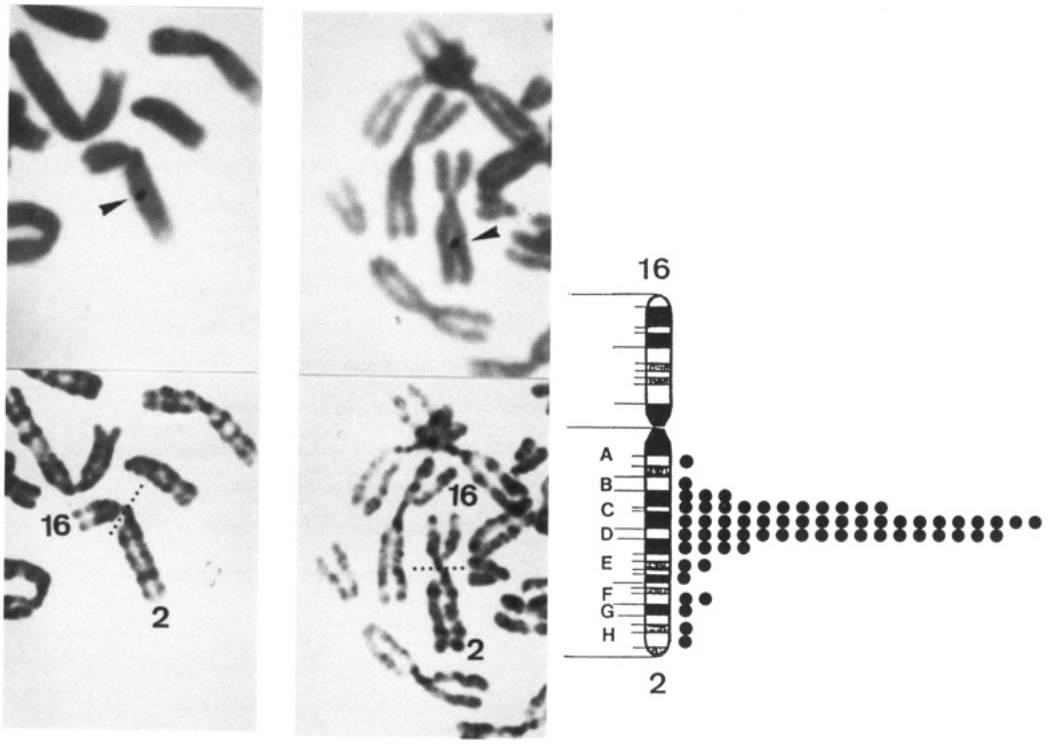

FIGURE 6: Assignment of *Dpp4* locus to mouse chromosome 2 by *in situ* hybridization. Two partial WMP mouse metaphases, showing the specific site of hybridization to chromosome 2. (Top) Arrowheads indicate silver grains on Giemsa-stained chromosomes following autoradiography. (Bottom) chromosomes with silver grains were subsequently identified by R-banding. To the right is a diagram of WMP mouse Rb(2;16) chromosome, indicating the distribution of labeled sites.

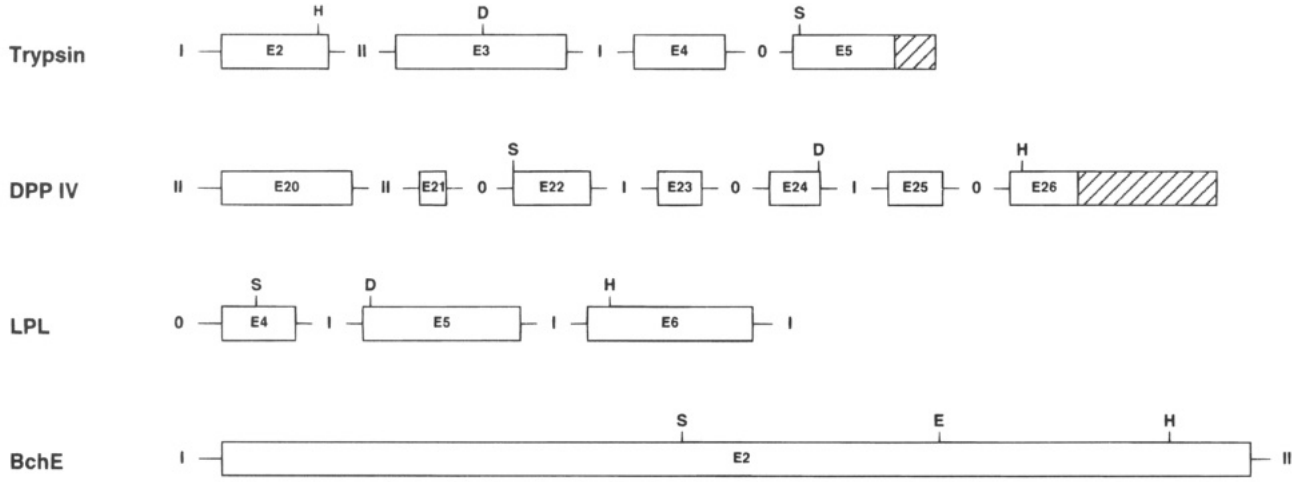

FIGURE 7: Comparison of DPP IV gene structure with that of other serine hydrolases. Gene segments analyzed correspond to the catalytic domains of trypsin (Craik et al., 1984), human lipoprotein lipase (Deeb & Peng, 1989), and human butyrylcholinesterase (Arpagauss et al., 1990). Exons are represented as open boxes, the noncoding regions as hatched boxes, and the introns (not to scale) as lines. The locations of the codons of the catalytic triad residues serine (S), histidine (H), and aspartic acid (D) or glutamic acid (E) as well as the phases of the intron (0, I, or II) (Rogers, 1985) are indicated.

PHA-activated T cells (Tanaka et al., 1992), as well as from rat liver (Hong & Doyle, 1987; Ogata et al., 1989). It should be pointed out that alternatively spliced forms of neutral endopeptidase, another ectopeptidase also encoded by small exons (D'Adamio et al., 1989), have been described (Iijima et al., 1992; Llorens-Cortes et al., 1990). However, one can question the physiological relevance of such observations made in a nonphysiological context. It seems very likely that both DPP IV and neutral endopeptidase molecules represent single isoforms resulting from unique splicing events. Such a contention implies that the soluble DPP IV identified in serum results from the proteolytic cleavage in the stalk region of the membrane-bound molecule.

Our data indicate that an alternative polyadenylation mechanism generates two mRNA species (3.3 and 4.9 kb) that differ in their 3′UTR. The high degree of homology observed between different regions of the 3′UTR of DPP IV mRNA in mouse, rat, and human species points toward a strong selective pressure. Duret et al. (1993) suggested that these highly conserved 3′ regions participate in post-transcriptional processes of gene expression (mRNA export, localization, translocation, or degradation) [for a review, see Jackson (1993)]. This raises the possibility that these distinct 3′UTRs could interfere with DPP IV mRNA expression.

Our observation that a single-copy gene in the [2C2−2D] chromosomal region encodes mouse DPP IV is consistent

with the mapping of the human DPP IV gene (Darmoul et al., 1992). Moreover, recent work has identified the ADA complexing protein as DPP IV (Morrison et al., 1993; Kameoka et al., 1993) and localized its gene on human chromosome 2, between MDH1 and IDH1, *i.e.*, in the segment 2p23−2q32 (Herbschleb-Voogt et al., 1981; Van Cong et al., 1981). The mouse map sublocation of DPP IV reinforces the definition of a large region of synteny with human chromosome 2 (i.e., 2q14−37) (Siracusa & Abbott, 1992; Copeland et al., 1993).

Wada et al. have recently isolated rat, bovine, and human cDNAs (Wada et al., 1992; Yokotani et al., 1993) that encode dipeptidyl aminopeptidase-related proteins named DPP X. The mapping of the corresponding genes to mouse chromosome 5 and to the human 7q region (Yokotani et al., 1993) supports the view that the DPP genetic loci are not clustered in mammalian genomes.

5′-Flanking Region of the DPP IV Gene. Since DPP IV expression is regulated in a tissue-specific fashion (Naquet et al., 1988; Hildebrandt et al., 1991; Marguet et al., 1992), we aimed at analyzing the gene 5′-flanking region in an attempt to identify DNA sequences that could control both the expression and the regulation of this prototypic member of a novel family of nonclassical serine proteases (Marguet et al., 1992). The putative mouse DPP IV gene promoter is highly similar in its organization to those presenting TATAA-box-deficient, G+C-rich promoter sequences (Ackerman et al., 1993). In particular, an imperfect symmetrical dyad is observed around the putative promoter's transcriptional initiation site, such as that described in genes of both mouse and human ADA (Valerio et al., 1985; Ackerman et al., 1993), mouse c-Ki-*ras* (Hoffman et al., 1987), human Ha-*ras* (Ishii et al., 1985), or mouse *Hox 1.4* (Galliot et al., 1989) [for a review, see Ackerman et al. (1993)]. The sequences upstream from the 5′-end of the mouse DPP IV cDNA very likely serve as a promoter region, which belongs to the TATAA-box-deficient, G+C-rich promoters group. However, we cannot exclude the presence of an additional exon upstream of exon 1, since no TATA- or CAAT-like sequences have been identified in the sequence analyzed upstream from the cDNA boundary.

Evolutionary Hypothesis for DPP IV-Related Enzymes. Functional and mechanistic convergence phenomena are both common in molecular evolution of enzymes, especially among the serine hydrolytic enzymes (Doolittle, 1994). Serine proteases sharing the ability to hydrolyze peptide bonds represent a good example for which distinct ancestors have evolved at different periods, as is the case for subtilisin, trypsin, and α/β type enzymes (Ollis et al., 1992). Comparative analysis of the catalytic triad (acid catalytic, histidine, nucleophile) from X-ray structures has revealed that these residues are arranged into the same geometry from completely different folds and sequence arrangements (Brady et al., 1990; Winkler et al., 1990; Derewenda & Derewenda, 1991; Schrag et al., 1992; Ollis et al., 1992; Tai et al., 1993).

Do the DPP IV-related enzymes appear as a new serine protease subfamily emmerging by convergent evolution, or is this subgroup related to one of the other subfamilies previously identified? At the present stage, we favor that the DPP IV-related family has diverged from a common ancestor of α/β hydrolases, a group of enzymes that has evolved to preserve the arrangement of the catalytic triad rather than the substrate binding sites, explaining the diversity of enzyme activities found among this subfamily's members (Ollis et al., 1992).

As a matter of fact, the sequence arrangement of the catalytic triad is similar between DPP IV-related enzymes and α/β hydrolases which regroup very diverse enzymes such as lipases, cholinesterases, dienelactone hydrolases, dehalogenases, and serine carboxypeptidases (Ollis et al., 1992). Close structural similarity was observed by comparing the short segments encoding the three catalytic residues as well as the respective topologies of the triads of lipases and DPP IV-related enzymes (Polgár, 1992; David et al., 1993).

In addition, the predictive secondary structure of DPP IV-related enzymes exhibiting an enzymatic domain made up by alternative α-helix and β-sheet segments (D. Marguet, unpublished observations) is strikingly similar to that provided by the tridimensional structure analysis of α/β hydrolases (Ollis et al., 1992).

Even though the exon/intron structure does not provide evidence for a divergent evolutionary process between DPP IV-related enzymes and the α/β hydrolases, the variations observed in the intron/exon junctions accumulated in these gene families may be attributed to insertions or deletions, or to the sliding of intron−exon junctions (Craik et al., 1983; Craik et al., 1984), occuring throughout evolution.

As the structural and mechanistic conservation in the absence of sequence similarity is the most remarkable feature of α/β hydrolases (Ollis et al., 1992), our observations suggest a possible evolutionary link between the DPP IV-related family and α/β hydrolases.

ACKNOWLEDGMENT

We wish to thank Marie Malissen and Jeannine Trucy for the gift of the genomic library; Hans Lehrach at ICRF for providing the YAC filter library; Elise Termine for oligonucleotide synthesis; and Giovanna Chimini, Pierre Golstein, Hai-Tao He, Anne-Odile Hueber, and Iciar Lazaro for helpful discussions and critical reading of the manuscript.

REFERENCES

Ackerman, S. L., Minden, A. G., & Yeung, C.-Y. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 11865−11869.

Ahmad, S., Wang, L., & Ward, P. E. (1992) *J. Pharmacol. Exp. Ther. 260*, 1257−1261.

Ameis, D., Stahnke, G., Kobayashi, J., Mclean, J., Lee, G., Büscher, M., Schotz, M. C., & Will, H. (1990) *J. Biol. Chem. 265*, 6552−6555.

Arpagauss, M., Kott, M., Vatsis, K. P., Bartels, C. F., La Du, B. N., & Lockridge, O. (1990) *Biochemistry 29*, 124−131.

Brady, L., Brzozowski, A. M., Derewenda, Z. S., Dodson, E., Dodson, G., Tolley, S., Turkenburg, J. P., Christiansen, L., Huge-Jensen, B., Norskov, L., Thim, L., & Menge, U. (1990) *Nature 343*, 767−770.

Breathnach, R., & Chambon, P. (1981) *Annu. Rev. Biochem. 50*, 349−383.

Breathnach, R., Benoist, C., O'Hare, K., Gannon, F., & Chambon, P. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 4853−4857.

Brenner, S. (1988) *Nature 334*, 528−530.

Cai, S.-J., Wong, D. M., Chen, S.-H., & Chan, L. (1989) *Biochemistry 28*, 8966−8971.

Callebaut, C., Krust, B., Jacotot, E., & Hovanessian, A. G. (1993) *Science 262*, 2045−2050.

Chimini, G., Boretto, J., Marguet, D., Lanau, F., Lauquin, G. J.-M., & Pontarotti, P. (1990) *Immunogenetics 32*, 419−426.

Copeland, N. G., Jenkins, N. A., Gilbert, D. J., Eppig, J. T., Maltais, L. J., Miller, J. C., Dietrich, W. F., Weaver, A., Lincoln, S. E., Steen, R. G., Stein, L. D., Nadeau, J. H., & Lander, E. S. (1993) *Science 262*, 57−66.

Craik, C. S., Rutter, W. J., & Fletterick, R. (1983) *Science 220*, 1125−1129.

Craik, C. S., Choo, Q.-L., Swift, G. H., Quinto, C., MacDonald, R. J., & Rutter, W. J. (1984) *J. Biol. Chem. 259*, 14255−14264.

D'Adamio, L., Shipp, M. A., Masteller, E. L., & Reinherz, E. L. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 7103−7107.

Darmoul, D., Lacasa, M., Baricault, L., Marguet, D., Sapin, C., Trotot, P., Barbat, A., & Trugnan, G. (1992) *J. Biol. Chem. 267*, 4824−4833.

David, F., Bernard, A.-M., Pierres, M., & Marguet, D. (1993) *J. Biol. Chem. 268*, 17247−17252.

Deeb, S. S., & Peng, R. (1989) *Biochemistry 28*, 4131−4135.

Derewenda, Z. S., & Derewenda, U. (1991) *Biochem. Cell Biol. 69*, 842−851.

Dessen, P., Fondrat, C., Valencien, C., & Mugnier, C. (1990) *Comput. Appl. Biosci. 6*, 355−356.

Doolittle, R. F. (1994) *Trends Biochem. Sci. 19*, 15−18.

Duret, L., Dorkeld, F., & Gautier, C. (1993) *Nucleic Acids Res. 21*, 2315−2322.

Erdös, E. G., & Skidgel, R. A. (1989) *FASEB J. 3*, 145−151.

Faisst, S., & Meyer, S. (1992) *Nucleic Acids Res. 20*, 3−26.

Fleischer, B. (1987) *J. Immunol. 138*, 1346−1350.

Fox, D. A., Hussey, R. E., Fitzgerald, K. A., Acuto, O., Poole, C., Palley, L., Daley, J. F., Schlossman, S. F., & Reinherz, E. L. (1984) *J. Immunol. 133*, 1250−1256.

Frischauf, A. M., Lehrach, H., Poutska, A., & Murray, N. (1983) *J. Mol. Biol. 170*, 827−837.

Frohman, L. A., Downs, T. R., Heimer, E. P., & Felix, A. M. (1989) *J. Clin. Invest. 83*, 1533−1540.

Galliot, B., Dolle, P., Vigneron, M., Featherstone, M. S., Baron, A., & Buboule, D. (1989) *Development 107*, 343−359.

Gilbert, W. (1985) *Science 228*, 823−824.

Hartel, S., Gossrau, R., Hanski, C., & Reutter, W. (1988) *Histochemistry 89*, 151−161.

Herbschleb-Voogt, E., Grzeschik, K.-H., Pearson, P. L., & Meera Khan, P. (1981) *Hum. Genet. 59*, 317−323.

Hildebrandt, M., Reutter, W., & Gitlin, J. D. (1991) *Biochem. J. 277*, 331−334.

Hoffman, E. K., Trusko, S. P., Freeman, N., & George, D. L. (1987) *Mol. Cell. Biol. 7*, 2592−2596.

Hong, W., & Doyle, D. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 7962−7966.

Hue, I., Trucy, J., McCoy, C., Couez, D., Malissen, B., & Malissen, M. (1990) *J. Immunol. 144*, 4410−4419.

Iijima, H., Gerard, N. P., Squassoni, C., Ewig, J., Face, D., Drazen, J. M., Kim, Y.-A., Shriver, B., Hersh, L. B., & Gerard, C. (1992) *Am. J. Physiol. 262*, L725−L729.

Irwin, D. M., Roberston, K. A., & MacGillivray, R. T. A. (1988) *J. Mol. Biol. 200*, 31−45.

Ishii, S., Merlina, G. T., & Pastan, I. (1985) *Science 230*, 1378−1381.

Jackson, R. J. (1993) *Cell 74*, 9−14.

Kameoka, J., Tanaka, T., Nojima, Y., Schlossman, S. F., & Morimoto, C. (1993) *Science 261*, 466−469.

Larin, Z., Monaco, A. P., & Lehrach, H. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 4123−4127.

Lehrach, H. (1990) in *Genome Analysis Volume 1: Genetic and Physical Mapping*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Li, Y., Camp, S., & Taylor, P. (1993) *J. Biol. Chem. 268*, 5790−5797.

Litt, M., & Luty, J. A. (1989) *Am. J. Hum. Genet. 44*, 397−401.

Llorens-Cortes, C., Giros, B., & Schwartz, J.-C. (1990) *J. Neurochem. 55*, 2146−2148.

Look, A. T., Ashmun, R. A., Shapiro, L. H., & Peiper, S. C. (1989) *J. Clin. Invest. 83*, 1299−1307.

Marguet, D., Bernard, A.-M., Vivier, I., Darmoul, D., Naquet, P., & Pierres, M. (1992) *J. Biol. Chem. 267*, 2200−2208.

Marguet, D., David, F., Vivier, I., Bernard, A.-M., Naquet, P., & Pierres, M. (1993) *Adv. Neuroimmunol. 3*, 209−215.

Mattei, M.-G., Philip, N., Passage, E., Moisan, J. P., Mandel, J. L., & Mattei, J. F. (1985) *Hum. Genet. 69*, 268−271.

Morrison, M. E., Vijayasaradhi, S., Engelstein, D., Albino, A. P., & Houghton, A. N. (1993) *J. Exp. Med. 177*, 1135−1143.

Mount, S. M. (1982) *Nucleic Acids Res. 10*, 459−472.

Naquet, P., MacDonald, H. R., Brekelmans, P., Barbet, J., Marchetto, S., Van Ewijk, W., & Pierres, M. (1988) *J. Immunol. 141*, 4101−4109.

Nausch, I., & Heymann, E. (1985) *J. Neurochem. 44*, 1354−1357.

Ogata, S., Misumi, Y., & Ikehara, Y. (1989) *J. Biol. Chem. 264*, 3596−3601.

Ogata, S., Misumi, Y., Tsujii, E., Takami, N., Oda, K., & Ikehara, Y. (1992) *Biochemistry 31*, 2582−2587.

Ollis, D. L., Cheah, E., Cygler, M., Dijkstra, B., Frolow, F., Franken, S. M., Harel, M., Remington, S. J., Silman, I., Schrag, J., Sussman, J. L., Verschueren, K. H. G., & Goldman, A. (1992) *Protein Eng. 5*, 197−211.

Polgár, L. (1992) *FEBS Lett. 311*, 281−284.

Rogers, J. (1985) *Nature 315*, 458−459.

Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) *Molecular cloning, a laboratory manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463−5467.

Schrag, J. D., Winkler, F. K., & Cygler, M. (1992) *J. Biol. Chem. 267*, 4300−4303.

Shipp, M. A., Richardson, N. E., Sayre, P. H., Brown, N. R., Masteller, E. L., Clayton, L. K., Ritz, J., & Reinherz, E. L. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 4819−4823.

Siracusa, L. D., & Abbott, C. M. (1992) *Mamm. Genome 3*, S20−S43.

Tai, M.-H., Chirala, S. S., & Wakil, S. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 1852−1856.

Tanaka, T., Camerini, D., Seed, B., Torimoto, Y., Dang, N. H., Kameoka, J., Dahlberg, H. N., Schlossman, S. F., & Morimoto, C. (1992) *J. Immunol. 149*, 481−486.

Tautz, D. (1989) *Nucleic Acids Res. 17*, 6463−6471.

Valerio, D., Duyvesteyn, M. G. C., Dekker, B. M. M., Weeda, G., Berkvens, T. M., van der Voorn, L., van Ormondt, H., & van der Eb, A. J. (1985) *EMBO J. 4*, 437−443.

Van Cong, N., Weil, D., Gross, M.-S., Foubert, C., Jami, J., & Frezal, J. (1981) *Ann. Genet. 24*, 141−147.

Vanderslice, P., Ballinger, S. M., Tam, E. T., Goldstein, S. M., Craik, C. S., & Caughey, G. H. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 3811−3815.

Vivier, I., Marguet, D., Naquet, P., Bonicel, J., Black, D., Li, C. X., Bernard, A.-M., Gorvel, J. P., & Pierres, M. (1991) *J. Immunol. 147*, 447−454.

Wada, K., Yokotani, N., Hunter, C., Doi, K., Wenthold, R. J., & Shimasaki, S. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 197−201.

Weber, J. L., & May, P. E. (1989) *Am. J. Hum. Genet. 44*, 388−396.

Winkler, F. K., D'Arcy, A., & Hunziker, W. (1990) *Nature 343*, 771−774.

Wu, Q., Li, L., Cooper, M. D., Pierres, M., & Gorvel, J.-P. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 676−680.

Yokotani, N., Doi, K., Wenthold, R. J., & Wada, K. (1993) *Hum. Mol. Genet. 2*, 1037−1039.