Alternative promoter usage and tissue specific expression of the mouse somatostatin receptor 2 gene

Jürgen Kraus*, Michael Wöltje, Nadja Schönwetter, Volker Höllt

Institut für Pharmakologie und Toxikologie, Universität Magdeburg, Leipzigerstr. 44, D-39120 Magdeburg, Germany

Received 21 April 1998

Abstract We have cloned the 5' upstream regulatory region of the mouse somatostatin receptor 2 gene. Its genomic organization is novel among all somatostatin receptor genes. It contains two previously unrecognized exons, separated by introns larger than 25 kb, and three tissue and cell specific alternative promoters. The first promoter in front of exon 1 is active only in AtT-20 tumor cells. The second promoter, located 5' to exon 2, is used in brain, pituitary, adrenals, pancreas, NG 108-15 and AtT-20 cells. Furthermore, it contains putative DNA elements for regulation by glucocorticoids, estradiol and cAMP. A third promoter, located in exon 3, is additionally used in lung, kidney and spleen.

© 1998 Federation of European Biochemical Societies.

Key words: Somatostatin receptor 2; Alternative promoter usage; Tissue specific gene expression

1. Introduction

Somatostatin 14 and its N-terminally extended form somatostatin 28 are widely distributed throughout the central nervous system and peripheral tissues. They act as neurotransmitters or neuromodulators, and in particular as inhibitors of neurotransmitter release and endocrine and exocrine secretion (for a review see [1]). The multiple physiological effects of somatostatin are known to be mediated via five distinct types of G protein coupled somatostatin receptors (sst1-sst5) [2-10]. A further degree of genetic diversification of ssts has been reported for the subtype 2, which is alternatively spliced to generate the two isoforms sst2A and sst2B with different Ctermini [11,12]. Sst2 is involved in inhibition of growth hormone release from the pituitary [13], pancreatic glucagon secretion [14] and gastric acid secretion from stomach [15]. At the cellular level, sst2 mediates somatostatin evoked inhibition of cAMP accumulation by coupling to adenylyl cyclase via G_i proteins [10,16]. As shown more recently, sst2 also couples to a tyrosine phosphatase, which has been proposed to be involved in inhibition of tumor cell growth observed after clinical administration of somatostatin analogs such as octreotide [17,18]. Little is known about the genomic structure of sst2 and sst genes in general and the molecular mechanisms controlling transcription in the various tissues. In fact, only genomic sequences immediately adjacent to the coding regions of sst genes have been the subject of investigation to date [19-23]. Here, we report the cloning of the 5' upstream regulatory region (5' URR) of the mouse sst2 gene. We provide evidence for the existence of two previously unrecognized exons and introns and three individual, alternative promoters located

within 50 kb upstream of the protein coding region. In RT-PCR experiments the usage of the three promoters was examined in various tissues and cell types.

2. Materials and methods

2.1. Cloning and mapping of genomic fragments of sst2

The exonic sequences of sst2 are contained on two independent bacteriophage P1 clones of a mouse embryonic stem cell library (Genome Systems, St. Louis, MO, USA). From P1 clone #24 exon 1 was obtained, while P1 clone #88 contained exons 2-4. For subcloning and mapping P1 DNA was first characterized by restriction enzyme analysis and Southern blot analysis. Then, appropriate fragments were subcloned in pBluesript SK⁻ (Stratagene, Heidelberg, Germany) according to standard methods [24]. As hybridization probes served randomly ³²P-dATP (Hartmann, Braunschweig) labeled fragments of a sst2B cDNA [11]. To determine the nucleotide sequence overlapping inserts of subclones were sequenced on both strands by the dideoxy chain termination method using the Sequenase II kit (USB, Braunschweig, Germany). The size of the first intron could not be determined, since exons 1 and 2 are contained on different, non-overlapping P1 genomic library clones. As revealed by hybridization, exon 1 is located at a distance of at least 15 kb from the 3' end of the P1 clone #24 insert, and exon 2 at least 10 kb from the 5' end of the P1 clone #88 insert. The first intron thus must be larger than 25 kb. The size of intron B was determined by hybridization of DNA isolated from P1 clone #88 with probes for the adjacent exons 2 and 3. A NotI fragment of about 40 kb contained both exons, whereas various fragments up to 25 kb in size obtained by digestion of P1 DNA with other restriction enzymes only contained either exon 2 or exon 3. The size of intron B is thus between 25 and 40 kb.

2.2. Primer extension experiments

Primer extension reactions were preformed with 1 μ g poly(A)⁺ RNA (RNeasy and Oligotex mRNA, Qiagen) and 750 nM oligonucleotide primer (sequences see Fig. 2). The antisense DNA was extended with 5 units Tth polymerase (Boehringer, Mannheim, Germany) under conditions suggested by the manufacturer with additionally 10 μ Ci α -32P-ATP. Cycling conditions were 30 s at 95°C, 30 s at 60°C and 1 min at 75°C for 30 cycles. The products were analyzed on a 8% denaturing polyacrylamide gel.

2.3. RT-PCR experiments

Total RNA was extracted from various mouse tissues using RNeasy columns (Qiagen). cDNA was transcribed with random hexamer and oligo dT primers and Superscript II RT (Gibco BRL). Nucleotide sequences of primers were as follows: P1 upstream primer: 5'-GCTGGAGGTAGTCATTGAGCTT-3', P2 upstream primer: 5'-CCCGGGCAAGCTCTCTCA-3', P3 upstream primer: 5'-AGGAA-GATCTCTAGGCAGCTTG-3', exon 3 downstream primer: 5'-TGATGGTCTTCATCTTGGCGT-3'. PCR was performed under the following conditions: 30 cycles of 1 min at 96°C, 1 min at 55°C and 2 min at 72°C. To check the integrity of the RNA/cDNA a 605 bp sequence of the glycerine aldehyde phosphate dehydrogenase gene (GAPDH) was amplified from the same preparations for 25 cycles of 1 min at 96°C, 1 min at 60°C and 3 min at 72°C. Primers were: upstream 5'-TGGCAGCTTTCTCCAGGCGGC-3' and downstream 5'-CCCACGGCAAGTTCAACGGCA-3'. PCR products were electrophoresed on 1% agarose gels, blotted and hybridized under stringent conditions (5×SSC, 20% formamide; o/n at 63°C) with a randomly labeled radioactive probe. The washed blots (0.2×SSC, 1%

^{*}Corresponding author. Fax: (49) (391) 6715869.

1	ATGAACCTCC	AACCAATCGT	CTCATTCTTG	GTTTGGACTG YY1	AAAGCCCATG	GTAATTTAGA	AAAGATACTA	TAACCTGAGG
81	ACAAATCAAG	ACCAGAAAAT	GAAAACATTT		AG CAGATTTT	TTTTTTGTCT	TAATATATTT ERE	CCAATACAGT
161	CATTCAGCAA	ACAGAAACAA	GTAACAACAA	CAGAAATCAG	ACAAAGCAAC	TGGAACTGCC		GACCTTAAAA
241	GTGCTAGTGC	GTTCTCTCAA	AGGACTTAGC	ACTCTCGGTT	TGCACAGGAA	GTTCTAATAC	CTTATTACTT	TGTCTAATAA
321	TCAGGTAATA	AAAAAGGTCT	GAGACACTCT	TGGGACCCTA	GACACACAGA	AATGAGAGCA	GGAAAAGACG	GCAAGGTCAG
401	CCAAGTCGGC	AAGCTCCAGC	TCTCTGCTCG	AAGGCGTCGG	AACCTTAGAA	GAGGCAAGGA	GACTGGCCTG YY1	TCTGCCTACA
481	GGAGGCATTT	AACTGCTTCT	GCCAAGCAGG	GTCTTATAAA	AATGACAGGA	AGCCAAGAAC		C TGACAAATA
561	ACGACAATCG	AGATGGTGAT	GTCAGGCCCA	ATAAAGCATT	TTGGCAAGTT	AAGAATGTGC	ACCTGTGAGA	ACTGAGGACG
641	ACTTGACAGA	CTCCTGTTTC	CCTGGCTCTT	TCTCGGCAAA	CTAATCGTGG	TGACTGAAGA	GTCTACCCTG	AGCTTTACAG
721	ATGCACGGAT	CCAACAACTG	CTGCTCTCCT	GGGCATCAGG	AAGACAAAAC	GACAGCGGTG	CTTACATGCA	GAAGTATTCT
801	GCGCACGATG	GAACAATGAG	CCAGAGAGGA GRE	TTTAGATGGG	GGAGAGGGGT	GCTCTGCAGA	TATACTGGTG	AGACCGGCCA
881	GGAAAGGAAA	GGACACCCAA		CGTGTGTGCT	TACTCATCCT	CAGAGCCTCA	GAGGCCTCTG	GCCACACTGC
961	CATCGAGCAT	GCTTTCCCCC		GTACC TTCGC	AGCTATAGCT	GTACACTGCC	ACGGCTGACC	TGTCCACCAG
1041	GTTCTCATCG	TGATGCCAGC		CTT CCCCATG	CCGAAATAGG	GCTCCTCTTT	CAAGTAGGGC	ATCTTCTGAG
1121	GATCCATGAA	GTTTAGCAAA	GTCACGTTGT	AGGCTGCTCT	GCTCTTAAGG	TCCACTTCAT	CATCACAGGA	CGGCCCCACG
1201	CCGGCCCTGG	GGAACTCTGC	CATGCACAGT	GGCACAGCAT	CTTCATTGGC	CTTCTCTCTG	ACAGCCAGTT	CTTCCAAGGC
1281	$\mathrm{TGG}\mathbf{ATGGTCT}$	CCAGCTGGAG	GTAGTCATTG	AGCTTTAGGA	AGGTCTGACA	TGCAGCGGCG	ATCTCAGCCT	CTGTGTACTT
1361	GACAGTGCAC	CCTTCACGGG	CCAGGGCACC	GTGAGAGTCT	GGTGTTCAAG	TACTTGTAGG	TGCAGCCTGG	GTCCCCGATG
1441	AGGATGCGAG	ATACTGGGGT	GAGCACATCT	TTGCCTTGGA	TCCTCACCAC	GTCCCGAAAC	AAGCAGCCAT	GCTTATCGAG
	AGGATGCGAG TGTGAGAAAG							
1521		GCCTCGGGCA	CCTCCTTATG	CAGCTCCTCT	GGTATGCTGC	CGGCCTCTCG	GAAAACCAGT	TTAGGGTATT
1521 1601	TGTGAGAAAG	GCCTCGGGCA	CCTCCTTATG ACCAATGAGG	CAGCTCCTCT GGGAGACAGg	GGTATGCTGC tacacttggc	cgcctctcg tatcgIn	GAAAACCAGT	TTAGGGTATT 5 kb)
1521 1601 1656	TGTGAGAAAG TCAGCTGCCA	GCCTCGGGCA CTAAAATAAA tcaattaacc	CCTCCTTATG ACCAATGAGG ctcactaaag	CAGCTCCTCT GGGAGACAGg ggagtcgact	GGTATGCTGC tacacttggc cgatcaaaaa	caggatggct	GAAAACCAGT ntron A (>25 gcatcaggac	TTAGGGTATT 5 kb) ccacagagaa
1521 1601 1656 1736	TGTGAGAAAG TCAGCTGCCA aactgggagc	GCCTCGGGCA CTAAAATAAA tcaattaacc ggggtttgtt	CCTCCTTATG ACCAATGAGG ctcactaaag ttctgatttt	CAGCTCCTCT GGGAGACAGg ggagtcgact tcatgttcta	GGTATGCTGC tacacttggc cgatcaaaaa gaaaccacca	caggatggct gggtgcactc	GAAAACCAGT ntron A (>25 gcatcaggac actgtctctg	TTAGGGTATT 5 kb) ccacagagaa tctgtctgtc
1521 1601 1656 1736 1816	TGTGAGAAAG TCAGCTGCCA aactgggagc agctaatcgg	GCCTCGGGCA CTAAAATAAA tcaattaacc ggggtttgtt tctctgcccc	CCTCCTTATG ACCAATGAGG ctcactaaag ttctgatttt tctctctctc	CAGCTCCTCT GGGAGACAGg ggagtcgact tcatgttcta tctctctctc	GGTATGCTGC tacacttggc cgatcaaaaa gaaaccacca ccctctctct	caggatggct gggtgcactc ctcctctctc	GAAAACCAGT ntron A (>25 gcatcaggac actgtctctg tcctctctct	TTAGGGTATT 5 kb) ccacagagaa tctgtctgtc ctctctctct
1521 1601 1656 1736 1816 1896	TGTGAGAAAG TCAGCTGCCA aactgggagc agctaatcgg tgtctgtctg	GCCTCGGCA CTAAAATAAA tcaattaacc ggggtttgtt tctctgcccc gggagtgtta	CCTCCTTATG ACCAATGAGG ctcactaaag ttctgatttt tctctctctc agaggtgaag	CAGCTCCTCT GGGAGACAGg ggagtcgact tcatgttcta tctctctctc gaagggaggc	GGTATGCTGC tacacttggc cgatcaaaaa gaaaccacca ccctctctct tacaggattg	caggatggct gggtgcactc ctcctctctc attatgactt	GAAAACCAGT ntron A (>25 gcatcaggac actgtctctg tcctctctct attaaactat	TTAGGGTATT 5 kb) ccacagagaa tctgtctgtc ctctctctct tggaacatcg
1521 1601 1656 1736 1816 1896 1976	TGTGAGAAAG TCAGCTGCCA aactgggagc agctaatcgg tgtctgtctg ctctctctgg	GCCTCGGCA CTAAAATAAA tcaattaacc ggggtttgtt tctctgcccc gggagtgtta cacagtcaaa	CCTCCTTATG ACCAATGAGG ctcactaaag ttctgatttt tctctctctc agaggtgaag aaataaatag	CAGCTCCTCT GGGAGACAGg ggagtcgact tcatgttcta tctctctctc gaagggaggc ctgcaacttg	GGTATGCTGC tacacttggc cgatcaaaaa gaaaccacca ccctctctct tacaggattg atgcgacaca	caggatggct gggtgcactc ctcctctctc attatgactt tctcccactt	gcatcaggac actgtctctg tcctctctct attaaactat cttgcgagtc	TTAGGGTATT 5 kb) ccacagagaa tetgtetgte ctetetetet tggaacateg ceteettget
1521 1601 1656 1736 1816 1896 1976	TGTGAGAAAG TCAGCTGCCA aactgggagc agctaatcgg tgtctgtctg ctctctctgg gctgaatcat	GCCTCGGCA CTAAAATAAA tcaattaacc ggggtttgtt tctctgcccc gggagtgtta cacagtcaaa gccacagcgc	CCTCCTTATG ACCAATGAGG ctcactaaag ttctgatttt tctctctctc agaggtgaag aaataaatag gcgtaggcgg	CAGCTCCTCT GGGAGACAGg ggagtcgact tcatgttcta tctctctctc gaagggaggc ctgcaacttg aacggttggg	description of the control of the co	caggatgget gggtgcactc ctcctctctc attatgactt tctcccactt gtcagaccag	gcatcaggac actgtctctg tcctctctct attaaactat cttgcgagtc caagcagtga	TTAGGGTATT 5 kb) ccacagagaa tctgtctgtc ctctctctct tggaacatcg cctccttgct ctgacgggct
1521 1601 1656 1736 1816 1896 1976 2056 2136	TGTGAGAAAG TCAGCTGCCA aactgggagc agctaatcgg tgtctgtctg ctctctctgg gctgaatcat cccgcagagc	GCCTCGGCA CTAAAATAAA tcaattaacc ggggtttgtt tctctgcccc gggagtgtta cacagtcaaa gccacagcgc tagaagagac	CCTCCTTATG ACCAATGAGG ctcactaaag ttctgatttt tctctctctc agaggtgaag aaataaatag gcgtaggcgg tgcatggccc	CAGCTCCTCT GGGAGACAGg ggagtcgact tcatgttcta tctctctctc gaagggaggc ctgcaacttg aacggttggg ggggcctcca cacctgcagg	cgatcaaaaa gaaaccacca ccctctctct tacaggattg atgcgacaca ctggggctgg gttgggagag	cagcatgate caggatgacte gggtgcacte ctcctctcte attatgactt tctcccactt gtcagaccag cggtggagce	gcatcaggac actgtctctg tcctctctct attaaactat cttgcgagtc caagcagtga cgcggagcct	TTAGGGTATT 5 kb) ccacagagaa tctgtctgtc ctctctctct tggaacatcg cctccttgct ctgacgggct gctggctgca
1521 1601 1656 1736 1816 1896 1976 2056 2136	TGTGAGAAAG TCAGCTGCCA aactgggagc agctaatcgg tgtctgtctg ctctctctgg gctgaatcat cccgcagagc aaattaaagc	GCCTCGGCA CTAAAATAAA tcaattaacc ggggtttgtt tctctgcccc gggagtgtta cacagtcaaa gccacagcgc tagaagagac ccaggtgccc	CCTCCTTATG ACCAATGAGG ctcactaaag ttctgatttt tctctctctc agaggtgaag aaataaatag gcgtaggcgg tgcatggccc cgcgcatgca GF	CAGCTCCTCT GGGAGACAG ggagtcgact tcatgttcta tctctctctc gaagggaggc ctgcaacttg aacggttggg ggggcctcca cacctgcagg	cgatcaaaaa gaaaccacca ccctctctct tacaggattg atgcgacaca ctggggctgg gttgggagag cgcctgcacg	caggatggct caggatgcactc gggtgcactc ctcctctctc attatgactt tctcccactt gtcagaccag cggtggagcc ggagacacgc	gaaaaccagac actgtctctg tcctctctct attaaactat cttgcgagtc caagcagtga cgcggagcct ttgctgctgt	TTAGGGTATT 5 kb) ccacagagaa tctgtctgtc ctctctctct tggaacatcg cctccttgct ctgacgggct gctggctgca cacccctttt
1521 1601 1656 1736 1816 1896 1976 2056 2136 2216	TGTGAGAAAG TCAGCTGCCA aactgggagc agctaatcgg tgtctgtctg ctctctctgg gctgaatcat cccgcagagc aaattaaagc ggaactgaag	GCCTCGGCA CTAAAATAAA tcaattaacc ggggtttgtt tctctgcccc gggagtgtta cacagtcaaa gccacagcgc tagaagagac ccaggtgccc ttagcctgac	CCTCCTTATG ACCAATGAGG ctcactaaag ttctgatttt tctctctctc agaggtgaag aaataaatag gcgtaggcgg tgcatggccc cgcgcatgca GF tctgcaacct	CAGCTCCTCT GGGAGACAG ggagtcgact tcatgttcta tctctctctc gaagggaggc ctgcaacttg aacggttggg ggggcctcca cacctgcagg E gtgccctgtt	cgatcaaaaa gaaaccacca ccctctctct tacaggattg atgcgacaca ctggggctgg gttgggagag cgcctgcacg cgcccggactg	cagcatgct caggatgcactc gggtgcactc ctcctctctc attatgactt tctcccactt gtcagaccag cggtggagcc ggagacacgc gaccgtggta	gcatcaggac actgtctctg tcctctctct attaaactat cttgcgagtc caagcagtga cgcggagcct ttgctgctgt ggcacggctt	TTAGGGTATT 5 kb) ccacagagaa tctgtctgtc ctctctctct tggaacatcg cctccttgct ctgacgggct gctggctgca cacccctttt tgcaaacttg
1521 1601 1656 1736 1816 1896 1976 2056 2136 2216 2296	TGTGAGAAAG TCAGCTGCCA aactgggagc agctaatcgg tgtctgtctg ctctctctgg gctgaatcat cccgcagagc aaattaaagc ggaactgaag tcctctgcc	GCCTCGGCA CTAAAATAAA tcaattaacc ggggtttgtt tctctgcccc gggagtgtta cacagtcaaa gccacagcgc tagaagagac ccaggtgccc ttagcctgac ttagcctgac ttaaaggtcc	CCTCCTTATG ACCAATGAGG ctcactaaag ttctgatttt tctctctctc agaggtgaag aaataaatag gcgtaggcgg tgcatggccc cgcgcatgca cgcgcatgca cgcgcatgca cggagctagt cggagctagt	CAGCTCCTCT GGGAGACAG ggagtcgact tcatgttcta tctctctctc gaagggaggc ctgcaacttg aacggttggg ggggcctcca cacctgcagg E gtgccctgtt ctaggtcacc	cgatcaaaaa gaaaccacca ccctctctct tacaggattg atgcgacaca ctggggctgg gttgggagag cgcctgcacg cgcccggactg	cagcatgate attatgactt attatgactt teteceaett gteagaceag eggtggagee ggagacaege gaccatggta caccagetge	gaalcaggac actgtctctg tcctctctct attaaactat cttgcgagtc caagcagtga cgcggagcct ttgctgctgt ggcacggctt ctggaatttg	TTAGGGTATT 5 kb) ccacagagaa tctgtctgtc ctctctctct tggaacatcg cctccttgct ctgacgggct gctggctgca cacccctttt tgcaaacttg ttggataaaa
1521 1601 1656 1736 1816 1896 1976 2056 2136 2216 2296 2376 2456	TGTGAGAAAG TCAGCTGCCA aactgggagc agctaatcgg tgtctgtctg ctctctctgg gctgaatcat cccgcagagc aaattaaagc ggaactgaag tcctctgccc agaacacgtc	GCCTCGGCA CTAAAATAAA tcaattaacc ggggtttgtt tctctgcccc gggagtgtta cacagtcaaa gccacagcgc tagaagagac ccaggtgccc ttagcctgac ttaaacgtcc ttaaacgtcc tcactgaaat	CCTCCTTATG ACCAATGAGG ctcactaaag ttctgatttt tctctctctc agaggtgaag aaataaatag gcgtaggcgg tgcatggccc cgcgcatgca GE tctgcaacct cggagctagt cggaagaaa	CAGCTCCTCT GGGAGACAGG ggagtcgact tcatgttcta tctctctctc gaagggaggc ctgcaacttg aacggttggg ggggcctcca cacctgcagg E gtgccctgtt ctaggtcacc gccacgcgcg	cgatcaaaaa gaaaccacca ccctctctct tacaggattg atgcgacaca ctggggctgg gttgggagag cgcctgcacg caccggactg caagtcgccg	cagcatcactc cagcatcactc attatgactt teteceaett gteagaceag eggagacacge gaccatcagetgaccag accagetgaccagetgacagetgacagetgacagetgacagetgacagetgacagetgacagetgacagetgacagetgacagagagactgt	GAAAACCAGT ntron A (>25 gcatcaggac actgtctctg tcctctctct attaaactat cttgcgagtc caagcagtga cgcggagcct ttgctgctgt ggcacggctt ctggaatttg ggagacactg	TTAGGGTATT
1521 1601 1656 1736 1816 1896 1976 2056 2136 2216 2296 2376 2456 2536	TGTGAGAAAG TCAGCTGCCA aactgggagc agctaatcgg tgtctgtctg ctctctctgg gctgaatcat cccgcagagc aaattaaagc ggaactgaag tcctctgccc agaacacgtc ccagaccgag	GCCTCGGCA CTAAAATAAA tcaattaacc ggggtttgtt tctctgcccc gggagtgtta cacagtcaaa gccacagcgc tagaagagac ccaggtgccc ttagcctgac ttaaaggtcc ttaaaggtcc tcactgaaat agcatcaagg	CCTCCTTATG ACCAATGAGG ctcactaaag ttctgatttt tctctctctc agaggtgaag aaataaatag gcgtaggcgg tgcatggccc cgcgcatgca cggagctagt cggaactagt cggaaagaaa tctgctgtct CRE	CAGCTCCTCT GGGAGACAG ggagtcgact tcatgttcta tctctctctc gaagggaggc ctgcaacttg aacggttggg ggggcctcca cacctgcagg E gtgccctgtt ctaggtcacc gccacgcgcg tatcctggcc	cgatcaaaaa gaaaccacca ccctctctct tacaggattg atgcgacaca ctggggctgg gttgggagag cgcctgcacg caccggactg caagtcgccg ccctcatagg acccgcaggt	cagcatcactc caggatgactc ctcctctctc attatgactt tctcccactt gtcagaccag cggtggagcc gaacagtgaa caccagctgc gaggagtcgt tttctccact	gaalcaggac actgtctctg tcctctctct attaaactat cttgcgagtc caagcagtga cgcggagcct ttgctgctgt ggcacggctt ctggaalttg ggagacactg ctgctcaccg	TTAGGGTATT 5 kb) ccacagagaa tctgtctgtc ctctctctct tggaacatcg cctccttgct ctgacgggct gctggctgca cacccctttt tgcaaacttg ttgggtaaaa acatcgcttc gtccaacagg
1521 1601 1656 1736 1816 1896 1976 2056 2136 2216 2296 2376 2456 2536 2616	TGTGAGAAAG TCAGCTGCCA aactgggagc agctaatcgg tgtctgtctg ctctctctgg gctgaatcat cccgcagagc aaattaaagc ggaactgaag tcctctgcc agaacacgtc ccagaccgag accatccaac	GCCTCGGCA CTAAAATAAA tcaattaacc ggggtttgtt tctctgcccc gggagtgtta cacagtcaaa gccacagcgc tagaagagac ccaggtgccc ttagcctgac ttaaaggtcc ttaactgaaat agcatcaagg cggagccaag	CCTCCTTATG ACCAATGAGG ctcactaaag ttctgatttt tctctctctc agaggtgaag aaataaatag gcgtaggcgg tgcatggccc cgcgcatgca GF tctgcaacct cggagctagt cggaaagaaa tctgctgtct CRE ctgccgtgac	CAGCTCCTCT GGGAGACAG ggagtcgact tcatgttcta tctctctctc gaagggaggc ctgcaacttg aacggttggg gggcctcca cacctgcagg E gtgccctgtt ctaggtcacc gccacgcgcg tatcctggcc gtcacgggcg	cgatcaaaaa gaaaccacca ccctctctct tacaggattg atgcgacaca ctggggctgg gttgggagag cgcctgcacg caccggactg caagtcgccg ccctcatagg acccgcaggt	cagcatggct caggatggct gggtgcactc ctcctctctc attatgactt tctcccactt gtcagaccag cggtggagcc ggagacacgc gaccgtggta caccagctgc gaggagtcgt tttctccact gaagtgtgcc	gaalcaggac actgtctctg tcctctctct attaaactat cttgcgagtc caagcagtga cgcggagcct ttgctgctgt ggcacggctt ctggaatttg ggagacactg ctgctcaccg caggggctgc	TTAGGGTATT 5 kb) ccacagagaa tctgtctgtc ctctctctct tggaacatcg cctccttgct ctgacgggct gctggctgca cacccctttt tgcaaacttg ttgggtaaaa acatcgcttc gtccaacagg cacccgccg ERE

Fig. 1. (Continued). Nucleotide sequences of the 5' region of the mouse sst2 gene. The sequences of exons are shown in bold, intronic sequences are shown in lower case letters. Putative binding sites for several transcription factors are also shown in bold. The dots above indicate transcription initiation sites. Both 5' splice sites follow the AG/GT rule. The 3' splice site of intron A has the classical NYAG/ but lacks the pyrimidine-rich stretch in front. The 3' splice site of intron B matches perfectly to the consensus sequence (Y)nNYAG/ ('j' is the junction point, N is A, C, G or T, and Y is C or T).

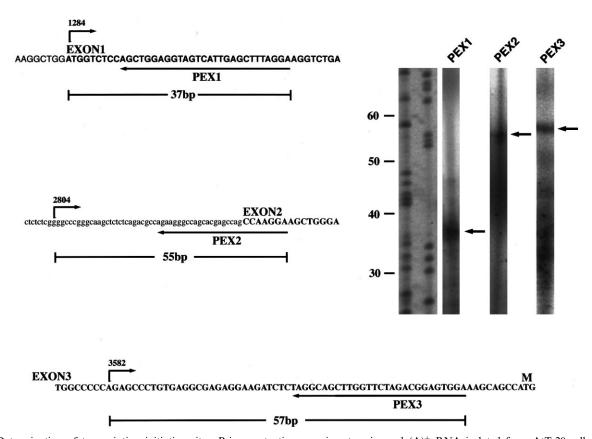


Fig. 2. Determination of transcription initiation sites. Primer extention experiments using $poly(A)^+$ RNA isolated from AtT-20 cells revealed three initiation sites which are depicted schematically along with sequences and primers used (PEX1, PEX2, PEX3). Short arrows indicate the respective extended fragments in the gel aligned to a co-electrophoresed sequencing reaction for size determination (C and G reaction of pBluescript SK $^-$ with an end-labeled M13 reverse primer).

SDS; 2 times 15 min at 68°C) were exposed to Kodak X-OMAT films

3. Results

3.1. The 5' upstream region of the mouse sst2 gene is divided by two large introns

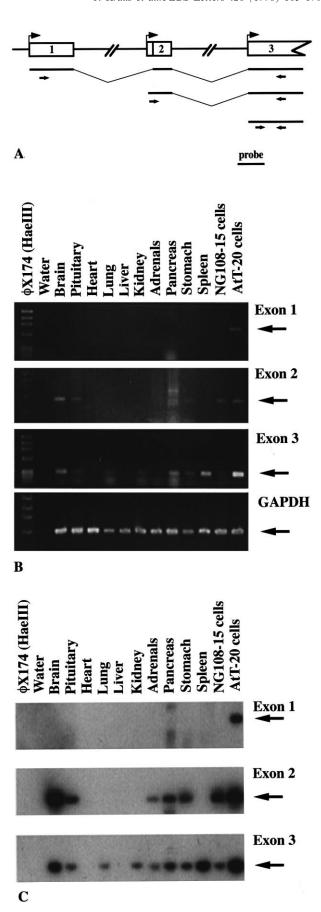
We previously reported the cloning of a cDNA encoding the mouse sst2B isoform, which has a long 5' untranslated region, containing 590 bp upstream of the first ATG codon [11]. Using fragments of this cDNA as probes genomic sequences of the upstream region were subcloned from two P1 library clones prepared from mouse 129/SvJ embryonic stem cells (Genome Systems, St. Louis, MO, USA). The nucleotide sequences (EMBL Nucleotide Sequence Database, accession numbers AJ005518, AJ005519, AJ005520) are shown in Fig. 1, Fig. 3A shows the upstream region of the gene schematically. The 5' URR of the sst2 gene comprises three exons, separated by two large introns: the first exon consists of 356 bp (nt 1284-1639). It is separated from exon 2 by at least 25 kb intronic sequences (intron A). Exon 2, which consists of 144 bp (nt 2852-2996) follows intron B, which is 25-40 kb in size. The sequences encoding the receptor protein start in exon 3 (1083 bp, nt 3557 onwards). The exon-intron junctions are similar to consensus splice sites [25]. Sequence analysis of the 5' flanking regions of the first three exons revealed several motifs similar to binding sites for transcription factors which typically occur in promoters (see Fig. 1): a classical cAMP response element (CRE) [26], several putative responsive elements for glucocorticoids and estrogen (GRE [27], ERE [28], respectively) and putative binding sites for Yin Yang 1 (YY1)[29]. However, no classical core promoter elements such as TATA or CAAT boxes are present in the sst2 gene sequences.

3.2. The sst2 gene has three distinct transcription initiation sites. To determine the transcription initiation site(s) of the mouse STR2 gene primer extension experiments were performed, which are shown in Fig. 2. The start of exon 1 is defined by an initiation site located at nt 1284. Additionally, a second initiation site is located 48 bp 5' to exon 2 (nt 2804) and a third transcription start site is located at the 26th bp of exon 3 (nt 3582). These multiple initiation sites suggest the existence of three individual promoters in the sst2 gene.

3.3. The three sst2 promoters are used cell and tissue specifically

Next, the transcriptional activity of these promoters was tested. For the RT-PCR experiments cDNAs were prepared from various mouse tissues and two cell lines. A fragment of

Fig. 3. Tissue and cell dependent usage of the sst2 gene promoters. A: Rationale of the experiments. One common primer located in exon 3 was used as the 3' downstream primer in combination with three different 5' upstream primers. Note that the specific upstream primer for promoter 2 transcripts is located within the intron, so that only transcript starting here and not spliced mRNA species starting in exon 1 are recognized. The upstream primer for the promoter 3 transcripts is located within exon 3. B: RT-PCR with cDNAs of various tissues and cells. Products of the three primer pairs specific for the sst2 promoters, P1 (810 bp), P2 (500 bp) and Pc (279 bp), and GAPDH control (605 bp) are indicated by arrows. C: Hybridization of the gels with a sst2 probe indicated in A.



the ubiquitous GAPDH gene cDNA was amplified as a control. The results of the experiments are depicted in Fig. 3. Fig. 3A shows the rationale of the experiments schematically with the location of the different primers used. The result of a typical experiment is shown in Fig. 3B. The specificity of the PCR products was demonstrated by hybridization of the bands with a radioactively labeled probe specific for exon 3 sequences (Fig. 3C). The experiments show that the three promoters are used in a tissue and cell specific manner: a PCR product of the first promoter was only found in cDNA generated from mouse pituitary AtT-20 cells. Specific PCR products from the second promoter located in front of exon 2 were detected in tissues of brain, pituitary, adrenals, pancreas and stomach, in NG 108-15 neuroblastoma x glioma hybrid cells and AtT-20 cells. In addition to the cells and tissues which contain P2 promoter transcripts, cDNA species starting exclusively from the third promoter were found in lung, kidney and spleen. No sst2 transcripts were detected in heart and liver. To demonstrate that these products were not amplified merely from incomplete spliced transcripts we performed additional RT-PCR reactions with primers located 5' of the three determined transcription initiation sites (nt 1226, 2756 and 3508). As expected, use of these upstream primers together with the common 3' primer did not result in PCR products (data not shown).

4. Discussion

Our results provide a more complete picture of the structure of the sst2 gene. The gene covers a range of at least 50 kb on the genome. The 5' URR comprises two exons and two long introns. The N-terminal part of the receptor protein including all of the transmembrane domains is coded by exon 3. A fourth short exon is critical for the generation of the two receptor variants sst2A and sst2B as reported earlier [11]. Interestingly, the gene contains three transcription initiation sites, arranged in such a way that transcripts may start at the 5' border of each of the first three exons. Consistent with the assumption that the gene then contains three promoters, a number of typical promoter elements are present within the sequences preceding the initiation sites: three putative binding sites for factor YY1 are present within the sequences in front of exon 1. This transcription factor is currently recognized as an important regulator protein whose name reflects its ambivalent function to serve in many cases as a repressor of the transcriptional machinery but also as a transcriptional activator under certain conditions [29]. The presence of such binding sites may explain why the first promoter is silent in most tissues. As shown earlier sst2 mRNA levels are also regulated by glucocorticoids and estradiol [30,31]. Potential response elements for such regulation are present on the first and second promoters. The second promoter is active in tissues in which somatostatin has essential physiological importance such as brain, pituitary and pancreas. This promoter also contains a classical CRE. It was observed earlier that sst2 mRNA is upregulated by agents that elevate the intracellular cAMP level [32], which may be accomplished via this element. Together, these observations suggest an important physiological role of the second promoter. Currently, the activities of all promoters in transfected cells and the putative transcription factor interactions are being investigated in our laboratory. Recently, a promoter was identified immediately upstream

of the sst2 protein coding region in the human gene. It contains an initiator element that binds a novel transcription factor and a TC-rich sequence, which was discussed as a possible transcription factor binding site [23]. The human sequences harboring these elements are highly homologous to the mouse gene. Using RNase protection the authors localized a transcriptional start site in the human gene which corresponds almost exactly to the intron B/exon 3 junction of the mouse gene. The TC-rich sequence in the human gene corresponds in the mouse gene to the pyrimidine-rich stretch typically occurring at 3' splice sites. The high homology to the murine splice junction includes the possibility that the human gene may also contain additional upstream exons, which is currently being tested by us. In the mouse gene transcripts start 26 bp within exon 3. Transcripts originating exclusively from this third promoter were found in lung, kidney and spleen (Fig. 3). Due to the start site of transcription of the third promoter within the exon the upstream primer P3 will always also recognize spliced mRNA species derived from exons 1 and 2. Thus, the question whether the third promoter is additionally active in those cells and tissues in which one of the other promoters is used or only in lung, kidney and spleen cannot be answered by RT-PCR and requires further investigation, e.g. the use of promoter-transgenic animals. Another novel feature is the observation of sst2 expression in lung. Previous studies performed with rat tissue showed only expression of sst4 in lung [19]. On the other hand, sst2 has been repeatedly found to be expressed in human lung carcinoma cells [33,34].

In the last few years the use of multiple promoters has been shown to be a frequently used mechanism creating diversity and flexibility in gene expression (for a review see [35]). Alternative transcripts of one gene can be generated that either are differentially regulated in different tissues [36] or developmental stages [37], or respond in different ways to various stimuli [38]. We have shown for the first time that a somatostatin receptor gene belongs to this group of genes featuring a complex 5' untranslated region and tissue dependent alternative promoter usage. Further investigations will reveal whether the other sst types have similar genomic structures.

References

- [1] Patel, Y.C., Greenwood, M.T., Panetta, R., Demchyshyn, L., Niznik, H. and Srikant, C.B. (1995) Life Sci. 57, 1249–1265.
- [2] Yamada, Y., Post, S.R., Wang, K., Tager, H.S., Bell, G.I. and Seino, S. (1992) Proc. Natl. Acad. Sci. USA 89, 251–255.
- [3] Kluxen, F., Bruns, C. and Lubbert, H. (1992) Proc. Natl. Acad. Sci. USA 89, 4618–4622.
- [4] Li, X., Forte, M., North, R.A., Ross, C.A. and Snyder, S.H. (1992) J. Biol. Chem. 267, 21307–21312.
- [5] Yasuda, K., Domiano, S., Breder, C.D., Law, S.F., Saper, C.B., Reisine, T. and Bell, G.I. (1992) J. Biol. Chem. 267, 20422– 20428.
- [6] Meyerhof, W., Wulfsen, I., Schonrock, C., Fehr, S. and Richter, D. (1992) Proc. Natl. Acad. Sci. USA 89, 10267–10271.
- [7] Yamada, Y., Reisine, T., Law, S.F., Ihara, Y., Kubota, A., Kagimoto, S., Seino, M., Seino, Y., Bell, G.I. and Seino, S. (1992) Mol. Endocrinol. 6, 2136–2142.
- [8] Bruno, J.F., Xu, Y., Song, J. and Berelowitz, M. (1992) Proc. Natl. Acad. Sci. USA 89, 11151–11155.
- [9] Demchyshyn, L.L., Srikant, C.B., Sunahara, R.K., Kent, G., Seeman, P., Van Tol, H.H., Panetta, R., Patel, Y.C. and Niznik, H.B. (1993) Mol. Pharmacol. 43, 894–901.
- [10] O'Carroll, A.M., Lolait, S.J., Konig, M. and Mahan, L.C. (1992) Mol. Pharmacol. 42, 939–946.

- [11] Vanetti, M., Kouba, M., Wang, X., Vogt, G. and Hollt, V. (1992) FEBS Lett. 311, 290–294.
- [12] Vanetti, M., Vogt, G. and Hollt, V. (1993) FEBS Lett. 331, 260– 266.
- [13] Raynor, K., Murphy, W., Coy, D., Taylor, J., Moreau, J., Yasuda, K., Bell, G.I. and Reisine, T. (1993) Mol. Pharmacol. 43, 838–844.
- [14] Rossowski, W. and Coy, D. (1994) Biochem. Biophys. Res. Commun. 197, 366–371.
- [15] Rossowski, W., Gu, Z., Akarca, U., Jensen, R. and Coy, D.C. (1994) Peptides 15, 1421–1424.
- [16] Reisine, T., Kong, H., Raynor, K., Yano, H., Takeda, J., Yasu-da, K. and Bell, G.I. (1993) Mol. Pharmacol. 44, 1008–1015.
- [17] Buscail, L., Delesque, N., Esteve, J., Saint-Laurent, N., Prats, H., Clerc, P., Robberecht, D., Bell, G.I., Liebow, C., Schally, A.V., Vayesse, N. and Susini, C. (1994) Proc. Natl. Acad. Sci. USA 91, 2315–2319.
- [18] Kubota, A., Yamada, Y., Kagimoto, S., Shimatsu, A., Imamura, M., Tsuda, K., Imura, H., Seino, S. and Seino, Y. (1994) J. Clin. Invest. 93, 1321–1325.
- [19] Rohrer, L., Raulf, F., Bruns, C., Buettner, R., Hofstaedter, F. and Schule, R. (1993) Proc. Natl. Acad. Sci. USA 90, 4196–4200.
- [20] Corness, J.D., Demchyshyn, L.L., Seeman, P., Van Tol, H.H., Srikant, C.B., Kent, G., Patel, Y.C. and Niznik, H.B. (1993) FEBS Lett. 321, 279–284.
- [21] Panetta, R., Greenwood, M.T., Warszynska, A., Demchyshyn, L.L., Day, R., Niznik, H.B., Srikant, C.B. and Patel, Y.C. (1994) Mol. Pharmacol. 45, 417–427.
- [22] Greenwood, M.T., Robertson, L. and Patel, Y.C. (1995) Gene 159, 291–292.
- [23] Pscherer, A., Dörflinger, U., Kirfel, J., Gawlas, K., Rüschoff, J., Buettner, R. and Schüle, R. (1996) EMBO J. 15, 6680–6690.

- [24] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: A Laboratorial Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- [25] Padgett, R.A., Grabowski, P.J., Konarska, M.M., Seiler, S. and Sharp, P.A. (1986) Annu. Rev. Biochem. 55, 1119–1150.
- [26] Comb, M., Birnberg, N.C., Seasholtz, A., Herbert, E. and Goodman, H.M. (1986) Nature 323, 353–356.
- [27] Scheidereit, C., Geisse, S., Westphal, H.M. and Beato, M. (1983) Nature 304, 749–752.
- [28] Walker, P., Germond, J.E., Brown Luedi, M., Givel, F. and Wahli, W. (1984) Nucleic Acids Res. 12, 8611–8626.
- [29] Shi, Y., Lee, J.S. and Galvin, K.M. (1997) Biochim. Biophys. Acta 1332, F49–F66.
- [30] Xu, Y., Berelowitz, M. and Bruno, J.F. (1995) Endocrinology 136, 5070–5075.
- [31] Kimura, N., Hatafuji, C., Konagaya, H. and Takahashi, K. (1986) Endocrinology 119, 1028–1036.
- [32] Patel, Y.C., Greenwood, M., Kent, G., Panetta, R. and Srikant, C.B. (1993) Biochem. Biophys. Res. Commun. 192, 288–294.
- [33] Zhang, C.J., Yokogoshi, Y., Yoshimoto, K., Fujinaka, Y., Matsumoto, K. and Saito, S. (1995) Biochem. Biophys. Res. Commun. 210, 805–815.
- [34] Fujita, T., Yamaji, Y., Sato, M., Murao, K. and Takahara, J. (1994) Life Sci. 55, 1797–1806.
- [35] Ayoubi, T.A.Y. and van de Ven, W.J.M. (1996) FASEB J. 10, 453–460.
- [36] Nishio, H., Takeshima, Y., Narita, N., Yanagawa, H., Suzuki, Y., Ishikawa, Y., Minami, R., Nakamura, H. and Matsuo, M. (1994) J. Clin. Invest. 94, 1037–1042.
- [37] Teerink, H., Kasperaitis, M.A.M., De Moor, C.H., Voorma, H.O. and Thomas, A.A.M. (1994) Biochem. J. 303, 547–553.
- [38] Iynedjian, P.B. (1983) Biochem. J. 293, 1-13.