

Short sequence-paper

## Molecular characterization and tissue-specific expression of a murine putative G-protein-coupled receptor

Herbert Mayer<sup>a</sup>, Johannes Breuss<sup>b</sup>, Sophie Ziegler<sup>b</sup>, Rainer Prohaska<sup>a,\*</sup>

<sup>a</sup> *Institute of Biochemistry, University of Vienna, Vienna Biocenter, Dr. Bohr-Gasse 9/3, A-1030 Vienna, Austria*

<sup>b</sup> *Department of Vascular Biology and Thrombosis Research, University of Vienna, A-1090 Vienna, Austria*

Received 6 March 1998; revised 12 May 1998; accepted 20 May 1998

### Abstract

We isolated by 5'- and 3'-RACE (rapid amplification of cDNA ends) clones from a murine brain cDNA library which encode a putative G-protein-coupled receptor. The composite nucleotide sequence revealed a coding region of 1197 nt; the deduced amino acid sequence of 399 amino acids showed 91.5% identity (95.7% similarity) when compared with the human homolog. An intron-like sequence, possibly involved in the regulation of expression, was found within the 5'-untranslated region. Northern blot analysis showed that the major 1.7-kb transcript is widely expressed, notably in brain and testis. In situ hybridization studies of tissue sections revealed high expression in neurons of the brain, epithelial cells of the lung, kidney and intestine, and in alveolar macrophages. © 1998 Elsevier Science B.V. All rights reserved.

*Keywords:* G-protein-coupled receptor; Membrane protein; Stomatin; In situ hybridization; Intron

The erythrocyte band 7 integral membrane protein, which is now known as stomatin (also termed protein 7.2b or band 7.2) [1–4], is not only expressed in red cells, but is also widely distributed in various tissues [5–7]. Because of its absence in the high Na<sup>+</sup>, low K<sup>+</sup> red cells of patients with overhydrated hereditary stomatocytosis, it has been proposed that it may play a role as a regulator of an ion channel [8,9]. In an effort to identify a stomatin-associated membrane protein, we have recently isolated by stomatin peptide-affinity chromatography a 40-kDa human erythrocyte membrane protein (p40) which has seven hydrophobic domains and several other features of a G-protein-coupled receptor [10]. This putative recep-

tor is, again, not only expressed in red cells, but also in other hematopoietic cells (thymocytes, macrophages and megakaryocytes) and in the neurons of the brain and spinal cord. In the present study we describe the cloning and characterization of the murine p40 homolog, and its tissue-specific expression.

Human p40 did not show a significant homology with known G-protein-coupled receptors, however, we noticed the presence of three murine homologs in the dbEST database (accession numbers W54092, W54529 and AA530541). On the basis of the corresponding nucleotide sequence information, we designed oligonucleotide probes: P1, 5'-AGAC-TACGGGGAGCACGGATGCAGG-3'; and P2, 5'-TGGCACCAATGGACAAGCAGGTCTC-3' (Fig. 1); and performed 5'- and 3'-RACE by PCR using a Marathon-Ready cDNA library from murine brain (Clontech 7450-1) and the Advantage cDNA PCR

\* Corresponding author. Fax: +43 (1) 79515-3114;  
E-mail: prohaska@bch.univie.ac.at

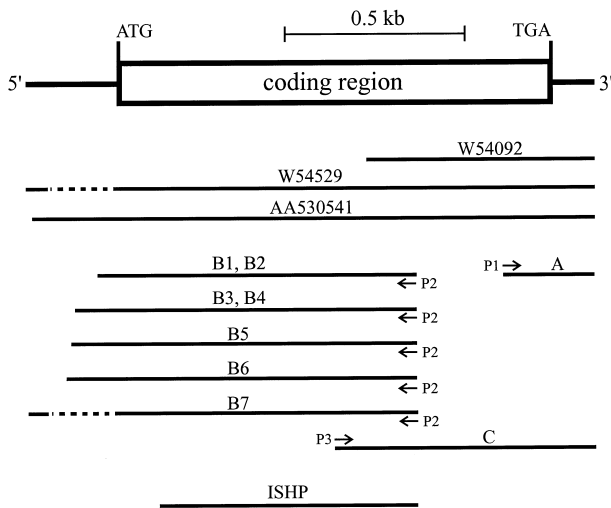


Fig. 1. Schematic model of the murine p40 cDNA. The coding region is depicted by the enclosed box, 5'- and 3'-UTRs are indicated by lines. The relative positions of three characterized EST-clones (W54092, W54529, AA530541) and of nine RACE-PCR clones (A, B, C) are shown. A dotted line is drawn within the clones which do not contain the intron-like sequence in the 5'-UTR, as described in the text. Additionally, the position of the cRNA probe used for in situ hybridization is indicated (ISHP). P1–P3: PCR-primers used for 5'- and 3'-RACE in combination with the appropriate cDNA anchor primers.

Kit (Clontech K1905-1). The primer P3: 5'-CTTGGCGGCAAAGTCTCCACTGATG-3' (Fig. 1) was designed according to the sequence information obtained by the P2 5'-RACE. Nine partially overlapping cDNA clones, seven containing the 5'-region (B) and two containing the 3'-region (A, C), were isolated and sequenced [11]. Additionally, the EST clones W54092 (from total murine fetus, clone ID 368170), W54529 (from total murine fetus, clone ID 337699), and AA530541 (from murine diaphragm muscle, clone ID 931450), obtained from the I.M.A.G.E. consortium [12], were also sequenced. Nucleotide sequence compilation, deduction of the amino acid sequence, and database searches were accomplished by the use of the GCG (Genetics Computer Group, Madison) software package available on EMBnet. The analysis of the combined sequences revealed an open reading frame of 1392 bp and a coding region of 1197 bp (399 amino acids), starting from the first ATG (Fig. 2). The start codon is located within an appropriate consensus sequence for the initiation of translation [13]. An in-frame stop

codon lies at position  $-195$  upstream of the start codon. The 3'-untranslated region (UTR) contains an imperfect polyadenylation signal (AATGAA) at position 1286 (relative to the start codon), which is followed by a poly(A) tail 19 bp downstream. The usage of imperfect polyadenylation signals has been shown before by the identification of various respective murine transcripts [14,15], and the 1.9-kb transcript of human p40 also originates from the use of a non-consensus polyadenylation signal [10].

When we compared the sequences of the 5'-UTR-containing clones, we noticed an insertion of 184 bp (position  $-200$  to  $-17$ ) in the EST clone AA530541 and, partially, in six out of seven 5'-RACE clones (Fig. 1). This inserted sequence has the structural characteristics of an intron [16]; its 5'-splice site reads ACgt and the 3'-splice site cagGT, which is preceded by a pyrimidine stretch and a putative branch acceptor. The EST clone W54529 does not contain this intervening sequence (Fig. 1). It is not clear, why the majority of our 5'-RACE clones are unspliced, however, the possibility exists that this putative intron has a regulatory function on expression, as described for a variety of mRNAs, including G-protein-coupled receptors [17–20]. Interestingly, a similar intron is also present in a homologous rat EST clone (accession number AA818760) and is highly conserved (93.5% identity, compared to mouse), stressing the functional importance of this sequence.

The deduced amino acid sequence was compared with the human p40 sequence (Fig. 3) and thereby showed that these proteins are highly homologous (91.5% identity, 95.7% similarity). All seven transmembrane domains identified by the TMpred-program [21] are conserved, except for four conservative exchanges, and all 13 cysteine residues are invariable. Also, the putative *N*-glycosylation sites (N-14 and N-337) and in general the phosphorylation sites for protein kinase C (S-100, S-194), casein kinase II (T-188), and tyrosine kinase (Y-298) are conserved; the PKC sites are threonine residues in human [10]. This structure prediction suggests that murine p40 is (like the human homolog) a member of the large superfamily of G-protein-coupled receptors. Because there is no significant sequence homology between p40 and the known G-protein-coupled receptors, it may be regarded as an orphan receptor [22]. Other

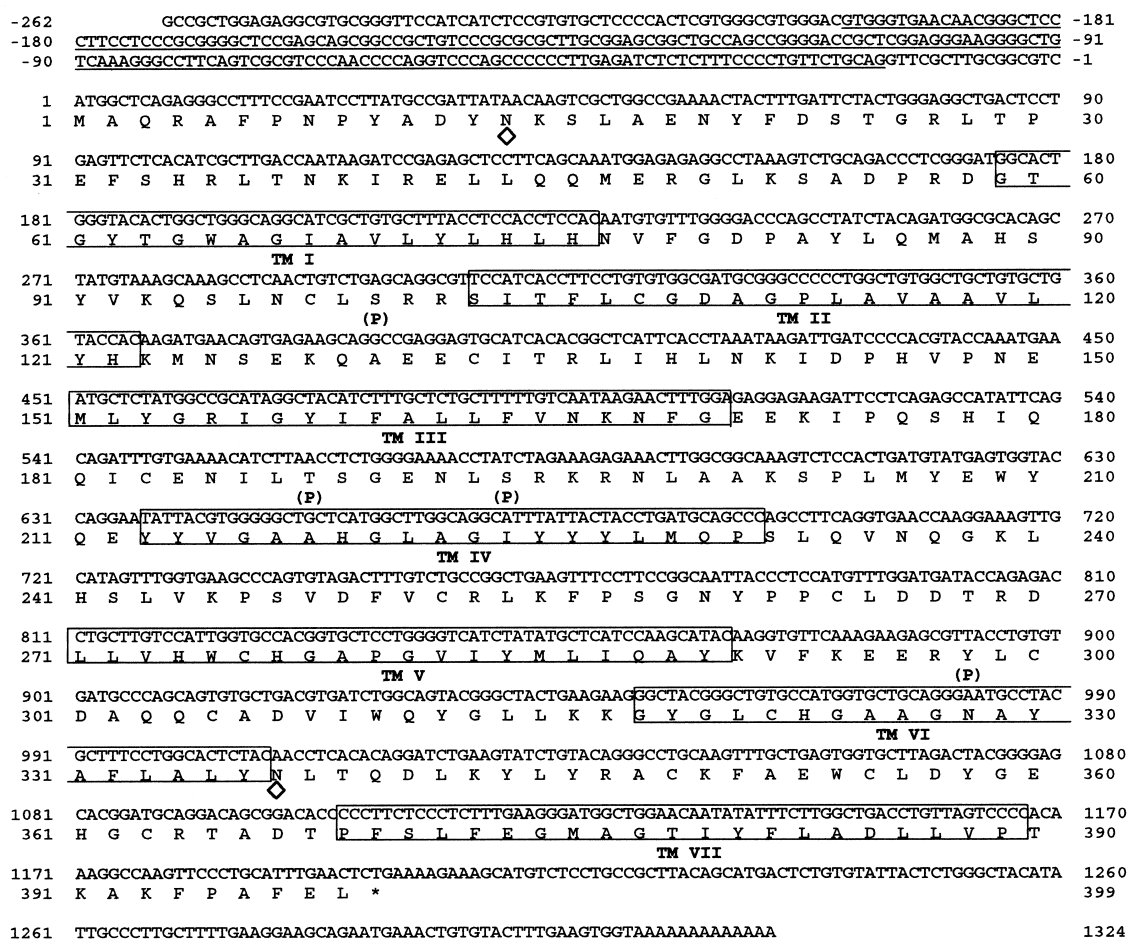


Fig. 2. Composite nucleotide and deduced amino acid sequence of murine p40 cDNA. Nucleotides and amino acid residues are numbered starting from the first ATG in the open reading frame. The seven putative membrane-spanning domains are boxed and labeled TM I–VII. The translation termination codon is indicated by an asterisk. A putative intron in the 5'-untranslated region is underlined; the imperfect polyadenylation signal is double-underlined. Putative glycosylation sites are marked by diamonds, putative phosphorylation sites by (P). The EMBL/GenBank/DBJ accession numbers are Y16518 and Y11550.

conserved features include a putative G-protein-binding site at the C-terminal end of TM III [23], all the charged residues within the transmembrane domains, and a repeat-structure within the seven helices, with the consensus GXAGhhY (h being a hydrophobic residue) [10]. Using the TBLASTN program [24], we searched the dbEST database and identified two EST clones coding for a highly homologous protein from the rat (accession numbers H33811 and AA818760), showing about 99% identity when compared with the murine p40 amino acid sequence.

Northern blot analysis of multiple murine tissues (Clontech 7762-1) revealed the high expression of a major 1.7-kb transcript, which is in accordance with

the size of the characterized cDNA (Fig. 2), in various tissues, notably in testis and brain (Fig. 4). Additionally, a minor 5-kb transcript was found in brain and testis. Human p40 mRNA is also found in two forms, 1.9 and 4.8 kb; however, the major transcript is the 4.8-kb isoform [10]. The human isoforms originate from the use of two polyadenylation signals in the 3'-UTR; therefore, it is likely that the murine 5-kb mRNA may also contain a long 3'-UTR. When we compared the intensity of the murine isoforms after normalization to the actin signals (Fig. 4), the amount of the 1.7-kb testis mRNA was four times higher than that of brain, alternatively, the 5-kb testis signal was only half that of the brain. It is,

therefore, possible, that the 3'-UTR plays a role in the regulation of p40 expression in various tissues.

In order to detect tissue-specific and cell type-specific p40 expression we hybridized cryosections (6 µm thick) of murine tissues with a 0.7-kb antisense-RNA (see Fig. 1) transcribed from the p40 coding region, and with the corresponding sense-RNA as a control. The probes were labeled with digoxigenin and detected by the specific antibody (Boehringer Mannheim) conjugated with alkaline phosphatase (NBT/BCIP staining). Prominent signals were identified in the neurons of the brain, in the epithelial cells of various tissues, and in macrophages, indicating a wider distribution of p40 in the murine tissues than in human ones [10]. In cerebellum, a strong signal in the Purkinje cell layer as well as in the granular layer was observed (Fig. 5A). High expression was also found in the bronchial epithelium and alveolar mac-

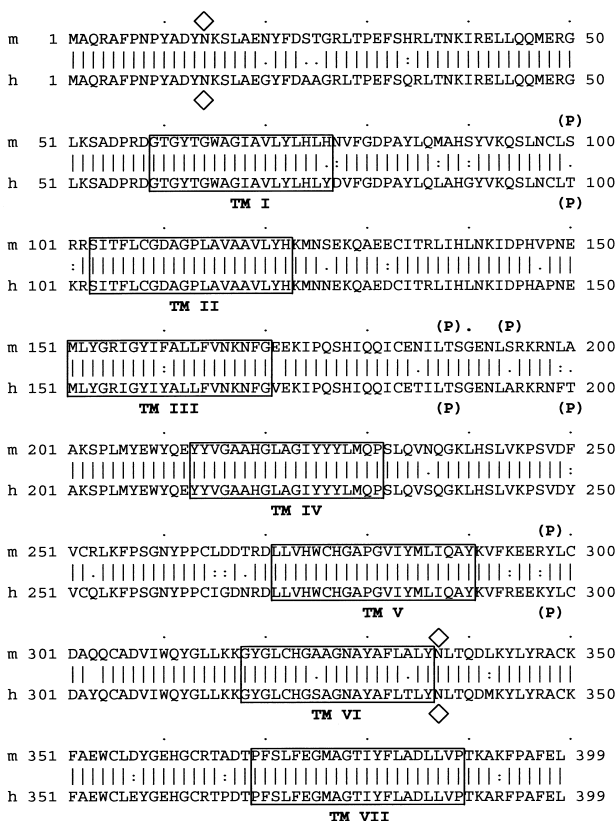


Fig. 3. Homology between murine and human p40. The amino acid sequences of murine (m) and human (h) p40 were aligned. The transmembrane regions (TMI–VII), including the charged residues, are highly conserved. Putative glycosylation sites are marked by diamonds, putative phosphorylation sites by (P).

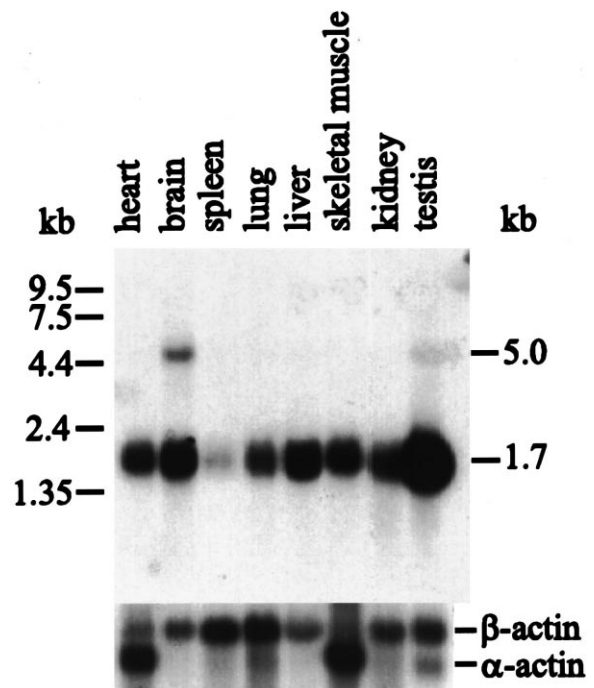


Fig. 4. Northern blot analysis. Blots containing 2 µg poly(A)<sup>+</sup> RNA per lane (Clontech) from eight murine tissues were hybridized with a <sup>32</sup>P-labeled p40 cDNA probe. The major transcript of 1.7 kb is highly expressed in testis, brain, heart, skeletal muscle, liver, kidney and lung. Additionally, a minor 5-kb transcript is found in brain and testis. Hybridization with an actin probe as a standard is shown below. RNA size markers are depicted on the left side.

rophages of the lung (Fig. 5B), in the epithelial cells of the proximal tubules of the kidney (Fig. 5C), and in the enterocytes and crypt cells of the small intestine (Fig. 5D). No signals were detectable in the corresponding sense controls (data not shown).

In conclusion, we have isolated and characterized the cDNA coding for murine p40 and analyzed its tissue-specific expression. Based on its structure and distribution, we postulate that murine p40, like its human homolog, may function as a G-protein-coupled receptor for a peptide or neurotransmitter. Because of the wider distribution of the murine receptor, notably in epithelial cells, it may have a broader functional spectrum than the human protein. Further biochemical and functional studies will provide new insight into the nature of the ligand and the receptor-associated G-protein(s).

This work was supported by the Austrian Science Foundation (FWF), Grant P12743 (to R.P.).

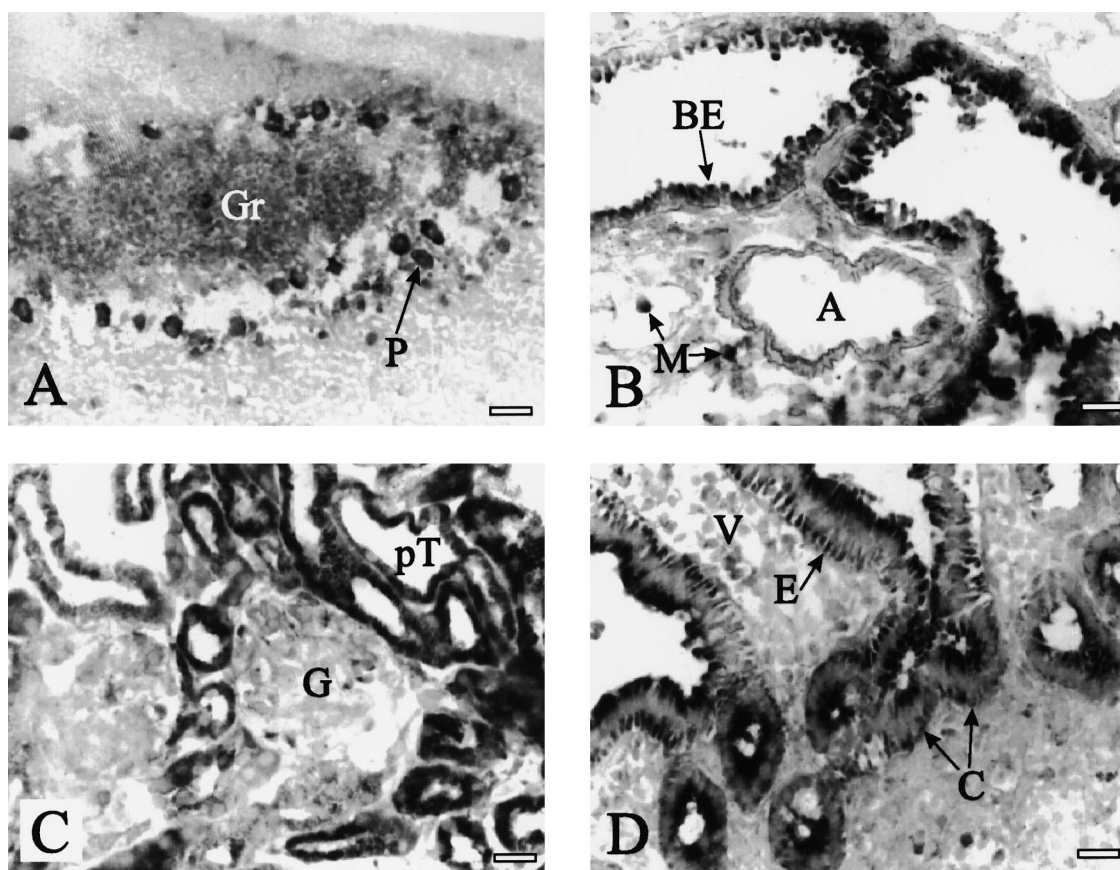


Fig. 5. In situ hybridization analysis. Fixed cryosections of tissues from adult mice were hybridized with the digoxigenin-labeled p40-antisense RNA probe (A–D). Strong signals were observed in the Purkinje cells and the granular layer of the cerebellum (A), bronchial epithelium and macrophages of the lung (B), proximal tubules of the kidney (C), enterocytes and crypt cells of the small intestine (D). A, artery; BE, bronchial epithelium; C, crypt cells; E, enterocytes; G, glomerulus; Gr, granular layer; M, alveolar macrophages; P, Purkinje cells; pT, proximal tubules; V, villus. Scale bars: (A) 100  $\mu\text{m}$ ; (B–D) 50  $\mu\text{m}$ .

## References

- [1] C.M. Hiebl-Dirschmied, G.R. Adolf, R. Prohaska, Isolation and partial characterization of the human erythrocyte band 7 integral membrane protein, *Biochim. Biophys. Acta* 1065 (1991) 195–202.
- [2] D. Wang, W.C. Mentzer, T. Cameron, R.M. Johnson, Purification of band 7.2b, a 31-kDa integral phosphoprotein absent in hereditary stomatocytosis, *J. Biol. Chem.* 266 (1991) 17826–17831.
- [3] C.M. Hiebl-Dirschmied, B. Entler, C. Glotzmann, I. Maurer-Fogy, C. Stratowa, R. Prohaska, Cloning and nucleotide sequence of cDNA encoding human erythrocyte band 7 integral membrane protein, *Biochim. Biophys. Acta* 1090 (1991) 123–124.
- [4] G.W. Stewart, B.E. Hepworth-Jones, J.N. Keen, B.C.J. Dash, A.C. Argent, C.M. Casimir, Isolation of cDNA coding for an ubiquitous membrane protein deficient in high  $\text{Na}^+$ , low  $\text{K}^+$  stomatocytic erythrocytes, *Blood* 79 (1992) 1593–1601.
- [5] I. Unfried, B. Entler, R. Prohaska, The organization of the gene (EPB72) encoding the human erythrocyte band 7 integral membrane protein (protein 7.2b), *Genomics* 30 (1995) 521–528.
- [6] P.G. Gallagher, M. Romana, J.H. Lieman, D.C. Ward, cDNA structure, tissue-specific expression, and chromosomal localization of the murine band 7.2b gene, *Blood* 86 (1995) 359–365.
- [7] W. Schlegel, I. Unfried, R. Prohaska, Cloning and analysis of a cDNA encoding the BALB/c murine erythrocyte band 7 integral membrane protein, *Gene* 178 (1996) 115–118.
- [8] W.M. Lande, P.V.W. Thiemann, W.C. Mentzer, Missing band 7 membrane protein in two patients with high  $\text{Na}$ , low  $\text{K}$  erythrocytes, *J. Clin. Invest.* 70 (1982) 1273–1280.
- [9] G.W. Stewart, A.C. Argent, B.C.J. Dash, Stomatin: a putative cation transport regulator in the red cell membrane, *Biochim. Biophys. Acta* 1225 (1993) 15–25.
- [10] H. Mayer, U. Salzer, J. Breuss, S. Ziegler, A. Marchler-Bauer, R. Prohaska, Isolation, molecular characterization, and tissue-specific expression of a novel putative G pro-

- tein-coupled receptor, *Biochim. Biophys. Acta* 1395 (1998) 301–308.
- [11] J. Sambrook, E.F. Fritsch, T. Maniatis, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.
- [12] G. Lennon, C. Auffray, M. Polymeropoulos, M.B. Soares, The I.M.A.G.E. consortium: an integrated molecular analysis of genomes and their expression, *Genomics* 33 (1996) 151–152.
- [13] M. Kozak, Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs, *Nucleic Acids Res.* 12 (1987) 857–872.
- [14] M.P. Brocard, D. Rousseau, J.J. Lawrence, S. Khochbin, Two mRNA species encoding the differentiation-associated histone H1<sup>o</sup> are produced by alternative polyadenylation in mouse, *Eur. J. Biochem.* 221 (1994) 421–425.
- [15] S. Anand, F.D. Batista, T. Tkach, D.G. Efremov, O.R. Burrono, Multiple transcripts of the murine immunoglobulin epsilon membrane locus are generated by alternative splicing and differential usage of two polyadenylation sites, *Mol. Immunol.* 34 (1997) 175–183.
- [16] R.M. Stephens, T.D. Schneider, Features of spliceosome evolution and function inferred from an analysis of the information at human splice sites, *J. Mol. Biol.* 228 (1992) 1124–1136.
- [17] S. Boularand, M.C. Darmon, J. Mallet, The human tryptophan hydroxylase gene. An unusual splicing complexity in the 5'-untranslated region, *J. Biol. Chem.* 270 (1995) 3748–3756.
- [18] D. Shire, C. Carillon, M. Kaghad, B. Calandra, M. Rinaldi-Carmona, G. Le Fur, D. Caput, P. Ferrara, An amino-terminal variant of the central cannabinoid receptor resulting from alternative splicing, *J. Biol. Chem.* 270 (1995) 3726–3731.
- [19] A. Damert, B. Leibiger, I.B. Leibiger, Dual function of the intron of the rat insulin I gene in regulation of gene expression, *Diabetologia* 39 (1996) 1165–1172.
- [20] K.M. Curnow, L. Pascoe, P.C. White, Genetic analysis of the human type-1 angiotensin II receptor, *Mol. Endocrinol.* 6 (1992) 1113–1118.
- [21] K. Hofmann, W. Stoffel, Tmbase – a database of membrane spanning protein segments, *Biol. Chem. Hoppe Seyler* 347 (1993) 166.
- [22] S. Watson, S. Arkinstall, *The G-Protein Linked Receptor Facts Book*, Academic Press, London, 1994.
- [23] T.M. Savarese, C.M. Fraser, In vitro mutagenesis and the search for structure–function relationships among G protein-coupled receptors, *Biochem. J.* 283 (1992) 1–19.
- [24] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, *J. Mol. Biol.* 215 (1990) 403–410.