# XL $\alpha$ s, the extra-long form of the $\alpha$ -subunit of the Gs G protein, is significantly longer than suspected, and so is its companion Alex

Joel Abramowitz, Dagoberto Grenet, Mariel Birnbaumer, Hector N. Torres\*, and Lutz Birnbaumer<sup>+</sup>

Transmembrane Signal Transduction Group, Laboratory of Signal Transduction, Division of Intramural Research, National Institute of Environmental Health Sciences, National Institutes of Health, Department of Health and Human Services, Research Triangle Park, NC 27709

Contributed by Lutz Birnbaumer, December 31, 2003

Because of the use of alternate exons 1, mammals express two distinct forms of  $\operatorname{Gs} \alpha$ -subunits: the canonical 394-aa  $\operatorname{Gs} \alpha$  present in all tissues and a 700+-aa extra-long  $\alpha$ s (XL $\alpha$ s) expressed in a more restricted manner. Both subunits transduce receptor signals into stimulation of adenylyl cyclase. The XL exon encodes the XL domain of XL $\alpha$ s and, in a parallel ORF, a protein called Alex. Alex interacts with the XL domain of XLas and inhibits its adenylyl cyclase-stimulating function. In mice, rats, and humans, the XL exon is thought to contribute 422.3, 367.3, and 551.3 codons and to encode Alex proteins of 390, 357, and 561 aa, respectively. We report here that the XL exon is longer than presumed and contributes in mice, rats, and humans, respectively, an additional 364, 430, and 139 codons to XL $\alpha$ s. We called the N-terminally extended XL $\alpha$ s extra-extra-long Gs $\alpha$ , or XXL $\alpha$ s. Alex is likewise longer. Its ORF also remains open in the 5' direction for  $\approx$ 2,000 nt, giving rise to Alex-extended, or AlexX. RT-PCR of murine total brain RNA shows that the entire XXL domain is encoded in a single exon. Furthermore, we discovered two truncated forms of XXL $\alpha$ s, XXLb1 and XXLb2, in which, because of alternative splicing, the Gs $\alpha$ domain is replaced by different sequences. XXLb proteins are likely to be found as stable dimers with AlexX. The N-terminally longer proteins may play regulatory roles.

he murine  $\alpha$ -subunit of the Gs G protein is encoded in the Gnas complex locus at the distal end of chromosome 2. This locus not only contains the promoter responsible for initiation of transcription at exon 1 of Gnas1 (46.3 codons), but it encodes at least four additional promoters, of which three drive transcription using the sense strand as template beginning at exons (5' to 3' direction) Nesp1, XL, and 1A, all upstream of Gnas1 exon 1; the fourth drives transcription using the antisense strand as template, beginning at exon Nespas1. Transcripts beginning with exons 1A and 1 of Nespas do not appear to be translated and code for RNAs of unknown function (reviewed in 1). The Nesp transcript codes for Nesp55 (neuroendocrine secretory protein of 55 kDa), which may play an important role in modulating central nervous system neurotransmission (2). Fully processed transcripts beginning with the XL exon give rise to the extra-long Gs $\alpha$  (XL $\alpha$ s) (3).

By analogy to the human Gs $\alpha$  (4), mouse (and rat) Gs $\alpha$  is presumed to be encoded in 13 exons, and, as happens with human and rat Gs $\alpha$ , mouse Gs $\alpha$  comes in at least two easily distinguishable flavors, long and short (394 and 379 aa, respectively), that depend on the inclusion or not of the 45-nt exon 3, of which each in turn can arise without (default) or with an added Ser, because of use of an alternative splice donor site at exon 4 (4, 5). Processing of the XL transcript has been reported to join its XL exon to exon 2 of Gs $\alpha$ , thus giving rise to XL $\alpha$ s, formed of the XL domain followed by the canonical GTPase and helical domains of the Gs $\alpha$ -subunit, in both its long and short variants. The XL domain is listed at 367 aa for the rat (GenBank accession no. X84047), corrected from an initial report of 498 aa (3, 7), 551 aa for the human (GenBank accession no. NM\_080425), and 422 aa for the mouse (Gen-Bank accession no. AF116268).

In addition to the ORF coding for the XL portion of  $XL\alpha s$ , the XL exon was found to have a concurrent ORF that runs along the XL ORF and terminates with a stop codon at the exact end of the exon (6). The protein encoded in this ORF was termed Alex. It interacts with the protein encoded in the XL domain of XL $\alpha$ s, whose activity it appears to suppress (8). Alex has been reported as having lengths that vary among species, as did those of the XL domain of  $XL\alpha s$ , with human, mouse, and rat Alex having 357, 390, and 561 aa, respectively. The binding of a recombinant human XL domain protein comprising the C-terminal 374 aa (stop codon at -7 from the end of the XL exon; GenBank accession no. NM\_080425) to an N-terminal truncated Alex of 323 aa was shown to be lost in humans bearing a particular polymorphism associated with prolonged bleeding times and metabolic and neurologic abnormalities. This finding provided evidence for the physiological relevance of the extra-long forms of  $Gs\alpha$  (8). Strong developmental and metabolic phenotypic changes found in mice harboring mutations in the locus encoding Gs $\alpha$  and XL $\alpha$ s have provided further evidence for roles of XLas and Gnas1 in normal mammalian development (9, 10).

Fig. 1 illustrates the intron-exon structures of the murine Gnas complex locus, the processed mRNAs as obtained from GenBank supercontig NT\_039211, region 3831406-3893830, and incorporates some of the features just described. In addition, the figure also incorporates the very interesting findings that transcription off the various promoters is under epigenetic control, being either exclusively from the maternal allele, as in the case of Nesp55, or from the paternal allele, as in the cases of XL $\alpha$ s, the antisense Nespas, and the transcript beginning with exon 1A. Gnas1 (the Gs $\alpha$ -subunit proper) mRNA is transcribed biallelically in most tissues (denoted by "b"). However, it should be noted here that in some tissues, such as proximal renal tubules, white and brown adipose cells, anterior pituitary cells, and thyroid, it is transcribed mostly from the maternal allele (for reviews see refs. 1 and 11; see also refs. 13–15).

The present report had its origin in that, upon analysis of the mouse genome upstream of the presumed translation initiation site of the XL exon, the first in-frame stop codon is located 2.4 kb from this presumed translation initiating ATG. This raised the possibility that, inadvertently, the so far known XL $\alpha$ s is a truncated form of the real translation product.

Abbreviations: XLas, extra-long Gsa; XXLas, extra-extra-long Gsa; AlexX, Alex-extended. Data deposition: The sequences reported in this paper have been deposited in the GenBank

database (accession nos. AY518305–AY518308 and AY519501–AY519504). \*Present address: Instituto de Biologia Molecular e Ingenieria Genetica, Vuelta de Ob-

<sup>&</sup>quot;Present address: instituto de Biologia Molecular e Ingeniería Genetica, vuelta de Obligado 2490, Buenos Aires, Argentina.

<sup>&</sup>lt;sup>†</sup>To whom correspondence should be addressed. E-mail: birnbau1@niehs.nih.gov.



**Fig. 1.** Intron–exon organization of the murine Gnas complex locus at the distal end of chromosome 2, orientation of the transcription units, and allelic expression characteristics as known before the present studies. (A) Intron–exon organization. Neither exons (boxes) nor introns (interbox distances) are to scale; however, their relative positions reflect their positions in the genome. Exact parameters can be obtained from GenBank genomic contig NT\_039211, region 3831406–3893830. p (paternal), m (maternal), and b (biallelic) next to arrows depicting promoter location and direction indicate a transcribed allele, where p transcripts are maternally silenced (imprinted) and m transcripts are paternally imprinted. *(B)* Main fully processed transcripts and their translation products. Two, 1A and Nespas, appear to be noncoding. \*, Stop codon.

PCR analysis of reverse transcripts synthesized by using adult mouse total brain RNA as template revealed that, indeed, the mRNA is longer than it had been thought to be. We propose that the XL $\alpha$ s protein synthesized (henceforth referred to as XXL $\alpha$ s, for extra-extra-long Gs $\alpha$ ) is likely to have an N-terminal extension that in the mouse is of 364 aa. We also found that Gs $\alpha$  amino acids 221–280, which in the human gene are encoded in exons 9 and 10, are encoded in a single exon in the mouse.

# **Materials and Methods**

Standard methods of molecular biology and of database searching for homologous and orthologous nucleotide similarities were used (ref. 16; see also www.ncbi.nih.gov). Brain RNA was isolated from adult male FVB/N mice by using RNeasy kits (Qiagen, Valencia, CA). Reverse transcripts were prepared from total RNA by using the SuperScript First-Strand Synthesis System (Invitrogen) with reagents and protocols supplied by the manufacturer. PCR was performed by using High Fidelity polymerase and reagents supplied by Roche Molecular Biochemicals. Primers used for PCR are shown in the figures and/or their legends. Amplifications were carried out as follows: 94°C for 2 min, followed by 35 cycles at 94°C for 30 s, 65°C for 1 min, and 72°C for 2 min, followed by 72°C for 7 min. The resulting PCR products were cloned into pCR2.1-TOPO (Invitrogen).

DNA sequencing was performed by using the dideoxynucleotide chain termination method. Templates consisting of plasmid DNA were subjected to the cycle PCR method by using protocols supplied by the National Institute on Environmental Health Sciences DNA sequencing facility. Briefly,  $\approx 100$  ng of plasmid DNA was used for each sequence reaction containing 8 µl of BigDye Terminator Ready Reaction mix (Applied Biosystems) and 1 µM of forward or reverse primer in a final volume of 20 µl. Cycling conditions were as follows: 1 cycle of 96°C for 2 min, followed by 25 cycles of 94°C for 10 s, 50°C for 5 s, and 60°C for 4 min. The sequence reactions were then cleaned by using a Qiagen DyeEx purification column. Purified sequence reactions were run on a PE-Applied Biosystem 3100 Automated Capillary DNA Sequencer. Resulting DNA sequences were compared with known database sequences by using both the GCG DNA Analysis Wisconsin package and National Center for Biotechnology Information (NCBI) BLAST to determine differences between the RNA (i.e., cloned PCR product) sequences and genomic sequences.

## **Results and Discussion**

ORFs Associated with the Coding Sequence of the XL Domain and Alex Protein Have the Potential of Encoding a Significantly Longer Protein. As depicted in Fig. 2, translation of the genomic DNA sequence of the murine Gnas complex locus reveals that the ORFs coding for the XL portion of the mouse XL $\alpha$ s, as well as that coding for Alex, remain open in the 5' direction for an additional 364 and 335 codons, respectively, for a total length of 785.3 and 726 codons. Analysis of the human and rat genome databases of the NCBI reveals a similar situation for rats and humans. Figs. 2 and 3A present the alignments of the amino acid sequences deduced from the mouse, human, and rat genomic DNA sequences of the extended forms of XL and Alex (referred to as XXL, for extra-extra-long, and AlexX, for Alex-extended).

Using mouse brain RNA as template, we tested by RT-PCR for the existence of RNA transcripts of the predicted size and composition required for translation of XXL $\alpha$ s and AlexX, shown in Fig. 3. Fig. 4*A* presents the PCR strategy used, with predicted product sizes. Fig. 4*B* shows the PCR products obtained. These products were cloned into pCR2.1-TOPO and sequenced. The results confirmed the existence of an mRNA coding for XXL $\alpha$ s/AlexX. Importantly, lane 1, next to the sizing standards, shows a positive band of ~1,450 bp that upon sequencing proved to be the expected 5' 1,092-nt extension of the XL exon plus the first 347 nt of the XL exon (from primer XXF to XL5R, Fig. 2).

The consecutively overlapping PCR products shown in lanes 2-5 (Fig. 4) proved the existence of extra-extra-long forms of XLb cDNA (GenBank accession no. AK030489) coding for XXLb1 and of XL $\alpha$ s coding for XXL $\alpha$ s. The XXLb1 nucleotide sequence predicts a protein of 876 aa [363 longer than that predicted by the AK030489 cDNA, which is part of a collection of 60,770 full-length cDNAs cloned by the RIKEN genome project (17)], which included numerous transcripts derived from the Gnas complex locus (18). Unexpectedly, PCR products derived from primers spanning from the XL domain to Gnas1 exon 2 and beyond (X3F to ex2R and ex12R, and XF to ex2R) comprised two products, each differing in 95 nt, that by sequence analysis constitute the 5' end of the XLb exon spliced to the splice donor site of Gnas1 exon 2. We refer to the protein encoded in these mRNAs as XXLb2 (Fig. 3C). Exon XLb splices to exon 2 out of frame and leads to a second form of the XXLb protein, XXLb2, encoded in exons XXL, XLb (first 95 nt), and Gnas exons 2-12. XXLb1 in turn is encoded in exons XXL and XLb. The nucleotide sequence predicts the length of XXLb2 to be 827 aa. The nucleotide sequence of XXL $\alpha$ s, XXLb1, XXLb2, and AlexX have been deposited in the GenBank database with accession numbers AY519501, AY519503, AY519504 and AY519502, respectively.

The Human Gs $\alpha$ -Subunit Is Encoded in 13 Exons, but the Mouse Orthologue Is Encoded in only 12. Throughout the text we have referred to the murine Gs $\alpha$  gene as a gene encoded in 12 exons, not 13 as is commonly done in the literature. The reason for this is that upon direct sequencing of the intron-exon bound-

\* K In-frame Stop for 786.3-codon XXL and/or 422.3-codon-XL KIF FHP S LAG C S V aaaatttgaa ttttttaaaa aaaaattttt tttcacccta gtttggctgg gtgctccgtc In-frame Stop for AlexX \* lwac s t a k v а t icq d n t. s t a i s pl k S S KS R v K F L H R R G P KFILKSR Met-1 of 785.3-codon XXL

LHGN NMS GQH DIPP EVG domain of XXlas, XXLb1 or XXLb2 FNC MGM EOP ttacggggcc ccaaatttat tttgaaaagt cgccaccgtg ttatgggcat gttcaactgc ctccacggca ataatatgtc aggacaacac gatatccccc ctgaagtcgg ggagcagccc Primer XXF 5'-TATGGCAT GTTCAACTGC CTCCA ->

k n l w k p d d d l p p v k w 1 g 1 a q p k r p pnrtpnrl Met-1 of 726-codon AlexX

NN F p s p s p t r P D E 0 E D LEA PGA AAPG AGA GPA EEMA TE D D S S E D gagcaagaac ctttggaagc cccaggggca gctgcccccg gtgctggggc tggcccagcc gaagaaatgg cgaccgaacc ggactccgaa ccgtctaaca atgagcccgt ccccgacgag Ss Qr W V G G a P rE PA k s E V R r sv E I S d t p n s K t D I r p s r Q A F E h v P P E k a p a S P C e t t 1 k G S G D S Q E D Y S P P E M actggcagtg agatcagtgg acceccagaa gacteeaaat etgacateea aageeeetge caggeetteg aggaagteeg agtgggtgga gactacagee caceteegga ggaageeatg Q ja i D t N s L ak Qr s F r r h E T s PA P P p h s s a w e S L G g p p w s W P T L E d g p h G T P q a S G f s p Q P s Q P G g p G T I E ccattegaga cacaacagee cageetggga gatteetgge ceaceetgga geageeagga ceatetggga ceceateagg eetecaagee tteaaceeag egattetgga geeegggaee F Y a Q a r G A a q P h qk} PEE 1 C h s s E F N t SIJ p ret P T GA p рр i s P C y p MP S G Τ. S D 0 A н L D cccactggcg cgageccagg cctgggagee tataceceee caccagaaga agetatgeea tttgagttea aegageetge ccagggagaa catageeage etceettgea agteecagae s R r q e G r h s PRA 1 f PA r pgt PGN t s o I R d 1 k t 1 a e k t a 1 1 k n 1 p ] т. L P E A D G D R А v E N A G P E D v s D E S v ettgegeeag gaggteegga ageattggte eccagagete tteeegegga geeegggaae ateagatttg aaaaegetgg etteegagaa gaetaeagee eteeeetga agaatetgtg v GE n s g E F G a Ĩ G D a h S P P P G d sressh LPRVIPO k s a I G I aas GGEF rq srs PTV AVP F G 0 S A ccattccagg tcgggtggaga agaattcggg ggcgatagcc cacccccagg actcccgcga gtcatcccac aaatcggcat tggcggcgag ttccccgacg tcgcggtccc gagtgcgctc a s l p p r t r l p s g s e a p l t d h s a r l s d l l l t s h a t a p r w r s C L A P A E N A P P L W V R G A I D R P F R E A V R S P P N F A C D S P P M E I tgeetegete eegeegagaa egegeeteee etetgggtee gaggegeeat tgacagaeea tteegegagg etgteagate teeteetaae ttegeatgeg acageeeeee gatggagate a s S E a q C A E i p q P P l V sek ERE q p 1 A A E KS R

		Met-1 of 422.3-codon	XL domain of XIGS (AF)	L16268) OF XLD (AK030489)		
		MEGSPTT	A T A V E G	KVPSPERGDG		
gataagagcg agtgcgcaga gagaccccca	gTTGAGCGAG AAGCAGCCGA	GATGGAAGGA AGCCCTACCA	CCGCCACTGC GGTGGAAGGA	AAAGTCCCCT CTCCGGAGAG AGGGGACGGA		
lpp slkq w	Primer XF 5'-A	GATGGAAGGA AGCCCTACCA	CC ->			
Met-1	of 391-codon Alex (A	F116268)				
m p	s q p l l p k	pslpdl	mlellr I	prcsqIaram		
SSTO PEAMDA	K P A P A A Q	AVS TGSD	AGAPTD	SAML TDS QSD		
TCTTCCACCC AGCCTGAAGC AATGGATGCC	AAGCCAGCCC CTGCTGCCCA	AGCCGTCTC TACCGGATCTG	ATGCTGGAGC TCCTACGGAT	TCCGCGATGC TCACAGATAG CCAGAGCGAT		
pektgapaer	laissrI	l k n s k k	pglsapI	lte gapa spa		
AGED GTA PGT	PSDL QSD	PEELEEA	PAVRAD	PDGGAAPVAP		
GCCGGAGAAG ACGGGACAGC CCCAGGAACG	CCTTCAGATCT CCAGTCGGA	TCCTGAAGA ACTCGAAGAAG	CCCCAGCTGT CCGCGCCGAT	CCTGACGGAG GGGCAGCCCC AGTCGCCCCA		
pll psps lka	a e I q p p s	q p p r q s	lppr ps l	ppg qpls prw		
ATPA ESESEG	SRDPAAE	PASEAVP	ATTAES	ASGA APV TQV		
GCCACTCCTG CCGAGTCCGA GTCTGAAGGC	AGCAGAGATC CAGCCGCCGA	GCCAGCCTCC GAGGCAGTCC	CTGCCACCAC GGCCGAGTCT	GCCTCCGGGG CAGCCCCTGT CACCCAGGTG		
Primer X5R <- GCT CAGACTTCCG	TCGTCTCTAG-5'					
spą prąs lpp	wrs lppq	qpl spp	rspl pqq	spllepIrpl		
E P A A A A V S A T	LAEPAAR	A A P I T P K	EPTTRA	V P S A R A H P A A		
GAGCCCGCAG CCGCGGCAGT CTCTGCCACC	CTGGCGGAGC CTGCCGCCCG	GGCAGCCCCT ATCACCCCCA	AGGAGCCCAC TACCCGGGCA	GTCCCCTCTG CTAGAGCCCA TCCGGCCGCT		
eqs lapq q c q	pll gglp	lqqpmq	vhws qep	qhs qllp plq		
GAVP GAPAMS	ASARAAA	ARAAYAG	PLVWGA	RSLSATPAAR		
GGAGCAGTCC CTGGCGCCCC AGCAATGTCA	GCCTCTGCTA GGGCAGCTGC	CGCTAGGGCAG CCTATGCAG	GTCCACTGGT CTGGGGAGCC	AGGTCACTCT CAGCTACTCC CGCCGCTCGG		
h p f l p a q q l p	pqqpllp	a g s l l a	q q p l p p p	aqpildppap		
ASLPARAAAA	ARAASAA	RAVAAGR	SASAAP	SRAH LRP PSP		
GCATCCCTTC CTGCCCGCGC AGCAGCTGCC	GCCCGGGCAG CCTCTGCTGC	CCGCGCAGTC GCTGCTGGCC	GGTCAGCCTC TGCCGCCCCC	AGCAGGGCCC ATCTTAGACC CCCCAGCCCC		
tst sgaersr	lltr llr	gll rgrl	pglaaa	gtr hrla sar		
DKYE RGREIQV	ADPPTP	R P P P R P T	AWPSCCF	RYEA SSGICE		
GACAAGTACG AGCGGGGCCG GAGATCCAGG	TTGCTGACCC GCCTACTCCG	CGGCCTCCTC CGCGGCCGAC	TGCCTGGCCT AAGCTGCTGC	AGGTACGAGG CATCGTCTGG CATCTGCGAG		
ssp pvm s r k k	g p p a a s s	gfc get	a a l a c p g	atr sgat q sa		
IESSSDESEE	GATG CFQ	WLLRRNR	RPGLPR	SHTV GSN PVR		
ATCGAGTCCT CCAGTGATGA GTCGGAAGAA	GGGGCCACCG GCTGCTTCCA	GTGGCTTCT GCGGCGAAACC	GCCGCCCTGG CCTGCCCCGG	AGCCACACGG TCGGGAGCAA CCCAGTCCGC		
tss pepseaa	s v y p s v p	dhd psa	pgrpril	wrr ganr cak		
NFFTRAFGSC	FGLSECT	RSR SLSP	GKAKDP	MEER RKQ MRK		
AACTTCTTCA CCCGAGCCTT CGGAAGCTGC	TTCGGTCTAT CCGAGTGTAC	CCGATCACGA TCCCTCAGCC	CCGGGAAGGC CAAGGATCCT	ATGGAGGAGA GGCGCAAACA GATGCGCAAA		
Primer XL3F AGCTGC TTCGGTCTAT CCGAGTGTA ->						
k p l r c e s r s a	q i r n a a s	sst sn w	rrrr wtt	cvh taccf*		
EAIEMREQKRA	A D K K R S K	LIDKQLE	EEKMDYN	1 C T H R L L L G		
GAAGCCATTG AGATGCGAGA GCAGAAGCGC	GCAGATAAGA AACGCAGCAA	GCTCATCGAC AAGCAACTGG	AGGAGGAGAA GATGGACTAC	ATGTGTACAC ACCGCCTGCT GCTTCTAGgt		

aatqcq.... non-coding sequence - splices to exon XLb (AK030489) or to exon 2 of Gnas1 (AF116268)

Fig. 2. The genomic sequence of mouse distal chromosome 2 encodes extended XL and Alex domains, XXL and AlexX, that are potentially much longer than suspected. Analysis of the mouse genomic DNA sequence (GenBank accession no. NT\_039211, region 3844806-3847406), which encodes the XL and Alex domains of the Gnas1 complex locus, shows the existence of ORFs of 785.3 and 725 codons that run in parallel and encode in their 3' end the murine XL and Alex domains of the 422.3-codon XL and 391-codon Alex domains reported in GenBank's AF116268 record. Shown are the nucleotide sequence and, above, the amino acids encoded in the two ORFs. Nucleotides in uppercase are from AF116268 (cDNA), as well as from NT\_039211 (genomic sequence); nucleotides in lowercase are from NT\_039211 only; nucleotides in italics are from intronic sequences. Amino acids in uppercase indicate the XL domain and its potential N-terminal extension, referred to as XXL; amino acids in lowercase indicate Alex and its potential N-terminal extension, referred to as AlexX (Fig. 4). Also shown, in bold and uppercase characters, are the forward (F) and reverse (R) PCR primers XXF, XF, X3F, and X5R (Fig. 4), which were used to test for the existence of the XXL/AlexX domains.

aries of the mouse  $Gs\alpha$  gene, we found that the intron that separates exons 9 and 10 in the human Gs $\alpha$  gene is absent from the murine gene (Fig. 5). As a consequence, the coding sequence of mouse  $Gs\alpha$  has only 12 exons. The exon 9 sequences of DNA obtained from 129Sv, FVB/N, BALB/c, and C57BL/6J mice have been deposited in the GenBank database with accession numbers AY518306, AY518305, AY518307, and AY518308.

It is curious that the full-length sequence of the proteins

encoded in the XXL exon has gone undetected for so long. We therefore searched the databases for ESTs and partial cDNAs that had as their origins the genomic region upstream of presumed Met-1 of XL $\alpha$ s. These searches returned one rat cDNA (GenBank accession no. AF093569) encoding a protein of unknown function, but whose N-terminal 70 aa are those of the rat AlexX shown in Fig. 3. Its sequence then diverges with alternating stretches of XXL and AlexX, ending with Alex. In addition, we found four ESTs, BY135711, BY785179,

A	
mus_XXL rat_XXL hum_XXL	1 MGMFNCLHGNNMSGQHDIPPEVGEQPEQEPLEAPGAAAPGAGAGPAEEMATEPDSEPSNNEPVPDETGSEISGPPEDSKSDIQSPCQAFEEVRVGGDYSPPPEEAMPFETQQPSLGDFWP 1 M
mus_XXL rat_XXL hum_XXL Consensus	121 TLEQPG.PSGTPSGLQAFNPAILEPGTPTGASPGLGAYTPPPEEAMPFEFNEPAQGDHSQPPLQVPDLAPGGP.EALVPRALPAEPGNIRFENAGFREDYSPPPEESVPFQVGGEEFGGD 121 TLEQPG.PSGTPGLQAFNPAILEPGTPTGASPGLGAYTPPPEEAMPFEFNEPAQGDHSQPPLQVPDLAPGGP.EALVPRALPAEPGNIRFENAGFREDYSPPPEESVPFQVGGEEFGGD 
mus_XXL rat_XXL hum_XXL Consensus	239 SPPPGLPRVIPQIGIGGEPPTVAVPSALCLAPAENAPPLWVRGAIDRPFREAVRSPPNFACDSPPMEITRPLLEIGRASIGVDDDTAVNMDSPPIASDGPPIEVSGAPDKSECAERPPVE 224YIS
mus_XXL rat_XXL hum_XXL <b>Consensus</b>	359 REAAEMEGSP.TTATAVEGKVPSPERGDGSSTQPEAMDAKPAPAAQAVSTGSDAG.APTDSAMLTDSQSDAGEDGTAPGTPSDLQSDPEELEEAPAVRADPDGGAAPVAPA. 43 - QTCN-ISST-E-ALE-PTVHIV-ES-T-SSK-DPTHAVL-IGPE-V-GV-TM-TP-ASEDAG-DET-ATESEDNREPA 56 EAADA-EGYA-SA-T-ARAAA-ADS.GAT-EST-AD-DSGAF 361 - AAEAAVPSG-GSPA
mus_XXL rat_XXL hum_XXL <b>Consensus</b>	468TPAESESEGSRDPAAEPASEAVPATTAESASGAAPVTQVEPAAAVSATLAEPAARAAPITPKEPTTRAVPSARAHPAAGAVPGAPAMSASARAAAAAAAAAAAYAGPLVWGAR 458 AAAABEPAAEPA - PAA - PAA - PAAE
mus_XXL rat_XXL hum_XXL Consensus	578 SLSATPAARASLPARAAAAARAASAARAVAAGRSASAAPSRAHLRPPSPEIQVADPPTRPPPRPTAWPDKYERGRSCCRYEASSGICEIESSSDESEEGATGCFQWLL.RRNRRPGI 578
mus_XXL rat_XXL hum_XXL Consensus	595 PRSHTVGSNPVRNFFTRAFGSCFGLSECTRSRSLSPGKAKD.PMEERRKQMRKEAIEMREQKRADKKRSKLIDKQLEEEKMDYMCTHRLLL 785.3 codons# 595
	# spliced to Gnasl Exon 2 or to Exon XXLb and then to Gnasl Exon 2 (see B below)
В	Exon XXL Exon XXLb (AK030489)
mus_XXLas mus_XXLb1 mus_XXLb2	776 YMCTHRLLLL   splices to Gnasl Exon 2 776 YMCTHRLLLL   GRKVVPSDTE GRYRPEASAS ASDRRLDRRG REVSPELLGW ALRGSPGSIV RDRGGLGPSG CAPPPRLARL LRLSQLVVGV 776 YMCTHRLLLL   GRKVVPSDTE GRYRPEASAS ASDRRLDRRG RE- splices to Gnas Exon 2
	Evon XXLb Gnasl Evon 1 Evon 2 Gnasl Evons 3 trough 12

		Exon XXLb	1	Gnasl Exon 1	Exon 2	Gnasl Exons 3 trop	ugh 12	
Gnas1	1			MGCLGN // HRLLLL	GAGESGKSTI VKQMRILHVN GFNGE	GGEE // HLRQYELL*	394	aa
XXLas	786		approx.	30 kb	GAGESGKSTI VKQMRILHVN GFNGE	GGEE // HLRQEYLL*	1133	aa
XXLb1	866	CWCPFSVFAC	A*		· · · · · · · · · · · · · · · · · · ·	·	876	aa
XXLb2	817		appr	ox. 30 kb	VLESLTKAPL*		827	aa

DNAS PN

mus

mus mus

С	
mus_AlexX	$1\ {\tt MSPSPTRLAVRSVDPOKTPNLTSKAPARPSRKSEWVETTAHLRRKPCHSRHNSPAWEISGPPWSSODH.LGPHQASKPSTQRFWSPGPPLARAQAWEPIPPHQKKLCHLSSTSLPRETIA$
rat_AlexX	1
hum_AlexX	1 MMA-PRS-DP-FRSST-H-G-L-FM-ALQ-P-LALHL-SPV-SMLRA-HSE-SVVPDDTALPLAAV
consensus	IK-VUFF-IIK-5-KB-IAUD-KK-C-3K-N5FANB-3C-W55-UC-VR-F5-Y-N5F-FVANB
mus AlexX	120 SLPCKSOTLROEVR.KHWSPELFPRSPGTSDLKTLASEKTTALPLKNLCHFRSVEKNSGAIAHPODSRESSHKSALAASSROSRSVRSASLPPRTRLPSGSEAPLTDHSARLSDLLLTS
rat_AlexX	120GGRR
hum_AlexX	114 NFSYRLAQVLQGLKP-G-QRP-EARRLSLMDL-TTRGFP-LTAPPQSAA-PK
Consensus	121SQTL-QSPEL-PRSPS-LLA-ETAL-LLCHEKGAHPEHKSAASSRQSR-RVRSASLPPRTRLPSGS-APHRLSD-LLTS
mus AlexX	239 HATAPRWRSPDPCLRLAEPPLGSTTTPLSIWTAPOSOVMARPSKSREPOIRASAORDPOLSEKOPRWKEAL, PPPLRWKEKSPLRREGTDL, PPSLKOWMPSOPLL, PKPSLPDLMLEL
rat AlexX	238 -T-F-QPSS-EK-PNCISLRS-ERTKPIRTQ
hum_AlexX	234 R-AGRSALTRLTR-PE-ERRQP-MREYPLGQP-RI
Consensus	241P-WRSPDPRLA-PPLGSTT-P-S-WTAPQSARPS-S-EPQ-R-S-QRDP-LKQ-RLK-PDPPQLPLP
Mug Mleyy	
rat AlexX	357 HTTOFFRSVIL-KSE-SOPC-PIFFRLIKM-A-M-GONOTEGOPO-P-PS-KT-N-PP-P-SOP-OP-SOPS-D-PSOLT-KOLT-KPS-KR
hum AlexX	331
Consensus	361P-Q-L-PGP-P-Q-LP
mun Many	
mus_AlexA	402 PEOPLOPPRSFLEGOPLIDETREDBOSIDAPQUQUPLIAQUPLIQUPROVINSGBPGRSQLIDPPLGHFFLEMPULPROQPLIPAQUILPAQULAQUPL PPAOPTLIDPPAPRSKLIJIKILIKGI
hum AlexX	380 TRT-OPI-TRPITPOOMPG
Consensus	481 -PG-PP-P-P-P-GPIPPG-PG-P-LPLPP-QPLQPLPP-AI-DPPA-RSRLRLLRGL
580 LRGRL	PGLTSTSGAEAAAGTRHR, LASARSSPPVMSRKKGPPAASSGF, CGETAALACPGATRSGATQSATSSPEPSEAASVYPSVPDHDPSAPGRP, RILWRRGANRCAKKPLRCESRS
hum AlexX	555
Consensus	601 L-R-PGA-AAAA-R-R-RW-R-A-RCAKKP-RRS-ABP
mus_AlexX	697 AQIRNAASSSTSNWRRRRWTTCVHTACCF* 726 aa
rat_AlexX	712 741 aa
Consensus	570
COND DATE OF	The second se

Fig. 3. Amino acid sequence alignments of the extended mouse, rat, and human XXL (A) and AlexX (B) domains as deduced on the basis of RT-PCR analysis of mouse brain transcripts and mouse, rat, and human genomic DNA sequences accessible at www.ncbi.nih.gov. Amino acid sequences are based on: (i) the RT-PCR analysis reported in this study (Fig. 4); (ii) GenBank cDNA records X84047 (rat XLas), AF116268 (mouse XLas), AK030489 (mouse XLb), and NM\_080425 (human XLαs); and (iii) orthologous sequences found in genomic contigs NT\_039211 (mouse), NT\_011362 (human), and NW\_047679 (rat). (A) Alignment of XXL amino acid sequences; (B) Amino acid alignment comparing variations in sequence in the region coded for by exons lying between the end of the XXL exon and Gnas1 exon 2; (C) Alignment of AlexX amino acid sequences. The key to alignments follows. First line: master sequence to which the other sequences are compared. Second and third lines: -, identical to the master sequence; uppercase letter, different from the master sequence. Consensus sequence (bold): uppercase letters, amino acids identical in all sequences compared; -, different by at least one from the master sequence. \*, Stop codon.

BX387357, and BX444740, that had template RNAs derived from the region immediately surrounding the ATGs of mouse XXL and AlexX. It would seem that the secondary structure of the RNA has prevented the cloning of the full-length cDNAs of the XXL/AlexX domains. The sequence alignments shown in Fig. 3 indicate a remarkable divergence in sequence identity for the XXL and AlexX domains (as was already true for XL and Alex) when the comparison is made among species. This



Fia. 4. scripts derived from the 5' end of the extended XXL/AlexX ORFs depicted in Fig. 3. (A) Diagrams of proven and potential XL-related transcripts and approximate location of primers used to test for their existence. Primers XXF, XF, X3F, and X5R are defined in Fig. 3. Other primers were as follows: XLb1R, <-3'-CCAAT GGAAC ACTAA ACGCG GTAC-5'; Ex2R, <-3'-CTCAG ACCGT TT-TCG TGGTA TGGTA ACAC-5'; Ex12R, <-3'-ACCTG TGACT CTTGT AGGCG GCACA-5', Blue diagonal numbers indicate codons of ORFs. (B) Agarose gel electrophoresis of PCR products: XXL and AlexX exist, and XXL splices to Gnas1 exon 2 either directly, giving XXL $\alpha$ s, or via an intermediary alternative exon, XLb, giving rise to XXLb2. Note that the PCR analysis shows the presence of a splice variant that gives rise to XXLb2, whose mRNA differs from that of XXL $\alpha$ s in that it incorporates 95 nt of the XXLb exon. Lanes: 1-5, PCR products obtained by using reverse-transcribed single-stranded cDNAs as templates; 6-12, PCR products obtained by using the reaction products of a mock single-stranded cDNA synthesis reaction containing the RNA template but obtained by omitting the reverse transcriptase. Absence of products indicates absence of contaminating DNA that could have been used as template in the reactions shown in lanes 2-6.

divergence contrasts with the almost absolute conservation of the Gs $\alpha$  domain sequences and raises questions as to the functions (if any) that the XXLb and AlexX proteins may have.

# A Murine gene

forward PCR primer: 5'-CCTC TTTGT TCTAG GACCT GCTTC --> in intron 8-9

0 Intron 8-9...tccggttc cagCATGTTC GATGTGGGCG GCCAGCGCGA E R R K W I Q C F N D V T A I I F TGAGCGCCCC AAGTGGATCC AGTGCTTCAA TGATGTGACT GCCATCATCT Ex9 V S v N М R E D A S S N 0 180 TCGTGGTGGC CAGCAGCAGC TACAACATGG TCATTCGGGA GGACAACCAG nt E N K S N N 0 A L ACTAACCGCC TGCAGGAGGC TCTGAACCTC TTCAAGAGCA TCTGGAACAA

CAGgtgcgtg gggt...Intron 9-10

reverse PCR primer <-- GTTGG ACATC CAGAC GTCGA ACG-5' in intron 9-10

# B Human gene

			PI F D	V G G	DA3
Intr	on 8-9	.ctccccacca	gCATGTTTGA	CGTGGGTGGC	59
QRDE	RRK	WIQ	CFNd		nt
CAGCGCGATG	AACGCCGCAA	GTGGATCCAG	TGCTTCAACG	gtaggatgct	
gtgggcttgg	ctgttcgtaa	agaacgcttt	gcttctgtgt	tgttagggat	Int
cagggtcgct	cagggtcgct	getcacgete	ttggctttgc	tctctttggt	9-10
DVT	AII	FVV	ASSS	Y N M	
taagATGTGA	CTGCCATCAT	CTTCGTGGTG	GCCAGCAGCA	GCTACAACAT	Ex10
VIR	EDNQ	TNR	LQE	ALNL	121
GGTCATCCGG	GAGGACAACC	AGACCAACCG	CCTGCAGGAG	GCTCTGAACC	nt
FKS	I W N	N R			
TCTTCAAGAG	CATCTGGAAC	AACAGgtttg	tggagtgacc.	Intron	10-11

**Fig. 5.** Nucleotide sequence of the region coding for  $Gs\alpha$  amino acids Met-221 through Arg-280, which in *Homo sapiens* are encoded in exons 9 and 10, reveals them to be encoded in a single exon in *Mus musculus*. (A) The genomic region of interest was amplified by PCR (forward and reverse primers as shown) using genomic DNA from an FVB/N mouse and sequenced by the dideoxynucleotide chain termination method. Shown is the complete sequence of the amplified DNA and the deduced amino acid sequence encoded in the fully spliced RNA. This sequence (only partially shown) spans from the middle of exon 7 to the middle of the intron that follows the exon 9–10 coding region. The sequence shown for the murine gene was identical for DNA from 1295V, FVB/N, BALB/c, and C57BL/6J mice. (*B*) Human Gnas1 nucleotide and deduced amino acid sequences (in blocks of 10): uppercase, coding (exonic); lowercase italic, noncoding (intronic).

Given that the XL and Alex domains interact (6, 8), the existence of the XXLb proteins raises the possibility that the availability of XXL $\alpha$ s for signal transduction can be regulated by the relative abundance of the message coding for the adenylyl cyclase-stimulating XXL $\alpha$ s protein versus message coding for the Alex-binding, but adenylyl cyclase-indifferent, XXLb proteins. Whether cells at steady state accumulate AlexX in preference over XXL $\alpha$ s or XXLb will depend not only on the levels of mRNA but also on the relative turnover of these proteins. Finally, the data predict the existence of AlexX/XXLb dimers, whose function and subcellular locations also need to be explored. In Fig. 6 we summarize the expression and splice patterns uncovered in this report (Fig.



Fig. 6. Update of splice patterns (A) and translation products (B) under control of the XXL promoter.

64) and present an updated list of mRNAs not previously recognized in the available GenBank database (Fig. 6B). XXL $\alpha$ s, AlexX, XXLb1, and XXLb2 replace the XL $\alpha$ s, Alex,

- 1. Weinstein, L. S. (2001) J. Clin. Endocrinol. Metab. 86, 4622-4626.
- Ischia, R., Lovisetti-Scamihorn, P., Hogue-Angeletti, R., Wolkersdorfer, M., Winkler, H. & Fischer-Colbrie, R. (1997) J. Biol. Chem. 272, 11657–11662.
- Kehlenbach, R. H., Matthey, J. & Huttner, W. B. (1994) *Nature* 372, 804–809.
  Kozasa, T., Itoh, H., Tsukamoto, T. & Kaziro, Y. (1988) *Proc. Natl. Acad. Sci.* USA 85, 2081–2085.
- Bray, P., Carter, A., Simons, C., Guo, V., Puckett, C., Kamholz, J., Spiegel, A. & Nirenberg, M. (1986) *Proc. Natl. Acad. Sci. USA* 83, 8893–8897.
- Klemke, M., Kehlenbach, R. H. & Huttner, W. B. (2001) EMBO J. 16, 3849–3860.
- 7. Kehlenbach, R. H., Matthey, J. & Huttner, W. B. (1995) Nature 375, 253.
- Freson, K., Jaeken, J., Van Helvoirt, M., de Zegher, F., Wittevrongel, C., Thys, C., Hoylaerts, M. F., Vermylen, J. & Van Geet, C. (2003) *Hum. Mol. Genet.* 12, 1121–1130.
- Cattanach, B. M., Peters, J., Ball, S. T. & Rasberry, C. (2000) Hum. Mol. Genetics 9, 2263–2273.
- 10. Skinner, J. A., Cattanach, B. M. & Peters, J. (2002) Genomics 80, 373-375.

and XLb mRNAs depicted in Fig. 1 and are expressed in addition to the Gnas1 (Gs $\alpha$ ), 1A RNA, Nesp55, and Nespas RNAs shown.

- 11. Weinstein, L. S., Yu, S., Warner, D. R. & Liu, J. (2001) Endocr. Rev. 22, 675–705.
- Hayward, B. E., Kamiya, M., Strain, L., Moran, V., Campbell, R., Hayashizaki, Y. & Bonthron, D. T. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 10038–10043.
- Hayward, B. E., Moran, V., Strain, L. & Bonthron, D. T. (1998) Proc. Natl. Acad. Sci. USA 95, 15475–15480.
- Peters, J., Wroe, S. F., Wells, C. A., Miller, H. J., Bodle, D., Beechey, C. V., Williamson, C. M. & Kelsey, G. (1999) *Proc. Natl. Acad. Sci. USA* 96, 3830–3835.
- Li, T., Vu, T. H., Zeng, Z. L., Nguyen, B. T., Hayward, B. E., Bonthron, D. T, Hu, J. F. & Hoffman, A. R. (2000) *Genomics* 69, 295–304.
- Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Lab. Press, Plainview, NY), 2nd Ed.
- Fantom Consortium and Riken Genome Exploration Project Phase I and II Teams (2002) Nature 420, 563–573.
- Holmes, R., Williamson, C., Peters, J., Denny, P. & Wells, C., RIKEN GER. Group, and GSL Members (2003) *Genome Res.* 13, 1410–1415.

ZNAS PNAS PI