The FASEB Journal express article 10.1096/fj.03-1378fje. Published online April 1, 2004.

Gene structure and organization of the human β-secretase (*BACE*) promoter

Kumar Sambamurti,* Rachel Kinsey,* Bryan Maloney,† Yuan-Wen Ge,† and Debomoy K. Lahiri†,‡

*Medical University of South Carolina, Charleston, SC, USA; †Departments of Psychiatry and ‡Medical and Molecular Genetics, Institute of Psychiatric Research, Indiana University School of Medicine, Indianapolis, IN, USA

Corresponding author: Debomoy K. Lahiri, 791 N. Union Dr., Indianapolis IN 46206, USA. E-mail: dlahiri@iupui.edu

Present address for Rachel Kinsey: GSBS, University of Manchester, Manchester, M13 9PL, UK.

ABSTRACT

The first step in the generation of the amyloid-β peptide (Aβ) deposited in the brains of patients with Alzheimer's disease (AD) is the processing of the larger Aβ precursor protein (APP) by an integral membrane aspartyl protease named the β-site APP-cleaving enzyme (BACE). We present the genomic organization of the BACE gene. BACE mRNAs are synthesized as nine exons and eight introns from a 30.6 kb region of chromosome 11q23.2−11q23.3. Regulation of BACE may play an important role in regulating the levels of Aβ produced and is therefore likely to play an important role in AD. Herein, we report the cloning and detailed analysis of 3765 nucleotides of the promoter region of BACE and 364 nucleotides of the 5′ untranslated region of the BACE mRNA (5′ UTR). Characteristic "CAAT" and "TATA" boxes are absent within 1.5 kb of the transcription start site (TSS). The promoter region and 5′ UTR contain multiple transcription factor binding sites, such as activator protein (AP)1, AP2, cAMP response element binding protein (CREB), estrogen responsive element (ERE), glucocorticoid responsive element (GRE), "GC" box, nuclear factor (NF)-κB, signal transducer and activator of transcription (STAT)1, stimulating protein (SP)1, metal-regulatory elements, and possible Zeste binding sites. Limited interspecies similarity was observed between the human sequence and corresponding genomic DNA from the rat and mouse sequences, but several transcription factor-binding sites are conserved. Thus, the BACE gene contains basal regulatory elements, inducible features and sites for regulated activity by various transcription factors. These results identify the important regions for functional analysis of the binding domains and neuron-specific expression (1). Such a study will allow us to further examine the possible role of changes in the promoter of BACE in AD pathogenesis.

Key words: dementia • aging • genomic • Alzheimer's disease • amyloid

The amyloid plaque that is invariably deposited in Alzheimer's disease (AD) is composed of an ~42-amino acid peptide (amyloid-β peptide, Aβ), which is derived by cleavage of the larger Aβ precursor protein (APP) by proteases referred to as β and γ secretase (2). β-Secretase cleaves APP on the amino side of Aβ, producing a large secreted derivative (sAPPβ) and an Aβ-bearing C-terminal fragment (CTFβ) of 99 amino acids that is subsequently cleaved by γ-secretase to release Aβ. CTFβ accounts for <10% of the cleaved APP. However, APP is primarily processed within the Aβ sequence by the alternative α-secretase pathway to a large secreted fragment (sAPPα) and a C-terminal fragment of 83 amino acids (CTFα). The vast majority of Aβ secreted by the brain ends at amino acid 40 (Aβ 40) but ~10% is the longer Aβ 42 found in amyloid deposits. The combination of the limiting β-secretase processing (<10%) and Aβ 42-specific γ-secretase processing (<10%) makes Aβ 42 a minor product of APP metabolism, accounting for <1% of synthesized APP. However, studies on familial forms of AD (FAD) have demonstrated that Aβ 42 consistently increases in the families carrying the mutations. This has led to the amyloid hypothesis, which states that Aβ forms toxic aggregates that accumulate in AD and trigger the observed neurodegeneration and progressive dementia (3, 4).

The first posttranslational step in the generation of Aβ is the processing of APP by β-secretase to sAPPβ and CTFβ. Because <10% of APP is processed by this pathway, it is a rate-limiting step in the biogenesis of both Aβ 40 as well as Aβ 42 (5, 6). An FAD mutation on APP that replaces wild-type amino acids KM flanking the N-terminal end of Aβ with NL invariably increases the fraction of APP processed by β-secretase and increases Aβ 40 as well as Aβ 42. Thus, the enzyme β-secretase appears to play a key role in the pathogenesis of AD and is an important therapeutic target for lowering Aβ.

β-Secretase was identified by five groups as a novel membrane-bound aspartyl protease and is also called as the β site APP cleaving enzyme 1 (BACE1) or membrane aspartyl protease (Memapsin2) (7−11). These early studies demonstrated that BACE expression was primarily in the brain and the pancreas and was very low in other tissues. Furthermore, they determined that BACE was primarily located in the Golgi and endosome compartments, where APP processing to sAPPβ and CTFβ was previously mapped. Although a homologue of BACE1**,** called BACE2**,** was identified by bioinformatics almost immediately after the discovery of BACE1, knockout mice lacking *BACE1* show a complete loss of β-secretase activity, indicating that BACE1 is the major, if not only, β-secretase responsible for Aβ production in the brain. Moreover, because BACE1 expression is limited to neuronal cells, Aβ production is well correlated with BACE1 expression (12). (Unless stated otherwise, BACE1 is referred to as "BACE" in the remainder of this report.)

This neuron specificity in expression as well as potential for the regulation of Aβ levels by regulation of *BACE* expression have led us to undertake a detailed analysis of the human *BACE* transcriptional unit. Herein, we report the cloning and detailed analysis of 3765 bp of the promoter region of the human *BACE* gene and 364 bp of the 5′ untranslated region of the *BACE* mRNA (5′ UTR).

MATERIALS AND METHODS

Isolation of the human *BACE* promoter-containing (BACEP) genomic clones

To amplify the human *BACE* 5′ flanking DNA, we performed a basic local alignment search tool (BLAST) (13) search of the high-throughput genome sequencing (HTGS) database against the 5′ end of BACE cDNA (GenBank accession number BC036084) (14). Initial studies identified an unordered sequence of genomic sequences, which were obtained in two BAC clones, CMB9-8M6 and RP11-794I-11. By aligning the two sequences against the *BACE* cDNA, we were able to align CMB9-8M6 with nucleotides 715–805 of the *BACE* cDNA clone. Because the databases were being constantly expanded, we repeated the search several times and identified RP11-677N11 (GenBank accession number AP003731), a BAC clone of chromosome 11, with the *BACE* sequence that lined up with the 5′ UTR region of the *BACE* cDNA starting from DNA base 9. Based on the sequence surrounding the 5′ region of the *BACE* gene, we designed several primers for amplification of the putative promoter (summarized in Table 1). Using the primers BACE1F (GGAATTCAGATCTTTGTTAGGGAGGTCTTCTTC—underlined bases indicate *Bgl*II site used in later cloning) and BACE1R (GGAATTCGTCGACTGGTGGCTTCTCAGGAGAG—underlined bases indicate *Sal*I site used in later cloning) and the Failsafe PCR kit (Epicentre, Madison, WI), we successfully amplified a 2.4 kb fragment containing the *BACE* promoter region in one of the buffer systems (buffer K) provided (Fig. 1). We further optimized this system by gradient PCR to obtain an optimum yield of the PCR product tagged with a *Bgl*II site on one side and a *Sal*I site on the other.

Cloning of the BACEP fragment into SEAP-based vector

The promoter sequence was unidirectionally subcloned into a promoterless secreted alkaline phosphatase (SEAP) reporter pBL12/PL/SEAP vector (15) to produce plasmid p2.4BACE-SEAP (Fig. 1). This construct was transiently transfected into Chinese hamster ovary (CHO), human embryonic kidney (HEK293), and rat N2A cells. The activity obtained was detectable but very low compared with the cytomegalo virus (CMV) promoter-driven SEAP-positive control. To further extend this promoter, we added a *Bgl*II site at the 5′ end and PCR-amplified an overlapping region at the N-terminus using primers BACE2F (GAATTCAGATCTCTCACTGCAACCTCTATCTC) and BACE2R (GAAAGCAAAGGAATCATTAG). The PCR product of 2.2 kb was digested with *Aat*II and *Bgl*II and inserted into the *Aat*II/*Bgl*II sites of p2.4BACE-SEAP to produce p4.1BACE-SEAP (Fig. 1).

Large-scale plasmid DNA amplification and purification

Plasmid DNA was prepared and purified by the Genelute Plasmid Maxi-Prep (Sigma, St. Louis, MO) kit as per the manufacturer's instructions. RNA-free plasmid DNA was used for restriction enzyme digestion, subcloning, sequencing, and transfection in different cell lines.

DNA sequencing

The 5′ flanking region of the BACE gene (4.1 kb) in plasmid p4.1BACEP-SEAP was sequenced via primer walking strategy using synthetic oligonucleotides either in sense or antisense orientation (Table 2). The sequencing strategy for the entire region has been summarized in

Figure 4. DNA sequencing was performed by BigDye cycle sequencing and read on the ABI 3700 (Applied Biosystems, Foster City, CA).

Analysis of DNA sequence

The sequence was assembled with the *BioEdit* package (16) and separately by the *CAP Sequence Assembly Machine* (17). Both assemblies were in agreement with each other. The contig generated was used for further analysis. The sequence was examined with the TESS utility (18) and *MatInspector* (19) for possible transcription factor binding sites and was likewise used for BLAST probing of GenBank (13) and examined with the EMBOSS analysis suite (20) for repeats. The BACEP sequence was also aligned by *ClustalX* 1.8 (21, 22) with rat and mouse genomic sequences 4.1 kb upstream to the translation start for these species' respective *BACE* genes (14), and *ClustalX* or BLAST was also used to generate alignments of internal repeats.

RESULTS

Molecular cloning of the BACE promoter region

The *BACE*-containing BAC clone RP11-677N11 was identified by BLAST searching of the RP1146 human genomic library, and the *BACE* promoter was PCR-amplified with either oligomers BACE1F and BACE1R or BACE2F and BACE2R (Table 1; Fig. 1). Two clones of different sizes were further analyzed. The larger clone, p4.1BACEP-SEAP, contained a 4.1 kb fragment of the BACE 5′ flanking region. The smaller clone, p2.4BACEP-SEAP, contained a 2.4 kb fragment of the BACE 5′ flanking region that was coterminous at the 3′ end with the 4.1 kb fragment. Restriction enzyme digest and DNA sequencing (Fig. 5) (GenBank accession number AY542689) have permitted the construction of a partial restriction enzyme map for the sequence (Fig. 2).

Genomic organization of *BACE*: structure of intron-exon regions

Reported *BACE* cDNA clones were either ~2.5 kb or 5.9 kb long. BLAST alignment of the *BACE* cDNAs reveals that these clones vary in the length of their 3′ UTR. The two *BACE* mRNAs are synthesized as nine exons and eight introns from a 30.6 kb region of chromosome 11q23.2-11q23.3 (Fig. 2). Additional introns were not observed in this region of BACE. However, note that a third mRNA has been seen in Northern blots of *BACE* (8, 10). Search of the EST database suggested that longer cDNAs exist that terminate further to the 3′ end of the genomic sequence, although boundaries have not been defined. The presence of multiple mRNAs that terminate at different sites is considered to be due to different possible polyadenylation sites being processed. (8, 10)

In addition to BACE proteins of 501 amino acids, three alternatively spliced forms of BACE of 432, 457, and 476 amino acids have also been described in literature. These forms arise by alternative and adventitious splicing of exons 3 and 4 and deletion of 25, 44 or 69 amino acids between the two conserved active site domains of BACE. These alternatively spliced BACE forms were previously found to be inactive against the APP substrate (23).

The boundaries and sizes of the nine coding exons are highly conserved, as would be expected, between mice and humans. In contrast, there is greater variation in the size of introns (Table 3), as routinely observed for several other genes. Notably, the first intron is quite large (13.6 kb in mice and 18.6 kb humans). Brief analysis of the first intron for possible transcription factor binding sites with TESS and *MatInspector* reveals thousands of possible sites, including nearly 1200 putative sites shared with the BACE promoter region. Notably, the intron has at least five predicted ORFs, two in forward and three in reverse orientation. One of these putative ORFs has a predicted 72% homology with a 375 amino acid segment of human neuronal thread protein (data not shown). Furthermore, compared with a 5′ UTR of 446 bp, the 3′ UTR is very large in both mice (2148 bp) and humans (3880 bp) (Table 3).

Transcription start site

The *BACE* open reading frame (ORF) is synthesized in mRNAs that fall into two size categories 2.2–2.5 kb (8, 24) and 5.9 kb (9). Among the multiple cDNAs cloned and sequence reported in literature, three presented the sequence starting from its 5′ UTR (8, 9, 24). These clones begin 449, 452, or 457 bp from the translation start site. BLAST search of the EST library available at NCBI revealed no longer clones. This suggests that the TSS is 457 bp from the translation start site and is the convention for determining the +1 position in this report (Fig. 2*B*). Note that while we were preparing this report, primer extension experiments have indicated that an additional TSS may exist 691 bp upstream of the initial "ATG" codon (25).

Chromosomal localization and genomic alignment

MegaBLAST database search of the GenBank human genomic sequence (NCBI Build 34) with the BACEP sequence as a probe demonstrated matches with the previously published human *BACE* mRNA (8) and with human chromosome 11q clones CMB9-94B1 (GenBank accession number AP000892) and CTD-2336M4 (GenBank accession number AP001822). Genomic MegaBLAST against the human genome placed the sequence at the complement of positions 20730273–20734562 on chromosome 11 and between chromosome bands 11q23.3 and 11q23.2 (Fig. 3). A 163 bp region from the GenBank human *BACE* genomic sequence is not found in our BACEP sequence. Eleven substitutions and two single-base insertions also appear in our sequence in comparison with the genomic sequence (Table 4). None of these single-base differences correspond to previously characterized SNPs. Alignment with the "variant a" prototype GenBank mRNA sequence (GenBank accession number NM_012104) revealed that the BACEP sequence ends 81 bp upstream of the human *BACE* gene "ATG" translation start (Fig. 2*B*). While this paper was in preparation, a shorter segment of the *BACE* gene 5′ flanking region has been published (25) with which our BACEP sequence has 98% homology (Table 4).

Structural characterization of the 5′ flanking region

A total of 4129 bp of *BACE* 5′ flanking promoter containing DNA was sequenced, including the majority of the 5′ UTR and the region's characteristics were mapped (Fig. 4 and 5). The 5′ end of the sequence has one tandem repeat of two 159 bp segments with 84% mutual homology (Fig. 6*A*); the sequence also has an inverted repeat of two 274 bp segments with 85% mutual homology, separated by 1.8 kb of sequence (Fig. 6*B*). The upstream portion of this inverted repeat also takes part in a triple direct repeat, which has 74% homology over aligned length.

Each element of this triple repeat is also part of an inverted repeat with a 271 bp segment near the 5′ end of the BACEP sequence. The homology of this quadruple repeat is 67% (Fig. 6*C*). Analysis of the entire BACEP region revealed an overall GC content of 47%. A 50 bp running average of GC content varies between 14% and 90% along the length of the sequence, reaching maximum levels at the 3′ end of the sequence (Fig. 7).

Transcription factor binding sites

Whereas TATA-like sequences appear in the far-upstream region of the sequence, canonical "CAAT" and "TATA" boxes are not found within 1.5 kb of the TSS (Fig. 5). Sequence analysis of the promoter region and 5′ UTR reveals the presence of multiple potential transcription factor binding sites. In the 5′ UTR (+1 to +364), one ACE-like sequence (26), three activator protein (AP)2 sequences, one AP4 sequence, one apolipoprotein AI regulatory protein (ApoAI RP)-1 sequence, one hypoxia-inducible factor helix-loop-helix region (bHLH) site, four GC boxes, two IL-6 responsive element sequences, one olfactory neuron-specific factor (ONSF) sequence, and eight stimulating protein (SP)-1 binding sites are found (Table 5). The immediate promoter region and 5′ flanking region (−3765 to −1) has, among other sites, one ACE-like sequence; one each of AP1, -2, -3, and -4 sequences; one ApoAI RP sequence; two bHLH sites; one CCAAT/enhancer-binding protein (C/EBP) site; one cAMP response element binding protein (CREB) site; more than 30 possible GATA binding sites; one GC box; two glucocorticoid receptor sites; one hypoxia-induced factor (HIF) site; three heat shock factor (HSF) sequences, thirteen octamer-binding sites; four potential NF-κB sites; two progesterone response element sites; four sites involved in retinoic acid recognition; eight Smad3 binding sites; one upstream stimulatory factor (USF) site; and eleven putative Zeste binding sequences (Table 5).

Comparison of DNA sequence among species

Comparing the human sequence with the corresponding genomic DNA from the rat and mouse sequences showed varying amounts of interspecific similarity, approaching or exceeding 90% within ±100 bp of the TSS but rapidly falling to <50% for most of the sequence before −150 bp, although there is a brief segment of >80% homology between −2000 bp to −1850 bp (Fig. 8). This suggests that the region from −150 bp and downstream may play an important role in *BACE* gene expression and that it could be regarded as a common proximal region for all mammalian *BACE* genes.

The 163 bp gap found when aligning the BACEP sequence with the human genomic sequence did not appear when aligning BACEP with mouse and rat genomic sequences. Further investigation revealed that this 163 bp segment has no homologue in the rodent sequences, as a gap of the same length appears in the same location when aligning mouse and rat genomic sequences from the database-derived human genomic sequence (data not shown). Several of the putative transcription factors found by *TESS* search of the human BACEP sequence also turned up in mouse and rat sequences, notably, AP2, a sequence similar to copper-responsive element-binding sequence (CLS), CREB, MZF1, GATA bindng factor (GATA)-1, SP1, and USF. The rat *BACE* sequence also preserves an upstream inverted NF-κB binding site in a location similar to the human sequence (Table 6).

DISCUSSION

The first and rate-limiting step in APP processing to Aβ is cleavage by BACE at the N-terminal end of the Aβ sequence to produce CTFβ of 99 amino acids (27). This cleavage occurs preferentially in neurons and accounts for only a small fraction of processed APP. However, Aβ accumulates to very high levels in the AD brain and apparently plays an important role in its pathogenesis (3). Consistent with its expression in the brain, *BACE* mRNA is preferentially expressed in neurons and in the pancreas (8). Thus, BACE is an important target for drug development and for understanding the processing of APP and brain-specific deposition of the Aβ peptide. In addition to this protein, a homologue known as BACE2 (ALP56, DRAP) has also been discovered. This gene is on region of chromosome 21 critical for Down's syndrome (DS) (28). BACE2 is not measurably expressed in brain tissue (29) except in the brains of those with DS (30), and its primary physiological activity is currently uncharacterized, although it may play a role in amyloid myelopathies (31, 32).

Characterization of the 5′ flanking region of the *BACE* gene as reported herein is important to understand its expression in different cell types and brain tissue regions. For this, we isolated the BACE genomic clone by screening a large RP1146 (BAC-based) genomic library. We used the published 5′ UTR sequence of BACE mRNA (8) to probe the human genomic sequence database (14) and observed matching sequences on chromosome 11. Structural characterization of the 5′ flanking region of the *BACE* gene was carried out by DNA sequencing, which reveals several interesting features.

DNA sequencing reveals that our clone consists of 4129 bp region, including the majority of the 5′ UTR as determined by comparison with published mRNA sequences (GenBank accession number NM_012104) and with primer extension experiments (25). The proportion of GC-enrichment varies between 14% and 90% along the length of the sequence, reaching maximum levels at the 3′ end of the sequence. Notably, we observed a 163 bp gap in our BACEP sequence in comparison with the human genomic sequence. Alignment with the published mRNA sequence revealed that the BACEP sequence ends 81 bp upstream of the human *BACE* gene "ATG" translation start. It displays characteristics of a housekeeping gene. For example, canonical "CAAT" and "TATA" boxes are not found within 1.5 kb of either possible transcription start site, a trait that is very similar to the *APP* promoter in many species (33–35). The promoter region and 5′ UTR contain multiple potential transcription factor binding sites, including AP1, AP2, one GC box, and ten SP1 binding sites in the 5′ UTR as well as a CREB-binding site, a glucocorticoid receptor binding site, and possible Zeste binding sites upstream of the TSS. We also detected NF-κB and several potential CLS sites in the promoter. Of particular note, three putative GATA-1 sites were found in the 5′ UTR at locations homologous to those found in the rat promoter (36). For two of these sites in the rat sequence (homologous to sequences beginning at +222 and +327 in the BACEP sequence), it has been shown that deletion will increase expression in a reporter assay (36), although this is yet to be repeated for the human sequence.

An active SP1 site has been experimentally identified and characterized in the human *BACE* promoter (25). This sequence begins in our BACEP sequence at −1158. Our sequence differs from the characterized SP1 sequence by only a single G→A transition ("GGGCGG" vs.

"GGGCAG"), and flanking regions are 98% homologous within 50 bases of either side of the site. Therefore, the previously reported SP1 activity is consistent with our sequence.

Overall, the implication of our studies related to expression and regulation of the *BACE* gene should be considered in the following perspective. BACE is the first enzyme that has been clearly identified as one of the secretases involved in the processing of APP in the mammalian brain. It is present in large quantities in the neurons. Like APP, BACE is a type I integral membrane glycoprotein. The BACE protein has a signal sequence, followed by a short prodomain, a catalytic domain, a single transmembrane domain, and a short cytoplasmic tail; the prodomain of BACE can be cleaved by furin in the secretory pathway (7). It is unlikely that BACE evolved primarily to make the potentially pathogenic Aβ peptide; it may have some other physiological function(s). This possibility is now strengthened by our results showing that the structure of BACE promoter resembles with that of a housekeeping gene because it lacks the characteristics of type II promoter, such as "CAAT" and "TATA" boxes, within 1.5 kb of the TSS.

We determined that region of greatest homology with characterized *BACE* promoter regions of other species begins around −150, which agrees with previously published results (36). In addition, we determined some homology upstream of previously published alignments, specifically between −2000 to −1850. Of particular interest is that an MZF1 putative zinc finger protein-binding site that spans the human TSS is preserved in rat and mouse sequences. Others have reported that mutating the MZF1 site did not affect basal expression in the rat gene (4). Although the BACEP sequence contains several putative cytokine binding to sites, mentioned herein, cell culture induction studies have shown that *BACE* mRNA production is not increased by treating neuronal or astrocytoma cell cultures with tumor necrosis factor (TNF)-α, interleukin (IL)-1β, IL-6, transforming growth factor (TGF)-β, or several other cytokines (37). However, the *APP* regulatory region has been shown to interact with cytokines such as IL-1α, TGF-β1, and TNF-α (38, 39). This suggests that *BACE* and *APP* regulation are distinct pathways and that it may be possible to control expression of one gene without significantly altering expression of the other except as BACE and APP directly interact. Alternatively, these cytokine-binding sites may normally be blocked and activating them may be part of an overall inflammation cascade that leads to large-scale Aβ cleavage from APP.

Overall, the *BACE* promoter displays inducible characteristics in that it contains potential binding sites for several important transcription factors, such as AP1, AP2, CRE, GRE, and NF-κB. Four of the five NF-κB sites predicted in BACEP appear in some segment of the triple direct repeat, suggesting a possible regulatory role for these repeats. The activity of the majority of these putative sites remains to be investigated. This is an important supplement to other recent study of regulation of BACE at the protein level. For example, BACE activity in the cell was shown to be regulated by several components of ordered lipid microdomains (lipid rafts), including caveolins (40, 41) and GPI anchored proteins (42). Whether some of these components interact with the regulatory region of the *BACE* promoter needs to be studied.

In addition to a possibly active promoter region of more than 3 kb in length, the *BACE* gene's first intron (of eight) is nearly 61% of the total length of the gene. As we have reported, it

contains thousands of putative transcription factor binding sites, including many that also appear in the BACEP sequence. We found five possible ORFs, although none have been shown to actually be expressed. It may be possible that such a large intron plays a role in expression and regulation of the *BACE* gene and may itself consist of several functional domains.

Taken together, our data show that the *BACE* gene contains basal regulatory elements, inducible features, and sites for regulated activity by various transcription factors. This study enables us to identify important regions for functional analysis of binding domains and neuron-specific expression (1). This work is also an important supplement to recent study of regulation of BACE at the protein level. These studies will allow us to further examine the possible role of changes in the *BACE* promoter and AD pathogenesis.

ACKNOWLEDGMENTS

This work was supported by grants from the National Institutes of Health (AG18379 and AG18884) and the Alzheimer's Association (IIRG2056 and IIRG3062).

REFERENCES

1. Ge, Y-W., Maloney, B., Sambamurti, K., and Lahiri, D. K. (April 1, 2004) Functional characterization of the 5′ flanking region of the *BACE* gene: identification of a 91 bp fragment involved in basal level of *BACE* promoter expression. *FASEB J.* 10.1096/fj.03-1739fje

2. Lahiri, D. K., Farlow, M. R., Sambamurti, K., Greig, N. H., Giacobini, E., and Schneider, L. S. (2003) A critical analysis of new molecular targets and strategies for drug developments in Alzheimer's disease. *Curr. Drug Targets* **4,** 97-112

3. Hardy, J., and Selkoe, D. J. (2002) The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science* **297,** 353-356

4. Hardy, J. A., and Higgins, G. A. (1992) Alzheimer's disease: the amyloid cascade hypothesis. *Science* **256,** 184-185

5. Citron, M., Oltersdorf, T., Haass, C., McConlogue, L., Hung, A. Y., Seubert, P., Vigo-Pelfrey, C., Lieberburg, I., and Selkoe, D. J. (1992) Mutation of the β-amyloid precursor protein in familial Alzheimer's disease increases β-protein production. *Nature* **360,** 672-674

6. Cai, X. D., Golde, T. E., and Younkin, S. G. (1993) Release of excess amyloid β protein from a mutant amyloid β protein precursor. *Science* **259,** 514-516

7. Lin, X., Koelsch, G., Wu, S., Downs, D., Dashti, A., and Tang, J. (2000) Human aspartic protease memapsin 2 cleaves the β-secretase site of β-amyloid precursor protein. *Proc. Natl. Acad. Sci. USA* **97,** 1456-1460

8. Vassar, R., Bennett, B. D., Babu-Khan, S., Kahn, S., Mendiaz, E. A., Denis, P., Teplow, D. B., Ross, S., Amarante, P., Loeloff, R., et al. (1999) B-secretase cleavage of Alzheimer's

amyloid precursor protein by the transmembrane aspartic protease BACE. *Science* **286,** 735-741

9. Sinha, S., Anderson, J. P., Barbour, R., Basi, G. S., Caccavello, R., Davis, D., Doan, M., Dovey, H. F., Frigon, N., Hong, J., et al. (1999) Purification and cloning of amyloid precursor protein β-secretase from human brain. *Nature* **402,** 537-540

10. Yan, R., Bienkowski, M. J., Shuck, M. E., Miao, H., Tory, M. C., Pauley, A. M., Brashier, J. R., Stratman, N. C., Mathews, W. R., Buhl, A. E., et al. (1999) Membrane-anchored aspartyl protease with Alzheimer's disease β-secretase activity. *Nature* **402,** 533-537

11. Hussain, I., Powell, D., Howlett, D. R., Tew, D. G., Meek, T. D., Chapman, C., Gloger, I. S., Murphy, K. E., Southan, C. D., Ryan, D. M., et al. (1999) Identification of a novel aspartic protease (Asp 2) as β-secretase. *Mol. Cell. Neurosci.* **14,** 419-427

12. Zhao, J., Paganini, L., Mucke, L., Gordon, M., Refolo, L., Carman, M., Sinha, S., Oltersdorf, T., Lieberburg, I., and McConlogue, L. (1996) B-secretase processing of the β-amyloid precursor protein in transgenic mice is efficient in neurons but inefficient in astrocytes. *J. Biol. Chem.* **271,** 31407-31411

13. Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., and Wheeler, D. L. (2003) GenBank. *Nucleic Acids Res.* **31,** 23-27

14. Wheeler, D. L., Church, D. M., Federhen, S., Lash, A. E., Madden, T. L., Pontius, J. U., Schuler, G. D., Schriml, L. M., Sequeira, E., Tatusova, T. A., et al. (2003) Database resources of the National Center for Biotechnology. *Nucleic Acids Res.* **31,** 28-33

15. Berger, J., Hauber, J., Hauber, R., Geiger, R., and Cullen, B. R. (1988) Secreted placental alkaline phosphatase: a powerful new quantitative indicator of gene expression in eukaryotic cells. *Gene* **66,** 1-10

16. Hall, T. A. (1999) BioEdit; a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **41,** 95-98

17. Huang, X. (1996) An improved sequence assembly program. *Genomics* **33,** 21-31

18. Schug, J., Overton, G. C. TESS: Transcription Element Search Software on the WWW. CBIL-TR-1997-1001-v0.0, 1-10. 2-26-1998. Computational Biology and Informatics Laboratory, School of Medicine, University of Pennsylvania, University of Pennsylvania.

19. Quandt, K., Frech, K., Karas, H., Wingender, E., and Werner, T. (1995) MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. *Nucleic Acids Res.* **23,** 4878-4884

20. Rice, P., Longden, I., and Bleasby, A. (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* **16,** 276-277

21. Jeanmougin, F., Thompson, J. D., Gouy, M., Higgins, D. G., and Gibson, T. J. (1998) Multiple sequence alignment with Clustal X. *Trends Biochem. Sci.* **23,** 403-405

22. Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G., and Thompson, J. D. (2003) Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res.* **31,** 3497-3500

23. Tanahashi, H., and Tabira, T. (2001) Three novel alternatively spliced isoforms of the human β-site amyloid precursor protein cleaving enzyme (BACE) and their effect on amyloid β-peptide production. *Neurosci. Lett.* **307,** 9-12

24. Strausberg, R. L., Feingold, E. A., Grouse, L. H., Derge, J. G., Klausner, R. D., Collins, F. S., Wagner, L., Shenmen, C. M., Schuler, G. D., Altschul, S. F., et al. (2002) Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. *Proc. Natl. Acad. Sci. USA* **99,** 16899-16903

25. Christensen, M. A., Zhou, W., Qing, H., Lehman, A., Philipsen, S., and Song, W. (2004) Transcriptional regulation of BACE1, the β-amyloid precursor protein β-secretase, by SP1. *Mol. Cell. Biol.* **24,** 865-874

26. Thiele, D. J. (1988) ACE1 regulates expression of the Saccharomyces cerevisiae metallothionein gene. *Mol. Cell. Biol.* **8,** 2745-2752

27. Esler, W. P., and Wolfe, M. S. (2001) A portrait of Alzheimer secretases--new features and familiar faces. *Science* **293,** 1449-1454

28. Acquati, F., Accarino, M., Nucci, C., Fumagalli, P., Jovine, L., Ottolenghi, S., and Taramelli, R. (2000) The gene encoding DRAP (BACE2), a glycosylated transmembrane protein of the aspartic protease family, maps to the down critical region. *FEBS Lett.* **468,** 59-64

29. Bennett, B. D., Babu-Khan, S., Loeloff, R., Louis, J. C., Curran, E., Citron, M., and Vassar, R. (2000) Expression analysis of BACE2 in brain and peripheral tissues. *J. Biol. Chem.* **275,** 20647-20651

30. Motonaga, K., Itoh, M., Becker, L. E., Goto, Y., and Takashima, S. (2002) Elevated expression of β-site amyloid precursor protein cleaving enzyme 2 in brains of patients with Down syndrome. *Neurosci. Lett.* **326,** 64-66

31. Vattemi, G., Engel, W. K., McFerrin, J., Pastorino, L., Buxbaum, J. D., and Askanas, V. (2003) BACE1 and BACE2 in pathologic and normal human muscle. *Exp. Neurol.* **179,** 150-158

32. Vattemi, G., Engel, W. K., McFerrin, J., Buxbaum, J. D., Pastorino, L., and Askanas, V. (2001) Presence of BACE1 and BACE2 in muscle fibres of patients with sporadic inclusion-body myositis. *Lancet* **358,** 1962-1964

33. Salbaum, J. M., Weidemann, A., Lemaire, H. G., Masters, C. L., and Beyreuther, K. (1988) The promoter of Alzheimer's disease amyloid A4 precursor gene. *EMBO J.* **7,** 2807-2813

34. La Fauci, G., Lahiri, D. K., Salton, S. R., and Robakis, N. K. (1989) Characterization of the 5′-end region and the first two exons of the β-protein precursor gene. *Biochem. Biophys. Res. Commun.* **159,** 297-304

35. Song, W., and Lahiri, D. K. (1998) Functional identification of the promoter of the gene encoding the Rhesus monkey β-amyloid precursor protein. *Gene* **217,** 165-176

36. Lange-Dohna, C., Zeitschel, U., Gaunitz, F., Perez-Polo, J. R., Bigl, V., and Rossner, S. (2003) Cloning and expression of the rat BACE1 promoter. *J. Neurosci. Res.* **73,** 73-80

37. Satoh, J., and Kuroda, Y. (2000) Amyloid precursor protein β-secretase (BACE) mRNA expression in human neural cell lines following induction of neuronal differentiation and exposure to cytokines and growth factors. *Neuropathology* **20,** 289-296

38. Lahiri, D. K., and Nall, C. (1995) Promoter activity of the gene encoding the β-amyloid precursor protein is up-regulated by growth factors, phorbol ester, retinoic acid and interleukin-1. *Brain Res. Mol. Brain Res.* **32,** 233-240

39. Lahiri, D. K., Chen, D., Vivien, D., Ge, Y. W., Greig, N. H., and Rogers, J. T. (2003) Role of cytokines in the gene expression of amyloid β-protein precursor: Identification of a 5′-UTR-Binding nuclear factor and its implications in Alzheimer's disease. *J. Alzheimers Dis.* **5,** 81-90

40. Ikezu, T., Trapp, B. D., Song, K. S., Schlegel, A., Lisanti, M. P., and Okamoto, T. (1998) Caveolae, plasma membrane microdomains for alpha-secretase-mediated processing of the amyloid precursor protein. *J. Biol. Chem.* **273,** 10485-10495

41. Nishiyama, K., Trapp, B. D., Ikezu, T., Ransohoff, R. M., Tomita, T., Iwatsubo, T., Kanazawa, I., Hsiao, K. K., Lisanti, M. P., and Okamoto, T. (1999) Caveolin-3 upregulation activates β-secretase-mediated cleavage of the amyloid precursor protein in Alzheimer's disease. *J. Neurosci.* **19,** 6538-6548

42. Sambamurti, K., Sevlever, D., Koothan, T., Refolo, L. M., Pinnix, I., Gandhi, S., Onstead, L., Younkin, L., Prada, C. M., Yager, D., et al. (1999) Glycosylphosphatidylinositol-anchored proteins play an important role in the biogenesis of the Alzheimer's amyloid β-protein. *J. Biol. Chem.* **274,** 26810-26814

Received December 18, 2003; accepted February 23, 2004.

Table 1

Primers for BACE 5' flanking region genomic clones[a]

BACE1F	5'-GGAATTCAGATCTTTGTTAGGGAGGTCTTCTTC-3'
BACE1R	5'-GGAATTCGTCGACTGGTGGCTTCTCAGGAGAG-3'
BACE2F	5'-GAATTCAGATCTCTCACTGCAACCTCTATCTC-3'
BACE2R	5'-GAAAGCAAAGGAATCATTAG-3'

[a]At end of primer name: F, forward orientation; R, reverse orientation.

Table 2

Primers used for sequencing BACE 5' flanking region containing promoter and 5' UTR (TSS is +1)[a]

Oligonucleotide	Position	Sequence
BSP2R	201	5'-CTACATCGGCACGGCGG-3'
BSP3R	−1603	5'-TGCTGGGATTACAGGCGTGAG-3'
BSP4F	−1121	5'-CCAGGAGTTCGAGACCAGC-3'
BSP5R	−309	5'-CGGTGATTCTACACGGGACAG-3'
BSP6R	−2059	5'-GAAGACCTCCCTAACAAACCCT-3'
BSP7R	−582	5'-TGCGCCTGCGATCTGCG-3'
BSP8F	−289	5'-CTGTCCCGTGTAGAATCACCG-3'
BSP9F	−3045	5'-CCAAATGTTACCAATCTGAGAGG-3'
BSP10F	−2395	5'-GGCTAACATGGTGAATTCCCG-3'
BSP11R	−3067	5'-CCTCTCAGATTGGTAACATTTGG-3'
BSP12R	−2411	5'-CGGGAATTCACCATGTTAGCC-3'
BSP13F	−3320	5'-CTGGTCTTGAACTCCTGGC-3'
BSP14F	−1823	5'-CATCTCTATCATATCTACAATTCCC-3'
BluFron (pBLCAT3 forward primer)		5'-GTAAAACGACGGCCAGTGCC-3'
CAT-R (pBLCAT3 reverse primer)		5'-CCTGAAAATCTCGCCAAGCT-3'

[a]At end of primer name: F, forward orientation; R, reverse orientation.

Table 3

Organization of exon and intron regions of the BACE gene (length of regions in bp)[a]

Region	Human	Mouse
5' UTR	446	446
Exon 1	261	261
Intron 1	18569	13648
Exon 2	96	96
Intron 2	1524	1649
Exon 3	217	217
Intron 3	1115	1237
Exon 4	146	146
Intron 4	682	888
Exon 5	136	136
Intron 5	1239	1150
Exon 6	104	104
Intron 6	663	408
Exon 7	150	151
Intron 7	239	215
Exon 8	174	177
Intron 8	677	600
Exon 9	247	247
3' UTR	3880	2148

[a]Rat genomic sequence is incomplete and is not included here.

Table 4

Specific differences among BACE 5' flanking region sequences (TSS is +1)

Location	BACEP	Genomic	Reference 25
–3409	C	T	—
–3386	C	T	—
–3356	C	T	—
–3346	C	T	—
–2245	gap	163bp	—
–2213	A	gap	—
–2062	A	A	C
–2061	A	C	A
–1756	T	T	C
–1660	G	G	A
–1659	G	A	G
–1616	T	T	C
–1615	T	C	T
–1361	G	G	gap
–1342	T	T	C
–1322	A	A	gap
–1320	A	gap	A
–1153	A	A	gap
–1152	A	G	A
–1147	C	C	T
–1112	A	A	gap
–1011	T	T	A
–1006	C	C	A
–921	T	T	C
–787	A	A	T
–687	C	C	T
–654	A	A	G
–637	C	C	gap
–556	G	G	A
–555	G	A	G
–483	G	G	A
–482	G	A	G
–343	A	A	T
–337	A	A	G
–329	C	C	T
–259	C	C	T
–175	G	G	A
–68	T	T	C
136	C	C	T
151	G	G	A
190	A	A	G
211	A	A	G
346	T	C	C
352	C	C	A

Table 5

Transcription factor sequences in BACE 5' flanking region (TSS is +1)

Factor	Promoter/Upstream Region	Proximal	5'-UTR
ALS	−3257		64
AP-1	−1673		
AP-2		−58	94, 173, 305
AP-3	−858		
AP-4			33
ApoAI RP-1	−774		86
bHLH	−1162, −202		247
C/EBP	−2659		
CLS	−2908, −2860, −2495, −433		
CREB	−3448		
ERE	−2650, −1555		
GATA	−3752, −3750, −3268, −3267, −2948, −2821, −2740, −2739, −2550, −2544, −2527, −2526, −2448, −2447, −2445, −2248, −2247, −1974, −1972, −1844, −1837, −1822, −1753, −1737, −1633, −1631, −1629, −1317, −1315, −1313, −1232, −867, −847, −846, −845, −744, −713, −277		206, 222, 327
GC box	−1483,		98, 155, 161, 194
GRE	−2714, −2562		
HIF-1	−1903		
HSF-1	−919, −918, −404		
IL-6 RE-BP	−3742, −3700, −2336, −2177, −1587, −1453, −1178, −1043, −852, −504, −395,	−120, −63, −54	160, 232
IRF-1	−2471, −2241, −1151, −319		
IRF-3	−2891, −1803, −913		
MZF1		−2	
N-Oct-3	−3140, −3128, −3071, −2939, −2802, −2660, −2613, −2024, −1993, −1988, −1806, −1800		
NF-1	−3516, −3187, −2846, −2491, −2413, −1906, −1712, −1531, −1303, −1298, −1211, −857, −687, −686, −598, −544, −177	−35, −34, −7	57
NF-IL-2A	−3011, −3009		
NF-IL6	−3166		
NFκB	−2403, −1521, −1110, −921, −342		
Oct-R	−3154		
ONSF	−3291		13
PRE	−3000, −2155		
RAR-β	−3446, −2648, −1553		
RTR	−3333		
SF1	−634		
Smad3	−3297, −1902, −1653, −986, −686, −204, −176, −118		
SP-1	−1620, −1570, −1211, −1158, −169	−152	4, 8, 106, 138, 147, 157, 162, 229
SREBP	−3440		
SRF	−3109, −2063, −2033, −1281, −837, −538		
STAT6	−1951, −931		
TFIID	−2802, −2771, −1701		
USF	−3692		
YY1	−3115, −3037, −2540, −2515, −2223, −1538, −447		
Zeste	−3629, −2474, −2279, −2212, −1670, −1388, −1372, −968, −733, −436, −407		

Table 6

Selected transcription factor binding sites in human, mouse, and rat sequences (TSS is +1)

Transcription Factor	Consensus Sequences (18, 19)	Human	Mouse	Rat
ALS	TGNNGCTG	−3257, 64	−2739,−2635, −2307, −2274, 159	−3285, −2395, −2346, −2289, −1145, 123, 213, 242
AP-1	TGAGTCA	−1673		
AP-2	CCCAGGG, CCCGCGC	−58, 305, 94, 173	−61	−1128, 3
AP-3	TGTGGWWW	−858	−2938	−1967, −442
C/EBP	TCCTAAT	−2659	−1545	
CLS	WWWTTTGCKCR	−2908, −2860, −2495, −433		−698
CREB	TGACCTCA	−3448	−1787	
ERE	TGACCT	−2650, −1555	−2656, −2328, −1785, −965	−3088, −2806, −1390
GATA-1	CNGATVBMV	206, 222, 327	136, 202, 221, 325	127, 199, 218, 321
GC box	GGCCGC	194	190, 338	−3171, 397
Glucocorticoid Receptor	TGAACT, AGAWCAGW	−2562, −2714	−3432, −3405, −3310, −151, −3056, −363	−3494, −2051, −1703, −1608, −562, −88
MZF1	YCTCCCCAG	−2	−3	−3
NFκB	GGGRHTYYHC	−2403, −1521, −1110, −921, −342	−875	−2328
PRE	TGTTCT	−3000, −2155	−1, 157	−2500
SP-1	CCCAGCC, GGCGGG	−1620, −1211, 4, 157, 229, −1570, −1158, −169, −152, 8, 106, 138, 147, 162	105, 148, −3637, −2963	−2397, −2348, −1263, 63, 211, 202
TFIID	TATAAA, TATATA	−2771, −2802, −1701, −1701	−3709, −1742	−3479, −529, −1900, −1900
USF	CATGTG	−3692	−3624, −2544	−2931, −1969, −1462, −481
Zeste	CACTCC, TGAGAT, TGAGCG, TGAGTC	−3629, −1372, −968, −733, −2474, −2279, −1388, −436, −2212, −1670, −407	−3466, −3074, −1438, −837, 296	−3046, −2121, −3515, −1998, −1571, −1419, −958, −493, −1839, 355

Fig. 1

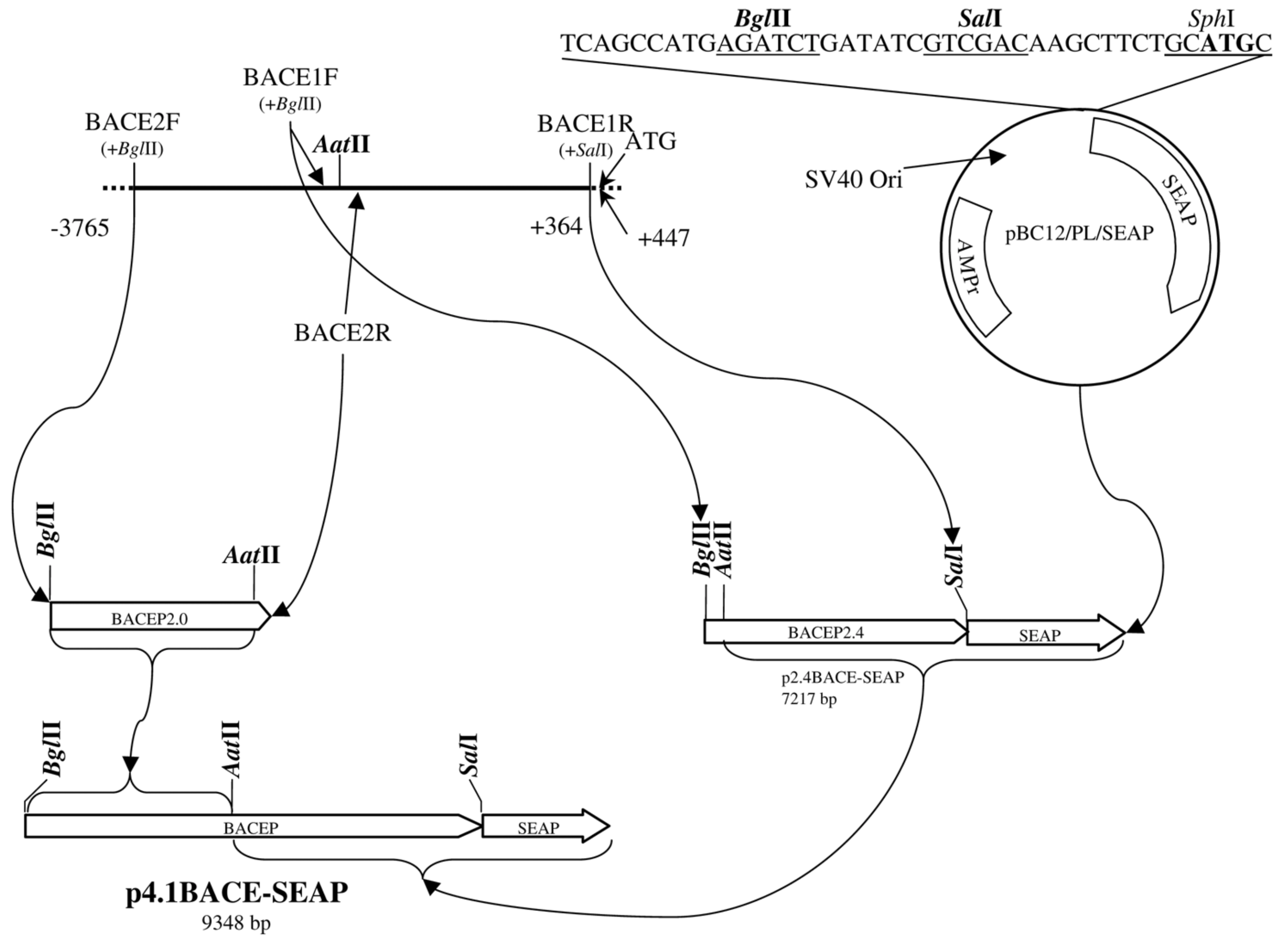

Figure 1. **Cloning of p4.1BACEP-SEAP and p2.4BACEP-SEAP from the human genomic library.** A 2.4 kb fragment of the human BACE 5' flanking region was isolated from the RP1146 library via PCR with primers that provided *Bgl*II and *Sal*I restriction enzyme sites. This was cloned into pBC12/PL/SEAP to produce p2.4BACEP-SEAP. A second fragment was isolated by PCR from the RP1146 library and cloned into the *Aat*II and *Bgl*II sites of p2.4BACEP-SEAP to produce p4.1BACEP-SEAP.

Fig. 2

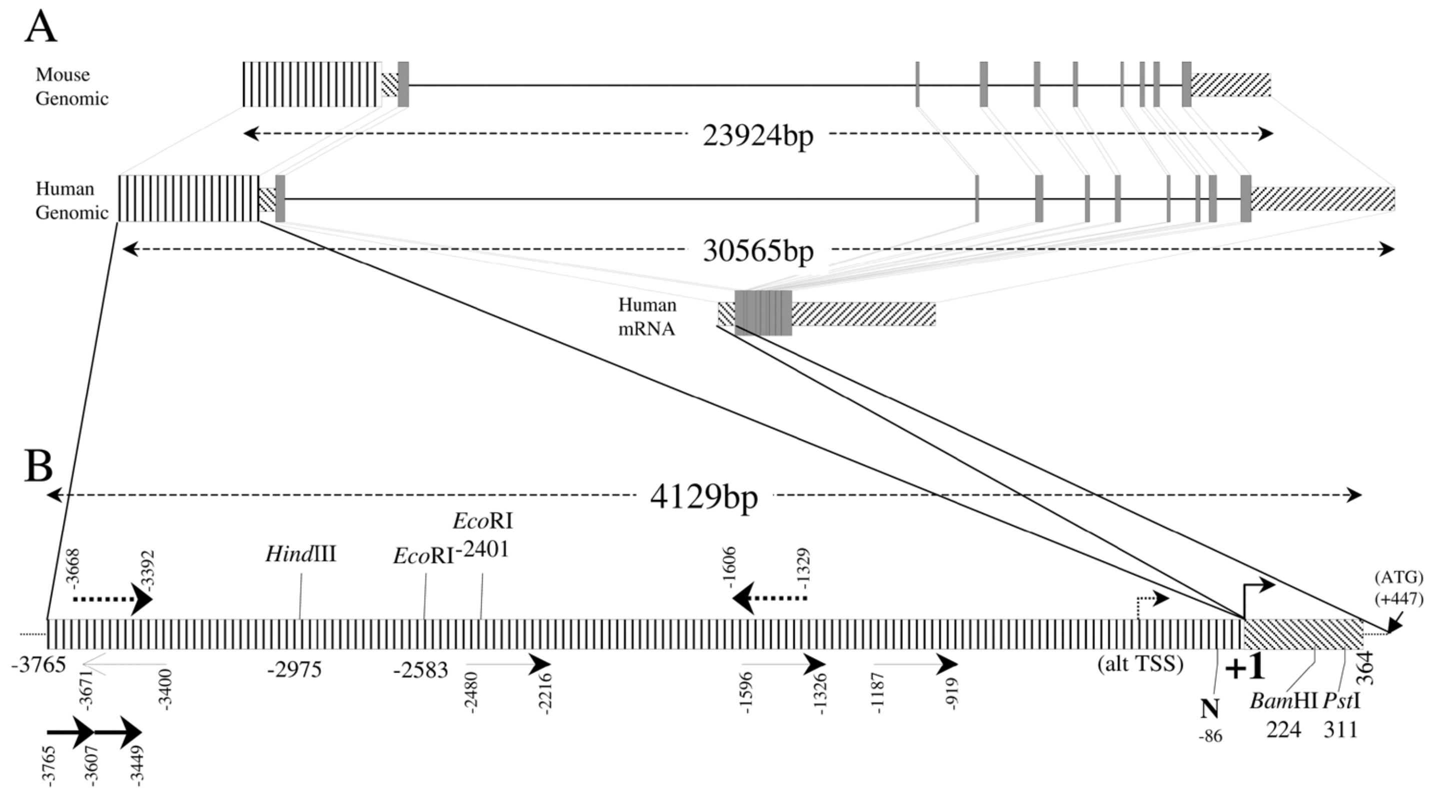

Figure 2. Diagram of *BACE* genomic organization and partial restriction map of the 5'-flanking region. ***A)*** *BACE* genomic organization for mouse, human, and human mRNA. Vertical hash indicates 5' flanking region containing the promoter. Rising cross-hatch indicates 5' UTR. Gray boxes indicate exons. Black line indicates introns. Falling cross-hatch indicates 3' UTR. ***B)*** Partial restriction map of *BACE* 5' flanking region. Vertical hash represents the promoter region and rising cross-hatch the 5' UTR. TSS is at +1, and arrow indicates the direction of transcription. Thick line arrows at 5' end of sequence represents a tandem repeat, and thick dotted lines represent an inverted repeat. Narrow lines represent triple repeat, and a narrow dashed line at the far upstream end represents a fourth sequence with lesser homology to the triple repeat. Solid bent arrow indicates TSS derived from mRNA sequences. Dotted bent arrow indicates alternate TSS from primer extension.

Fig. 3

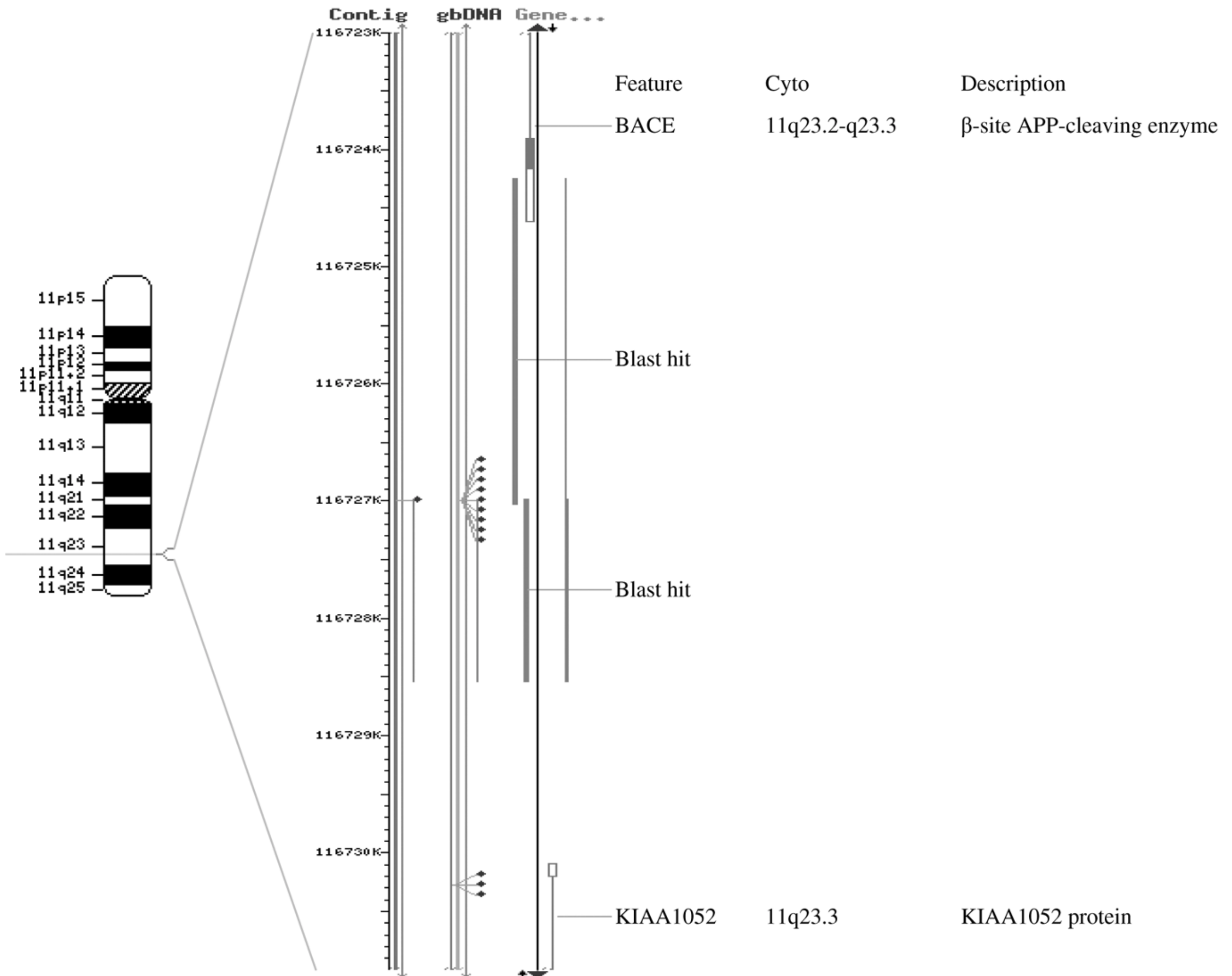

Figure 3. Genomic BLAST search. The genomic BLAST homology search against the human genome at NCBI revealed that the BACEP sequence was most homologous to a region immediately upstream of the *BACE* gene on chromosome 11, between approximately 116728.55 kb and 116724.25 kb. “gbDNA” indicates nine sequenes in GenBank homologous to this region but not used to assemble the genomic contig. The BACEP sequence produced two genomic BLAST hits, corresponding to two segments of genomic contig separated by a highly repetitive 163 bp gap (described in the text). The open bar at the beginning of the “BACE” feature corresponds to the 5' UTR of the *BACE* gene in the genomic contig, and the closed bar represents the first exon of the *BACE* gene.

Fig. 4

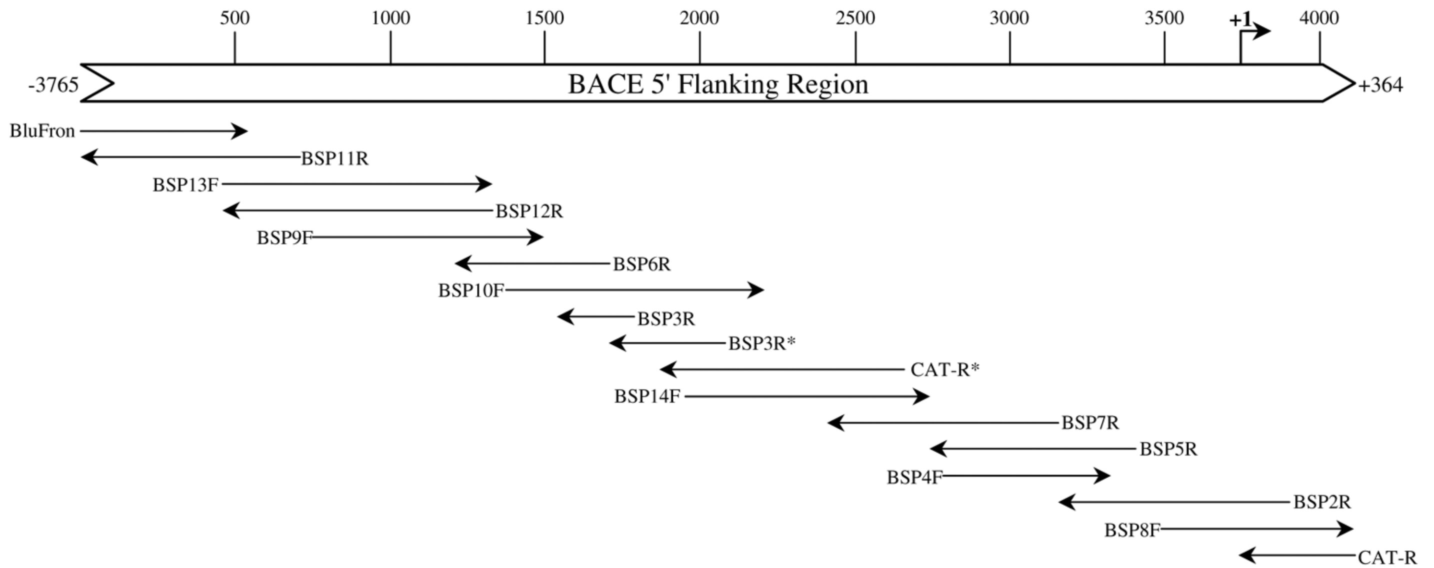

Figure 4. DNA sequencing strategy for *BACE* 5' flanking region. Primers complementary to forward and reverse sections of pBLCAT3 were used to generate initial sequences, which were then used to generate further primers. Primer names are adjacent to their locations, with respective arrows indicating either a forward-orientation primer or a reverse-orientation primer. Individual sequence results are indicated beneath the complete contig diagram. BACEP sequence is covered from end to end in both orientations. *Primers BSP3R and CAT-R were also used to sequence part of the p2.4BACE-SEAP clone, and these two arrows indicate the position of that sequence data.

Fig. 5

```
-3765                                         CTCAC TGCAACCTCT
      ATCTCCCAGG TTCAAGCGAT TCTCCTGCCT CAACCTCCCA AGTAGCTGGG
      ATTACAGGCA TGTGCCATCA CACCTAGCTA AATTTTTTTT TTTGAGACAG
              USF
      TCTTGCTCTG TCACCCAGGC TGGAGTGCAG TGGTGCAATC TCGGCTCACT
                             Zeste
      GCAACCTCCA CCTCCCGGGT TCAAGCGATT CTCCTGCCTC AGCCTCCCAA
      GTAGCTGTGA TTATGGGCAC CCACCACCAT GCCTGGCTAA TTTTTTTGTA
-3500 TTTTTAGTAG AGACAGGATT TCACCATGTT GGCCAGGCTG GTCTCAAACT
      CCTGACCTCA GGTGATCCAC CCACCTCGGC CTCCCAAAGC GCTAGGATTA
        CREB/ERE
      CAGGCTTGAT GACTATTTTT GGTTTTGCTT AAAGAGATGG GGCCTTGCTC
      TGCCTCCAGG CTGGTCTTGA ACTCCTGGCC ACAAGCGATC CTCCCACCTC
      AGACTCCCAA GTGGCTTGGA TTACAGGCGT GAACTATCAC ACCCAGCTTC
                                                        ALS
      ATTTGTTTTG AATCCAAGGT TTTGCCATAG TACACACTGC TGAAATATTG
      TTGTTTTTGT TGGAAAGGGT CCTAGAAATG CCATTGCCCA AAGAGTTTGC
      ATTTTTAACA TATTAATAGG CATTGCAAAT TACTCTCCAA AAATGGAATA
      ATTTACAGTC CAAACAATAC CTTCACCAAT ACCAAATGTT ACCAATCTGA
      GAGGAGACGA AAGAAAGAAA AAGGAACAAA GAATCTACTG TTTTACTTTG
                                              IRF-1&2
-3000 TGTTCTTTCA ACTATTAACG AAGTTAAGCT TTTTTCAAGG TTTTTTATGT
        PR                       HindIII
      TTATCATCCA TTTGAATTTC CTCTGTGAAT TAACTGTTCA CATTCTTTGT
                                                      CLS1
      TCATTTTTCT TTTTCTTTTC TTTTTTGTGT TTTGTTGTTG TTTTTCCTCA
                                                     CLS1
      CTCTTTCCCC TTACCCTTTG TTCATTTTTC TATCAGGTTG TTTGGGATTT
                                                         TF
      ATATTGATTT GTAGAAGCTC TTTGCACATT ATAAACATTA GCCCTTTGTC
      IID                              TFIID
      TCTATTGTGT GATAAATGTT GTTTCTTAAT CTGTGACTGA TCTTTTAACT
                                            GRE
      TTGTAAATGT CAGTTTTCTC AGTAGTTGCA ACAGTAATAA TCCTAATAGC
                                                   C/EBP
      TGACCTTTAT TGAGCACTTA CAATATGCTG GACACCATTC TAAATGCTTT
        ERE
      ACATGGATTG TCTCATCGAA TTCTTACCTT TTGAGGTTTG AACTATTAAT
                        EcoRI                 GluR
      ATCATTATCT CCATTTTACA AATGATAAAA TCAAGTCTCA AAAAGGGGCC
-2500 AGGCAAGTTG GCTCACGCCT GTAATCTCAT CACTTTGGGA GGCTGAGGCA
             CLS1             Zeste IFN-RF
      GGCAGATAAC TTGAGGTAAG GAGTTAGAGA CCAGCCTGGC TAACATGGTG
      AATTCCCGTC TCCACTAAAA ATACAAAAAT TAGCTGGGCG TGCTGGCGCG
      EcoRI/NF-κB
      TGCCCGTAGT CCCAGCTACT TGGGAGGCTG AGGCAGGAGA ATCGCTTGAG
      GAGGCAGAGG TTTCAGTGAG CTGAGATCAT GCCACTGCAT TCCAGCCTGG
                             Zeste
      GTGATAAAGT GAGACTGTTT CAAAAAAAAA AAAAAAGGAC TCAAAAAGTA
           IFN-RF                            Zeste
      AGTAACCAGA GTGGGCTCTC CCAGTTACTA ATTGAACCAG AATTAGAACA
                                                         PR
      CTGGTGGTAT TGGAATACTT CTGATTCTAC AGAAGTTTTA AATATTTAAG
      TCATCAAATC CATCCAGTTT TTGGGTTTGT TTACCTTATA GGGTTTGTTA
      GGGAGGTCTT CTTCAATATA AGATCATGAA AATATTCTAT TTTTCCTGTA
-2000 GTTTTATTTT AATGTTAAAT TCTTTAATCT ATCTGGAGTT TATTTTCAGA
      GGAATGAAAT TGGATTTGGT TTTTTTTCCT TCCAAATGGG CTGCCAGACG
      TCCCAACAAC AGTTTTTGAA TAATGCGTCC ATTTTCCTCT AATACGAATG
      CCATCTCTAT CATATCTACA ATTCCCATAG ATACATTTCT CTATTTTGAT
      TTCGATTTTT TTAACTTGTG TCTAATCATC TTATTGATTA GGGTTTTCTA
      TCCATATCAT ATAATCTGCA ACTAATGATT CCTTTGCTTT CCAACATATA
      TAACACTTTT TCCATTTTCT TGTCTTTTGA GTCAATCAGG CCGGTCTGAA
                                   AP-1
      CCGTGTAGGA TAAGAGTAGA GATAATGATC GGCTGGGCGC GGTGGCTCAC
                                        Sp1
      GCCTGTAATC CCAGCACTTT GGGAGGCCGA GGCGGGAGGA TCACGAGGTC
                                        Sp1             ERE
      AGGAGATCGA GACCATCTTG GCTAACACGG TGAAACCCCG TCTCTACTAA
                                         NF-κB
-1500 AAATACAAAA AATTAGCGGG GCGTGGTTGC AGGCGCCTGT AGTCCCAGCT
      ATTCGGGAGG CTGAGGCAGG AGAATAGCCT GAACCCGGGA GGCAGAGCTT
      GCAGTGAGCT GAGATCCCAC CACTGCACTC CAGCCTGGGC GACAAAGCGA
      GACGCCTTCT CAAAAAAAAA AAAAAAAAAG AGTAGAGATA ATGATCTTTC
      CTTTCCAAAG CCTGCCTTTA ATGGGAGTAA CGTAGCCTGT TATTACTACA
      TTTGTTGTTG TTGTCGTTAT CTTGTTTAAA AAGTTTCCTG GCTGGGTACA
                                                  Sp1
      GTGGTCCATG CCTATAATCC CAGCATTTTG GGAGGCCGAC GTGGGCAGAT
      CACTTGAGGC CAGGAGTTCG AGACCAGCCT GGCCAACATG GTGAAATCCC
      IFN-RF                                        NF-κB
      ATCTCTACTA AAAATACAAA AATTAGCTGG GTATGGTGGC GCATGCTTGT
      AATCCCAGCT ACTCAGAGGC TGAGGCATGA GAACTGCTTG AACCCGGGAG
-1000 GCGGAGGTTG CAGACTTGAG ACCATGCCAC TGCACTCCAG CCTGGGTGAC
                                         Zeste
 -950 AGAGCAAGTC TCCTTTTCAA AAGAAAAATG AAGATTCCCT TTCATTTTTT
                                        NF-κB
 -900 AAAAAAATTA AATCAGAGTA ACTGCTGAAT TTTATCAATG GCTTTCCACA
                                                     AP-3
 -850 TTTGATAAAC AATATATTGT TAAAACCATA CCGGCTTTTC TCCTTTAAGC
 -800 TATTAGAACT CTAACGCTTT TGGTTCAAGG CTTTAAGCTC TTACCAGTTA
 -750 ATATAGAGAT ACTTTTCGGA GTGTTAACAC TTTAAAGCTA TCAACAGCCA
                         Zeste
 -700 TTTCTCCTCA GTCTGGCTGA TGGTGCAGAA AATCAGAGGA GCCATAAGGA
 -650 TTCAAGTAGA GCCTCAAGGG CACAAACAGG TTCAGATGGG AGAAGACCCA
 -600 CTGGCTCTTA CCTGCCCGCA GATCGCAGGC GCAGGAGAAA ATAGGACAAC
 -550 TGCACTTTCC CCTTTTGGAC CCTGCTCCGC TGGTTTGCAG CCTGGAAAAA
 -500 CTCTTCCTCC CCTTTCGTCC TGCTCCCTCA ATCTCTGCTC GTGAAAATCC
 -450 TACCCATTTT TAAACGCTCA AATTAAATAC CACCTTCTCC CACTGAGTCT
                  Zeste/CLS2                       Zeste
 -400 TCCCAGATCC CCTAGATGTC CCTCCAAATC CCTACTCCTT AATGCCACAT
 -350 CGTATAGTGT GAAACCCCAC CGCGGCAGGA ATCACTTAAC TGTCCCGTGT
                NF-κB                  IRF
 -300 AGAATCACCG TTTATTTCTC TCTGTATCTC CAGCGCCTGG CAGTGTGCCT
 -250 GGCATACAGT GGGTGCTCCT TCCATGCTGA AAGAAAGACT GACAGACGGG
 -200 AGGTGTGCCC CTCTCCATCC GTCTGGCCCT TCCCGCCAGG GCCTTGCAGG
                                         SP-1
 -150 GCGGACTCCA CCTCGGCAGA GGGCATCCCA GACCCCTCTC CAGCCCCGGA
      SP-1
 -100 AGCCGGATTG CCTGCCATGG GAAGACTACA CTTCCCAGCG ATCCCAGGGA
                     NcoI                           AP-2
  -50 AAAGCGAAAA CCTTTTGGCT TTGACAGCCG CCGCCACAAG TCTTTCCGCC
    1 TCCCCAGCCC GCCCGGGAGC TGCGAGCCGC GAGCTGGATT ATGGTGGCCT
         SP-1 SP-1
   51 GAGCAGCCAA CGCAGCCGCA GGAGCCCGGA GCCCTTGCCC CTGCCCGCGC
                   ALS                                AP-2
  101 CGCCGCCCGC CGGGGGGACC AGGGAAGCCG CCACCGGCCC GCCATGCCCG
          SP-1                                SP-1      SP-1
  151 CCCCTCCCAG CCCCGCCGGG AGCCCGCGCC CGCTGCCCAG GCTGGCCGCC
           SP-1  SP-1            AP-2                 GC Box
  201 GCCGTGCCGA TGTAGCGGGC TCCGGATCCC AGCCTCTCCC CTGCTCCCGT
           GATA-1              BamHI   SP-1
                               GATA-1
  251 GCTCTGCGGA TCTCCCCTGA CCGCTCTCCA CAGCCCGGAC CCGGGGGCTG
  301 GCCCAGGGCC CTGCAGGCCC TGGCGTCCTG ATGCCCCCAA GCTCTCTCTC
        AP-2        PstI          GATA-1
  351 CTGAGAAGCC ACCA 364
```

Figure 5. Nucleotide sequence of *BACE* 5' flanking region. DNA was sequenced by the primer walking strategy (**Figure 4**). The thymidine at +1 represents the major transcription start site. The positions of some of the restriction enzymes that were used for subcloning are indicated in italics. The putative sites for binding selected transcription factors are bold if in forward orientation and underlined if in reverse orientation. GenBank accession number is AY542689.

Fig. 6

```
A
-3765 CTCACTGCAACCTCTATCTCCCAGGTTCAAGCGATTCTCCTGCCTCAACCTCCCAAGTAGCTGGGATTACAGGCATGTGC -3686
-3607 CTCACTGCAACCTCCACCTCCCGGGTTCAAGCGATTCTCCTGCCTCAGCCTCCCAAGTAGCTGTGATTATGGGCACCCAC -3528

-3685 CATCACACCTAGCTAAATTTTTTTTTTTGAGACAGTCTTGCTCTGTCACCCAGGCTGGAGTGCAGTGGTGCAATCTCGG -3606
-3529 CACCATGCCTGGCTAATTTTTTTGTATTTTTAGTAGAGACAGGATTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCT -3449

B
-3668 TTTTTTTTTTGAG-A-CAGTCTTGCTCTGTCACCCAGGCTGGAGTGCAGTGGTGCAATCTCGGCTCACTGCAACCTCCACCTCCCGGGTTCAAGCGATTCTCCTGCCTCAGCCTCCCAA -3551
-1329 AAAAAAAAAAACTCTTCCGCAGAGCGAAACAGCGGGTCCGACCTCACGTCACCACCCTAGAGTCGAGTGACGTTCGAGACGGAGGGCCCAAGTCCGATAAGAGGACGGAGTCGGAGGGCT -1448

-3550 GTAGCTGTGATTATGGGCACCCACCACCATGCCTGGCTAATTTTTTTGTATTTTTAGTAGAGACAGGATTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTGACCTCAGGTGATCCAC -3431
-1449 TATCGACCCTGATGTCCGCGGACGTTGGTGCGGGGCGATT-AAAAAACATAAAAATCATCTCTGCCCCAAAGTGGCACAATCGGTTCTACCAGAGCTAGAGGACTGGAG—CACTAGGAGG -1567

-3430 CCACCTCGGCCTCCCAAAGCGCTAGGATTACAGGCTTGA -3392
-1568 GGCGGAGCCGGAGGGTTTCACGACCCTAATGTCCGCACT -1606

C
-2480 GTAATCTCAT CACTTTGGGA GGCTGAGGCA GGCAGATAAC TTGAGGTAAG GAGTTAGAGA CCAGCCTGGC TAACATGGTG AATTCCCGTC TCCACTAAAA -2381
-1596 GTAATCCCAG CACTTTGGGA GGCCGAGGCG GGAGGATCAC ..GAGGTCAG GAGATCGAGA CCATCTTGGC TAACACGGTG AAACCCCGTC TCTACTAAAA -1499
-1187 ATAATCCCAG CATTTTGGGA GGCCGACGTG GGCAGATCAC TTGAGGCCAG GAGTTCGAGA CCAGCCTGGC CAACATGGTG AAATCCCATC TCTACTAAAA -1088
-3400 GTAATCCTAG CGCTTTGGGA GGCCGAGGTG GGTGGATCAC CTGAGGTCAG GAGTTTGAGA CCAGCCTGGC CAACATGGTG AAATCCTGTC TCTACTAAAA -3499

-2380 ATACAAAAA. .TTAGCTGGG CGTGCTGGCG CGTGCCCGTA GTCCCAGCTA CTTGGGAGGC TGAGGCAGGA GAATCGCTTG A.....GGAG GCAGAGGTTT -2288
-1498 ATACAAAAAA .TTAGCGGGG CGTGGTTGCA GGCGCCTGTA GTCCCAGCTA TTCGGGAGGC TGAGGCAGGA GAATAGCCTG AACCCGGGAG GCAGAGCTTG -1400
-1087 ATACAAAAA. .TTAGCTGGG TATGGTGGCG CATGCTTGTA ATCCCAGCTA CTCAG.AGGC TGAGGCATGA GAACTGCTTG AACCCGGGAG GCGGAGGTTG -991
-3500 ATACAAAAAA ATTAGCCAGG CATGGTGGTG GGTGCCCATA ATCACAGCTA CTTGGGAGGC TGAGGCAGGA GAATCGCTTG AACCCGGGAG GTGGAGGTTG -3599

-2287 CAGTGAGCTG AGATCATGCC ACTGCATTCC AGCCTGGGTG ATAAAGTGAG AC..TGTTTC AAAAAAAAAA AAAA -2216
-1399 CAGTGAGCTG AGATCCCACC ACTGCACTCC AGCCTGGGCG ACAAAGCGAG ACGCCTTCTC AAAAAAAAAA AAAA -1326
-990  CAG..ACTTG AGACCATGCC ACTGCACTCC AGCCTGGGTG ACAGAGCAAG TCTCCTTTTC AAAAGAAAAA TGAA -919
-3600 CAGTGAGCCG AGATTGCACC ACTGCACTCC AGCCTGGGTG ACAGAGCAAG AC..TGTCTC AAAAAAAAAA ATTT -3671
```

Figure 6. Sequences of BACEP internal repeats. ***A)*** Sequence from both segments of a 160 bp (−3765/−3606 and −3607/−3449) tandem repeat (83.6% homology between segments). ***B)*** Sequence from both 277 bp (−3668/−3392 and −1606/−1329) segments of an inverted repeat (85% homology between segments). ***C)*** *ClustalX* alignment of all three segments (−2480/−2216, −1596/−1326, −1187/−919) of a triple repeat (74% homology among segments) with the limited homology far upstream sequence (−3671/−3400) (dropping overall homology to 67% among four segments) on the bottom.

Fig. 7

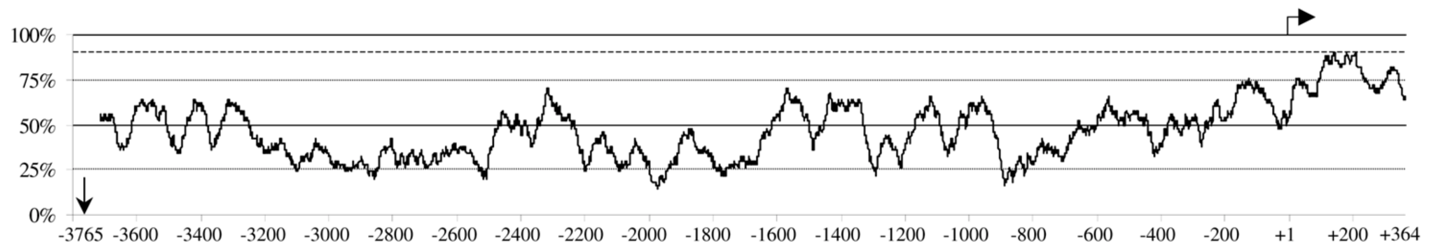

Figure 7. GC content of BACEP. Chart displays a running 50-base average of percent GC content for the 4129 bp fragment of BACEP. Bent arrow indicates TSS.

Fig. 8

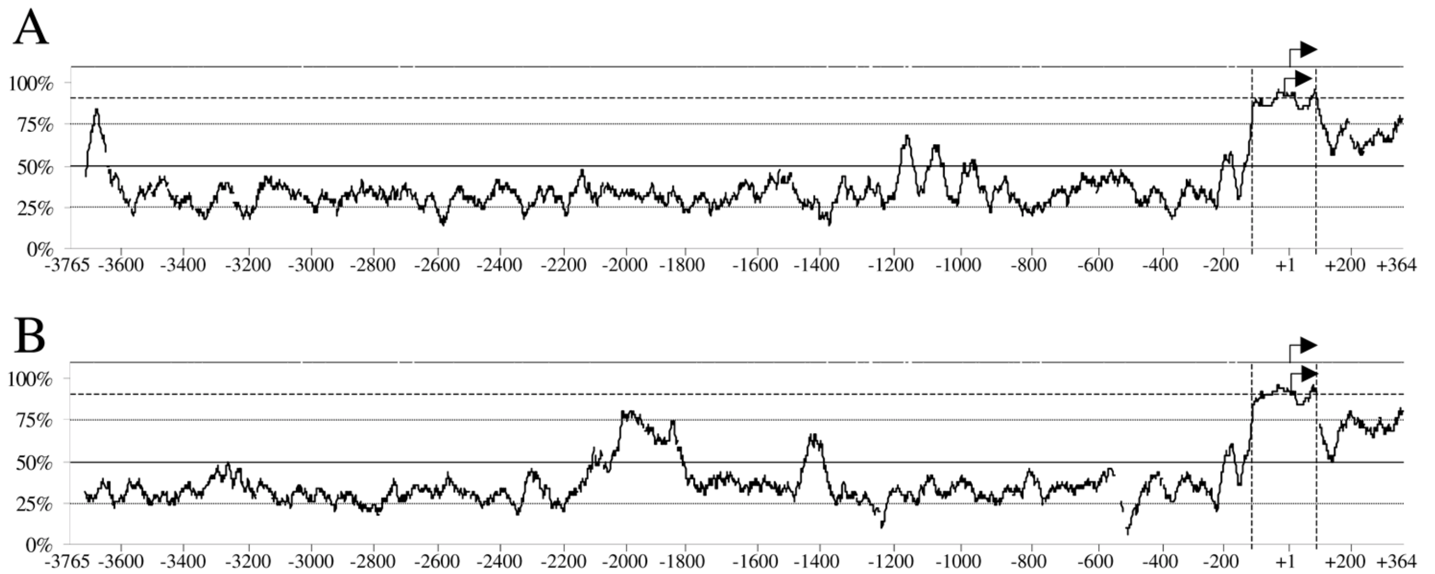

Figure 8. Alignment of the BACEP sequence against mouse and rat sequences. ***A)*** Homology of rat sequence vs. human. Variable height line indicates percent homology of rat sequence to human sequence in a 50-base window. Black gapped horizontal line above 100% indicates the human sequence. Arrow on uppermost line indicates human +1. Arrow on variable line indicates rat +1. Dashed vertical line indicates the borders of the "proximal region" of greatest homology. ***B)*** Homology of mouse sequence vs. human. Variable height line indicates divergent homology of mouse to human. Arrow on uppermost line indicates human +1. Arrow on variable line indicates mouse +1. Dashed vertical line indicates the borders of the "proximal region" of greatest homology. In most cases, mouse and rat share the same homology as the human sequence. Greater than 90% homology in a 50-base window only appears in either comparison very near the +1 transcription start site.