

Accelerated Evolution of Crotalinae Snake Venom Gland Phospholipase A₂ Isozymes

中島, 欽一
九州大学理学研究科化学専攻

<https://doi.org/10.11501/3081158>

出版情報：九州大学，1994，博士（理学），課程博士
バージョン：
権利関係：



**Structures of Genes Encoding TATA Box-binding
Proteins from *Trimeresurus flavoviridis* and
Trimeresurus gramineus.**

In *T. gramineus* and *T. flavoviridis* venom gland PLA2 isozyme genes, the protein-coding regions have been evolving at much higher rate than the noncoding regions including introns. It was noted that the evolutionary rates for introns are almost equivalent to those for all the noncoding regions, indicating that the evolutionary rates are almost equal between introns and other noncoding regions in venom gland PLA2 isozyme genes. To gain a further insight into the evolutionary feature of venom gland PLA2 isozyme genes, it is of interest to compare the evolutionary rates of introns of venom gland PLA2 isozyme genes with those of introns of general (nonvenomous) genes. TATA box-binding protein (TBP) is commonly present in all cell types and its gene could be assumed to have evolved under neutrality. Therefore, the author determined the entire nucleotide sequences of the genes encoding TBPs of *T. gramineus* and *T. flavoviridis*.

TATA box-binding protein (TBP) is a cellular protein that recognizes and binds to TATA box, and is required for all transcription by RNA polymerases I, II, and III of genes (Schultz et al., 1992; Cormac and Struhl., 1992; Sharp., 1992; Comai et al., 1992). Complementary DNAs of TBPs from various organisms have been cloned and sequenced (Apsit et al., 1993; Lichtsteiner and Tjian., 1993; Rowland et al., 1994; Marsh et al., 1994). The nucleotide sequences of TBP genes of *Arabidopsis thaliana* (Gasch et al., 1990), *Acanthamoeba castellanii* (Wong et al., 1992), *Schizosaccharomyces pombe* (Hoffmann et al., 1990), *Saccharomyces cerevisiae* (Horikoshi et al., 1989), and archaeobacteria (Rowlands et al., 1994; Marsh et al., 1994) have been determined. However, the whole nucleotide sequences of TBP genes of higher animals have not been worked out although the nucleotide sequences of cDNAs for human (Kao et al., 1990) and mouse TBPs (Tamura et al., 1991) have been determined and the loci of introns of mouse TBP gene have been assigned (Sumita et al., 1993). This chapter describes the

cloning and the nucleotide sequences of the genes encoding *T. gramineus* TBP (TgTBP) and *T. flavoviridis* TBP (TfTBP).

EXPERIMENTAL PROCEDURES

Materials. Restriction endonucleases and other enzymes were obtained from Takara Shuzo (Kyoto). [α - 32 P]dCTP (3,000 Ci/mmol; 1 Ci = 37 GBq), [α - 35 S]dATP (1,000 Ci/mmol), [γ - 32 P]ATP (5,000 Ci/mmol) and [35 S]L-Met (1,000 Ci/mmol) were from Amersham. Other reagents were of reagent grade.

Labeling of DNAs and Oligonucleotides. DNA fragments were labeled with [α - 32 P]dCTP (3,000 Ci/mmol) by the random priming method (Feinberg and Vogelstein, 1983). Oligonucleotides were labeled at the 5'-OH end with [γ - 32 P]ATP (5,000 Ci/mmol) and T4 polynucleotide kinase (Sambrook et al., 1989).

Screening of cDNA Library. A *T. gramineus* venom gland cDNA library made in λ gt10 (Chapter 3) was screened by the plaque hybridization method with human TBP cDNA (this cDNA clone was a generous gift from Dr. Arnold J. Berk; Kao et al., 1990). Hybridization was carried out in 1 mM EDTA/0.5 M NaHPO₄, pH 7.2/7% SDS (Church and Gilbert., 1984) at 42°C overnight. One positive clone was obtained, and the insert was subcloned into pUC118. The sequence was determined by the dideoxy chain termination method (Sanger et al., 1977; Hattori and Sakaki., 1986).

Screening of *T. gramineus* and *T. flavoviridis* Genomic DNA Libraries. Two genomic libraries (from *T. gramineus* and *T. flavoviridis*) made with EMBL3 (Chapters 2 and 3) were screened with full-length TgTBP cDNA using the standard method (Sambrook et al., 1989). *Trimeresurus gramineus* phage clones Tg-1, Tg-2, and Tg-7 contain fragments A and B, fragment C, and fragment D, respectively (see text). *Trimeresurus flavoviridis* phage clones Tf-17, Tf-6, and Tf-7 contain fragment E, fragments F and G, and fragments H and I, respectively (see text). These fragments were subcloned into pUC19 or 118. The DNA sequences were determined by the dideoxy chain termination method (Sanger et al., 1977; Hattori and Sakaki., 1986).

Primer Extension Analysis. Total cellular RNAs from *T. gramineus* and *T. flavoviridis* venom gland were fractionated (Chirgwin et al., 1979). An antisense 18-mer oligonucleotide primer (5'-GCTCCAGAAGTTCTCCCG-3', nucleotide positions 84 to 101 of TgTBP gene as well as TfTBP gene) was used to map the initiation site. The [³²P] end labeled primer (1 pmol) was annealed to 50 µg of total venom gland RNAs and extended with reverse transcriptase (Sambrook et al., 1989). The products were loaded onto 6% polyacrylamide-7 M urea sequencing gel and electrophoresed.

RESULTS AND DISCUSSION

Screening of *T. gramineus* Venom Gland cDNA Library. The *T. gramineus* venom gland cDNA library (Chapter 3) was screened with human TBP cDNA (Kao et al., 1990). One positive clone was obtained from 5 x 10⁵ plaques. This clone was 1778-bp long and contained a single open reading frame (Fig. 15) which encodes a protein highly homologous in the C-terminal half to TBPs of other organisms. Its nucleotide sequence contained two in-frame stop codons (TAG, nucleotides -120 to -118 and nucleotides -51 to -49) and the sequence matched well the consensus sequence flanking the ATG initiator codon (Kozak., 1986). The RNA derived from the cloned cDNA was translated in a reticulocyte lysate and [³⁵S]methionine-labeled protein gave a single band of 33-kDa in SDS-PAGE which corresponds to the size expected from the deduced amino acid sequence (data not shown). Therefore, we judged that this clone synthesizes TgTBP.

Screening of *T. gramineus* and *T. flavoviridis* Genomic DNA Libraries. *Trimeresurus gramineus* and *T. flavoviridis* genomic DNA libraries (1.6 x 10⁶ plaques each) constructed with EMBL3 as a vector (Chapters 2 and 3) were screened with full-length TgTBP cDNA (Feinberg and Vogelstein, 1983) and 13 and 20 positive clones were obtained, respectively. As shown in Fig. 16, *T. gramineus* phage clones Tg-1, Tg-2, and Tg-7 contained fragments A and B, fragment C, and fragment D, respectively. *T. flavoviridis* phage clones Tf-17, Tf-6, and Tf-7 contained fragment E,

-211 CGGCTCAGGTCGGGAGAAGCTTCTGGAGCGGTCTGGAACCTTTATTTATATACTTGATGTTTTATCAAATATGTCAATACAGACTGCAAACTAGTTTTTGGGATGCTTCAAGGAAAGCTA
 -91 CTTCTTAAGGCAGATGTTGTGAAGCTTGATTTCTGGATTTAGTTTCTGGAAGTAATCCTCATTGACTGGACACTTACTTGATTGCTACAATGGATCAAAACAACAGTTTACCACCTA
 (1) M D Q N N S L P P Y
 30 TGCTCAGGGCTTAGCATCCCCACAGAGTGAATGACTCCAGGCATCCCTATTTTCAGTCCTATGATGCCCTACGGCACTGGGCTCACACCCCAACCTGCTCAGAGTACCAACAGTTTATC
 (11) A Q G L A S P Q S A M T P G I P I F S P M M P Y G T G L T P Q P A Q S T N S L S
 150 TATCCTGGAAGAACAACAGAGGCAGCAGCAGCAGCAGCAAGCAGCAGCAGCAGTCCACTTCACAACCAACACAAGCACCATCTGGTCAAACCCACAGCTCTTCCACTCACAGACTCT
 (51) I L E E Q Q R Q Q Q Q Q A A A Q Q S T S Q P T Q A P S G Q T P Q L F H S Q T L
 270 CACGACTGCTCCTTTACCAGGAACCACACCTCTCTACCCCTCTCCAATGACTCCCATGACTCCTATAACGCCTGCCACCCCTGCTTCCGAAAGCTCCGGGATTTGTACCGCAGCTCCAGAA
 (91) T T A P L P G T T P L Y P S P M T P M T P I T P A T P A S E S S G I V P Q L Q N
 390 TATTGTGTCTACAGTAAATCTTGGTTGCAAACTTGACCTAAAAACCATTGCCCTTCGTGCTCGAAATGCTGAA TACAATCCCAAGCGTTTTCGTGCTGTAATTATGAGAATAAGAGAGCC
 (131) I V S T V N L G C K L D L K T I A L R A R N A E Y N P K R F A A V I M R I R E P
 510 ACGTACCACAGCCCTCATATTCAGCTCTGGAATAATGGTGTGACGGGAGCTAAAGTGAAGAGCAGTCCAGACTGGCTCGGAGAAAGTATGCTAGAGTTGTACAGAAGTTGGGCTTTCC
 (171) R T T A L I F S S G K M V C T G A K S E E Q S R L A A R K Y A R V V Q K L G F P
 630 TGCAAAATTCCTGGATTTTAAATTCAAAATATGGTTGGCAGTTGTGATGTAATTCCTATCAGACTTGAAGGATTGGTTCTTACACATCAACAGTTTAGCAGTTATGAACCAGAATT
 (211) A K F L D F K I Q N M V G S C D V K F P I R L E G L V L T H Q Q F S S Y E P E L
 750 GTTTCCTGGTTTAACTACAGAATGATCAAAACCAAGAATTGTACTACTCATATTTGTTTCTGGAAAGGTAGTTTAACTGGTGCAAAAGTCAGAGGTGAAATTTATGAAGCATTTGAAAA
 (251) F P G L I Y R M I K P R I V L L I F V S G K V V L T G A K V R G E I Y E A F E N
 870 TATTTATCCTATTTTAAAGGATTCAGAAAAACAACGTAACAGTTCCGTGTGCTTATTCATGAAGCTGGAGTACATTTTAAAGGTGTTAATTATGGCACAGTAATAACACACATCAG
 (291) I Y P I L K G F R K T T * (302)
 990 GACTTGGAATCCATCTTCCAGATAGTGCCTCTTCCCTGTTTTTCACGGAAGGGAACATATCACTGTGTGTTTCGTCTGCACTAAGTTACTGTGCATTGCTTTTCCGTGTTCTCACAC
 1110 AAGAAAAACAGTTTATTTATATTTCTGGCTCTTAAACAGCTAATTTTAAATCTGAGAGTCTTATGGTGGCAAGCTGGCTAACTGTATTATCCAACATTTAAACCCCTGGCGTTACA
 1230 GTATTTCTTCAACCATGGATTTTATAATCAGAAATGTTAATTTTATGCTTTTAGACCTACTTCCACATTCCCTACATTTCCAGTGTTCAGCAGGACAGAACTTTGCTAGTGAAGACAAA
 1350 TGGCTGTACATATTTCTATTTTGAGAAGCAATTCTAAATGTATAATTACTTGACTAAATAATTTGAATAGTAAATTTGTGGTCAATCTTTGCTTTTGCTGTTTAAATGTGCTAACTGC
 1470 ATGCATATAGCTAAAATAACTTGATTTTATAAGATTTTATTACAATCCAGTATTAAGTAATGTTATTTGAAATGTT(A)₁₇ 1567

FIG. 15. The nucleotide sequence and deduced amino acid sequence of TgTBP. The 1778-bp fragment encoding TgTBP is numbered +1 corresponding to the first nucleotide of the ATG start codon. The amino acid positions are numbered (in parentheses) starting at Met⁺¹. In-frame stop codons in the upstream sequence of the ATG start codon are underlined. The putative polyadenylation signal is bolded.

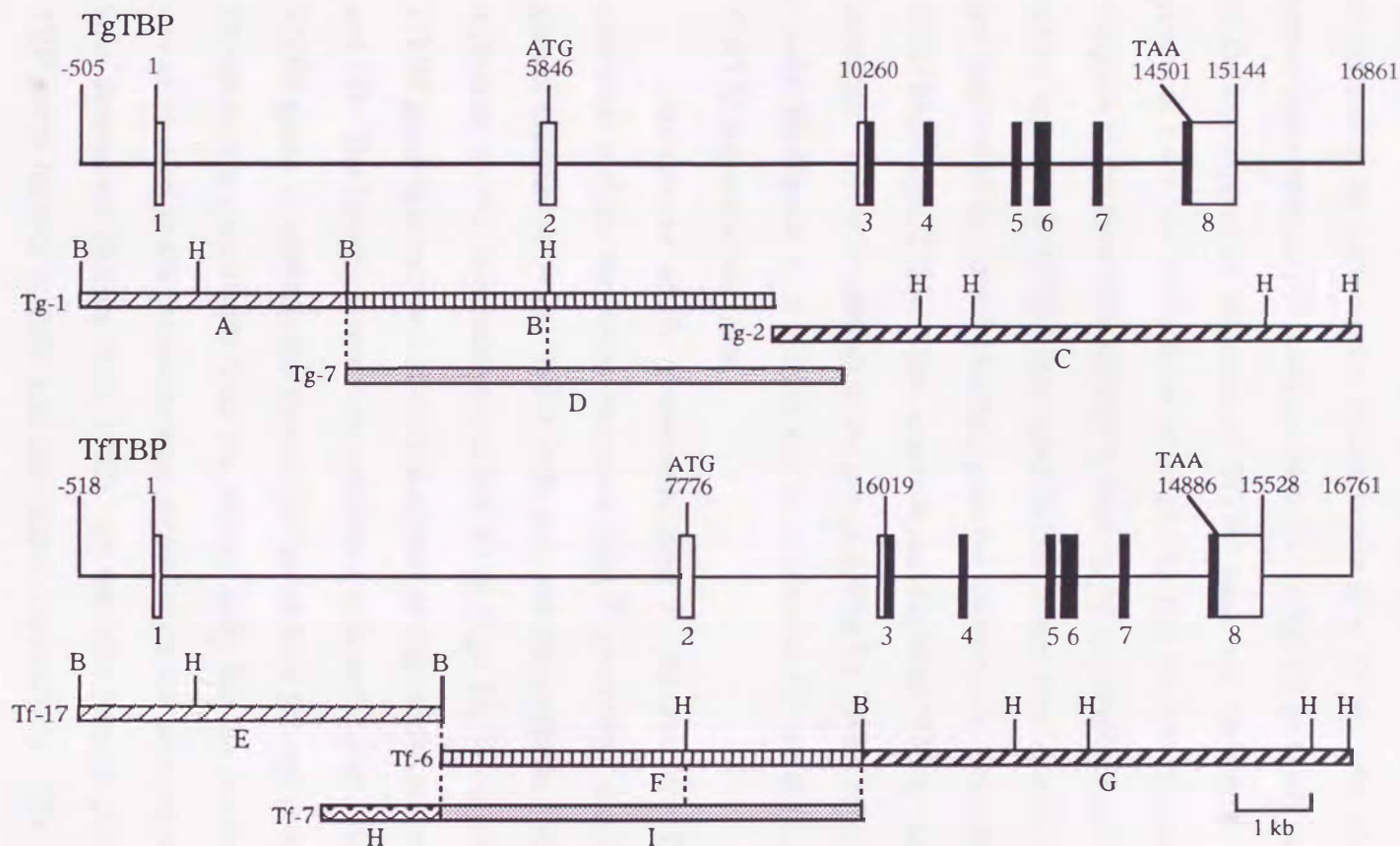


FIG. 16. The schematic representation of the structures of *TgTBP* and *TftBP* genes. *Trimeresurus gramineus* phage clones Tg-1, Tg-2, and Tg-7 contain fragments A and B, fragment C, and fragment D, respectively. *Trimeresurus flavoviridis* phage clones Tf-17, Tf-6, and Tf-7 contain fragment E, fragments F and G, and fragments H and I, respectively. These fragments were subcloned into pUC 19 or 118. The DNA sequences of *TgTBP* and *TftBP* are numbered +1 corresponding to their transcription start site determined by the primer extension analysis. Exons are indicated by boxes, and the conserved 180-amino acid C-terminal regions are indicated by black boxes. Partial restriction maps are indicated. The symbols are: B, *Bam* HI; H, *Hin* dIII.

fragments F and G, and fragments H and I, respectively. These fragments were subcloned into plasmid vector pUC19 or 118 and sequenced (Fig. 17).

Determination of Transcription Initiation Site. Primer extension analysis was carried out for RNAs extracted from *T. gramineus* and *T. flavoviridis* venom glands to determine the transcription initiation site (Fig. 18 a). An 18-mer oligonucleotide primer complementary to positions 84 to 101 of TgTBP gene whose sequence is identical to the corresponding positions of TfTBP gene was employed. The both extended products were 101 nucleotides in length, so that the transcription initiation site was assigned to the thymidine residue bolded in the sequences shown in Fig. 18 b. In the region upstream from the transcription initiation site, there are GC-rich sequences which are assumed to be potential binding sites for transcription factor Sp 1 (Jones and Tjian., 1985; Kadonaga., 1986). They, however, have neither "TATA" nor "CAAT" consensus sequence. This is in contrast to the genes coding for TATA box-binding proteins of *S. pombe* (Hoffmann et al., 1990) and *A. castellanii* (Wong et al., 1992) which have a "TATA" consensus sequence.

Structures of *T. gramineus* and *T. flavoviridis* TBP Genes. Primer extension analysis for RNAs extracted from *T. gramineus* and *T. flavoviridis* venom glands (Sambrook et al., 1989) indicated the transcription initiation site to be the thymidine residue numbered nucleotide +1 in Figs. 16, 17, and 18. Both TgTBP and TfTBP genes spanned over 15-kb and contained eight exons and seven introns (Figs. 16 and 17). The loci of introns were conserved with each other. Similarly to TgTBP and TfTBP genes, *A. thaliana* and mouse TBP genes have ten and seven exons, respectively. However, it is conceivable from the present study that the mouse TBP gene may have one more exon in the 5' untranslated region, since the transcription start point has not been determined (Sumita et al., 1993). On the other hand, *S. pombe* and *A. castellanii* TBP genes have only three and two exons, respectively. The TBP genes of higher eukaryotes appear to have greater numbers of exon than lower eukaryotes. TgTBP encoded a protein of 302 amino acid, whereas TfTBP encoded a protein of 300 amino acid which is 2 amino acids shorter in the Q-run region than TgTBP (Fig. 17). Two

TfTBP -518 GGATCCGCATTTTGAATAGAGATGGGAATCGCTGATAGAAAATCCATACAACAGGCAGAACCAAGTCACA
TgTBP -505A.....A.....T.....T.....

-445 CCATAGCTTCTCTGAACACTGAAAGCCAGCCTGGCTGAAATAAGTAGTACAGAGTACAGAGAGGA----GTAATTCTAC
-432C.....GTACC.....

-370 CTGCCTGTCCCGGATCTTTGCATGTTCTGTGCTCTCCTTCATCGATCCCTTAAACTCTACTGCGTTTTAGCAGTCCGCC
-352T.....A.....

-290 CACCTCGTTGCACCTGGGCGCTTAGAACTCGAGGGATGGGTCTGATGGGTTTCCCCCCTCCTCCACGTTTACCTTTT
-272T.....T.....---

-210 AGGTCTTCATAATTTTCTTTTACTGCTGCCCTCCCTCCACCCACCTACATGAATTAATATATTATTGCATGAATTAAT
-195AA.....-----CCC.....G.

-130 ATATTATCCCGTGCCTGGCTGGGCGCGCGCTGCAAACGATTTCTTCGAGGCAGAACTTTAGAACGCAACTTCGCA
-129-.....G...G.....

-50 CCGCCGCCCTCCCTGCCGCCCTTAAGAAATGTGGTGGCCGAGTAAGTGC**TGGTGAAGGCTATCGGCGCCTGGCTTGATG**
-50 ...T.....G.....A.....

31 ACGCAGAAGAGGCGCGCGCGGAATGAGCCGAGCGCGGGGACGGCTCAGGTGCGGAGAACTTCTGGAGCGGTAAGTCG
31T.....G.....

111 CCGGGCCCGGCAGCCTTATTCTCGCGGTGCCTCTGTCTGCCGCCCGGG--AGAGAGTCTCTTTTTAGGGGGCT
111G.....CCCC.G.....

188 TCTCAAGAGCGGAGAGATCTGGCCGAAAGCGGTGCAAACC-TTAAGCACCCCACTCCCGGTCACTTCGGGACGTAGAG
191G.....C.....

267 TTAGCGTGTCCGTGAGTGTGCTTGGGCCCCGAGGTGCGAGTTCACCTTTTGTCAAATGGAACCGCTGGAAGAGGCGCT
271 ...T.....T.....-.....-.....

347 TTGCTTTCTCTCCCTCCCTCGCTTTTCTCTGCCAGGTGGTTTCGCCAACCTTGAAGAGTCAATGCTATTGGCGTGC
349C.....T.....T.....

427 CTGTAACCTCTGTATATTAACAACAGACCGATCCTTGGGCATGTTTG-CTGGAACATCTCATCCTCTGTCCAGTAGGAC
429-.....T.....

506 TGATTGCAAAGAAGTGGGCGCGCTCGTCCGTGAAAAGAAAAGAAAA--TATCCCTGGGTGAACAAAGAGGTGTAGAG
508T.T.T.TA.....

584 TAGGTATGAATGTGTCTCTGAAACTCACTCGGTGTCATTGAATTGACTTGC-GCTGC-TTGAGATGGTTGAAGACTTT
588T.A....C.....

662 TGCTAACCGGATGTATGGCACGGGCTCTAAGAATGGAATCGTTTGAATTTTAGACTTTGCACCTAATGCAAAGTCA
668A.....C.....G.....-.....CT.

742 CCCCATATTTAGGTGGCAATGTGTGTGATTCCGTTTATTCAAAAATTGGTGAGACAGTCTGTACGTTTGGGAAATTTT
747A.....T.....

822 CTGTTAAGTGAGCTCTGGAATCTTGTCCAAAATTCTTCCTTAATGGCCCCAGTAGCCATAAATCTGTTTCAATGGAT
827T.....

902 GTATAGTAGTTCAAGTATTTAATGTAAATGGATCAACCAAAGAAAGAAATCTGGTAATGTCTTTACATTGTTGTTCTAA
907 C.....A.....C.....

982 GGTATTAATGGTAAAGGAATATTAAGTTCATTGTATCAGACAAATCTGTTTGTGCTTTGAGAACCCCATTTGCAATT
987T.C.....CG.....---

FIG. 17. Nucleotide sequences of TfTBP and TgTBP genes. Exons are italicized. The corresponding amino acid sequences are indicated directly below. Dashed marks indicate the deleted nucleotides, and dots and asterisks represent the positions identical to TfTBP gene in the nucleotide and amino acid sequences, respectively. The translation initiation site ATG is underlined, and the transcription initiation site and potential polyadenylation signal ATTAAG are bolded.

1062 TCCCCCTAAATGTACTTCAAAGATGAAGTGGTATTTATAGGCATCATATTCCTTACTTAAGCTTATATTCAAGTTAGGAT
 1063A.....T.....G.....
 1142 CTGCTCTGGTGGCTGCTTATGTGTACTGGGAGGAAAGCAAAATCCTATACCACTTCTGTTGTGATCCCTTATCCATTTTG
 1143
 1222 GGAATCTACTATACATTAGAAGCATTCTTGTATCTTCTACAGTATTTAGAAACCACTGCAGATAATTGGTCAAAATG
 1223
 1302 TAGAGATTCTGTGGAAATGAAGGGGTAATTATCACCAGTGAGCTTAGTCATGCTTGGTTTTCTCATGAAACTGAGAAT
 1302
 1382 AAGAGAATAAATTTATATATTGTTTTATTTATTTATTTATTTATTTATTTAAATTTATATACCGCCCTATCTCCCTT
 1382AA.G.-----
 1462 ACGGGACTCAGGGCGGTGAACAGGCAATAATAAAATCACAGTAAATATTACATAAAAGCCTGCTCTAAATATAAAATC

 1542 CCTAACAAAACCAATCAAAACTACGCCAGGCTGGCCGCATGGAACAGATGAGTTTCAGGGCTCGGCGGAAGAGGCGCA

 1622 GGTCTGCAATTGTACGTAGGTCTGGGGGCGAGTCATTCCAGAGGCGAGAGCTGCCACAGAGAAGGCTCTCCCCCTGGGG

 1702 GTCGCCAGCCGGCATTGGCTGGCCAACGGTACCCTAAGGAGCCCTCTCTGTGGGAGCGGACTGGTCGAAGGGAGATATT

 1782 TGGTAGGAGGCGGTCTTGCAAATACCTAGGTCCATTGCCATGGAGCGCTTTAAAGTTGTTACTAACACCTTGAAGCGCA

 1862 CCCGGAAGGCCACCGGCAGCCAGTGCAGCTCGCGCAGGAGAGGTGTTACATGGGAGCATCTAGAGGCCCCCAAATCCCC

 1942 CGCGCAGCCGATTCTGAACTAGCTGGAGCCTCCAGGTGCACCTCAAGGGGAGCCCCATGTAGAGAGCATTGCAATAGTC

 2022 CAGGCGGGAACAACGAGGGCTTGAGTGACCGTGCGTAGGGCGTCCCGATCCAGAAAGGGTGCAACTGGCGTATCAGGC

 2102 GAACCTGATAAAAAGCCCCCTGCCACGGCCGTACATGGTCGTCTAAAGACAGCCGAGCATCCAGGAGGACGCCAGGT

 2182 TGCGGGCCCTCTCCGCGGAGCTAATGACTCGCCCCGATAGCCAGAGATGGAACCAACTGGCTATGACGTGACGCCGGA

 2262 AACCACAGCCACTCGGTCTTGACGGGTTGAGACGAAGCCTGTCCACCCCATCCAGACTCGCACGGCCTCCAGGCACCG

 2342 GGACATCACTTCCGCGGCATCGTTGGGGTGGTCCGGGGTGAGATGTACAGCTGGGTGTCATCTGTGTACTGATGAAACT

 2422 CCACCCGAAACCACGGATGATCTCACCCAGTGGCTTCATATAAATGTTGAACAGGAGAGGCGAGAGGACCGACCCCTGT

 2502 TGTACCCACAAAGTGAGGCGCTCGAGGTTGACCTCTGCCCCCTGCCAACACAAATTGCGTCCGGTCTGAGAGGTAGGA

 2582 GGAGAACCTCCGTAAACGGTGCCCCCAACCCCAACCCCTAACCAGTGCAACAAGATACCATGGTCGATGGTATCAA

 2662 AAGCCGCTGAGAGGTCTAGCAGGACCAGGATAGAGGAACAGCCTCTATCCGAGCCCTCCAGAGATCATCAACCAATGCG

 2742 ACCAATGCCGTTTCCGTGCTGTGCCAGGCCTGGCGGAGTCTAGATAGACAGTTTCTTCCAGGTGCTGAGGGAGCTGTC

FIG. 17 continued

2822 GGAAGTGTAGAGTTAATTGTGGCAGTTTAATATTTGCTTTGATTTTATTTGTGTTTGATAAAAAATTAGAGAAATAGATGA
 1414 AA.....C.....

 2902 GATTTCTTAATCCCATTTAACTACAGACATATTAGTTTAAATATTAATGCCAGTACATACATTCTTCCAAATATTGATA
 1494

 2982 AAATGTAACCAAAGTGAACAGGTCTTTTATAATTTAGTCTCTGCATGTATAAATCTATGGTTAATGTTGCTTATAG
 1574

 3062 ACACAAAACCAAAAACCATACATAAACTACAAAACGAGAAATAAGTCCCTTTGAATTTAGAATATTTTGTGCTGATGGGTC
 1654-----.....A.....

 3142 GTTGGTTTCTCTGAGCTTGGTTGTTTCTTGAGAGGTTTCCTTACCAAACCTAGTAACCTCAGTGCCTGCCCTATCTG
 1726C.....C.....

 3222 TGTGGTGGGAGTGGTCTAGAGGTCTTTGGTTGAGCCCTTTACTGTTATCTTGTGTTGGCAGTTCCTTGATTGGGGTAT
 1806

 3302 TGATTACTTCTTGATTGTTGGCCTGATGTTAATCCTGGTGTTAATCTATGCTCTTTGGGGTATTGATTGCTGGTGGGCAG
 1886

 3382 GTATTTTGGTCTTTTGTTCCTTTTGCCTTGTAATTGGTTAAGACTTATTGTCATAGTTTCAACTGGCAAATGACCG
 1966T.....C.....A

 3462 TCCTAGACCAAGCAAGTCCAAAAATTCTAGAGAATCTTAGAAATTGGCCATCAATAGACATAAACAACATCTACATGC
 2046

 3542 TGTGCAAAAGAAACAGCCAACAAGCTAAAAAGGGAACAAAAAACCC-AAAACATCTCCCCACCTGCAATCATCACTCTGA
 2126AT.....C.....C.....

 3621 TGAGCATAGATTAATACCAAGATTAACATCAGACCAACAATCAAGAAGTAAACAATACCCCAATTAAGGAAGTGCAAAAC
 2206C.....

 3701 AAGATAACAGCGAAAAGCCCAACCAAGAATCTCCCAAGACCACTCCCACCAGCAGGGCATGCCATCAGAAATATACATTG
 2286T.....C.....

 3781 AGAGCAAACCTCCAGTTCTTACTAGCACTGATGATATTGGTTTGGGTACTGAAACCTCTGCCAGAAAACAACCAAGGTCA
 2366

 3861 GAGGTACCAAAAACAGTCATTTCAAGCCTGAACTGGAAATATTCTCCTTTTTTGGAAATTTGCCGTTTGTACACTTTC
 2446A.....

 3941 TTGTTCACTTTTACAGGTAGGATATTTAGACCTTTCCATTCTGATTACATTTTTT-GCAGATAGAAATATTTATGCAA
 2526T.....G...

 4020 GAGATTTTAAAGTACTGATTTATTTATGGTTTTTAGCTAGATCATTATTGAAGTTATCCATCCAATTGAATAAGCATTA
 2606

 4100 AATTTGATTTTTCATTTTCACTGCTGCACTGTTAAATCATAATTCTT-AAAAAAATGACTTCTCCTACCGGTAGTAGTG
 2686A.....TT.....

 4179 GGAATGAATATTATTATAAATTGGCTGTTCTAATGGGGATGACTAAACAACTACAGAATCTGTGATTTATTTAGTAAAT
 2766G.....ATGACT..A.....GG.

 4259 GCAAATAGAGAATTCTTGCTACCTTTGTAACAAATTAGAATTGATAACATAGTCTTGTTTATATAACTTGTCTG
 2846AA.....G.....

 4339 GGTGGGATAAAGTGACAAGGATCCTCTTT-----GTGCTAAATATGCAATGAAAGATTTCTTGATTCTTTATCACCT
 2926G.....TCCATT.....

 4412 TCTGCTAAAAGGAACACCACAAAGAGACTGACTGAACCTAGTCATGATAAGCTAGAATTCTTCAGAATACTAGCTAAGAGT
 3006

 4492 GACATACAGATGAGTTATTTAATTGAACTATGTACAGTAATCCAAAGGCCTAGTTAATCAAGTCTTGTTTATTAATG
 3086C.....

FIG. 17 continued

4572 TGTTAGAATCTTAGACTAAGATGGAGCCACATAGTCCATTAATCAACTCATGATTTGTGTAGGAATTCAAATTTAATCAT
 3166 ...G.....C.....T.....G.....
 4652 CTCCAATCATGGTTATTTAGGCTTATCTTAAACACATTTGGTGAAGCGAATCAGTTTCCCGGTAAACTGTTCTTTCT
 3246
 4732 GTTAGGAAAAGATTTTGTAAGTCAAAATATGTCTTCCTGTAACCTAAATCTGTTATTCTATGTGCAATATTCTGGACTGT
 3326TT.....T.....
 4812 AAGTCTTAGTTTTCTTCTGTGATATACTTTTCTAGGTTTGAAGAGTGGTATTGCATTGTTTGTGGTCCTCTCATGCTC
 3406A.....A.....
 4892 CTTCAACCTTTCTTTATTGGTCTTAATTCTAGCCTTGAATCATCCTTATTATGCTTCTGTGAACCAATTTAATCAGTCT
 3486
 4972 AAATCATTTCTTAAGTGGCTGTCCAGTATGGGTACATAAAATCAAAGTGGAGTTGCCTCATGGAGAATAAAGTATGT
 3566
 5052 AGAACACTTCCCATAGGTTACTACATTTCTATAAATGCATACTCAATATCAGGCCTATCCAATCTATGGCCCATGGGACC
 3646TAT.....C.....CTG.
 5132 ATGCAGTGGCTCCACAAGATAATAAACTTTTAACATTATTATGTGATTTACACACATTAATTATATAT-TTTATATGTGG
 3726T.....A.....
 5211 CCTGTGACAATTCCTCTTCACTTAATGCAACCCAGGCAAGCCAAAAGGGTGGACACCATGTTCAATATTCAATTTGCCCT
 3806
 5291 TTTGCTGTATATTTGTTCTTT-CTGGAGTGAATCATTCTGGGGTAATTCAGCAGTAGCAAAAAATACATTTTTTAGAT
 3886A..T..T.....G.....T..CC.....
 5370 CAGCACTGGTTCAGTGTGTTCTGGGAGTGATCTCACCCCAAAGCAGACCTTTTGAGGCCAGAACGCCACGCTCTTGAAACA
 3966A..T.....
 5450 GTTACTGCTAAGAATATTTTGATATTTGTTTTACTGAGCCTTTGCATTAGGTTTATCAATTACTTGGAGTAATCAGGTA
 4046C.
 5530 TTAACCTACTAGCCAGTGTCTTCTTAATTATTTGGTAATTGAAATCATCCAGTGCATCCAGCCCTCTCTACATTTTTGTC
 4126A.....
 5610 TTTAAAGTCATTACACATATTTGGGAAGAAATAATTTAGTTGCATGTATGCATACTTGCCATTTCTTTCATTTTCAAAG
 4206A.....T.....T....
 5690 GCCTTTTAAACCAAATTCGGTACCAAATATTAACATTAACATTTATTTACTATTTCAATTATTTATGGCATTACAAA
 4286
 5770 AAAAATCTTCGCACTTTTGTCAAATGCATTCTTCAGTAAACATGTTGAGGGGCTAAGGGTGGAGGAATGACTAGGTTA
 4366A.....A.....A.....C
 5850 CCATTCCAATGAAGCCCTGTTTTAGGTGTTTCATCATCCTATTTCCTATTTGCTCTTTCTTTTGGAAGTCTTGTGCAA
 4446
 5930 AAATCAGTGGCACTTAACACAGCACAGGTGGACTCATTGTGAGGTTCTGTGCTTGAGACCTTTTATTATGATCAGTT
 4526 ...-.....T.....AA.....C.....CA.....
 6010 TTCCAAATGTTAATAAACAGTTTTAGCCACTAATCCATTTTGAAGCCAGTACTAGTCTTCAAAATGGGGAAGTGA
 4605G.....C.....
 6090 GAGCTGAGGTGATGAACACACTCTGGCTTGGGAGGCAGTAAGTTGAGGCAACAATTGAAACAATGCGCAGGCTGTGGCT
 4685A.....A.....
 6170 GGTGAATTCAGGAAGGTTGCAGCAGACCACTCCATGAAACACGGTTCAAGAATCGCTGCTATTAAAGTAGTTGGCACAAA
 4765G.....A.....A.....-.....
 6250 ATAAATAAAACCGTAACAATATTTGTTTACCTTACTTCGGCTGTGTTTACCTCACCTTTTCTAGCTAAATGTATGCA
 4844A.....T.....-----A.....

FIG. 17 continued

6330 AAGAAAAATGGTTGTATGCATGATATAAGTGAAAGGACTGGGCAAGAATCCTTTCTGCCTAGAAATTGCTTCAAGAGAA
 4912G.....-.....A.....
 6410 GGTAGATGACAGTCGTGAAATAAAAAAATAGGACAAAACAAATTTAAGTATTTTAAATTTACTCAGTAAGTTTGTGTTT
 4991A.....-.....
 6490 TTTTCTTTGGAAACAGTTTGGTAATGGTTAATAGAAATTTATTCATCTTTTATAGGTTAGTTCAGGGATGTCAAACCTCA
 5070 ...-.....A.....C.....-
 6570 ATTTTCATTGAGGGCCGCATCAGCATTGCGTTGCCCTCAGAGGGGTGGGGGGTGGCCAGGGTGGTGTGCCTACTGGGGG

 6650 GGCATGGCAGCTCTACATTGTTTCGCGTCAAGGGTGCCTGTGGGGGCTGGGCAGGAGTGGGCAGATCTGAGGGTCTCTG

 6730 GCAGAAGGGGGTGGGACCGGAGCAAGCGCAAGGCATGGTCCATTTCTTCATTCCTCACAAGGGTGACACCTCCCCAGCA

 6810 TGCACCCAGGGGCGGGCGACACGCAGGGGCTGGTTCCTTGCTGGTTCAGGCGGCCTTGTGGAGCTAAGACTGAATGCG

 6890 TTTACGATAGATGGCCACCACAAAATTGTACATTGGCTCCCGCCGGCCTGGGCAGGGCAAGGCAGTGGGGGCCGTTCC

 6970 CTCACGTGTTCCAGGATGGCTTCGCAGGCCACATGAAACCACTTCGCGGGCGCATCCGACCCCGGGCCTTGTGTTGA

 7050 CATGCTGCGGTTTGGTTGGTTGGTTGTTCTCAGTAGTTACAATGGACTATAAAAGTAACAAGCTGTGACAAACCGATTT
 5133 -----.....G.T.....
 7130 GCTGTTCCCTGTTATGCTTCCAACTGTAACATATATCTCTTAATTCAGGGCATTATTGAGTTAAA-GTTAAAAGTTAATT
 5196T.....G.
 7209 TTTTTTAATGTGTATACCCTTTGTTTTCAAAGTTCAGTCTTGGGTAATGTAATTTTTTTT-TTTGGCCATTTTCTGGG
 5276AA.....G....
 7288 AACCCCTCCCCCCCCCCCCCAACAGGTACTCTCCTTATGTACTTTTGTCTATGATTTTAAATGTTATATTGCACTAATA
 5356 GG-AAAAA...A..A..-.....G.....
 7368 CATTT---AAATGAAAGCTGCAAAGATTTATCTTTTCGGATTTTAAATTGACATTTTGTGTATATTAACCTATATAACT
 5434ATTT.....G.A.....AT..ACT.....
 7444 AATTACTGAATCCCATTCCTGTGGTGTGGAACAACAACCTTGATAACATTAGTTGTCTGTGAAAAAATAGTTACTCC
 5514C..C.....
 7524 AGCACAGCAGTAATGTATGTTATGAGTTTTACTTGAACACAGTAATCTGAGACTTTTTTATCTTTTCAGGTCTGGAACC
 5594A.....-.....
 7604 TTTATTTTATATACTTGA-TGTTTTATCAAATATGTCAATACAGACTGCAAACTAGTTTTTGGGATGCTTCAAGGAAAGC
 5673A.....
 7683 TACTTCCTTAAGGCAGATGTTGTGAAGCTTGATTCTGGATTAGTTTCTGGAAGTAATCCTCAATTGACTGGACACTTA
 5753A.....T.....
 7763 CTTGATTGCTACAATGGATCAAAACAACAGTTTACCACCCTATGCTCAGGGCTTAGCATCCCCACAGTAATATATATAT
 MetAspGlnAsnAsnSerLeuProProTyrAlaGlnGlyLeuAlaSerProGln
 5833
 * * * * *
 7843 ATAACTAATGAACCTATTGAAAATCAAATTTTCAAACAGTTTTTAGAGACCTGTAGTGTGAATGGGTTGTGGATTC
 5913
 7923 CACTTTATTTTAACTTTCTTTGCTGTTCTTTTCCCTGAATTTTAAAGCAGCCTTGTTTATAATCTCTTTATAGCTC
 5993T..
 8003 AATATTTGATTGGTTGTTACTTGCCATGCCAGCCTCTCTGCAGGTTAGGACAAGTTTTTGTAAATCTTATCTCTTATTT
 6072C.

FIG. 17 continued

8083 TTTCAATGACATGAGTGATGGATATGAAAGCAGTAAGAAGCTGTTTATTGTTGTACAATTTTCATTATTGAAACTCTTGCC
 6152C.....
 8163 ATGAACAGAAATAATTGCTTGACAGCTAACTCTTCAGGTTTATCTGCCTCAGATATACATATGGACATATTCTTGAT
 6232C.....A....G.....T..T..
 8243 TTATGTCCCTCTATATTAAGGCTTTACAACCTTCTTATTGGAACCAGAACAAAAAGCATTCTCCTCCTTTATTCTTTG
 6312 ..T.....T.....G...
 8323 AATTTTT-CTAATA--TTTATTAAAGATTATTTGATCAAATAAAATTAATATAAAATCTAATACAAACATA----A
 6392 T.....T.....GT.....T.....G.....TATTA.
 8395 AAAAGAAAAA-GTGAGAGACCGCC-AAGAACTAAACAGTAGGAAAAGGAACTTCCCCCTTCATCACCAGCAAGTAT
 6472A....A....C...C.....A.....
 8473 TATTACAGCATTAAACAGCTTATTACTGTCTATTAATATAGTATAGTAAGTGTGAGAGGGATCTCGGAGTCCTAGTGGAC
 6552C..G.....A.....
 8553 AATCACTTAAATATGAGCCAGCAGTGTGCTGCAGCTGCCAAAAAGCCAACACAGTGTAGGCTGCATTAACAGAGGGAT
 6632T.....
 8633 AGAATCAAGATCACGTGAAGTATTAATACCCCTTTACAATGCCTTGGTAAGGCCACACTTGAAATACTGCATCCAGTTT
 6712 G.....TT.....T.....A.....
 8713 GGTACCCGCGATGTAAAAAGATGTTGCGACTCTAGAAAGAGTACAGAGAAGCAACAAAGATGATAAGAGGACTGGAG
 6792TT.....T.-.....
 8793 GCTAAACATATAAGGAACGGTGTGGAATTGGGTATGTCTAGTTTAAATGAAGAGAAGGACGAGGGGGACATGATAGC
 6871G.....C.....A.....
 8873 AGTGTTCATATCTCAGGGGCTGCCACAAAGGAGAGGGAGTCAAGCTATTCTCAAAGCATCTGAGGGAAAGACAAGCT
 6951T.G...A.....A.....A
 8953 ATGGATGGAACTAATCAAGGAGAGAAGCAACCTAGAACTAAGGAGAACTTGTGACAGTTAGAACAAATTAACCAATGG
 7031A.....C.....G...
 9033 AACAACTGCCTCAGAGTGGTGAATGCTCCAACACTGGAAGTTTTTAAGAAGAGACTGGACAGCCATTGTCCGAAAT
 7111T.....CA.....
 9113 GGTATAGGTTTC-T-CCTGAGCAGGGGTTGGACTAGATGACCTCCAAGGCCCTTCCAACATTGTTGTTGTTGTTGTT
 7191C.G.....A....A.....C...A..A..
 9191 GTTGTATTATTATTATTATTATTATTATCATCATCATCATCATCAATATATCTTCTTTCCATACCTCAA
 7271 A..A.....T.....
 9271 CCCCTTTCTATCAATTAATCTCTAAATCTTCAACCCTCCTTATTCAATCAGCAGAGATCTAGCAACAAATACCCCAAAT
 7309C.....T.....
 9351 ACCTAATAACATAAGTATATAATATCCCTTTCCATTTAGTCTGAATCCCTAAATCTATATTTGCTTCATTCTGTGAG
 7388C.....G.....
 9431 TTGAGTGTCTCTCTAATTGTAATCTTTAGTTCCCTCTAGTATCCAAGTTTTAAAGTTTTATGGGTTTTTTTTTTTAA
 7468T.....CT.....AAAAA..
 9511 AAAAAACAACTGAAGAAGCATTTCACAGGAAGGATTCTTTATAATGGATTCCA-----
 7548 ...T.....T.....AAATCTTATTGATCTTTATC

 7628 TGAAGCCTTAGTCTGGGGTAGGCAACCTTGGCTCTTTTATAACTCGTGGACTTCACTCCAGAAATCCTGAGCCAGCA

 7708 TGGCTAGCTCAGGAATTCTGGGAGTTGAAGTCTACAAGTCATAAAGAGCCAAGGTTGCTACCTGCCTTAGTCCATTTT

 7788 AACTTCTTAAATCAAAGTTTTAACCATATTTTGTGTTGTCTCTCCCTCCCCGCAAAACAAAAGTCTCTAAAC

FIG. 17 continued

 7868 CTGTAGTTAGCGCTTCTTCTGCTAAGGATTGAAGGAAATCATTTTGGTGCTTTAAAAACGCGCTGCTCCAAGGGTCCC

 7948 AATAGTTGAAAAACAGTTAACAACAATTTCCAGAAGCAATTAGAAAGGAAGTATGTGAGCCTATTATCTTGAAATGTGC

 8028 TGACCAGAGTTTGAACGAGATGGCTATTCTCTCATCTCCAGAGCAGTAAACCTGCTGGAGACTGCTGTCTTGTTGTTTC

 8108 CTTGCAAGGAATTGCTGCCTGGAGCACTTGTAGGTAATCTCACATCCTCTCCTTGCTGGAGGGAAATATCAAAAGCCAG

 8188 TTTGCTTACTGGCTCTTCATTCCACTGCAGTCATCTGATCCACTTTTTTCAGACCTTCAGTTCATTTCTTCCCCAGAAGT

 8268 CCCCTTTACAGTTTCTGTTTGTGGCCCCCTAGAATACAAGTAATCCCGAGTTATGAGCATTCCCTTAGCAAATGTTTG

 8348 AAGTTATGCGATAAGCTCCCCAAGTGTATGAAAAAGTCCCGAAATTATTAATGTTGCCACACCACGGCAGTTGTTCA

 8428 CCCCTTCAAAGTGGTTGCTGAAGCAACCGTTTGCAAAGTGCTTCTTGAATCCTGGTCTGTTCTGTTGTTGCAACGTG

 8508 TTCCTGTGGACTATTTCTTCTTTGGACTTGAGTTTGTGTTTGTGTTTGGACTTATTGACAGGTTTGTACCCAAA

 8588 TAAAGGGATTTTGTGTTTAAAGTCTCCAGCCTATTGTATCAGCAGGGGTTGGTCCTTGTCCTTAACAAGGAC

 8668 GGCAATAGGCTGGGATAGAACACCCCCGAGTCTACCTAGGCCAGTGATGGCGAACCTTTAGAGGCCAAGTCCCCAAC

 8748 TGCAACCCAAAACCCACTTATGCATCGTCGGTCCTGGGTGGGTGTAGCCCATGGGTGGGTGTGGCCATTGGGTGGGCG

 8828 TGGCTGCAGTGGTAACAGCAGACTGGGTGGGACTAAAATGTTCAATCAATCAATTCCAACCTTGCAAGATCTTTTTTTC

 8908 CCGCGAGAAGTTGAAATTGACTGATTGAACCTTTTTTTTTTTTACACGGCGGCGACACCTAAAGCGAAAAGGTTTTAT

 8988 AAAACCCAGCTAAAGGCCTAGGACTGGGGCGACGGTGCACGTGCCCACAGAGAGGGCTCTGAGTGCCGCCTCTGGCACCC

 9068 GTGCATAGGTTGCCACCACTGACCTAGGCACTAGGCCTTGGATACTCAACTTGCGAAGATCGAGATCCTTCGCTTTCTT

 9570 -----CGTGAGATCAGTGGCGTGAG
 9148 GGCTGCTCCAGGCTTTCGGTGCATTCGCAGAGAATGGAGGCCTAAAGTAATGCAAGAACA.....T....

 9589 ATCTCATGCTACTGATCTCAGGTACTTTAGGCCTCTGCAACAGATAACAACAGAGCGGAAGCCGGGGAGGCTTCTGCG
 9228C....

 9669 TGCCTCTGGAGGCCTGATCCAGGTGTTAGAAGTCTTTCCGGGCCACTGCTGCCAGAAAAGAGGCCAGGGGAAACCTCC
 9306T..G.....C....

 9749 ACAGACCTCTGGAGGCCTGATCCCGGCATCAGAGGCCTTTTTTAGGCTGCTGATGACTGGAAAAGAGGCTGGGGGACAC
 9386T.....

 9829 TTCTTTGGACCTCTGGAGGCCAATCTGAGCCTCAGACGCGTTTTCCGGGCTGCTGCCGCCAGAAAGGAGCCTGGAGGA
 9465C.....G.....

 9909 TGCTTTCATGGGCCTCCGGAGATCCGTTCTGGGCCTTGAGGCCTCCCTCTGCTCTTGCCAGGAGCAGGGAAAGAAACAG
 9545C.....G..T....A.....

FIG. 17 continued

9989 AGAGGCTGGAGAGGAGACAGATCTTCACAAGGTAGGTGACCTGGCACAGCAGGATTGGGCAGGGGAGCAATTGTGGGGAT
 9625A.....
 10069 CCCAGGGGAGGGAGCATGAGGTAAATGAGGACCCAGATTGTTGGGGGCAATATGGTGGTCTGTAAGGCCTGCTTAAAG
 9705 T.....A.....G.....CT.....
 10149 AGGGCTGTAAAGCGCTATGAAGTGATATATAAGTGTATTGCTATTGCAAGATATCCAGAATACAACAGCCATTCTTTAA
 9785A.....
 10229 ATAACACTATGCATTCCTAATCTTGTCGTGATAGTTCATGTATGTTGTAGACTTTTGATTTTTTTAAATTTGATTAT
 9865T..GC.....C...A.....C.....-.....
 10309 TTTGTAGAGTGCAATGACTCCAGGCATCCCTATTTTACGTCTATGATGCCCTACGGCACTGGGCTCACACCCCAACCTG
 SerAlaMetThrProGlyIleProIleLeuArgProMetMetProTyrGlyThrGlyLeuThrProGlnProA
 9944
 * * * * *
 10389 CTCAGAGTACCAACAGTTTATCTATCTGGAAGAACAACAGAGGCAGCAGCAG-----CAAGCAGCAGCACAGCAGTCC
 laGlnSerThrAsnSerLeuSerIleLeuGluGluGlnGlnArgGlnGlnGln - - GlnAlaAlaAlaGlnGlnSer
 10024CAGCAG.....
 * * * * * GlnGln * * * * *
 10463 ACTTCACAACCGACACAAGCACCATCTGGTCAAACCCACAGCTCTTCCATTACAGACTCTCAGACTGCTCCTTTACC
 ThrSerGlnProThrGlnAlaProSerGlyGlnThrProGlnLeuPheHisSerGlnThrLeuTheTheAlaProLeuPr
 10104C.....
 * * * * *
 10543 AGGAACCAACCTCTCTACCCCTCTCCAATGACTCCCATGACTCCTATAACGCCTGCCACCCCTGCTTCCGAAAGCTCCG
 oGlyThrThrProLeuTyrProSerProMetThrProMetThrProIleThrProAlaThrProAlaSerGluSerSerG
 10184
 * * * * *
 10623 GGATTGTACCGCAGCTCCAGTAAGTGCAGGTCTACTCTGCTCACAGATCCTGTACTTTTCATGATGGTACTTTTAAAGA
 lyIleValProGlnLeuGl
 10264C.....A.....
 * * * * *
 10703 GGGAAATTTTATATTGCTATTTTTTTAGAGCAAGATGGGCAACCTGCGTTCTGAGTTCATGCTGCCAAATGGCTATGTTTC
 10344G.....C.....
 10783 AGTAGCATATGTAGCTATGTTGAATGCTTGGCAAAAAAAAAAAAA--ACACCTTTGAAAACAGGTCCTTGCAGGCATTT
 10424T...G.....C...C..CAC.....AT.....
 10861 ATTCAGAACGCCAATGCATTAAGTTGGATTAAAAATACTTTCTTTCAGTT---TACAATTTATTGATGTCCTCTGAGGT
 10504GTT..AC.....
 10938 TTTTGTAGTTAGATTTTGACGGGTATTGTGGCTTGTAAGTAGTGTCTACATAGTTACTGTTGTAATAAGTTTGTGTT
 10584A.....G.....
 11018 GTTGTAGTTGCGAAGTCATGTCTGACCCATCTCGACCCCTATGGACGACGTTCCCTCCAGGCCTTCCTGTCTCTACCATC
 10664G...C.....A.A...C.....C.T.....
 11098 TTCTAGAGTCCATTTAAGCACAGCCTACTGCTTCAGTGACTCCATCCAGCCACCTCATTCTGTGTCATTCCCTTCTTCT
 10744G.....T..T.....A.....TG..C.....
 11178 ATTGCCCTCAATCTTTCC-CAGCATTAGGCTCTTCTGCAGTGAGTCCTTCCTTCTCATTAGGTGGCCAAAGTACTTGAGT
 10824C..T.....G.....CT.....-----.....C.....
 11257 TTCAGCTTCAG-ATCTGGCCTTCTAAAGAGCATCAGG-GTTGATCTAGGGTGGCTTTATTGAAGTTTAAAGATTGCCT
 10888A.....G.....GTT..A.....C.....A...
 11335 AATATATCTGCAAATCTCTCTAGGAATATTGTGTCTACAGTAAATCTTGGTTGCAAACTTGACCTAAAAACCATTGCC
 nAsnIleValSerThrValAsnLeuGlyCysLysLeuAspLeuLysThrIleAlaL
 10968
 * * * * *

FIG. 17 continued

11415 TTCGTGCTCGAAATGCTGAATACAATCCCAAGTAAGATATAATGGAGAGATAACAACCTAATCTTAGATTAATCAGTGAT
 euArgAlaArgAsnAlaGluTyrAsnProLys
 11048T.....
 * * * * *
 11495 GATTGTTTGCTATTTTCTGCTAGCTCAGTGGCTTCTGCTTACTGCAAAGGAGATTCTGAGTATGCTTTT
 11128T.....T.....T.....
 11575 AAATGTAGCCAAGAAAAA-TCTTTGTGGCTGCAGTAATAGCCTTAAAGAACATGCCACTATGTAGTCACGTTGTGTC
 11208A.....T.....
 11654 AAGCCATTTTATACTAAAACTTTGCAAATCTTTTGTAAAGTATGAAATAACTTATTTTCTGCTTTTGAAGTATTTT
 11288A.....
 11734 AGATTTACTCTGTTTGTGCTGCAAAAGAGTTTGTGAGCTCCAAATTTCCCTAGAAATGCTCAGAGCTTCCCATTTTATTT
 11368-.....G.....G.....A.....
 11814 TGGAATTTACGAGGGTTCATGCTTGAATGGTTCATATTTTAAATTTGTACACCACTACATACGTCGGTTCATTTTAA
 11447
 11894 GACTTGCTATATAAGCATATAATCATTGTTGTGATATATGGTAGTTTGTGATAAAT-AAAAGGTAACATTTGACAGCAC
 11527 .G.....-.....C...G.T.....G....
 11973 TAGTAGATCCCATCTCTAATTGGCCATAAAGAATGGTTTACAAAGCTGGCTGGGTTTCATACCTCATGCCAAGCAACCTAC
 11606G.....TT.....
 12053 ATTGTGACTTTGCCTTCTGGTAGTAATGTGGGATGAAGATCTTGTAAACCGTTTGTAAAAGAAGGCATAGGAACAAGG
 11686C.....
 12133 GCAAAGCTTTGCAATTAACAAAAAGTTTGTGAAAGAAATGAA-GTATTACATGAAAACAAGAAGCAATTGTTTCTACTTA
 11766G.....TT.....
 12212 AGCATTACATAAGCAGTCTCTTCTAGAATTAATTTGATGAAAAAGTACTGTCTATTCTACTGAATTCAACTTAGACATG
 11846A...T.....
 12292 ACCCTCAAACCAATTTATTC-AATGTTTGGCTAT-AAAGAAGTTTGAGAAATGTATTTCCATAAACAAACAAAGGTA
 11926C.G.....C.....C.....
 12370 ACTAACTAATATAATTATTATATGATACATTACATTGTATTTGTAATTGTAGATGTTTTAATATTTTTTTGAAGTGAAGA
 11998 G.....T.....C.....
 12450 ATGGAGATTAGTGCAACTTCAACTACTAATGTACGTTTATGATATTTGTTGTTATTGTTGTTGTTATAAAAGTGTCTTC
 12078-.....G.....
 12530 TTGTTTCAGCGTTTGTGCTGTAATTATGAGAATAAGAGAGCCACGTACCACAGCCCTCATATTAGCTCTGGAAAAAT
 ArgPheAlaAlaValIleMetArgIleArgGluProArgThrThrAlaLeuIlePheSerSerGlyLysMe
 12151 ..---.....
 * * * * *
 12610 GGTGTGCACGGGAGCTAAAAAGTATGCAAGCCCAAGCCGAAGCAATTATCTGCAGCCATTAACTTATTAGCATTGGCC
 tValCysThrGlyAlaLysSe
 12228C.....
 * * * * *
 12690 CAACAGATATGGTTATTTGACA-----GCTAATAGGTATGTAATTTACAACTGTGT--ATATC
 12303C.....TAGGTACTCTGATAAAGCCA.....C.....GTG....
 12747 ATTCAGAGTTTCCTTGATGAAATAACAACGCATCTTATACATTTTTCAGTGAAGAGCAGTCCAGACTGGCTGCGAGAAAGT
 rGluGluGlnSerArgLeuAlaAlaArgLysT
 12383G.....
 * * * * *
 12827 ATGCTAGAGTTGTACAGAAGTTGGGCTTCTGCAAAATCTTGGATTTTAAATTCAAAATATGGTTGGCAGTTGTGAT
 yrAlaArgValValGlnLysLeuGlyPheProAlaLysPheLeuAspPheLysIleGlnAsnMetValGlySerCysAsp
 12463
 * * * * *

FIG. 17 continued

12907 GTAAATTCCTATCAGACTTGAGGGATTGGTCTTACACATCAACAGTTTAGCAGGTAGGCCAAAAATGATGTTATCT
ValLysPheProIleArgLeuGluGlyLeuValLeuThrHisGlnGlnPheSerSe
12543A.....G.T.....
* * * * *

12987 TAAATGTGTATTAAAAACTTTTTTATTTGTGTTTCTCAGTTTATTTACTCATAGTTTTTGGAGTCATTTATGTA
12623C-.....

13067 CTTTCTCTGAATGCAAAGAGGGACAGCTGTAGGCAATTCTGATGTATCTGTGATGTATAGGTAATTCAACTATTTTA
12702CA.....A.....

13147 AATCTGTATAGTTCTTTAATATTCTGTGTTTGATCTTCAATACTGAAACAAAAAGCTTTGTTATTTAAACAAAAA-TC
12782T...T.....C.....AC..

13226 GATTATCTGTGATATGTATTGGTGTAAATTTAAATCCTTAAGAATAGTAAGATAAGACACATTGCTAATTGAGCAGCTCT
12862T.....

13306 AAAAGTGAAATCTGAATTCACACTAGCCCAGCATCTTTAAATGTTCTTGGCCTTACAGTATAGAATTGTTCAATAATT
12942
13386 CTTATGTGCAGTCTGTATATATTTTACCTGAAAGTTTGTGTATATACCCACTTGAGGAATTAGCTAATAATGAAAA
13022
13466 GTACAGATAGGCCCTGGGGGTACATAAGACATGACTTAGTATCATCTCAGCTTAATCAACGTGTGGCAGAAGAAAAA-T
13102A.G.....A..

13545 GCATGTCTGTTCTCAGCAAATACTTTACATTTTACTTATATTGAACTTTCTTTCAGTTATGAACCAGAATTGTTTCCT
rTyrGluProGluLeuPhePro
13182
* * * * *

13625 GGTTTAATCTACAGAAATGATCAAACCAAGAATTGTACTACTCATATTTGTTTCTGGAAAGGTAGTTTAACTGGTAAGCC
GlyLeuIleTyrArgMetIleLysProArgIleValLeuLeuIlePheValSerGlyLysValValLeuThrGlyLysA
13262
* * * * *

13705 -ATATTTGAAATTACTGCACTGACTATAGCAAACCTGAAAGATACCCCAAAATATTTCAAATAGCTAATTAACAAATACA
13342 C.....C..C.....

13784 AATATATCGATGCATTGGTTTGTTCCTTTTTTGTACTTTATTATAATATACTGTTTGGGTGCATTTTATACCCAAATATT
13422T.....G....

13864 TTTATTTGTCCTGTAGTGTTAACATGTTGCAGCAATTAATGTGTTAATGTGCCTTTAAAG-CCCCAGAAAATACAGTT
13502A.....

13943 GGTACACAATGCAGCCATGTCATGGGCTGCTGGATGCAGTGAACATACTATTCCTGTGCTGCAGGCCAGCCGTGGCTGC
13582T.....T....C....G.....C.....A.....

14023 CTGTAAGTTTTCGGGTGCAATTTAGAATGCAAGTTTGTCTGATGAAGTCTTATTTGGCTTGGTGTGAGGACCTGTAG
13662C.....A.....

14103 GACTGCCTCTTTCAACAGAATCCGATTGTTGGTAGCTCCCTGTTACTATGGAATGGAGGAGAAATGCTTTTCTACTACG
13727T.....

14183 GCTTCTGTTTGTGCAATAGACTTCCATTATCTGGTGGCCTTTGGGAAGCAGGTTAAAACTCTTCCAGAGGGCTTTTTA
13807C.....G.....

14263 ACTAGTGCCATAGGAATTGTTTCATATATTGCTCCAGTTTAAATAAATAAACTAACCAGAGAGAAGCAACCTAGAAC
13880
14343 TAAAGAGAACTTCTGACAGACAATTAATCAGTGAACAACCTGCCTCCAGAAGCTGTGAATGCTCCAACAATTGAAG
13960C.....C.G....

14423 TCTTTAAGAAGAGATTGGACAGACATTTGTCTGAAATGGTATAGGGTTTCTGCCTGAGC-GGGAGGTGGACTAGATGAC
14040A...GA.....

14502 CTCCAAGTCCCTTCCAACCTTTTTATTATTATTATTAATGTTAAGTATTGTTTATCACCAGATAGTGATGACGAGGG
14120A.....A.....C.....

FIG. 17 continued

16295 ACCTTTTCCAACAGGATGAATGTGAAAATTCTAGATTCTTGAATGAGTGAATGTCTTCTCAAATATGAATGGCATATTCC
 15956C.....G.....C.....

 16375 AGCCAACATACTGAAGAGTATTTTTTAAAAATGTGCTTCATTGCTTTTACTAACTGTGTTTCAAGGATGCTAATCT
 16036G.....

 16455 TTTACTTCAATTCTCATCAAAGAACTAAAGAATCGGTGCCTGGATGGGATCATCTCACTATAGCTC-ATAGTTATATA
 16116G...C..T....T..C.....

 16534 ACGGAATAAACTATACATTTTCGTATTCTAATTCATTCCATACCTGGTTTGGCCAAAACCTTCCTGGTTTGTTAAATTA
 16193C...G.....C...C.....A.....

 16614 TATTTCAAAGTAATTAGTCCTTGAACCAGACTATAACCAAAATAAGCACAATGAAGAATAAGTGATGAGGAGTCTCTCC
 16273G.....C.....-.....

 16694 CATAAAGCAACCAATGGTAACACAAGAGCAAATTCACCTAAAGCTTATTCAGACAGGCTCTTAAGATC (17279 bp)
 16352A.....CTTTACCAAGGG

 16432 TGGCAATGGGAAAGGTTTCAGGAATGGGTGGAGGAAAATAAACAGTGGAAACAACAATTTATCATTGATTCTTTACTAT

 16512 AGTAAATTTATCATGTCTATCATGATTGATTTTTTTTTTTTCATTTCAAAGTTCTAGCAAAATAATCACCTCTCTGGTCT

 16592 CAGATTGATTTTATATTCTTATTCTATAACTGATGTGGAAGTTCAAGCAATGGCGATCATTTGTTATCAGTTGAATATTAA

 16672 AAGGTGAGAAAAATTTATGATAACGGAAGAACTTATCTCACTGCTTAAGCTTTGCAGACATTAAACAGAATTGGAATG

 16752 TCCTGCTTCCAGATGAACCTCTCAACATAGTGGTTTCCTAGATCACAATTCTCAGCACATGTGTAGATAGCCAGTGTTC

 16832 CCAGTCAATACTTTCTCCAAGACTGTCGAC (17366 bp)

FIG. 17 continued

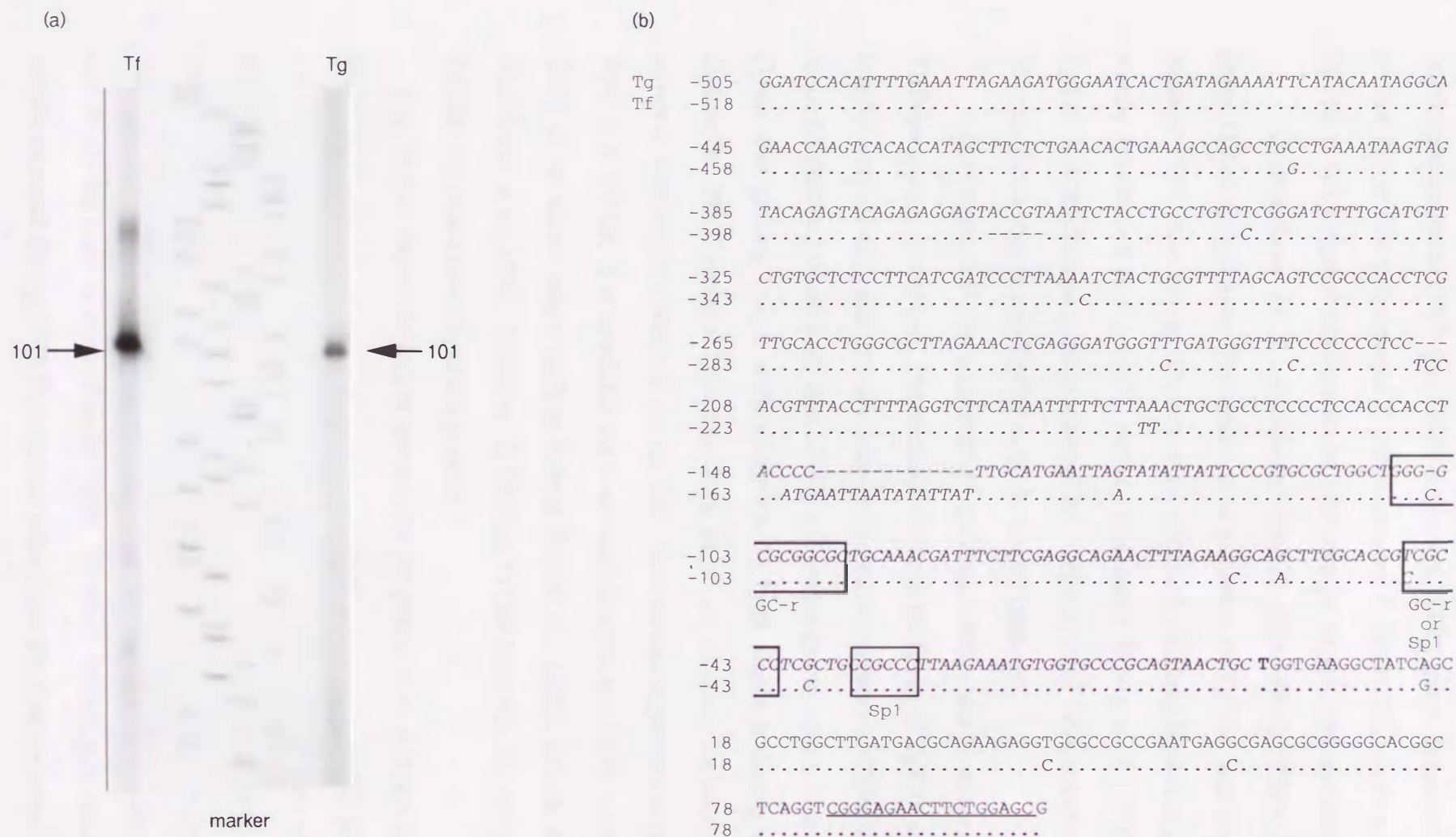


FIG. 18. Determination of the transcription initiation site of two snake TBP genes. (a) Primer extension analysis. The products were loaded onto 6% polyacrylamide-7 M urea sequencing gel and electrophoresed. Lanes Tf and Tg indicate the products from *T. flavoviridis* and *T. gramineus* venom gland total RNAs, respectively. Products of 101 base length are indicated by arrows. (b) The sequences of the 5' flanking region and exon 1 of two snake TBP genes. The numbers are relative to the transcription initiation site (the bolded T in the sequence). The sequences of the 5' flanking regions, italicized; G+C-rich region (GC-r) and possible Sp1-binding site, boxed. The locus of the antisense primer sequence are underlined.

silent mutations were found in the coding regions of the TgTBP and TfTBP genes (nucleotide positions 11031 and 13448 in terms of TfTBP gene). Southern blot analysis of *Bam* HI- or *Hin* dIII-digested *T. gramineus* and *T. flavoviridis* genomic DNAs with TgTBP cDNA indicated that there is one TBP gene per haploid genome (data not shown).

Comparison of *T. gramineus* and *T. flavoviridis* TBPs with Those from Other Organisms. The amino acid sequence of TgTBP was compared with those of TBPs from mouse (Tamura et al., 1991), *A. thaliana* (Gasch et al., 1990), *S. pombe* (Hoffmann et al., 1990), and *A. castellanii* (Wong et al., 1992) (Fig. 19). TgTBP is highly homologous to mouse TBP. TBPs of other three eukaryotes are 24 to 55 amino acids shorter than TgTBP in the N-terminal domain.

In contrast with the conserved C-terminal domain, the N-terminal domain of TBPs are greatly diversified. The numbers of Gln in the Q-run of TgTBP and TfTBP are 6 and 4, respectively, and they are intermediate between those of TBPs from *Xenopus laevis* (amphibian) (Hashimoto et al., 1992) and mouse (mammalian). It appears that the Q-run was getting longer with evolution by base slippage during chromosome replication. The C-terminal 180 amino acids which are necessary for basal transcription, however, were highly conserved among TBPs from various organisms as pointed out by Apsit et al (1993). It is noted that the C-terminal domains are 100% conserved among TBPs of vertebrate origin such as human (Kao et al., 1990), mouse, and *Xenopus* (Hashimoto et al., 1992). However, TgTBP and TfTBP have Ala-282 (the numbering in TgTBP) but others have Gly at this position.

Fig. 19 also shows the loci of introns and the positions of β -sheets and α -helices found in the crystal structure of *A. thaliana* TBP2 (Kim et al., 1993; Nikolov et al., 1992). The structure of *A. thaliana* TBP consists of curved ten-strand antiparallel β -sheets, generating a saddle-shaped structure with four α -helices lying on its upper surface. The C-terminal domains of TgTBP and TfTBP are split by five introns as in the case of mouse TBP (Fig. 19). The loci of introns of TBPs from other eukaryotes are not completely the same as those of snake TBPs. However, three highly conserved loci of introns are noted among TBPs from various eukaryotes. The first conserved intron site (4

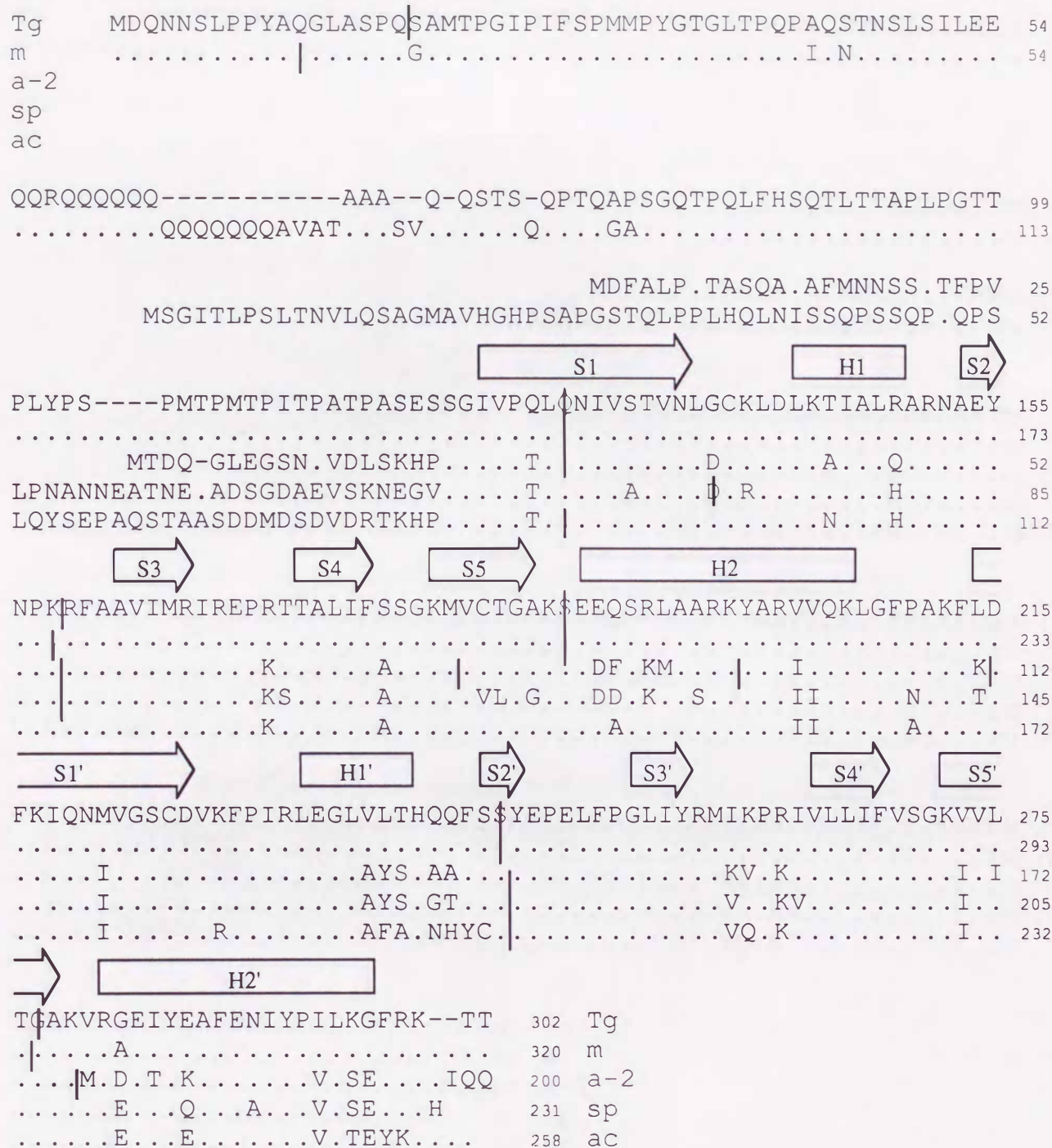


FIG. 19. Aligned amino acid sequences of TBPs whose genomic structures have been determined. Tg, *T. gramineus*; m, mouse; a-2, *A. thaliana* TBP2; sp, *S. pombe*; ac, *A. castellanii*. β -strands S1 to S5 and S1' to S5' and α -helices H1, H2, H1', and H2' are indicated in the upper region of the amino acid sequences according to the crystal structure of *A. thaliana* TBP2. The loci of introns in each amino acid sequence are displayed by vertical line (|). Dashes indicate the gaps for alignment purpose and dots represent the amino acid residues identical to those of TgTBP.

of 5 TBPs) is located in the middle of β -strand 1 (S1). According to the crystal structure, this site is in the boundary of two subdomains. This is consistent with the idea that most of introns are located in the boundaries of modules (Gö., 1981; Blake., 1981). The second conserved intron site (4 of 5 TBPs) is located in the loop forming a stirrup which is in the boundary between S2 and S3. It is known that introns are often located in the loop regions of protein structures (Branden and Tooze., 1991). The third conserved intron site (5 of 5 TBPs) is in S2'. Since it is unlikely that introns were inserted by chance into the same positions of genes, it seems that previously existing introns in the TBP genes of lower eukaryotes such as *S. pombe* (Hoffmann et al., 1990) and *A. castellanii* (Wong et al., 1992) were lost from the C-terminal domain during evolution (Gö., 1983). It is worthy to note that *Trimeresurus* and mouse TBPs have only one intron in the N-terminal domain but at different positions in spite of their quite similar sequences (Fig. 19).

Recently it has been reported that archaebacterium *Pyrococcus woesei* TBP consists of 191 amino acids with homology to the conserved C-terminal domains of eukaryotic TBPs and has no N-terminal domain (Rowlands et al., 1994). On the other hand, eukaryotic TBPs have the N-terminal domains of the varying lengths (Fig. 19). Thus, it appears that the N- and C-terminal domains of eukaryotic TBP genes have evolved independently and fused together during evolution.

The Accelerated Substitutions in the Protein-coding Regions Are Responsible for Darwinian Evolution of Venom Gland PLA2 Isozymes Genes to Acquire New Physiological Activities.

In *T. gramineus* and *T. flavoviridis* venom gland PLA2 isozyme genes, the protein-coding regions have been evolving at much higher rate than the noncoding regions including introns and the evolutionary rates for all introns are almost equivalent to those for all the noncoding regions. For example, when gPLA2-I and gPLA2-VI were compared, K_N values for all introns and for all the noncoding regions were computed to be 0.0716 and 0.0635, respectively. This holds for all the pairs of genes compared between species, indicating that the evolutionary rates are almost equal between introns and other noncoding regions in venom gland PLA2 isozyme genes. To gain a further insight into the evolutionary feature of venom gland PLA2 isozyme genes, the evolutionary rate of introns were compared with those of introns of *T. gramineus* and *T. flavoviridis* TATA box-binding protein (TBP) gene, which are assumed to be a general gene ubiquitously present in all cell types. Based on the data obtained in the present study, it became evident that Darwinian evolution is universal in crotalinae snake venom gland PLA2 isozyme genes.

EXPERIMENTAL PROCEDURES

Data Analysis. For the nucleotide sequences of *T. gramineus* and *T. flavoviridis* TBP genes which contain eight exons and seven introns (Chapter 4), the numbers of nucleotide substitutions per site (K_N) for the noncoding regions including introns and the numbers of nucleotide substitutions per synonymous site (K_S) and per nonsynonymous site (K_A) for the protein-coding regions were computed according to the method of Miyata and Yasunaga (1980) with corrections for multiple substitutions

(Kimura, 1983). The K_N , K_S , and K_A values were compared with those of *T. gramineus* and *T. flavoviridis* PLA2 isozyme genes (Chapters 2 and 3)

RESULTS AND DISCUSSION

Accelerated Substitutions in the Protein-coding Regions are Responsible for Darwinian Evolution of Venom Gland PLA2 Isozyme Genes. The K_N , K_S and K_A values were computed from the nucleotide sequences of *T. gramineus* TBP (TgTBP) and *T. flavoviridis* TBP (TfTBP) genes (Table 6). Those values are compared with those for a pair of venom gland PLA2 isozyme genes, gPLA2-VI and BP-I, as an example (Fig. 20). It is noted that K_N values for venom gland PLA2 isozyme genes are comparable to that of TBP genes (Tables 3 and 6 and Fig. 20). This indicates that introns of PLA2 isozyme genes have evolved at similar rates to introns of TBP genes which could be assumed to have evolved under neutrality (Kimura, 1983; Nei, 1987). Some functional constraints in the introns of PLA2 isozyme genes that had once been considered because of their high homology compared with exons could be excluded. Thus, the highly accelerated substitutions in the protein-coding regions are responsible for Darwinian evolution of venom gland PLA2 isozyme genes to acquire new physiological activities.

The fact that the nucleotide sequences of *T. gramineus* venom gland PLA2 isozyme genes showed the same characteristics as found for those of *T. flavoviridis* venom gland PLA2 isozyme genes reveals that accelerated evolution is rationally assumed to be universal for crotalinae snake venom gland PLA2 isozyme genes. Recently, it was reported that in the genes encoding mojave toxins from *Crotalus scutulatus scutulatus* (crotalinae), which are heterodimeric neurotoxic PLA2s, the noncoding regions including introns are much more conserved than the protein-coding regions (John et al., 1994). This further supports the consequence which has been drawn from the present study.

Mechanism of Acquisition of Diverse Physiological Activities. Some mechanisms have been suggested for proteins to acquire diverse functions, that is,

Table 6. K_N/K_S and K_A/K_S values for *T. gramineus* and *T. flavoviridis* TBP genes

Pair of genes	K_N	K_S	K_N/K_S	K_A	K_A/K_S
TgTBP ~ TfTBP	0.0365	0.0095	3.84	0.000	0.000

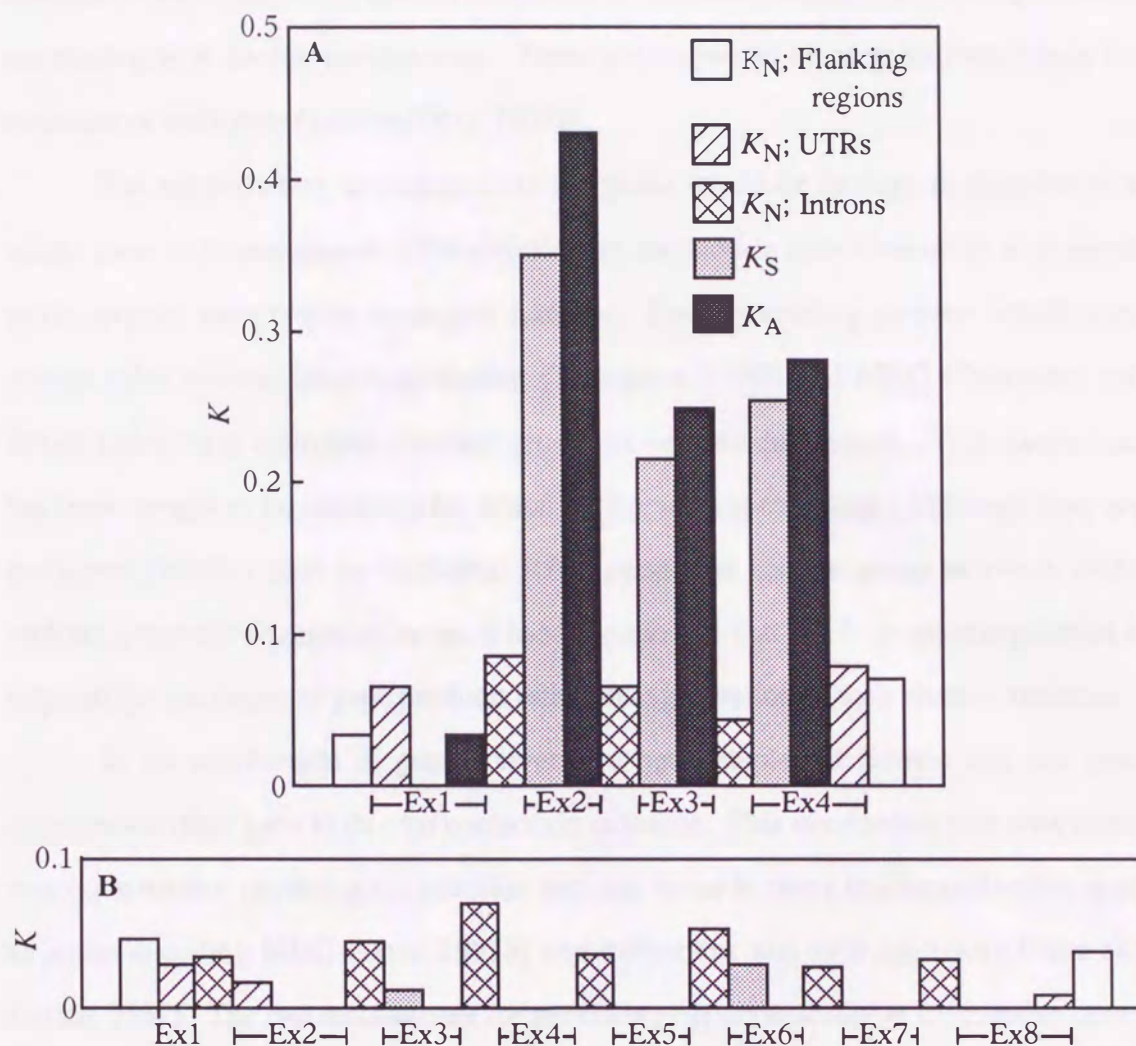


FIG. 20. Schematic representation of sectional K_N , K_S , and K_A values for a pair of gPLA2-VI and BP-I (A) and those for *T. gramineus* and *T. flavoviridis* TBP genes (B). Ex designates exons and introns are parts sandwiched between exons. Each K is expressed by different patterns as indicated at right upper region of A. In PLA2 isozyme genes (A), K_S for Ex1 is zero. In TBP genes (B), Ex1 contains the UTR only. In TBP genes, K_A for Ex3 and Ex6 and K_S and K_A for Ex2, Ex4, Ex5, Ex7, and Ex8 are all zero. It is also evident from A that in PLA2 isozyme genes, K_S and K_A values are much greater than K_N values.

diversification of gene expression, alternative splicing, and posttranslational modification. Furthermore, three mechanisms, gene duplication, gene conversion, and natural selection, have often contributed to increment of functional diversities of proteins while associating with former mechanisms. These are known to be of great importance for evolution of multigene families (Ohta, 1991a).

The simplest way to acquire new functions would be through duplication of a whole gene with subsequent differentiation by nucleotide substitutions in new genes while original gene retains necessary function. Genes encoding proteins which have critical roles such as immunoglobulins (Tonegawa, 1983) and MHC (Steinmetz and Hood, 1983) form multigene families generated by gene duplication. This mechanism has been thought to be important for adaptive (Darwinian) evolution. Although there are multigene families such as ribosomal RNA genes and histone genes in which many uniform genes are arranged in series, it has been thought that this large gene population is required for increment of gene products rather than generation of their diverse functions.

In the mechanism of gene conversion, heteroduplex is formed and one gene converts the other gene to its own nucleotide sequence. This mechanism also contributes to acquire diverse physiological activities and was found in many multigene families such as genes encoding MHC (Ohta, 1991b) and kallikreins and their inhibitors (Ohta and Basten, 1992). The two mechanisms for generating hypervariability at the reactive center of serine proteases and their inhibitors are gene conversion followed by natural selection and natural selection for point mutation (Ohta and Basten, 1992). Natural selection is thought to help the survival of genes with desired variability that occurred by gene conversion. If the number of synonymous substitutions at the variable region exceeds that at the conserved region, gene conversion between loci is thought to be operative. In the case of kallikrein and MHC multigene families, K_S values at the variable region of kallikreins which contains an active site that interacts with the inhibitors and at the antigen recognition site of MHC are larger than those in the remaining regions (Ohta and Basten, 1992; Hughes and Nei, 1988).

In the case of venom gland PLA2 isozyme genes, K_S values of exons, especially the protein-coding regions, exceed K_N values in the noncoding regions including introns. Moreover, it is found that the noncoding regions of PLA2 isozyme genes have been evolving at normal rate because their K_N values are comparable to those of introns of TBP genes from *T. flavoviridis* and *T. gramineus* snakes. Therefore, there might be a mechanism that includes gene conversion specific for the protein-coding regions of exons. The most significant constructional difference between mammalian PLA2 genes and snake venom PLA2 genes is that the former is a single copy gene and the latter is a multigene family. Thus, it could be inferred that mammalian PLA2 genes have not been able to evolve in a similar manner as snake venom gland PLA2 isozyme genes and that formation of multigene family and gene conversion are important mechanisms for accelerated substitutions of venom gland PLA2 isozyme genes.

If the number of nonsynonymous substitutions exceeds that of synonymous substitutions at a certain region, positive Darwinian selection for point mutation is thought to have operated (Ohta and Basten, 1992; Hill and Hastie, 1987; Hughes and Nei, 1988). Since the K_A values are close to or larger than K_S values in the protein-coding regions of venom gland PLA2 isozyme genes, it is highly likely that positive Darwinian selection occurred in these regions so as to acquire diverse physiological activities.

ACKNOWLEDGEMENTS

The author wishes to express his most sincere thanks to Professor Motonori Ohno, in our laboratory, for his invaluable suggestions and discussion throughout this work.

The author is deeply grateful to Professor Yoshiyuki Sakaki, Dr. Masahira Hattori, Institute of Medical Science, University of Tokyo, and Yasuyuki Fukumaki, Institute of Genetic Information, Kyushu University, for their excellent suggestions and discussion and for use of facilities of their laboratory.

The author also wishes to thank Dr. Yasuyuki Shimohigashi, Dr. Tomohisa Ogawa, Mr. Makoto Nakai, in our laboratory, and Dr. Hideki Sumimoto, Kyushu University School of Medicine, for their discussions and encouragements, and Mr. Ikuo Nobuhisa and Mr. Masanobu Deshimaru, in our laboratory, for their valuable experimental assistance.

REFERENCES

- Apsit, V., Freeberg, J. A., Chase, M. R., Davis, E. A. and Ackerman, S.: *Nucleic Acid Res.* **21**, 1494 (1993)
- Arita, H., Hanasaki, K., Nakano, T., Oka, S., Teraoka, H. and Matsumoto, K.: *J. Biol. Chem.* **266**, 19139 (1991)
- Blake, C. C. F.: *Nature* **291**, 616 (1981)
- Bline, N. and Stafford, D. W.: *Nucleic Acid Rec.* **3**, 2303 (1976)
- Branden, C. and Tooze, J.: *Introduction to Protein Structure*. Garland Publishing, New York and London, 1991
- Breathnach, R., Benoist, C., O'Hara, K., Gannon, F. and Chambon, P.: *Proc. Natl. Acad. Sci. USA* **75**, 4953 (1978)
- Brunie, S., Bolin, J., Gewirth, D. and Siglar, P. B.: *J. Biol. Chem.* **260**, 9742 (1985)
- Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. and Rutter, W. J.: *Biochemistry* **15**, 5294 (1979)
- Coden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Keding, C. and Chambon, P.: *Science* **209**, 1406 (1980)
- Comai, L., Tanese, N. and Tjian, R.: *Cell* **68**, 965 (1992)
- Cormack, B. P. and Struhl, K.: *Cell* **69**, 685 (1992)
- Church, G. M. and Gilbert, W.: *Proc. Natl. Acad. Sci. USA* **81**, 1991 (1984)
- Dhillon, D. S., Condrea, E., Maraganore, J. M., Henrikson, R. L., Benjamin, S. and Rosenberg, P.: *Biochem. Pharmacol.* **36**, 1723 (1987)
- Dijkstra, B. W., Drenth, J. and Kalk, K. H.: *Nature* **289**, 604 (1981a)
- Dijkstra, B. W., Kalk, K. H., Hol, W. G. J. and Drenth, J.: *J. Mol. Biol.* **147**, 97 (1981b)
- Dijkstra, B. W., Renetseder, R., Kalk, K. H., Hol, W. G. J. and Drenth, J.: *J. Mol. Biol.* **168**, 163 (1983)
- Feinberg, A. and Vogelstein, B.: *Anal. Biochem.* **132**, 6 (1983)

- Fucharoen, S., Fucharoen, G., Fucharoen, P. and Fukumaki, Y.: *J. Biol. Chem.* **264**, 7780 (1989)
- Fukagawa, T., Matsumoto, H., Shimohigashi, Y., Ogawa, T., Oda, N., Chang, C.-C. and Ohno, M.: *Toxicon* **30**, 1331 (1992)
- Fukagawa, T., Nose, T., Shimohigashi, Y., Ogawa, T., Oda, N., Nakashima, K., Chang, C.-C. and Ohno, M.: *Toxicon* **31**, 957 (1993)
- Francis, B., Gutierrez, J. M., Lomonte, B. and Kaiser, I. I.: *Arch. Biochem. Biophys.* **284**, 352 (1991)
- Gasch, A., Hoffmann, A., Horikoshi, M., Roeder, R. G. and Chua, N.-H.: *Nature* **346**, 390 (1990)
- Gô, M.: *Nature* **291**, 90 (1981)
- Gô, M.: *Proc. Natl. Acad. Sci. USA* **80**, 1964 (1983)
- Hanasaki, K. and Arita, H.: *Biochem. Biophys. Acta* **1127**, 233 (1992)
- Hashimoto, S., Fujita, H., Hasegawa, S., Roeder, R. G. and Horikoshi, M.: *Nucleic Acid Res.* **20**, 3788 (1992)
- Hattori, M. and Sakaki, Y.: *Anal. Biochem.* **152**, 232 (1986)
- Hayashida, K., Okubo, H., Noguchi, M., Yoshida, H., Kenji, K., Matsuo, H. and Sakaki, Y.: *J. Biol. Chem.* **260**, 14224 (1985)
- Heinrikson, R. L.: *Methods in Enzymology* **197**, 201 (1991) New York: Academic Press
- Heinrikson, R. L., Kreuger, E. T. and Keim, P. S.: *J. Biol. Chem.* **252**, 4913 (1977)
- Hill, R. E. and Hastie, N. D.: *Nature* **326**, 96 (1987)
- Hoffmann, A., Horikoshi, M., Wang, C. K., Schroeder, S., Weil, P. A. and Roeder, R. G.: *Genes Dev.* **4**, 1144 (1990)
- Horikoshi, M., Wang, C. K., Fujii, H., Cromlish, L. A., Weil, P. A. and Roeder, R. G.: *Nature* **341**, 299 (1989)
- Holland, D. R., Clancy, L. L., Muchmore, S. W., Ryde, T. J., Einspahr, H. M., Finzel, B. C., Heinrikson, R. L. and Watenpaugh, K. D.: *J. Biol. Chem.* **265**, 17649 (1990)
- Hughes, A. L. and Nei, M.: *Nature* **335**, 167 (1988)

- Ishizaki, J., Hanasaki, K., Higashino, K., Kishino, J., Kikuchi, N., Ohara, O. and Arita, H.: *J. Biol. Chem.* **269**, 5897 (1994)
- Ishizaki, J., Kishino, J., Teraoka, H., Ohara, O. and Arita, H.: *FEBS Lett.* **324**, 349 (1993)
- John, T. R., Smith, L. A. and Kaiser, I. I.: *Gene* **139**, 229 (1991)
- Jones, K. A. and Tjian, R.: *Nature* **317**, 179 (1985)
- Kao, C. C., Lieberman, P. M., Schmidt, M. C., Zhou, Q., Pei, R. and Berk, A. J.: *Science* **248**, 1646 (1990)
- Kihara, H., Uchikawa, R., Hattori, S. and Ohno, M.: *Biochem. Int.* **28**, 895 (1992)
- Kim, J. L., Nikolov, D. B. and Burley, S. K.: *Nature* **365**, 520 (1993)
- Kimura, M.: *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, Cambridge, U.K.) (1983)
- Kini, R. M., Kawabata, S. and Iwanaga, S.: *Toxicon* **24**, 1117 (1986)
- Kini, R. M. and Evans, H. J.: *J. Biol. Chem.* **262**, 14402 (1987)
- Kini, R. M. and Evans, H. J.: *Toxicom* **27**, 613 (1989)
- Kadonaga, J. T., Jones, K. A. and Tjian, R.: *TIBS*. **11**, 20 (1986)
- Kozak, M.: *Nucleic Acid Res.* **12**, 857 (1984)
- Kramer, R. M., Hession, C., Johansen, B., Hayes, G., MacGray, P., Chow, E. P., Tizard, R. and Pepinsky, R. B.: *J. Biol. Chem.* **264**, 5768 (1989)
- Kuhara, S., Matsumoto, F., Futamura, S., Fujita, A., Shimohara, T., Takagi, T. and Sakaki, Y.: *Nucleic Acid Res.* **12**, 88 (1984)
- Kusunoki, C., Satoh, S., Kobayashi, M. and Niwa, M.: *Biochim. Biophys. Acta* **1087**, 95 (1990)
- Lee, C. Y., Ho, C. L. and Botes, D. P.: *Toxicon* **20**, 637 (1982)
- Lichtsteiner, S. and Tjian, R.: *Proc. Natl. Acad. Sci. USA* **90**, 9673 (1993)
- Liu, S.-Y., Yoshizumi, K., Oda, N., Ohno, M., Tokunaga, F., Iwanaga, S. and Kihara, H.: *J. Biochem. (Tokyo)* **107**, 400 (1990)
- Li, Y. S., Liu, K. F. and Wan, Q. C.: *Toxicon* **24**, 875 (1986)
- Lomonte, B. and Gutiérrez, J. M.: *Toxicon* **27**, 725 (1989)

- Mancheva, I., Keinschmidt, T., Aleksiev, B. and Braunitzer, G.: *Hoppe-Seyler's Z. Physiol. Chem.* **365**, 885 (1984)
- Marsh, T. L., Reich, C. I., Whitelock, R. B. and Olsen, G. J.: *Proc. Natl. Acad. Sci. USA* **91**, 4180 (1994)
- Maraganore, J. M., Merutka, G., Cho, Wonhwa., Welches, W., Kézdy, F. J. and Heirikson, R. L.: *J. Biol. Chem.* **259**, 13839 (1984)
- Maraganore, J. M., Poorman, R. A. and Heinrikson, R. L.: *J. Protein. Chem.* **6**, 173 (1987)
- Matsuoka, M., Itoh, H., Kozasa, T. and Kaziro, Y.: *Proc. Natl. Acad. Sci. USA* **85**, 5384 (1988)
- Miyata, T. and Yasunaga, T.: *J. Mol. Evol.* **16**, 23 (1980)
- Nei, M.: *Molecular Evolutionary Genetics* (Columbia Univ. Press, Irvinton-on-Hudson, NY) (1987)
- Nikolov, D. B., Hu, S.-H., Lin, J., Gasch, A., Hoffmann, A., Horikoshi M., Chua, N.-H., Roeder, R. G. and Burley, S. K.: *Nature* **360**, 40 (1992)
- Oda, N., Nakamura, H., Sakamoto, S., Liu, S.-Y., Kihara, H., Chang, C.-C. and Ohno, M.: *Toxicon* **29**, 157 (1991)
- Oda, N., Ogawa, T., Ohno, M., Sasaki, H., Sakaki, Y. and Kihara, H.: *J. Biochem (Tokyo)* **108**, 816 (1990)
- Ogawa, T., Oda, N., Nakashima, K., Sasaki, H., Hattori, M., Sakaki, Y., Kihara, H. and Ohno, M.: *Proc. Natl. Acad. Sci. USA* **89**, 8557 (1992)
- Ohara, O., Tamaki, M., Nakamura, E., Tsuruta, H., Fujii, Y., Shin, M., Teraoka, H. and Okamoto, M.: *J. Biochem (Tokyo)* **99**, 733 (1986)
- Ohno, S., Kawasaki, H., Imajoh, S., Suzuki, K., Inagaki, M., Yokokura, H., Sakoh, T. and Hidaka, H.: *Nature* **325**, 161 (1987)
- Ohta T.: *J. Mol. Evol.* **33**, 34 (1991a)
- Ohta, T.: *Proc. Natl. Acad. Sci. USA* **88**, 6716 (1991b)
- Ohta, T. and Basten, C. J.: *Mol. Phyl. Evol.* **1**, 87 (1992)

- Renetseder, R., Brunie, S., Dijkstra, B. W., Drenth, J. and Sigler, P. B.: *J. Biol. Chem.* **260**, 11627 (1985)
- Rowlands, T., Baumann, P. and Jackson, S. P.: *Science* **264**, 1326 (1994).
- Sambrook, J., Fritsh, E. F. and Mamiatis, T.: *Molecular Cloning: A Laboratory Manual* 2nd ed. (Cold Spring Harbor Lab. Press, Plainview, NY) (1989)
- Sanger, F., Nicklen, S. and Coulson, A. R.: *Proc. Natl. Acad. Sci. USA* **74**, 5463 (1977)
- Schultz, M. C., Reeder, R. H. and Hahn, S.: *Cell* **69**, 697 (1992)
- Scott, D. L., Achari, A., Vidal, J. C. and Sigler, P. B.: *J. Biol. Chem.* **267** 22645 (1992)
- Scott, D. L., White, S. P., Otwinowski, Z., Yuan, W., Gelb, M. H. and Sigler, P. B.: *Science* **250**, 1541 (1990)
- Seilhamer, J. J., Prouzanski, W., Vadas, P., Plant, S., Miller, J. A., Kloss, J. and Johnson, L. K.: *J. Biol. Chem.* **264**, 5335 (1989)
- Seilhamer, J. J., Randall, T. J., Yamanaka, M. and Johnson, L. K.: *DNA* **5**, 519 (1989)
- Sharp, P. A.: *Cell* **68**, 819 (1992)
- Shukla, S. D. and Hanahan, D. J.: *Arch. Biochem. Biophys.* **209**, 668 (1981)
- Steinmetz, M. and Hood, L.: *Chromosome* **17**, 727 (1983)
- Strathmann, M. and Simon, M., I.: *Proc. Natl. Acad. Sci. USA* **87**, 9113 (1990)
- Sumita, K., Makino, Y., Katoh, K., Kishimoto, T., Muramatsu, M., Mikoshiba, K. and Tamura, T.: *Nucleic Acid Res.* **21**, 2769 (1993)
- van den Gergh, C.J., Slotboon, A. J., Verheij, H. M. and de Haas, G. H.: *Eur. J. Biochem.* **176**, 353 (1988)
- Tamura, T., Sumita, K., Fujino, I., Aoyama, A., Horikoshi, M., Hoffmann, A., Roeder, R. G., Muramatsu, M. and Mikoshiba, K.: *Nucleic Acid Res.* **19**, 3861 (1991)
- Tanaka, T., Kimura, S. and Ota, Y.: *Nucleic Acid Res.* **15**, 3178 (1987)
- Teng, C.-M., Kuo, Y.-P., Lee, L.-G. and Ouyang, C.: *Thrombosis Res.* **44**, 875 (1986)

Tonegawa, S.: *Nature* **302**, 575 (1983)

Vishwanath, B. S., Kini, R. M. and Gowga, T. V.: *Toxicon* **25**, 501 (1987)

Vital, B. O. and Excel, B. J.: *J. Physiol. Lond.*, **212**, 34 (1971)

Verheij, H. M., Boffa, M. C., Rothen, C., Bryckaert, M. C., Verger, R. and de Haas,
G. H.: *Eur. J. Biochem.* **112**, 25 (1980)

Wong, J. M., Liu, F. and Bateman, E.: *Nucleic Acid Res.* **20**, 4817 (1992)

Yoshizumi, K., Liu, S.-Y., Miyata, T., Saita, S., Ohno, M., Iwanaga, S. and Kihara,
H.: *Toxicon* **28**, 43 (1990)

Zuker, M. and Steigler, P.: *Nucleic Acid Res.* **9**, 133 (1981)

