Biochemical and Biophysical Research Communications **285,** 152–157 (2001)
doi:10.1006/bbrc.2001.5096, available online at http://www.idealibrary.com on IDEAL®

Use of an *in Silico* Approach to Define the Gene Structure of Eukaryotic Adenylyl Cyclases

Brinda K. Rana, Linda Pan, and Paul A. Insel[1]
Department of Pharmacology, 0636, University of California at San Diego, 9500 Gilman Drive, La Jolla, California 92093-0636

Received May 22, 2001

The limited information available regarding the gene structure of adenylyl cyclases (AC), which catalyze the synthesis of cAMP, suggests a complex arrangement with many exons and large introns such that molecular techniques to define these gene structures are time- and labor-intensive. We report here the use of a computer-based approach involving the assembly of fragmented sequence data generated by the Human Genome Project and nucleic acid analysis software to decipher the gene structure of human and murine AC 6 and other human AC isoforms (ACs 3, 7, and 8). The results, which document 21 exons in human and murine AC 6, human AC 3, 18 exons in AC 8, and 24 exons in AC 7, show substantial conservation of exon organization in the AC family and in particular regions of the AC protein. Application of such *in silico* methods should prove useful to characterize genes for other ACs and protein families and data provided here should facilitate studies of polymorphisms in AC genes.

Adenylyl cyclase (AC), which converts ATP into cyclic-3′,5′-adenosine monophosphate (cAMP), is regulated by heterotrimeric GTP-binding proteins (G-proteins) and G-protein-coupled receptors (GPCRs) via neurotransmitters and hormones. Previous studies on the stoichiometric relationship of the GPCR signaling pathway have revealed that AC is the limiting component for efficacy of the system (1–3). Two classes of ACs have been identified in mammals: the transmembrane ACs (tmAC), of which nine isoforms have been identified, and a cytosolic form of AC, soluble AC (sAC), which differs in sequence, structure and function from the nine tmAC isoforms (4, 5). The nine tmACs, as well as sAC, are differentially regulated and expressed in various tissues [recently reviewed in (5)]. The AC isoforms have been subgrouped by regulation and sequence similarity. AC 1, 3, and 8 are stimulated by calcium-calmodulin; AC 2, 4, and 7 are activated by β,γ subunits and protein kinase C phosphorylation; AC 5 and 6 are inhibited by calcium; AC 9 is not grouped with any other tmACs. Despite differences in regulation and expression, all tmACs have similar protein structure: a N-terminal cytoplasmic region, two transmembrane domain spanned by six alpha helices, and two cytoplasmic regions (Fig. 1a) that form the catalytic site [reviewed in (6, 7)].

Of the tmACs, AC 6 is of interest because this isoform is subject to inhibitory regulation by a variety of factors (e.g., $G_{\alpha i}$, $G_{\beta/\gamma}$, Ca^{2+}, protein kinase C, nitric oxide). AC 6 is expressed in numerous tissues and has been localized to human chromosome 12q12-13 (8). Because of its importance in signaling in the heart and kidney, genetic variants of the AC 6 gene (*ADCY6*) might contribute to altered signaling and function during development, aging, and with renal and cardiovascular disease. In order to test this possibility and understand how AC 6 is regulated, it is necessary to define the organization and sequence of the *ADCY6* gene. Although the cDNA sequences for human *ADCY6* and the AC 6 gene from other species are available, little is known regarding the genomic structure of *ADCY6* and other AC genes. Inferring from the genomic information of murine *Adcy3* (9), one expects that AC genes will have numerous, large introns, making it very difficult, or impossible, to use PCR-based strategies to isolate major portions of these genes. In addition, because the genomic DNA (gDNA) sequence of *ADCY6* has not been available for any mammalian species, it has been difficult to infer its intron–exon organization via sequence comparison methods. Therefore, in this study, we tested the utility of an *in silico* method to determine the gDNA sequence and genomic structure of *ADCY6*. Based on our success in using this approach, we propose it as an efficient means for ascertaining the type of genomic information that was previously only obtainable through the use of more laborious molecular techniques.

[1] To whom correspondence should be addressed. Fax: 858-822-1007. E-mail: pinsel@ucsd.edu.

0006-291X/01 $35.00

MATERIALS AND METHODS

Sequences. *Homo sapiens ADCY6* (Accession No. NM_015270), *Mus musculus Adcy6* (Accession No. NM_007405), and *Fugu rubripes AC-VI* (Accession No. U72484), *Homo sapiens ADCY3* (Accession No. NM_004036), *Homo sapiens ADCY7* (Accession No. 001114), and *Homo sapiens ADCY8* (Accession No. 001115) cDNA sequences were obtained from GenBank. *Rattus norvegicus Adcy5* (Accession No. NM_022600) was used to search for the human genome database for the human homolog. Human genomic sequence of AC genes were obtained by selecting through using NCBI's BLAST 2.0 search engine (http://www.ncbi.nlm.nih.gov/BLAST) against the Human Genome Database database (htgs), which contains unordered working draft sequences of the human genome.

Partial gene organization of AC 3 and 7 were obtained using AceView (http://www.ncbi.nlm.nih.gov/AceView). *Drosophila melanogaster* genomic sequences were obtained from FlyBase (http://www.flybase.org).

Sequence analysis. Sequence analysis tools in Biology Workbench 3.2 (http://workbench.sdsc.edu) were used to determine the reverse complement of contig sequences (REVCOMP) and align sequences (ALIGN or CLUSTALW) (10).

PCR and sequencing PCR and sequencing were used to verify ambiguous human genome project draft sequences using primer pairs: 5′-GCACATAGCACCGCAGTTG-3′ and 5′-GCCAAGCAGTAGGTCATAGT-3′ encompassing the 3′ end of exon 13 to the 3′ end of exon 15; 5′-CAATGATGTGACCCTGGC-3′ and 5′-CCGTTCAGGTACTGCAGT-3′ encompassing the 3′ end of exon 6 to the 5′ portion of intron 8. The PCRs were performed under standard protocol with AmpliTaq Gold (PE Biosystems) for 35 cycles at 94°C for 45 s, 59°C 45 s, 72°C for 60 s with a 2 s extension per cycle following a preliminary incubation at 95°C for 7 min for both primer pairs.

RESULTS AND DISCUSSION

Determining the Gene Organization of ADCY6

A BLAST search of human *ADCY6* cDNA sequence against NCBI's Human Genome Database, which contains working draft sequences of the human and mouse genomes, revealed significant matches in >20 fragments with two chromosome 12 working draft sequences: (1) a sequence obtained from clone RP11-455I22 containing 60 unordered pieces of total length 216900 (AC021647) and (2) a sequence obtained from clone RP11-579D7 containing 17 unordered pieces 171945 (AC025557). These sequences are a conglomeration of multiple contigs in unknown orientation and order. Hence, for example, the BLAST search located *ADCY6* coding region nucleotides 1–863 in one of 60 contigs from the AC021647 sequence, matching it to nucleotides 36767–35928 of this contig, and coding region nucleotides 864–1024 to nucleotides 31617–31453 of another contig in the same sequence. Because both contigs are in the reverse orientation, they were reverse complemented for further analyses. Since the BLAST search was performed with cDNA sequences, which do not contain intronic sequence, intronic sequences were deduced by analyzing sequences from contigs upstream and downstream of the matches. Exons and introns spanning the coding region and partial 5′- and 3′ untranslated regions of human *ADCY6* were pieced together to reveal 21 exons of variable size (Table 1a). Except for the >1.6-kb intron 20, which was assembled from two nonoverlapping contigs and is therefore incomplete, the sequences of all introns and exons spanning the coding region of *ADCY6* and encompassing approximately 14.7 kb of nucleotide sequence have been obtained by this approach.

A similar approach was used to decipher the murine *ADCY6* gene structure (Table 1b). BLAST search of murine *Adcy6* cDNA sequence against the HTGS database revealed matches with contigs from the mouse chromosome 15 working draft sequence containing 33 unordered pieces obtained from the *Mus musculus* C57BL6/J strain clone RP23-34P24 (AC074028).

Wicker *et al.* (11) reported a GenBank sequence from human brain (Accession No. AB007883) that was identical to the cloned cDNA sequence of human AC6 with the exception of a 159 nucleotide gap in the coding region, suggesting alternate splicing. Our gene structural analyses reveal that this 159 nucleotide gap sequence corresponds to exon 14 and adheres to the GT/AG rule (12). PCR amplification and sequencing of the region of *ADCY6* containing exon 13, intron 14 and exon 14 verify these results. Other experimental evidence for splice variants of human adenylyl cyclases have not previously been reported; however, three alternatively spliced rat AC8 messages have been identified with altered structural and enzymatic properties (13) and two splice variants of AC 5 have been reported in dog and rat (14, 15).

In addition to the human and murine AC 6 gene, we obtained the exon–intron organization of the *Fugu rubripes* (pufferfish) AC 6 gene in a genomic sequence in GenBank containing 62-kb region around the *G6PD* locus. Figure 1b compares the exon organization of the AC 6 gene in three species (human, mouse and pufferfish). While intron size may vary, exon organization is known to be conserved in genes of vertebrates as distant from each other as lamprey and humans: 14 of the 16 exons of the fibrinogen gene are conserved in lamprey and human (16). Our comparison of the exon organization of *G6PD* in human and pufferfish (data not shown) also reveals conservation of exon structure (11 of 13 exons are of identical size and order). In contrast, we observed that 6 of the 21 coding exons of the pufferfish AC 6 gene vary in size compared to mammals, and three additional exons vary by one or two codons. Of interest, the exons that are conserved between pufferfish and mammals are those containing the region encoding the catalytic domains of AC 6: pufferfish exons 3 through 7 contain the coding region for C1 domain, which correspond to human and murine exons 2 through 6; pufferfish, human and murine exons 18 through 21 contain the coding region for the C2. In addition, in comparing human and mouse AC 6 sequences, we find that certain introns (e.g., intron 17) show nucleotide similarity close to that of the coding

TABLE 1

The Exon–Intron Boundaries for (a) Human *ADCY6* and (b) Murine *Adcy6*

Exon/intron	Exon size	Coding sequence	5′ intron 3′ sequence	Coding sequence	Intron size
			a		
		K Q		**L G**	
1	864	...AAG CAG	gtaggt...ctgcag	CTC GGT...	4312
		Q Q		**E R**	
2	150	...CAG CAG	gtgggg...ggctag	GAG CGG...	371
		V S		**I L**	
3	122	...T GTC AG	gtaggg...ctacag	C ATC CT...	93
		A A		**E N**	
4	112	...GCT GCG	gtgagg...attcag	GAG ATT...	178
		I S		**L V**	
5	128	...C ATC TC	gtaagc...tcccag	G CTG GT...	517
		A G		**R I**	
6	159	...G GCT GG	gtgagt...ggagag	C CGC AT...	47
		K R		**K E**	
7	142	...AAA CGG	gtcagg...gcacag	AAA GAG...	117
		Q M		**G I**	
8	129	...CAG ATG	gtgagg...ctccag	GGC ATT...	264
		D N		**R G**	
9	26	...A GAC AA	gtaagt...cctcag	C CGG GG...	221
		K K		**Y S**	
10	148	...AAG AAG	gtttga...cctcag	TAC TCC...	248
		P H		**S T**	
11	98	...C CCA CA	gtgaga...ccccag	C TCC AC...	178
		G S		**L F**	
12	88	...GGT TCT	gtacgt...ctctag	CTG TTC...	172
		N M		**F T**	
13	117	...AAC ATG	gtaagg...ttgcag	TTC ACC...	303
		P E		**Y F**	
14	159	...CCT GAG	gtgttc...ccgcag	TAC TTC...	292
		G L		**A S**	
15	179	...T GGC TT	gtgagt...taccag	G GCT TC...	925
		D C		**P A**	
16	36	...G GAC TG	gtaagt...ccttag	T CCA GC...	83
		L Q		**A T**	
17	130	...CTA CAG	gtgact...ccacag	GCA ACA...	319
		D E		**I I**	
18	264	...GAT GAG	gtactt...gttcag	ATT ATC...	739
		K I		***G L***	
19	205	...AG ATT G	gtaaga...cctcag	GG CTG A...	>1684
		I Q		**V T**	
20	125	...ATC CAG	gtgagg...ccccag	GTG ACC...	217
21	126+				

region (86.4% identical) while comparison of other human and murine *ADCY6* introns on average share approximately 54% identity.

We observed this similar conservation of exon organization in the C1 domain across phyla. For example, examination of the exon organization of several fly ACs obtained from genomic sequence data in FlyBase (http://www.flybase.org) revealed conservation of exon organization in the C1 domain. *Drosophila* also has multiple isoforms of ACs. These include orthologs of mammalian ACs (e.g., rutabaga AC) and unique ACs resulting from gene duplication events (17) that likely occurred in the arthropod or *Drosophila* lineage. The order, size and sequence of three exons of *Drosophila AC3* are conserved with exons 4, 5, and 6 of human *ADCY3* and two exons of a *Drosophila* calmodulin-sensitive AC, *AC76E,* are conserved precisely with the C1 coding exons 5 and 6 of *ADCY6.* Further investigation in other distantly related organisms may yield similar results.

In order to investigate whether exon organization was conserved after gene duplication events leading to the AC isoforms, we compared the exon organization of *ADCY6* with three additional *ADCY* genes, which code for members of different subfamilies of ACs (Fig. 2). Similar to the results obtained from the comparisons between species for *ADCY6,* the exon organization in the coding region for the C1 domain is well conserved among the four isoforms. This result is consistent with observations that the amino acid sequence similarity of

TABLE 1—*Continued*

Exon/intron	Exon size	Coding sequence	5′ intron 3′ sequence	Coding sequence	Intron size
		b			
		K Q		**L G**	
1	858	...AAG CAG	gtaggt...caacag	CTC GGT...	2815
		Q Q		**E R**	
2	150	...CAG CAG	gtgggg...gtctag	GAA CGG...	364
		V S		**I L**	
3	122	...T GTC AG	gtaggt...ccacag	C ATC CT...	105
		A A		**E N**	
4	112	...GCT GCG	gtgagg...ctttag	GAG AAT...	150
		I S		**L V**	
5	128	...C ATC TC	gtaagc...tctcag	G CTG GT...	883
		G R		**R I**	
6	157	...GGC CGG	gtaagt...ttacag	CGC ATC...	194
		K R		**K E**	
7	141	...AAA CGG	gtcaag...gcacag	AAA GAG...	>628
		Q M		**G I**	
8	129	...CAG ATG	gtaagg...ctccag	GGC ATG...	232
		K N		**R G**	
9	26	...A GAC AA	gtaagt...tatcag	C CGG GG...	223
		K K		**Y S**	
10	148	...AAG AAG	gtttgg...cctcag	TAT TCA...	249
		P Y		**S T**	
11	98	...C CCA TA	gtgagt...ctccag	C TCC AC...	128
		G S		**F F**	
12	88	...GGT TCT	gtaagt...tggcag	TTC TTC...	153
		N M		**F T**	
13	117	...AAC ATG	gtaaat...ttgcag	TTT ACC...	218
		P E		**Y F**	
14	159	...CCT GAG	gtgtcc...ccccag	TAC TTC...	198
		G L		**A S**	
15	179	...T GGC TT	gtaagc...tgccag	G GCT TC...	>927
		D C		**P A**	
16	36	...G GAC TG	gtaagt...tttcag	C CCA GC...	184
		L Q		**A T**	
17	130	...TTA CAG	gtgact...ccacag	GCA ACA...	318
		D E		**I I**	
18	264	...GAC GAG	gtgctn...attcag	ATC ATC...	>50
		K I		**G L**	
19	205	...AG ATC G	gtgagg...cttcag	GG TTG A...	668
		I Q		**V T**	
20	125	...ATA CAG	gtgagg...ccctag	GTG ACT...	197
21	126+				

Note. The numbers correspond to coding exons. The amino acid corresponding to the codons are shown above the nucleotide sequence.

transmembrane regions of ACs are less than 30% while those of the catalytic domains are greater than 50% with the C1 domain showing greater similarity than does the C2 domain (18, 19). Intron size is not conserved among the isoforms. In addition, although AC 3 and AC 8 are members of the same functional subfamily of ACs, there is little conservation of exon organization in these genes outside the catalytic domains. Higher conservation of exon organization outside the catalytic domain is observed between *ADCY3* and *ADCY6.* For example, exons 11 and 12 of *ADCY3* correspond to exons 12 and 13 in *ADCY6;* sequence alignment verifies that the sequences of these two exons are also more similar between the isoforms than sequences of other exons coding for the transmembrane region. The reason why only certain exons that code for regions of AC not implicated in catalytic function have been maintained through evolution is not clear. Further analysis of gene organization for additional members of the *ADCY* gene family should provide more definitive information regarding evolution of this gene family

In this study, we have defined for the first time the genomic organization of a human tmAC isoform, AC 6, through the assembly of fragmented nucleotide sequence information from the Human Genome Project. Using this technique, we have also determined the genomic structure of murine *ADCY6* and have compared the genomic organization of AC 6 from several species and with other AC isoforms. Our results provide new insights regarding genomic structure and

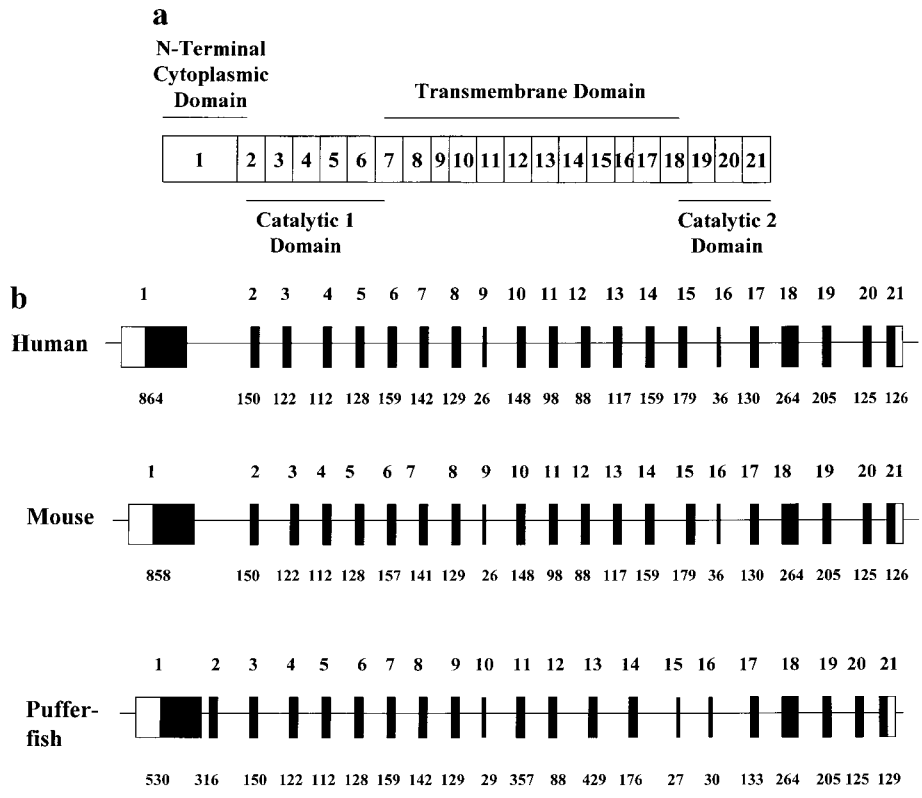

FIG. 1. (a) Human *ADCY6* coding exons corresponding to structural domains of AC 6. (b) Organization of exons spanning the coding region of *ADCY6* in human, mouse, and pufferfish. Shaded boxes represent translated regions of exons. Distances between exons are not representative of intron sizes.

organization of AC isoforms and also show the utility of a computer-based method for determining such information.

The *in silico* approach that we have used to discern the genomic organization and structure of AC isoforms offers a number of advantages relative to more tradi-

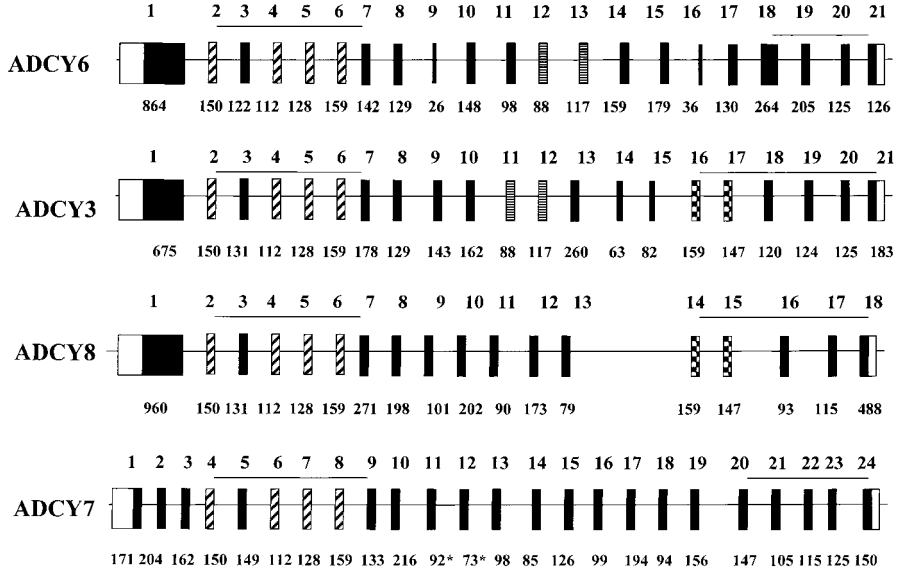

FIG. 2. Comparison of exon organization between *ADCY* isoforms. Exons under lines represent C1 and C2 domains. Diagonally striped boxes represent exons that have been conserved in all four isoforms represented here. Patterned boxes represent exons that are conserved between at least two isoforms. *Exon size may vary slightly due to gap in human genome sequence contig.

tional biochemical and molecular biological techniques. It is faster, less labor-intensive, and avoids exposure of personnel to potentially hazardous biological and chemical reagents. With the availability of genomes from humans and other species, one has the opportunity to utilize computer-based methods to predict both the organization of individual genes and gene family members and to help define appropriate strategies, such as PCR, to identify polymorphisms and other genetic variants. The application of such approaches for AC 6 and other AC isoforms may prove of interest. For example, we have used this method to obtain a contig (Accession No. AC025571) containing the first 1534 nucleotides of coding sequence and over 2 kb of the 5′ untranslated region of human *ADCY*5, whose complete coding sequence is not yet available, and have found that this portion of the coding region is intronless (data not shown). We thus believe that the application of *in silico* approaches should prove highly useful to define genomic sequence, structure and organization of ACs, and perhaps of other complex multigene families.

ACKNOWLEDGMENT

This work was supported by grants from NIH.

REFERENCES

1. Alousi, A. A., Jasper, J. R., Insel, P. A., and Motulsky, H. J. (1991) Stoichiometry of receptor-Gs-adenylate cyclase interactions. *FASEB J.* **5,** 2300–2303.
2. Post, S. R., Hilal-Dandan, R. Urasawa, K. Brunton, L. L., and Insel, P. A. (1995) Quantification of signaling components and amplification in the beta-adrenergic-receptor-adenylate cyclase pathway in isolated adult rat ventricular myocytes. *Biochem. J.* **311,** 75–80.
3. Milligan, G. (1996) The stoichiometry of expression of protein components of the stimulatory adenylyl cyclase cascade and the regulation of information transfer. *Cell Signal* **8,** 87–95.
4. Chen, Y., Cann, M. J., Litvin, T. N., Iourgenko, V., Sinclair, M. L., Levin, L. R., and Buck, J. (2000) Soluble adenylyl cyclase as an evolutionarily conserved bicarbonate sensor. *Science* **289,** 625–628.
5. Defer, N., Best-Belpomme, M., and Hanoune, J. (2000) Tissue specificity and physiological relevance of various isoforms of adenylyl cyclase. *Am. J. Physiol. Renal Physiol.* **279,** F400–F416.
6. Krupinski, J., Coussen, F. Bakalyar, H., Tang, W.-J., Feinstein, P. G. Orth, K., Slaughter, C., Reed, R. R., and Gilman, A. G. (1989) Adenylyl cyclase amino acid sequence: Possible channel- or transporter-like structure. *Science* **244,** 1558–1564.
7. Hurley, J. H. (1998) Structure, mechanism, and regulation of mammalian adenylyl cyclase. *J. Biol. Chem.* **274,** 7599–7602.
8. Haber, N., Stengel, D., Defer, N., Roeckel, N., Mattei, M. G., and Hanoune, J. (1994) Chromosomal mapping of human adenylyl cyclase genes type III, type V and type VI. *Hum. Genet.* **94,** 69–73.
9. Muglia, L. M., Schaefer, M. L., Vogt, S. K., Gurtner, G., Imamura, A., and Muglia, L. J. (1999) The 5′-flanking region of the mouse adenylyl cyclase type VIII gene imparts tissue-specific expression in transgenic mice. *J. Neurosci.* **19,** 2051–2058.
10. Subramaniam, S. (1998) Knowledge-based interaction potentials for proteins. *Proteins* **32,** 1–2.
11. Wicker, R., Catalan, A. G., Cailleux, A., Starenki, D., Stengel, D., Sarasin, A., and Suarez, H. G. (2000) *Biochim. Biophys. Acta* **1493,** 279–283.
12. Aebi, M., Hornig, H., Padgett, R. A., Reiser, J., and Weissmann, C. (1986) Sequence requirements for splicing of higher eukaryotic nuclear pre-mRNA. *Cell* **47,** 555–565.
13. Cali, J. J., Parekh, R. S., and Krupinski, J. (1996) Splice variants of type VIII adenylyl cyclase. *J. Biol. Chem.* **271,** 1089–1095.
14. Iwami, G., Akanuma, M., Kawabe, J., Cannon, P. J., Homcy, C. J., Ishikawa Y. (1995) Multiplicity in type V adenylylcyclase: Type V-a and type V-b. *Mol. Cell. Endocrinol.* **110,** 43–47.
15. Emala, C. W., Kumasaka D., Hirshman C. A., Lindeman K. S. (1998) Adenylyl cyclase messenger ribonucleic acid in myometrium: Splice variant of type IV. *Biol. Reprod.* **59,** 169–175.
16. Pan, Y. (1992) Studies on Fibrinogen Evolution, University of California, San Diego. [Thesis]
17. Cann, M. J., Chung E., Levin L. R. (2000) A new family of adenylyl cyclase genes in the male germline of *Drosophila melanogaster*. *Dev. Genes Evol.* **210,** 200–206.
18. Katsushika, S., Chen, L., Kawabe, J., Nilakantan, R., Halnon, N. J., Homcy, C. J., and Ishikawa, Y. (1992) Cloning and characterization of a sixth adenylyl cyclase isoform: Types V and VI constitute a subgroup within the mammalian adenylyl cyclase family. *Proc. Natl. Acad. Sci. USA* **89,** 8774–8778.
19. Hanoune, J., Pouille, Y., Tzavara, E, Shen, T, Lipskaya, L., Miyamoto, N., Suzuki, Y., and Defer, N. (1997) Adenylyl cyclases: Structure, regulation and function in an enzyme superfamily. *Mol. Cell. Endocrinol.* **128,** 179–194.