Genomic Organization of the Mouse and Human Genes Encoding the ATP Sulfurylase/Adenosine 5'-Phosphosulfate Kinase Isoform SK2*

(Received for publication, July 26, 1999, and in revised form, September 3, 1999)

Kiyoto Kurima‡, Bhawani Singh§, and Nancy B. Schwartz¶

From the Departments of ‡Pediatrics, §Chemistry, and ¶Biochemistry and Molecular Biology, University of Chicago, Chicago, Illinois 60637

Mammalian ATP sulfurylase/adenosine 5'-phosphosulfate (APS) kinase consists of kinase and sulfurylase domains, and catalyzes two sequential reactions to synthesize the universal sulfate donor, phosphoadenosine phosphosulfate (PAPS). In simpler organisms, the ATP sulfurylase and APS kinase reactions are catalyzed by separate enzymes encoded by two or three genes, suggesting that a fusion of separate genes during the course of evolution generated the bifunctional enzyme. We have characterized the genomic structure of the PAPS synthetase SK2 isoform genes for mouse (MSK2) and human (HSK2) and analyzed the possible fusion region. The MSK2 and HSK2 genes exhibit a common structure of 13 exons, including a 15-nucleotide alternatively spliced exon 8. Enzyme activities of several bacterially expressed exon assemblages showed exons 1–6 encode APS kinase, while exons 6–13 encode ATP sulfurylase. The MSK2 construct without the exon 6-encoded peptide showed no kinase or sulfurylase activity, demonstrating that exon 6 encodes sequences required for both activities. Exon 1 and its 5'-flanking sequence are highly divergent between the two species, and intron 1 of the HSK2 gene contains a region similar to the MSK2 promoter sequence, suggesting that it may be the remnant of a now-superseded regulatory region. The HSK2 promoter contains a GC-rich region, not present in the mouse promoter, and has fewer transcription factor binding sites in common with MSK2. These differences in the two promoter regions suggest that species-specific mechanisms regulate expression of the SK2 isoform.

Biosynthesis of the universal sulfate donor, phosphoadenosine phosphosulfate (PAPS), requires two sequential enzymatic reactions: the transfer of sulfate to ATP to form APS catalyzed by ATP sulfurylase (EC 2.7.7.4), followed by transfer of phosphate to APS yielding PAPS catalyzed by APS kinase (EC 2.7.1.25) (1). The ATP sulfurylase and APS kinase reactions are catalyzed by separate proteins in organisms such as bacteria (2–4), fungi (5, 6), yeast (7, 8), and plants (9–11). In contrast, the catalytic units for the ATP sulfurylase and APS kinase reactions are fused in animals; cDNAs encoding a single bifunctional protein have been isolated in the marine worm Urechis caupo (12), in Drosophila melanogaster (13), and in several mammalian species (14–17). Recombinant individual ATP sulfurylase or APS kinase domains of a mouse PAPS synthetase retain the ability to catalyze the respective reactions (18), implying that the ancestral PAPS synthetase gene was formed by fusion of separate genes encoding the independent functional units.

Variations in the structural organization of monofunctional and bifunctional PAPS synthetase genes in simpler organisms and animals, respectively, likely reflect differing histories of gene fusion/duplication events in evolution. For instance, the sulfate activation operon in Escherichia coli encodes ATP sulfurylase (cysD and cysN), followed by APS kinase (cysC) (3). In a symbiotic nitrogen-fixing bacterium, Rhizobium meliloti, there are only two genes; nodP is homologous to cysD, and encodes a subunit of ATP sulfurylase, while nodQ is homologous to cysN and cysC, encoding both a subunit of ATP sulfurylase and APS kinase (4). In a filamentous fungus, Penicillium chrysogenum, the ATP sulfurylase gene encodes the ATP sulfurylase domain plus a nonfunctional APS kinase-like domain at the C terminus (5), reminiscent of the cysN-cysC fusion in R. meliloti nodQ. A functional APS kinase is encoded by a separate gene in P. chrysogenum as well. In plants and yeast, the functional catalytic units are encoded by two separate genes with no combination of domains. In the fused PAPS synthetase structures, the structural orientation of the two functional domains is the reverse of the nodQ arrangement with APS kinase N-terminal of the ATP sulfurylase, suggesting an unrelated origin of gene fusion. Our previous activity and stability studies with rearranged recombinant enzymes showed that reversing the order of the domain organization resulted in either diminished activity or a thermally unstable enzyme (18). Such a structural or functional advantage could be a possible selective factor for gene fusion in a specific order. Thus far, the complete genomic structure of the fused PAPS synthetase genes has not been determined; further, this additional information is needed to gain insights into the mechanism of the gene fusion that brought two separate genes into a single gene encoding a bifunctional PAPS synthetase.

The PAPS synthesis pathway plays an important role in normal cartilage and skeletal development in mammals as demonstrated by the recent identification of PAPS synthetase isoforms (14, 16) and of mutations in one member of the PAPS

* This work was supported by Grants HD-17332 and AR-19622 from the National Institutes of Health, a grant from the Mizutani Foundation (to N. B. S.), and a Markey fellowship (to B. S.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

** The abbreviations used are: PAPS, phosphoadenosine phosphosulfate; APS, adenosine 5'-phosphosulfate; nt, nucleotide(s); kb, kilobase pair(s); bp, base pair(s); PCR, polymerase chain reaction.
synthesizes family, SK2, in murine brachymorphism and human spondyloepimetaphyseal dysplasia (16, 17). Brachymorphism is characterized by a dome-shaped skull, short thick tail, and shortened but not widened limbs (19), and is associated with limited PAPS availability due to a reduction of PAPS synthetase activity (20–22). Interestingly, severely reduced PAPS synthetase activity was found in brachymorphic cartilage and liver, while no reduction was observed in brachymorphic brain and skin (23). Spondyloepimetaphyseal dysplasia (Pakistani type) is characterized by short and bowed lower limbs, enlarged knee joints, and early onset of degenerative joint disease in the hands and knees (24). The presence of two PAPS synthetase genes and the tissue-specific defects in PAPS synthetase in mammalian mutants implicate coordinated mechanisms that control the expression of the PAPS synthetase genes.

To gain better insights into the origin and regulation of the PAPS synthetase gene, we determined the mouse and human genomic structures of one PAPS synthetase gene family member, SK2, whose importance in development is highlighted by both human and murine growth disorders due to mutations in this gene. Recombinant monofunctional ATP sulfurylase and APS kinase proteins and a fused protein with an internal deletion were produced to determine the distribution of exons for each catalytic domain. We also analyzed the 5′-flanking region sequences of the genes for potential transcription binding sites to guide future functional expression studies.

**MATERIALS AND METHODS**

**Analysis of the Mouse and Human SK2 Genomic Structure—** A mouse C57BL/6 genomic BAC library (GenomeSystem, St. Louis, MO) was screened with a 32P-labeled mouse SK2 DNA fragment following the manufacturer’s instructions. A human genomic BAC library, release I (GenomeSystem) was screened by PCR with a set of human SK2 primers derived from a cDNA sequence (accession no. AF173365) reported in the present study. Two mouse BAC clones, BACSK2#1 and BACSK2#5, and one human BAC clone, BAChSK2, were isolated and various plasmid subclones were prepared. Direct sequencing and PCR amplification with cDNA-derived primers were performed to determine the genomic structure, comprising exon-exon junctions and intron sizes. Search of publicly available human genomic sequences identified two BAC clone, CIT-HSP-306M19 (Caltech Genome Research Laboratory) and AC006191 (Human Chromosome 10 Sequencing Group, Sanger Center), both containing exons 1 through 9 of the HSK2 and at least 8 kb from the upstream 5′-flanking region of the gene. The sequence data obtained from BACHSK2, CIT-HSP-306M19, and AC006191 were used to determine the complete human genomic structure. The TESS (Transcription Element Search Software) program was used to search the sequence for high-quality matches to a data base of position-weighted nucleotide distribution matrices.

**Alternatively Spliced Mouse and Human SK2 Open Reading Frame cDNA—** Reverse transcription-PCR was performed on total RNA extracted from 3-day-old C57BL mouse cartilage. The first strand was synthesized by SuperScript II Reverse Transcriptase (Life Technologies, Inc.) using an antisense primer (SK5, 5′-GCAATTGATACAGACGACGAC-3′) complementary to the 3′-untranslated region of SK2 mRNA. The complete open reading frame cDNA was amplified with a 5′ end sense primer containing a NdeI site (SK3, 5′-GAGAGTTTCAATGACGGAGGAT-3′) and a 3′ end antisense primer containing a XhoI site (SK1, 5′-GAGAGATCTCGAGTTAGTTCTTCTCCAGGGACCTGTAATAATCTG-3′) using Expand polymerase (Roche Molecular Biochemicals). The sequence was compared with a previously reported HSK2 cDNA (accession no. AF091242) prepared from cartilage RNA (17). Expressed Mouse SK2 Constructs and Enzyme Assays—The original SK2 construct in a pET-15b bacterial expression vector was prepared as described (16) and used as the template DNA to prepare other constructs by PCR. MSK2 is the original SK2 (accession no. AF052453) prepared from liver RNA, while MSK2#8 is an alternatively spliced SK2 prepared from cartilage RNA, lacking the 15-nt/5-amino acid insert (amino acid residues 290–294) encoded by exon 8. Two APS kinase constructs were prepared. The functional APS kinase domain construct, APSK22–9 (residues 1–227) corresponds to the mouse SK1 APS kinase domain previously shown to catalyze the APS kinase reaction (18), and encodes a protein containing the first 72 bp of exon 6. A second APS kinase construct, APSK23–9 (residues 1–214), is encoded by exons 1–5 only. Two ATP sulfurylase constructs were prepared. ATPS24–11 (residues 215–621) is an ATP sulfurylase domain protein encoded by exons 6–13, while ATPS212–19 (residues 253–621) is another ATP sulfurylase protein encoded by exons 7–13. MSK2#8 (residues 1–214 and 253–621) is a MSK2 deletion construct lacking the sequence encoded by exon 6 (residues 215–252).

**RESULTS**

**Genomic Structure of the MSK2 and HSK2 Genes—** The genomic structure of one PAPS synthetase gene family member, SK2, was determined both in mouse and human. A mouse BAC clone, BACSK2#1, contained exons 2–13, while another BAC clone, BACSK2#5, contained the mouse exon 1 and 5′-flanking regions. One human BAC clone, BACHSK2, contained the whole gene. Two additional BAC clones, CIT-HSP-306M19 (Caltech Genome Research Laboratory) and AC006191 (Human Chromosome 10 Sequencing Group, Sanger Center) con-
Fig. 1. Structure of the mouse and human SK2 genes. A, exons are represented as closed boxes and introns as the lines connecting the boxes in the gene structure. Distribution of exons within the corresponding domains of SK2 protein is shown below. A black box represents the location of the alternatively spliced exon 8 region in the ATP sulfurylase domain (stippled). Exon and intron sizes are shown in Table I. B, comparison with the C. elegans SK gene. Exons are represented as open boxes with coding sequence nucleotide number indicated, and introns are represented as the lines connecting the boxes with intron base pairs. A black box represents the unique insert in exon 4 corresponding to the exon 8 insert found in mammalian SK2. Striped boxes indicate the interdomain regions in both molecules.

The MSK2 and HSK2 genes have identical exon organizations, consisting of 13 exons (Fig. 1A) with a 15-nt alternatively spliced exon 8 (Fig. 2; see below). The minimum sizes of the genes are 40 and 85 kb for the MSK2 and HSK2 genes, respectively. Nucleotide sequences of the intron/exon boundaries show that all introns in the MSK2 and HSK2 genes begin with a 5′-GT dinucleotide and conclude with 3′-AG termini (Table I). The size of exons and phases of splicing are also conserved between the species, with the exception of exon 1. The deduced amino acid sequences at the N termini encoded by exon 1, as well as the nucleotide sequences of the 5′-flanking regions, are significantly different between MSK2 and HSK2 (Fig. 4). By scanning the intron 1 sequence of the HSK2 gene, we identified a 800-bp sequence 25 kb downstream from exon 1, which is similar to the MSK2 5′-flanking region sequence (68% identity). However, this region does not appear to encode a peptide corresponding to the MSK2 N-terminal sequence, or to contain transcription factor binding motifs as found in the MSK2 promoter region, and thus may be the remnant of a former promoter.

Comparison of the exon structures to the functional domains of PAPS synthetases reveals that the APS kinase domain is encoded by exon 1 through exon 6, base 39, and the ATP sulfurylase domain is encoded by exon 6, base 40, through exon 13. The intervening sequence present in the functional APS kinase domain is encoded by exon 5, base 48 through exon 6, base 39. An alternatively spliced 15-nt sequence comprises one exon, designated as exon 8 (Fig. 2).

Exon distribution of several functional motifs was also analyzed. The GXXXXG motif essential for ATP binding (P-loop) (29) is found in the APS-kinase domain (30). The first base of the G50 codon is encoded by exon 2, and the rest of the P-loop codons reside in exon 3, forming a phase 1 splicing. Our previous study on mutational analysis of the P-loop in mouse SK1 showed that a mutation of glycine 50 to alanine did not affect any of the enzyme activities (30), indicating that the functionally essential residues of the P-loop are solely encoded by exon 3. Another functional unit, a PAPS-dependent enzyme motif, KAXAXXXFXTG, in mouse SK2 is found as 166KRARA-GEIKG175FTG176 in the kinase portion. The first base of the G175 codon is encoded by exon 4, and the rest of the motif is encoded by exon 5, forming a phase 1 splicing. The FISP motif (31), HXXH motif (phosphodiester cleavage) (32) and PP-loop (IVGRDPAG, pyrophosphate binding) (33) motif are each encoded entirely by exons 3, 11, and 12, respectively. Functionally essential residues in each of these form motifs have been identified.2

Alternatively Spliced SK2 Form Generated Mouse and Human—In previous studies of mouse and human SK2 cDNAs (16, 17), the deduced amino acid sequences showed high sequence identity (93%). However, HSK2 lacks five amino acids in the ATP-sulfurylase domain, and the first 9 or 11 N-terminal residues are dissimilar between MSK2 and HSK2. Amplification of MSK2 partial cDNA generated from cartilage RNA identified a MSK2 cDNA variant, referred to as MSK2Δ8 (Fig. 2). This cartilage cDNA lacks 15 nt at positions 869–883, predicting a deletion of the five amino acids, GVVP, from residues 290–294, in the ATP sulfurylase domain. Analysis of the genomic sequence shows that this 5-amino acid segment is encoded by an exon with typical flanking donor and acceptor sequences (Table I). PCR analysis of MSK2 gene expression showed that MSK2Δ8 is the major form in cartilage, while the original MSK2 is the major form in liver (data not shown). We have prepared liver HSK2 cDNA, and the sequence was compared with HSK2 cDNA isolated from a cartilage cDNA library (accession no. AF091242). The liver HSK2 cDNA is equivalent to MSK2 and contains the 15-nt exon 8 sequence, while the cartilage HSK2 protein is equivalent to MSK2Δ8, lacking the 5-amino acid sequence, GMALP from residue 289 to 293, encoded by exon 8. Both MSK2 and MSK2Δ8 proteins were bacterially expressed, and the ability of these proteins to catalyze the APS kinase, ATP sulfurylase, and overall PAPS synthetase reactions in vitro was compared. No significant difference between the two proteins was observed in any of the three assays (see below).

Recombinant Enzyme Assays—In order to determine the distribution of exons encoding each enzyme catalytic unit, several expression constructs with different exon compositions were prepared, and the expressed proteins subjected to enzyme assays (Fig. 3). Three assays were used to assess expression levels for the recombinants as well as the functionality of each mutant: the reverse sulfurylase, the forward kinase, and the overall reaction, which measures APS and PAPS simultaneously. The full-length proteins MSK2 and MSK2Δ8 were soluble proteins, and after purification and dialysis were assayed for the three activities (see “Materials and Methods” for

2 Deyrup, A. T., Singh, B., Krishnan, S., Lyle, S., and Schwartz, N. B. (1999) J. Biol. Chem. 274, 28929–28936.
The intact sulfurylase/kinase (MSK2 1–13) sequence encoded by exon 6 is required for the APS kinase activity. The ATP sulfurylase construct ATPS26–13 was soluble in IMAC5 buffer, but ATPS2 6–13 was also sub-

The results of the enzyme preparations was restored. The MSK2 Δ construct was designed in order to test whether additional sequences at the C terminus of APSK21–5 and N terminus of ATPS27–13 would increase enzyme solubility and restore the kinase and sulfurylase activities, respectively. MSK2 Δ was insoluble, and following urea extraction and dialysis into phosphate buffer, the recombinant protein was still unable to catalyze ATP sulfurylase or APS kinase reactions. This demonstrates a requirement for exon 6 sequences similar to that observed for the separate domain constructs.

**Figure 3. Schematic diagram of bacterially expressed mouse SK2 constructs and enzyme activities.** The APS kinase domain is depicted as a black box, the ATP sulfurylase domain as a stippled box, and the 5-amino acid insert as a white box. The striped boxes represent the interdomain region. The intact sulfurylase/kinase (MSK2Δ–13) is encoded by exons 1–13, while MSK2Δ lacks the 5 amino acid residues encoded by exon 8. APSK21–5 includes the APS-kinase domain and the first 13 amino acids encoded by exon 6, while APSK2Δ–5 is an APS kinase domain encoded by exon 1–5 only. ATPS27–13 is an ATP sulfurylase domain encoded by exon 6–13, while ATPS2Δ–13 lacks the region encoded by exon 6. MSK2Δ is MSK2 without the exon 6-encoded region. The results of in vitro enzyme assays for each construct are shown at the left. Y represents the presence of the activity, whereas an asterisk (*) indicates the lack of the activity. An asterisk (*) represents expressed protein that failed to be released into the reaction buffer and had to be solubilized as described. MSK2Δ exhibited normal sulfurylase (2.35 μmol of ATP/min·mg), kinase (227.9 pmol of PAPS/min·mg), and overall (13.6 nmol of PAPS/min·mg) activities. Alternatively spliced isoforms MSK2Δ–13 and ATPS2Δ–5 and APSK2Δ–5 were both soluble in IMAC5 buffer, but APSK2Δ–5 exhibited decreased binding to His·Bind resin. APSK2Δ–5 also failed to exhibit kinase activity, whereas an APS kinase construct with C-terminal amino acids encoded by exon 6 (APSK2Δ–5·5) showed APS kinase activity comparable to MSK2. This result shows that the first 13 residues of the linker sequence encoded by exon 6 are required for the APS kinase activity. The ATP sulfurylase construct ATPS2 Δ–13 was soluble, whereas ATPS27–13 was found in the bacterial pellet and had to be solubilized with 6 M urea followed by dialysis into phosphate buffer. When tested for sulfurylase activity, ATPS26–13 showed normal sulfurylase activity while ATPS27–13 was inactive. To ensure that the urea treatment did not cause irreversible denaturation, ATPS2Δ–13 was also subjected to urea treatment and reequilibration. Urea-solubilized ATPS2Δ–13 was initially inactive, but after dialysis into phosphate buffer, activity comparable to the non-urea-treated enzyme preparations was restored. The MSK2 Δ construct was designed in order to test whether additional sequences at the C terminus of APSK21–5 and N terminus of ATPS27–13 would increase enzyme solubility and restore the kinase and sulfurylase activities, respectively. MSK2 Δ was insoluble, and following urea extraction and dialysis into phosphate buffer, the recombinant protein was still unable to catalyze ATP sulfurylase or APS kinase reactions. This demonstrates a requirement for exon 6 sequences similar to that observed for the separate domain constructs.

**Table I**

| Exon Size (bp) | 3’ splice acceptor | 5’ splice donor (kb) | Type | Intron size (kb) |
|---------------|-------------------|-------------------|------|-----------------|
| **Mouse SK2 gene** | | | | |
| 1 | 33 (coding sequence) | | | |
| 2 | 115 | gttattttagAGACGAGCA | GCTAAGAgttagctaaacc | O | >5 |
| 3 | 236 | atctctagAGTCTGCTCT | TGGCAAAGtagttttat | O | 3.2 |
| 4 | 139 | gtatctctagAGATGCTG | GATAAAGAgttattttat | O | 1.2 |
| 5 | 119 | atctctctagGAGTTTACA | AGAGCAAGtagttttatg | O | 0.5 |
| 6 | 114 | tctctctagTGGTTTACA | TTCAACAGtagttttatg | O | 0.2 |
| 7 | 112 | gtgtgcttagCTGTAATC | ACTGGACGtagttttatg | O | 2.2 |
| 8 | 15 | atctctctagGGTGGITT | TCCGGCGtagttttatg | O | 5.5 |
| 9 | 206 | tctctctagAGATGCGTG | ATACAAAGtagttttatg | O | 1.6 |
| 10 | 136 | tctctctagAGATGCGTG | GATGGCGTAtagttttatg | O | 6.3 |
| 11 | 269 | tctctctagATGGCGTG | CACAGAGtagttttatg | O | 3.3 |
| 12 | 230 | tctctctagGTCGAGG | CCAGCAAGtagttttatg | O | 1.3 |
| 13 | 142 | tctctctagGACGACG | CCAGCAAGtagttttatg | O | 2.9 |

**Human SK2 gene**

| Exon Size (bp) | 3’ splice acceptor | 5’ splice donor (kb) | Type | Intron size (kb) |
|---------------|-------------------|-------------------|------|-----------------|
| 1 | 27 (coding sequence) | | | |
| 2 | 118 | tatttttagAGACGAGCA | GCTAAGAgttagctaaacc | O | 49 |
| 3 | 236 | gtctctagAGTCTGCTCT | TGGCAAAGtagttttat | O | 3.8 |
| 4 | 139 | gtatctctagAGATGCTG | GATAAAGAgttattttat | O | 0.7 |
| 5 | 119 | atctctctagGAGTTTACA | AGAGCAAGtagttttatg | O | 0.6 |
| 6 | 114 | tctctctagTGGTTTACA | TTCAACAGtagttttatg | O | 0.1 |
| 7 | 112 | gtgtgcttagCTGTAATC | ACTGGACGtagttttatg | O | 0.2 |
| 8 | 15 | atctctctagGGTGGITT | TCCGGCGtagttttatg | O | 5.9 |
| 9 | 206 | tctctctagAGATGCGTG | ATACAAAGtagttttatg | O | 2.8 |
| 10 | 136 | tctctctagAGATGCGTG | GATGGCGTAtagttttatg | O | 1.9 |
| 11 | 269 | tctctctagAGATGCGTG | CACAGAGtagttttatg | O | 1.7 |
| 12 | 230 | tctctctagGTCGAGG | CCAGCAAGtagttttatg | O | 1.0 |
| 13 | 139 | cctcttagAGACGACG | CCAGCAAGtagttttatg | O | 2.0 |
the HSK2 include NF-E2, LyF-1, MEF-2, Pit-1a, AP-1, TFE3-S, USF, CAC-binding protein, NFAT-1, and CCAAT-binding protein. Only a few detected motifs (Pit-1a, CCAAT-binding protein) were common to both MSK2 and HSK2 5'-flanking regions.

**DISCUSSION**

A number of eukaryotic multifunctional enzymes have been shown to consist of several monofunctional catalytic domains, each of which are encoded by a single gene in simpler organisms (18, 34–38). When expressed as monofunctional domains, each domain often retains its activity, suggesting that the genes encoding each catalytic unit were fused during the course of evolution (18, 39, 40). It is also common that these present-day multidomain enzymes have interdomain segments that are not homologous to any protein sequences and display no clear boundary of those domains (14, 36, 37, 41, 42). In addition, structure comparisons, as well as phylogenetic analysis of mult-

---

**Fig. 4.** Nucleotide sequence from the mouse (A) and human (B) SK2 5'-flanking regions and first exons. Translated exon 1 sequences are shown in **bold**, and the first methionine codon ATG is assigned +1. TATA-box sequences are **bold** and **italicized**. **Underlined** cis-elements delineate binding sites for the transcription factors indicated.
tidoxin proteins in different organisms, show a striking variety of gene fusion events. In the case of PAPS synthetases, the APS kinase domain is located at the N terminus and ATP sulfurylase at the C terminus linked by a short non-homologous intervening sequence, and the boundary of the two functional catalytic units has been established (18). The structural organization of genes encoding APS kinase and APS sulfurylase in E. coli and fungi suggests the simple gene fusion of adjacent genes, while the reverse domain structure in the present-day bifunctional PAPS synthetases implicates a historically unrelated gene fusion event. Elucidation and analysis of the genomic structures of the mammalian and Caenorhabditis elegans PAPS synthetase genes provide a possible mechanism for how these bifunctional enzymes have evolved.

The MSK2 and HSK2 genes are structurally very similar, consisting of 13 exons with some variation in intron sizes. Exon sizes and intron phases are also conserved between the species, with the exception of exon 1. The exact sizes of the genes have not been determined, but the minimum sizes are 45 and 80 kb for the MSK2 and HSK2 genes, respectively. We have identified two splicing variants in mouse and human SK2 CNAs, and the alternatively spliced 15 nt was identified as an exon in both genomes (Figs. 1 and 2). The SK2A8 proteins lack 5 amino acids, GVVPR in MSK2 and GMALP in HSK2 in the ATP sulfurylase domain. A similar 15-nt/5-amino acid insertion, MDGSY, is found in E. coli and fungi suggests the simple gene fusion of adjacent genes, while the reverse domain structure in the present-day bifunctional PAPS synthetases implicates a historically unrelated gene fusion event. Elucidation and analysis of the genomic structures of the mammalian and Caenorhabditis elegans PAPS synthetase genes provide a possible mechanism for how these bifunctional enzymes have evolved.

The MSK2 and HSK2 genes are structurally very similar, consisting of 13 exons with some variation in intron sizes. Exon sizes and intron phases are also conserved between the species, with the exception of exon 1. The exact sizes of the genes have not been determined, but the minimum sizes are 45 and 80 kb for the MSK2 and HSK2 genes, respectively. We have identified two splicing variants in mouse and human SK2 CNAs, and the alternatively spliced 15 nt was identified as an exon in both genomes (Figs. 1 and 2). The SK2A8 proteins lack 5 amino acids, GVVPR in MSK2 and GMALP in HSK2 in the ATP sulfurylase domain. A similar 15-nt/5-amino acid insertion, MDGSY, is found in D. melanogaster, and a larger 19–22-amino acid insertion in C. elegans PAPS synthetase at the same location. Consistent occurrence of a splice variant at this position among the different species implies some functional significance; however, no differences in the ability of MSK2 and HSK2 to catalyze ATP sulfurylase, APS kinase, or PAPS synthetase overall reactions were detected. Additionally, there is no apparent motif associated with the inserted sequences; therefore, theinsert may not influence enzyme function directly.

Relative to the exon distribution, the APS kinase domain is encoded by exons 1–6, while the ATP sulfurylase domain is encoded by exon 6–13. Based on the recombinant monofunctional constructs of the mouse SK1 protein, the first 236 amino acid residues constitute a functional APS kinase domain, while the ATP sulfurylase domain consists of residues 237–624 (18). The functional boundary between the APS kinase and ATP sulfurylase is thus located in exon 6 of the MSK2 gene. In fact, exon 6 encodes only the last 13 C-terminal amino acid residues of the APS kinase. Initially, because of the location of the functional boundary relative to the location of the exon 5/6 boundary and the phase 0 exon insertion of intron 5, we suspected that the exon 5/6 boundary might be the ancestral fusion site where the two separate genes were fused. However, the bacterially expressed protein encoded by exons 1–5, APSK2^{1–5}, failed to exhibit APS kinase activity, and APS kinase activity was restored only when an additional 13 amino acid residues encoded by exon 6 were added to the APSK2^{1–5} protein. The failure of the MSK2Δ6 protein to catalyze both APS sulfurylase and APS kinase reactions further supports the contention that the exon 5/6 boundary does not reflect the ancestral fusion site.

In fact, examination of the C. elegans PAPS synthetase gene, which consists of only six exons, with the large exon 4 encoding a part of both the APS kinase and ATP sulfurylase domains, further suggests random insertion of introns. Since there is no correspondence in the location of the introns between the C. elegans PAPS synthetase gene and the mammalian PAPS synthetase gene (Fig. 1B), it is plausible that introns were inserted randomly and independently into an intron-less ancient fused PAPS synthetase gene in different organisms. Alternatively, pre-existing introns could have been lost as well. Locations of introns are not necessarily coincident with the boundaries of catalytic units in other bi- or multifunctional enzymes (43, 44). Introns appear to be able to be located proximal to the functional boundary as seen in PAPS synthetases, or distal as seen in the UMP synthetase gene (43). Although introns appear to have been introduced randomly, it is interesting that most of the highly conserved functional motifs that this complicated bifunctional enzyme contains, including FISP, phosphodiester (HXXH) motif, and PP-loop, are all contained within single exons. An exception is Gly^{50} of the P-loop, which is encoded by exon 2 while the rest of the motif is encoded by exon 3. However, we have shown that Gly^{50} is not essential for kinase activity, and therefore may not be part of this common motif in PAPS synthetases.

Another interesting finding from the genomic sequence data for the two species is the lack of sequence similarity in exon 1 and its 5′-flanking regions. For instance, nine amino acid residues, MSGIKKQKT, are encoded by HSK2 exon 1, while 11 residues, MSANSKMKHKK, are encoded by MSK2 exon 1. The 5′-flanking region of the HSK2 gene is rich in GC sequence, characteristic of a housekeeping-type promoter with multiple Sp-1 binding sites, while the MSK2 gene contains several potential recognition sites for tissue-specific transcription factors, such as progesterone receptor, MyoD, PU.1, GC1, Pit-1a, Evi-1, TIN-1, and GRαβ. Even though the significance of these transcription factor binding motifs must await functional analysis, the highly heterologous promoter sequences of the two orthologous genes implies different origins and modes of regulation of the promoter. A stretch of 800 bp in intron 1 of the HSK2 gene is 68% identical to the proximal promoter region and exon 1 of the MSK2 gene but lacks a translatable or spliceable exon, suggesting it is a former exon/promoter, which was once an alternative or sole promoter. It is not unusual to have more than one promoter in many genes, and some promoters are preferentially used in a tissue-specific manner. We still cannot exclude possible SK2 transcript variants differing at the 5′ end sequence due to the usage of alternative promoters in either species. It is possible that the mammalian SK2 gene had two promoters, and that the HSK2 gene used the upstream promoter, while the MSK2 used the downstream promoter as the major promoter, creating dissimilar N-terminal SK2 so far identified in human and mouse. Alternatively, the ancestral mammalian gene may have had only one promoter, and an additional or new promoter was acquired by the primate line. A bacterially expressed mouse SK1 construct starting from the second methionine codon, which is equivalent to SK2 without the exon 1-encoded sequence, showed equivalent enzyme activity to the first methionine SK1 construct (30). Thus, it is likely that the sequence encoded by exon 1 is catalytically nonessential, allowing divergence of the region. The dissimilarity of the promoter sequence of the orthologous mouse and human genes also suggests different spatio-temporal expression patterns of the gene in the two species. Although no data are currently available for the SK1 isoform, the failure of SK1 to fully complement the SK2 mutation in the brachymorphic mouse (16) and human spondyloepimetaphyseal dysplasia (17) suggests differential roles for the members of this gene family as well. Analysis of expression patterns of the PAPS synthetase genes and functional analysis of its promoters will provide the specific roles of each PAPS synthetase in each species. In sum, we have provided the first report on the genomic structure, intron/exon mapping, alternative spliced forms, correlation between functional protein domains and genomic organization, and initial promoter analysis for the PAPS synthetase gene family.

Acknowledgements—We thank Mary Lou Spach for maintenance of the mouse colony, Glenn Burrell for manuscript preparation, and James Mensch for manuscript review.
REFERENCES

1. Schwartz, N. B., Lyle, S., Ozern, J. D., Li, H., Deyrup, A., and Westley, J. (1998) Chem. Biol. Int. 109, 143–151
2. Leyh, T. S., Taylor, J. C., and Markham, G. D. (1988) J. Biol. Chem. 263, 2499–2416
3. Leyh, T. S., Vogt, T. F., and Suo, Y. (1992) J. Biol. Chem. 267, 10405–10410
4. Schwedock, J. S., Liu, C., Leyh, T. S., and Long, S. R. (1994) J. Bacteriol. 176, 7055–7064
5. Foster, B. A., Thomas, S. M., Mahr, J. A., Renosto, F., Patel, H. C., and Segel, I. H. (1994) J. Biol. Chem. 269, 19777–19786
6. Borge-Walmsley, M. I., Turner, G., Bailey, A. M., Brown, J., Lehmbeck, J., and Clausen, I. G. (1995) Mol. Gen. Genet. 247, 425–429
7. Cherest, H., Keren, P., and Sardin-Kerjan, Y. (1987) Mol. Gen. Genet. 210, 307–313
8. Korch, C., Mountain, H. A., and Bystrom, A. S. (1992) J. Biol. Chem. 267, 173–178
9. Leustek, T., Murillo, M., and Cervantes, M. (1994) Plant Physiol. 105, 99–108
10. Jain, A., and Leustek, T. (1994) Plant Physiol. 105, 76, 229–234
11. Arz, H. E., Gisselmann, G., Schiffmann, S., and Schwenn, J. D. (1994) Biochim. Biophys. Acta 1218, 447–452
12. Rosenthal, E., and Leustek, T. (1995) Gene (Amst.) 165, 243–248
13. Jullien, D., Crozatier, M., and Kas, E. (1997) Mech. Dev. 68, 179–186
14. Li, H., Deyrup, A., Menegh, J. R., Jr., Domowicz, M., Konstantinidis, A. K., and Schwartz, N. B. (1995) J. Biol. Chem. 270, 29453–29459
15. Venkataraman, K. V., Fuda, H., Koonin, E. V., and Strott, C. A. (1999) J. Biol. Chem. 274, 2601–2604
16. MacRae, I. J., Rose, A. B., and Segel, I. H. (1998) J. Biol. Chem. 273, 28683–28689
17. Chen, Z., Dixon, J. E., and Zalkin, H. (1990) Proc. Natl. Acad. Sci. U. S. A. 87, 3097–3101
18. Duncan, K., Edwards, R. M., and Coggin, J. R. (1987) Biochem. J. 248, 373–386
19. Witkowski, A., Rangan, V. S., Randhawa, Z. I., Amy, C. M., and Smith, S. (1991) Eur. J. Biochem. 198, 571–579
20. Sakata, J., Abe, Y., and Uyeda, K. (1991) J. Biol. Chem. 266, 16774–16770
21. Faure, M., Cameron, J. H., and Jacquet, M. (1989) Eur. J. Biochem. 179, 345–358
22. Lin, T., and Suttle, D. P. (1992) Somat. Cell Mol. Genet. 19, 183–202
23. Lin, T., and Suttle, D. P. (1993) Somat. Cell Mol. Genet. 19, 193–202
24. Beaucamp, N., Hofmann, A., Kellerer, B., and Jaenicke, R. (1997) Protein Sci. 6, 2159–2165
25. Crawford, I. P., Clarke, M., van Cleemput, M., and Yanofsky, C. (1987) J. Biol. Chem. 262, 239–244
26. Lin, T., and Suttle, D. P. (1995) Somat. Cell Mol. Genet. 21, 161–175
27. Suchi, M., Mize, S., Kawai, Y., Tsuibo, T., Suni, S., Okajima, K., Hodgson, M. E., Ogawa, H., and Wada, Y. (1997) Am. J. Hum. Genet. 60, 525–539
28. Amy, C. M., Williams-Ahl, B., Naggert, J., and Smith, S. (1992) Proc. Natl. Acad. Sci. U. S. A. 89, 1105–1108