J Mol Evol (1996) 43:170–178

JOURNAL OF MOLECULAR EVOLUTION

Class III POU Genes: Generation of Homopolymeric Amino Acid Repeats Under GC Pressure in Mammals

Kenta Sumiyama,[1] Keiko Washio-Watanabe,[1] Naruya Saitou,[2] Toshiyuki Hayakawa,[1] Shintaroh Ueda[1]

[1] Department of Biological Sciences, Graduate School of Science, University of Tokyo, 7-3-1 Hongo, Bunkyo-Ku, Tokyo 113, Japan
[2] Laboratory of Evolutionary Genetics, National Institute of Genetics, 1111 Yata, Mishima 411, Japan

Received: 27 December 1995 / Accepted: 10 April 1996

Abstract. The class III POU transcription factor genes play an important role in the nervous system. Comparison of their entire amino acid sequences disclosed a remarkable feature of particular mammalian class III POU genes. Alanine, glycine, and proline repeats were present in the mammalian Brain-1 gene, whereas most of these repeats were absent in the nonmammalian homologue. The mammalian Brain-2 gene had alanine, glycine, proline, and glutamine repeats, which were missing in the nonmammalian homologue. The mammalian Scip gene had alanine, glycine, proline, and histidine repeats, but the nonmammalian homologue completely lacked these repeats. In contrast, the mammalian Brain-4 gene had no amino acid repeats like its nonmammalian homologue. The mammalian genes containing the characteristic amino acid repeats had another feature, higher GC content. We found a positive correlation between the GC content and the amino acid repeat ratio. Those amino acids were encoded by triplet codons with relatively high GC content. These results suggest that the GC pressure has facilitated generation of the homopolymeric amino acid repeats.

Key words: POU — Human Brain-1 — GC pressure — Amino acid repeats

Correspondence to: S. Ueda

Introduction

A large number of homeobox genes have been isolated and grouped into several families. The POU family gene is identified as a homeobox gene containing a conserved 150–160-amino-acid region termed ''POU domain,'' which was first discovered as a common sequence among four transcription factors, Pit-1, Oct-1, Oct-2, and unc-86 (Herr et al. 1988). The POU domain is required for high-affinity binding to the octamer DNA sequence, which is a well-studied motif present in the enhancers and the promoters of both ubiquitously and tissue-specifically expressed genes. The POU domain is divided into two regions separated by a short sequence called a linker: a classical homeodomain termed ''POU-homeo domain'' in the carboxy-terminal region and a specific domain termed a ''POU-specific domain'' in the amino-terminal region. At present, POU family genes are classified into five or more classes, based on the homology of the POU domain (for review, see Rosenfeld 1991; Verrijzer and Van der Vliet 1993; Duboule 1994).

POU family proteins act as transcription factors, which regulate tissue-specific gene expression at different stages of development. The class III POU members are especially thought to have an important role in neural system development. The mammalian class III POU genes (Brain-1, Brain-2, Brain-4, and Scip) are expressed in the central nervous system (He et al. 1989; Le Moine and Young 1992; Mathis et al. 1992). The Brain-2 gene is strongly expressed in the dorsomedial paracellular region of the paraventricular nucleus in which corticotropin-releasing hormone (CRH) is produced, and its prod-

uct binds to the DNA element in the promoter region of the CRH gene (Li et al. 1993). The Scip gene is expressed in specific neurons and myelinating glia, and its product binds to the promoter of myelin protein P0 gene, a Schwann cell surface adhesion molecule (He et al. 1991). The Brain-4 gene is expressed in rats from embryonic day 11.5 into adulthood (Le Moine and Young 1992) and is considered to be a candidate gene for deafness with fixation of the stapes (De Kok et al. 1995) Cf1a, a neuron-specific class III POU transcription factor of fruit fly, binds to the DNA element required for expression of the DOPA decarboxylase gene in selected dopaminergic neurons (Johnson and Hirsh 1990). POU-M1 of silk worm is expressed in neural system as well as in silk gland (Fukuta et al. 1993). Class III POU genes were found also in nematode and planarian (Bürglin et al. 1989; Orii et al. 1993).

Through comparison of the amino acid sequences among the class III POU genes so far published, we noticed the presence of amino acid repeats only in certain genes. Using newly determined complete nucleotide sequences of the human Brain-1 gene, we discovered a remarkable feature—that characteristic amino acid repeats were present only in the mammalian Brain-1, Brain-2, and Scip genes and were well conserved in both position and repeat number among each mammalian gene. It is known that the genome of warm-blooded vertebrates including mammals is a mosaic of very long (more than 200 kilobases) DNA segments, isochores, whereas the genome of cold-blooded vertebrates is compositionally uniform, and that the base composition of genes mainly depends on the GC content of the isochores harboring the sequences (Bernardi 1993 and references therein). Nucleotide changes are, therefore, considered to be under the influence of compositional constraints (GC/AT pressure). As a significant proportion of the nucleotide changes at the third codon position are silent substitutions, it is likely that the GC content at the third codon position reflects directly the degree of compositional constraints onto the isochores harboring the sequences. We found a strong correlation between the repeat numbers of the characteristic amino acids and the GC content at the third codon position of the class III POU genes.

Materials and Methods

Cloning of the Human Class III POU Genes. The class III POU probe was prepared by degenerate PCR using total brain cDNA from rat embryo as a template DNA. The primers were 5′-CCGAATTC-ACCTCTGACGACCTGGA-3′ and 5′-CCAAGCTTGGGTCAT-GCT(T/C)TT(T/C)TC(T/C)TT(T/C)TG-3′, which were designed to amplify the conserved POU domain region. PCR product was digested with restriction enzymes *Eco*RI and *Hin*dIII and cloned into a phagemid vector. A phagemid clone containing the POU domain of the rat Brain-2 gene was obtained, and its insert DNA was used as a probe for screening the human genomic library. Recombinant phage clones containing the human class III POU genes were isolated, and their nucleotide sequences were determined by the dideoxynucleotide chain termination method. Other DNA manipulation procedures were performed using standard protocols (Sambrook et al. 1989).

Table 1. List of genomic DNA (g), cDNA (c), and amino acid (a) sequences employed in this study

Species	Gene	Reference
Human	Brain-1 (g)	This study
	Brain-2 (c)	Schreiber et al. 1993
	Brain-4 (c)	de Kok et al. 1995
	Scip (c)	Faus et al. 1994
Mouse	Brain-1 (g)	Hara et al. 1992
	Brain-2 (g)	Hara et al. 1992
	Brain-4 (g)	Hara et al. 1992
	Scip (c)	Meijer et al. 1990
Rat	Brain-2 (a)	Li et al. 1993
	Brain-4 (c)	Le Moine and Young 1992
	Scip (c)	Monuki et al. 1990
Frog	XLPOU1 (c)	Agarwal and Sato 1991
	XLPOU2 (c)	Witta et al. 1995
	XLPOU3 (c)	Baltzinger et al. 1992
Zebrafish	ZFPOU1 (c)	Matsuzaki et al. 1992
Fruit fly	Cfla (c)	Billin et al. 1991
Silk worm	POU-M1 (c)	Fukuta et al. 1993
Planarian	DJPOU1 (c)	Orii et al. 1993
Nematode	ceh6 (a)	Bürglin et al. 1989

Sequence Analysis. The class III POU genes employed here were from five vertebrates and four invertebrates as listed in Table 1: human (*Homo sapiens*), mouse (*Mus musculus*), rat (*Rattus norvegicus*), frog (*Xenopus laevis*), zebrafish (*Brachydanio rerio*), fruit fly (*Drosophila melanogaster*), silk worm (*Bombyx mori*), planarian (*Dugesia japonica*), and nematode (*Caenorhabditis elegans*). The amino acid sequence of the human Brain-1 was deduced from the nucleotide sequence of genomic DNA determined in the present study. Amino acid sequences of human Brain-2, Brain-4, and Scip (also termed Oct-6 or Tst-1); mouse Scip; rat Brain-4 and Scip; frog XLPOU1, XLPOU2, and XLPOU3; zebrafish XFPOU1; fruit fly Cf1a; silk worm POU-M1; and planarian DJPOU1 are deduced from their cDNA sequences, while those of mouse Brain-1, Brain-2, and Brain-4 are from genomic sequences. As for rat Brain-2 and nematode ceh6, only their amino acid sequences are reported. These sequence data were obtained from DDBJ/EMBL/GenBank, PIR and SWISS-PROT databases. These amino acid sequences were initially aligned using a multiple sequence alignment program in CLUSTAL V (Higgins et al. 1992) and were further visually adjusted to increase similarity. Phylogenetic trees were constructed for 144 amino acid residues of the conserved POU domain containing POU-specific, linker, and POU-homeo domains using the neighbor-joining method (Saitou and Nei 1987), because the sequence homology outside the POU domain was too low to align their entire amino acid sequences among all the class III POU genes. The bootstrap probabilities were obtained by resampling 1,000 times using CLUSTAL V.

Results and Discussion

Human Brain-1 Gene

We isolated several independent clones of the human genomic DNAs containing the class III POU genes with the reduced washing condition using 0.15 M sodium

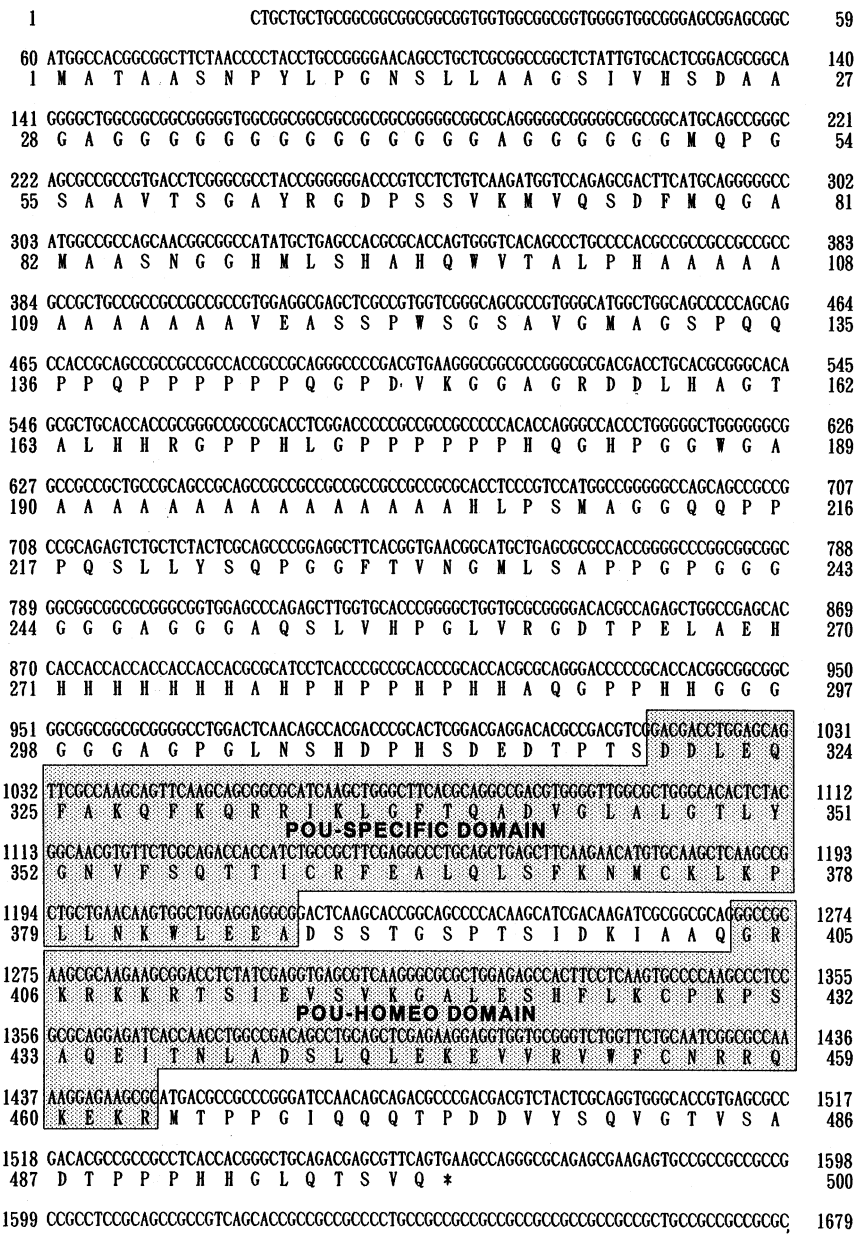

Fig. 1. Nucleotide sequence of the human Brain-1 genomic DNA and its deduced amino acid sequence. Translation start site (ATG) is predicted in concordance with that of the mouse Brain-1 gene (Hara et al. 1992). The POU-specific and POU-homeo domains are indicated with *shadowed boxes. Asterisk* indicates stop codon.

chloride and 0.015 M sodium citrate at 55°C. These clones were identified by partially determining their nucleotide sequences as the Brain-1, Brain-2, and Scip genes. Entire coding sequences of the human Brain-2, Brain-4, and Scip genes have already been published, whereas only the amino acid sequence of the POU domain was reported for the human Brain-1 gene. Therefore, we determined the complete nucleotide sequence of the human Brain-1 gene on both strands and deduced its amino acid sequence (Fig. 1), based on the similarities to the mouse Brain-1 gene (Hara et al. 1992). There was a discrepancy of a single amino acid residue between the previous sequence by He et al. (1989) and the present sequence: Position 433 of figure 1 was serine in He et al. (1989), whereas that was alanine in the present study. This discrepancy can be accounted for by polymorphism. However, the alanine residue of this position is conserved among the Brain-1 homologues of the distantly related species (mice and zebrafish). This alanine residue is also highly conserved among the class III POU genes. We therefore adopted our sequence for the basis of the comparison in the present study.

The human Brain-1 gene was intronless and GC rich (73.9% in the entire coding region), especially in the amino-terminal region of the coding sequence, like the mouse Brain-1 gene (Hara et al. 1992). This region encoded alanine, glycine, and proline repeats, and these characteristic amino acid repeats were well conserved

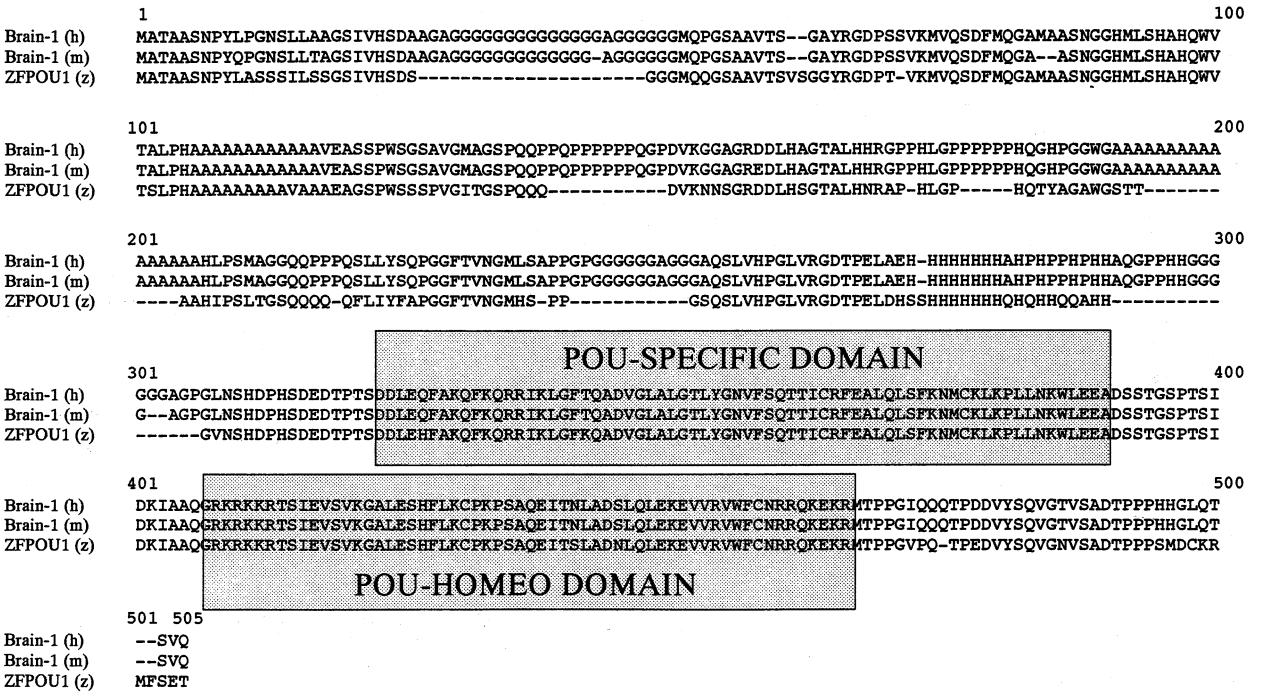

Fig. 2. Sequence alignment within the Brain-1 homologues. *h*, *m*, and *z* indicate human, mouse, and zebrafish, respectively. The POU-specific and POU-homeo domains are indicated with *shadowed boxes.* Gaps are represented by *hyphens.*

between humans and mice not only in position but also in repeat number.

Amino Acid Repeats Characteristic to Mammalian Genes

Whole amino acid comparison showed high sequence similarity among the human and mouse Brain-1 and the zebrafish ZFPOU1 genes, as shown in Fig. 2. The human and the mouse amino acid sequences were almost identical with each other. Variations in the Brain-1 between humans and mice were single amino acid difference in positions 10, 17, and 158, one gap involving two residues in positions 84–85, and two gaps caused by length variations in repeat number of amino acid repeats in positions 30–43 (14 and 13 glycine repeats in humans and mice, respectively) and positions 298–303 (six and four glycine repeats in humans and mice, respectively). In contrast, the zebrafish ZFPOU1 showed a striking feature quite different from the mammalian homologues (Brain-1): absence of some homopolymeric amino acid repeats (sequences without interruptions in the run of a single amino acid residue). Those missing repeats were alanine repeats in positions 191–206, glycine repeats in positions 30–43, 243–248, and 298–303, and proline repeats in positions 141–146 and 176–181.

There was the same situation in the Brain-2 homologues as that in the Brain-1 homologues. The mammalian Brain-2 genes also had several homopolymeric amino acid repeats like the mammalian Brain-1 genes. However, the nonmammalian Brain-2 homologue (frog XLPOU3) lacked all the amino acid repeats consisting of alanine (positions 173–177), glycine (positions 68–88), proline (positions 241–247), and glutamine (positions 129–151) in mammals, as shown in Fig. 3A.

In the Scip homologues, the situation was the same as those in the Brain-1 and Brain-2 homologues (Fig. 3B). The mammalian genes were highly similar to each other throughout the entire coding region including amino acid repeats such as alanine, glycine, proline, and histidine repeats (positions 27–37, 80–85, 113–118, 176–181, 449–454, 458–463, and so on), whereas the frog homologue (XLPOU1) had no amino acid repeats. In addition, an intraspecific length variation of alanine repeats in positions 27–37 both in humans and mice was found. In humans Faus et al. (1994) found nine repeats, while we found 11 repeats (data not shown). In mice there were eight and nine repeats (Meijer et al. 1990; Suzuki et al. 1990; Zimmerman et al. 1991; Hara et al. 1992).

In contrast to the Brain-1, Brain-2, and Scip genes, the mammalian Brain-4 genes so far reported have no characteristic amino acid repeats. Therefore, there were no gaps between the mammalian and nonmammalian homologues (Brain-4 an XLPOU2), as shown in Fig. 3C.

Wide Variation of GC Content in the Class III POU Gene Group

A high GC content was another characteristic feature for the mammalian Brain-1, Brain-2, and Scip genes, in ad-

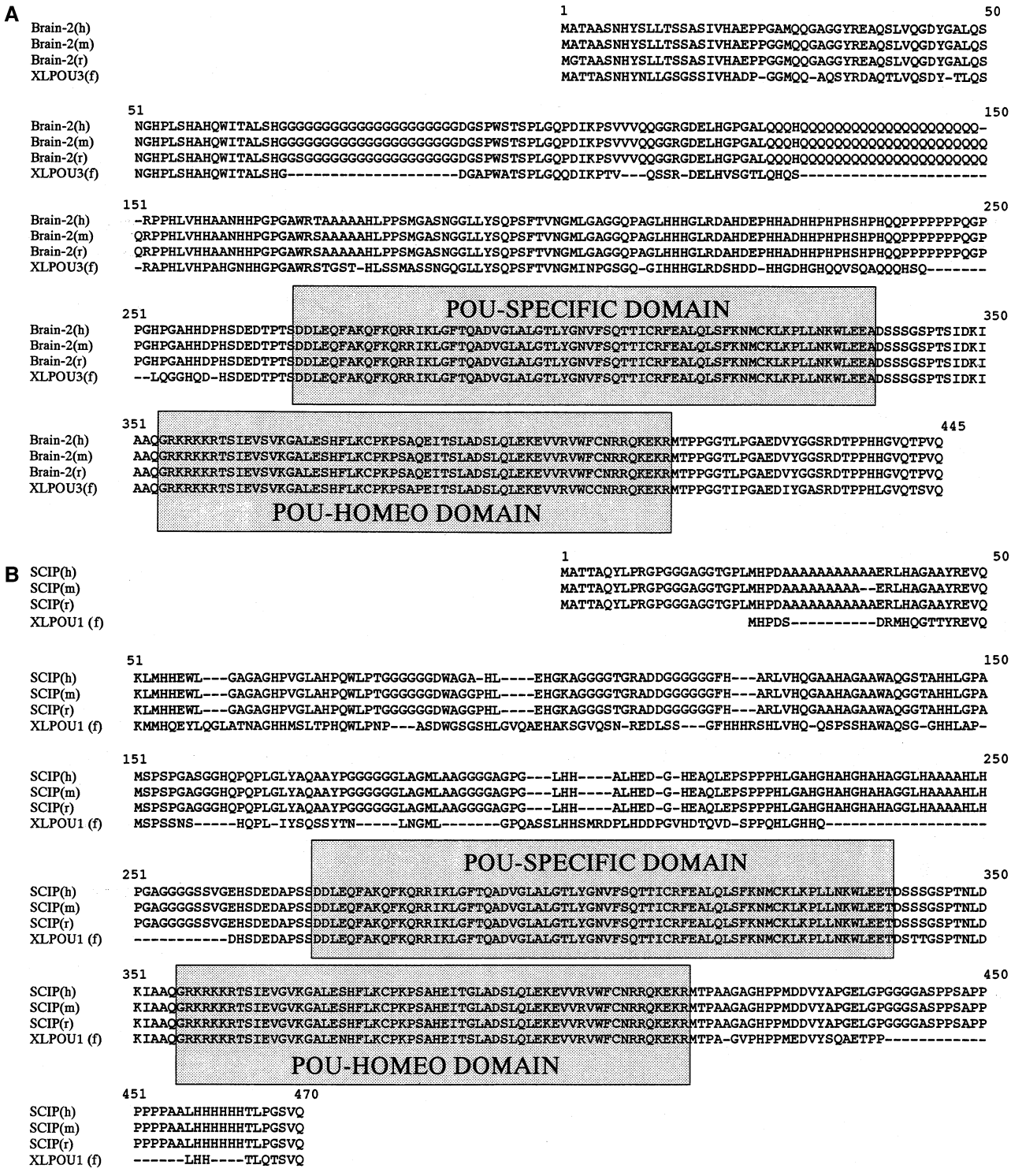

Fig. 3. Sequence alignment within the homologues of the class III POU genes. **A** Brain-2, **B** Scip and **C** Brain-4. *h, m, r,* and *f* indicate human, mouse, rat, and frog, respectively. The POU-specific and POU-homeo domains are indicated with *shadowed boxes.* Gaps are represented by *hyphens.*

dition to an extraordinary structure containing homopolymeric amino acid repeats. Table 2 shows the GC content at the first, second, and third codon positions of all the class III POU genes, together with those of the POU domain alone. A great variation in the GC content was observed especially at the third codon position, where GC pressure is directly reflected. The lowest GC content at the third position was 23.0% for the planarian DJPOU1, while the highest one was 91.4% for the human Brain-1. There was a wide variation of the GC content at the third position only among the mammalian genes, ranging from 64.6% for the rat Brain-4 to 91.4%

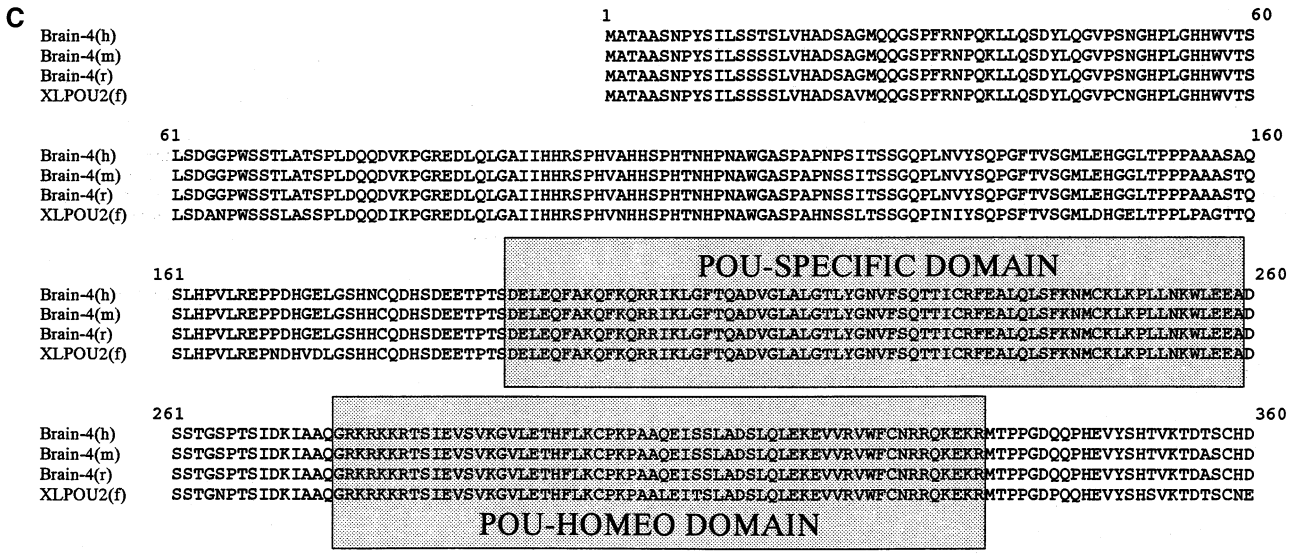

Fig. 3. Continued.

for the human Brain-1. This wide variation among the mammalian class III POU genes was not interspecific but intergenic.

To the contrary, the zebrafish ZFPOU1 and the frog XLPOU1 and XLPOU3 genes showed relatively lower GC content. There is a wide variation at the third codon position among genes of warm-blooded vertebrates including mammals (from 40% through nearly 100%), whereas the GC content in the other organisms was within a narrow range (Bernardi 1993 and references therein). Therefore, difference of the GC content between the nonmammalian genes (zebrafish ZFPOU1, frog XLPOU3, and frog XLPOU1) and their mammalian homologues (Brain-1, Brain-2, and Scip) might reflect the difference in genome structure between the warm- and cold-blooded vertebrates.

Table 2. GC contents (%) of class III POU genes

		POU domain			Entire coding region		
Species	Gene	First	Second	Third	First	Second	Third
Human							
	Brain-1	55.6	38.9	95.8	73.3	57.1	91.4
	Brain-2	50.0	39.6	84.0	70.5	49.6	86.3
	Brain-4	50.0	38.9	76.4	59.4	46.7	72.7
	Scip	56.9	38.9	96.5	76.7	56.5	89.4
Mouse							
	Brain-1	55.6	38.9	82.6	73.0	56.9	83.3
	Brain-2	49.3	39.6	79.2	70.2	49.3	77.8
	Brain-4	49.3	38.9	63.9	58.6	46.7	65.8
	Scip	57.6	38.9	91.7	77.3	56.4	84.7
Rat							
	Brain-4	49.3	38.9	63.2	58.6	46.7	64.6
	Scip	57.6	38.9	90.3	77.0	56.6	83.9
Frog							
	XLPOU1	46.5	38.2	68.1	56.2	42.4	67.6
	XLPOU2	48.6	38.2	56.9	55.0	44.2	61.6
	XLPOU3	46.5	41.0	72.9	57.7	46.2	71.0
Zebrafish							
	ZFPOU1	50.0	38.2	72.2	58.5	47.9	65.7
Fruit fly							
	Cf1a	56.3	38.9	84.7	65.1	48.2	74.0
Silk worm							
	POU-M1	53.5	38.9	74.3	63.4	47.7	78.1
Planarian							
	DJPOU1	39.6	38.2	25.7	37.3	36.4	23.0

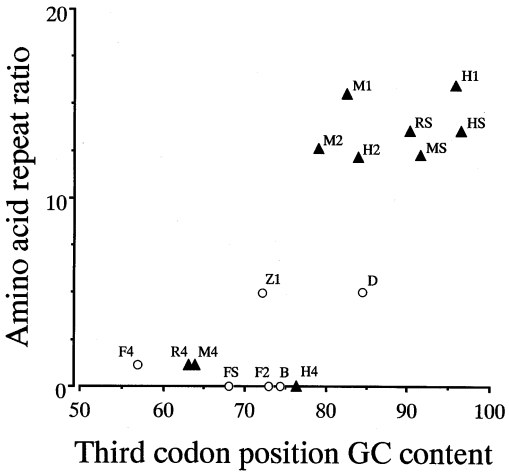

Fig. 4. Correlation between the GC content (%) and the amino acid repeat ratio (AARR). The genes employed here were human Brain-1 (*H1*), mouse Brain-1 (*M1*), zebrafish Brain-1 homologue, ZFPOU1 (*Z1*), human Brain-2 (*H2*), mouse Brain-2 (*M2*), frog Brain-2 homologue, XLPOU3 (*F2*), human Scip (*HS*), mouse Scip(*MS*), rat Scip (*RS*), frog Scip homologue, XLPOU1 (*FS*), human Brain-4 (*H4*), mouse Brain-4 (*M4*), rat Brain-4 (*R4*), frog Brain-4 homologue, XLPOU2 (*F4*), fruit fly Cf1a (*D*), and silk worm POU-M1 (*B*). *Triangles* and *circles* represent the mammalian and nonmammalian class III POU genes, respectively. Since the planarian DJPOU1 containing no amino acid repeats showed an extraordinary GC content (23.0%), it was excluded from this figure and calculation of correlation coefficient.

Correlation Between GC Content and Amino Acid Repeat Ratio

The mammalian Brain-1, Brain-2, and Scip genes had both the high content of characteristic amino acid repeats and the high GC content. We thus investigated whether there was a positive correlation between them or not. There were no amino acid repeats within the POU domain and the GC content at the third codon position of the POU domain was nearly equal to that at the third codon position of the entire region (see Table 2). Thus we used the GC content at the third codon position of the POU domain so as to make an independent comparison with the amino acid repeat ratio possible.

We defined the amino acid repeat ratio (AARR) in % as follows:

$$\text{AARR} = \text{NRAA/NAA} \times 100 \qquad (1)$$

where NRAA is the total number of identical amino acid residues consecutive more than three, and NAA is the total number of amino acid residues compared. As shown in Fig. 4, we found a clear positive correlation between GC content and AARR (correlation coefficient was 0.82).

One possible explanation for this correlation is that

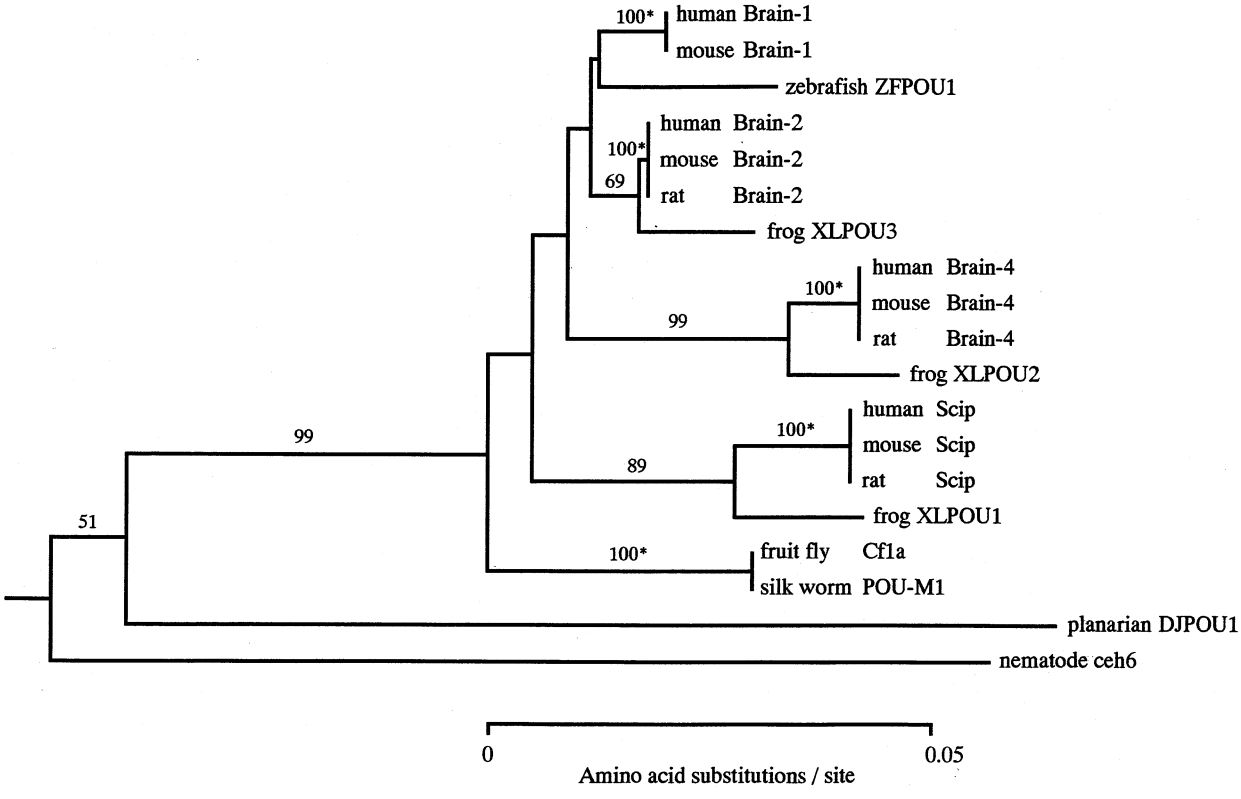

Fig. 5. A phylogenetic tree of the class III POU genes. The human Oct-1 was used as an outgroup. *Numbers* on the branches represent bootstrap probabilities (%). The probabilities are shown only for those greater than 50%. Since the POU domain amino acid sequences were highly conserved, those of Brain-1, Brain-2, Brain-4, and Scip were identical among their respective homologues of all the mammals compared in this study, as well as between fruit fly Cf1a and silk worm POU-M1, shown as *vertical lines* in this figure. *Asterisks* indicate the theoretically expected values, because of identical amino acid sequences.

the GC pressure has caused generation of amino acid repeats during the vertebrate evolution. There is a wide variation of the GC content among genes of warm-blooded vertebrates, whereas cold-blooded vertebrate genes show the relatively lower GC content within a narrow range (Bernardi 1993 and references therein). In fact, the class III POU genes from cold-blooded vertebrates (amphibians and fish) having low GC content had few repeats, whereas those from warm-blooded vertebrates (mammals) showed two phases opposite to each other: the higher GC-content genes containing a number of amino acid repeats such as Brain-1, Brain-2, and Scip, and the lower GC-content genes containing very few amino acid repeats such as Brain-4. Moreover, amino acid residues constituting the repeats common to the mammalian Brain-1, Brain-2, and Scip were alanine, glycine, and proline, whose codons were GC rich (GCN, GGN, and CCN, respectively).

Figure 5 shows a phylogenetic tree among the class III POU genes constructed using the conserved POU domain sequences. The phylogenetic relationship among Brain-1, Brain-2, Brain-4, and Scip was not clear because of low bootstrap probabilities (less than 50%) for the corresponding clusters. But it is likely that Brain-1, Brain-2, Brain-4, and Scip genes emerged before divergence among vertebrates. This suggests that the ancestral class III POU gene contained no or few amino acid repeats and that generation of homopolymeric amino acid repeats encoded by GC-rich trinucleotides occurred independently in the lineages of particular class III POU genes (Brain-1, Brain-2, and Scip) under the GC pressure. Those events occurred at least after emergence of mammals, probably after the divergence between cold- and warm-blooded vertebrates. It might result from the location effect in the mosaic genome of warm-blooded vertebrates (Bernardi 1993 and references therein). The human and mouse Brain-4 genes are mapped on Xq21, an evolutionarily conserved region of the X chromosome (Douville et al. 1994). Based on the above assumption, the mammalian Brain-4 gene retaining the ancestral feature of the class III POU gene (no or few amino acid repeats) might be confined in the lower GC region of the X chromosome. DXS995 has a relatively lower GC content (43.7%), which locates 20 kb distal to the Brain-4 gene in humans (Weissenbach et al. 1992; Huber et al. 1994; de Kok et al. 1995). The other mammalian class III POU genes with the homopolymeric amino acid repeats encoded by GC-rich trinucleotides and relatively higher GC contents locate on other chromosomes (Xia et al. 1993; Atanasoski et al. 1995).

Amino acid residues constituting the repeats of the class III POU genes were alanine, glycine, proline, glutamine, and histidine. Since alanine-, glycine-, proline-, and glutamine-rich sequences are identified as transcriptional activation domains (Mitchell and Tjian 1989; Duboule 1994 and references therein), occurrence of the GC-rich sequences encoding these characteristic amino acid residues could influence the transactivation ability. The occurrence of such amino acid repeats in the mammalian transcription factors must facilitate diversification of gene regulation mechanisms in the central nervous system and might generate novel brain organizations including neocortex. Interestingly, after the emergence of warm-blooded mammals the amino acid repeats of the class III POU genes have been well conserved both in amino acid residue and in their repeat numbers. These suggest that evolution of the neural system has been influenced by the GC pressure, at least in mammals. Sequence data of the class III POU genes from reptiles and birds could provide further evidence for generation of the homopolymeric amino acid repeats under GC pressure.

Acknowledgments. This study was supported by grants from the Ministry of Education, Science, Sports, and Culture of Japan, The Inamori Foundation, and NIG Cooperative Research Program.

References

Agarwal VR, Sato SM (1991) XLPOU1 and XLPOU2, two novel domain genes expressed in the dorsoanterior region of *Xenopus* embryos. Dev Biol 147:363–373

Atanasoski S, Toldo SS, Malipiero U, Schreiber E, Fries R, Fontana A (1995) Isolation of the human genomic Brain-2/N-Oct 3 gene (POUF3) and assignment to chromosome 6q16. Genomics 26:272–280

Baltzinger M, Payen E, Remy P (1992) Nucleotide sequence of XL-POU3 cDNA, a member of the POU domain gene. Nucleic Acids Res 20:1993

Bernardi G (1993) The isochore organization of the human genome and its evolutionary history—a review. Gene 135:57–66

Billin AN, Cockerill A, Poole SJ (1991) Isolation of a family of *Drosophila* POU domain genes expressed in early development. Mech Dev 34:75–84

Bürglin TR, Finney M, Coulson A, Ruvkun G (1989) *Caenorhabditis elegans* has sources of homeobox-containing genes. Nature 341: 239–243

De Kok YJM, Van der Maarel SM, Bitner-Glindzicz M, Huber I, Monaco AP, Malcolm S, Pembrey ME, Ropers HH, Cremers FPM (1995) Association between X-linked mixed deafness and mutations in the POU domain gene *POU3F4.* Science 267:685–688

Duboule D (1994) Guidebook to the homeobox genes. Sambrook and Tooze Publication at Oxford University Press, Oxford

Douville PJ, Atanasoski S, Tobler A, Fontana A, Schwab ME (1994) The brain-specific POU-box gene *Brn4* is a sex-linked transcription factor located on the human and mouse X chromosomes. Mamm Genome 5:180–182

Faus I, Hsu H, Fuchs E (1994) Oct-6: a regulator of keratinocyte gene expression in stratified squamous epithelia. Mol Cell Biol 14:3263–3275

Fukuta M, Hui C, Matsuno K, Nagata T, Takiya S, Suzuki Y (1993) Molecular cloning of a POU domain-containing factor involved in the regulation of the *Bombyx sericin-1* gene. J Biol Chem 268: 19471–19475

Hara Y, Rovescalli AC, Kim Y, Nirenberg M (1992) Structure and evolution of four POU domain genes expressed in mouse brain. Proc Natl Acad Sci USA 89:3280–3284

He X, Treacy MN, Simmons DM, Ingraham HA, Swanson LW, Rosen-

feld MG (1989) Expression of large family of POU-domain regulatory genes in mammalian brain development. Nature 340:35–42

He X, Gerrero R, Simmons DM, Park RE, Lin CR, Swanson LW, Rosenfeld MG (1991) Tst-1, a member of the POU domain gene family, binds the promoter of the gene encoding the cell surface adhesion molecule Po. Mol Cell Biol 11:1739–1744

Herr W, Sturm RA, Clerc RG, Corcoran LM, Baltimore D, Sharp PA, Ingraham HA, Rosenfeld MG, Finney M, Ruvkun G, Horvitz HR (1988) The POU domain: a large conserved region in the mammalian *pit*-1, *oct*-1, *oct*-2, and *Caenorhabiditis elegans unc*-86 gene products. Genes Dev 2:1513–1516

Higgins DG, Bleasby AJ, Fuchs R (1992) CLUSTAL V: improved software for multiple sequence alignment. Comput Appl Biosci 8:189–191

Huber I, Bitner-Glindzicz M, De Kok YJM, Van der Maarel SM, Ishikawa-Brush Y, Monaco AP, Robinson D, Maocolm S, Pembrey ME, Brunner HG, Cremers FPM, Ropers H (1994) X-linked mixed deafness (DFN3): cloning and characterization of the critical region allows the identification of novel microdeletions. Hum Mol Genet 3:1151–1154

Johnson WA, Hirsh J (1990) Binding of *Drosophila* POU-domain protein to a sequence element regulating gene expression in specific dopaminergic neurons. Nature 343:467–470

Le Moine C, Young III SW (1992) RHS2, a POU domain-containing gene, and its expression in developing and adult rat. Proc Natl Acad Sci USA 89:3285–3289

Li P, He X, Gerrero MR, Mok M, Aggarwal A, Rosenfeld MG (1993) Spacing and orientation of bipartite DNA-binding motifs as potential functional determinants for POU domain factors. Genes Dev 7:2483–2496

Mathis JM, Simmons DM, He X, Swanson LW, Rosenfeld MG (1992) Brain 4: a novel mammalian POU domain transcription factor exhibiting restricted brain-specific expression. EMBO J 11:2551–2561

Matsuzaki T, Amanuma H, Takeda H (1992) A POU-domain gene of zebrafish, ZFPOU1, specifically expressed in the developing neural tissues. Biochem Biophys Res Commun 187:1446–1453

Meijer D, Graus A, Kraay R, Langevelt A, Mulder MP, Grosveld G (1990) The octamer binding factor Oct6: cDNA cloning and expression in early embryonic cells. Nucleic Acids Res 18:7357–7365

Mitchell PJ, Tjian R (1989) Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science 245: 371–378

Monuki ES, Kuhn R, Weinmaster G, Trapp BD, Lemke G (1990) Expression and activity of the POU transcription factor SCIP. Science 249:1300–1303

Orii H, Agata K, Watanabe K (1993) POU domain genes in planarian *Dugesia japonica:* the structure and expression. Biochem Biophys Res Commun 192:1395–1402

Rosenfeld MG (1991) POU-domain transcription factors: pou-er-ful developmental regulators. Genes Dev 5:897–907

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Sambrook TH, Fritsche EF, Maniatis T (1989) Molecular cloning: a Laboratory Manual. 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Schreiber E, Tobler A, Malipiero U, Schaffner W, Fontana A (1993) cDNA cloning of human N-Oct 3, a nervous-system specific POU domain transcription factor binding to the octamer DNA motif. Nucleic Acids Res 21:253–258

Suzuki N, Rohdewohld H, Neuman T, Gruss P, Schöler HR (1990) Oct-6: a POU transcription factor expressed in embryonal stem cells in the developing brain. EMBO J 9:3723–3732

Verrijizer CP, Van der Vliet PC (1993) POU domain transcription factors. Biochim Biophys Acta 1173:1–21

Weissenbach J, Gyapay G, Dib C, Vignal A, Morissette J, Millasseau P, Vaysseix G, Lathrop M (1992) A second-generation linkage map of the human genome. Nature 359:794–801

Witta SE, Agarwall VR, Sato SM (1995) X1POU2, a noggin-inducible gene, has direct neutralizing activity. Development 121:721–730

Xia YR, Anderson B, Mehrabian M, Diep AT, Warden CH, Mohandas T, McEvilly RJ, Rosenfeld MG, Lusis AJ (1993) Chromosomal organization of mammalian POU domain factors. Genomics 18: 126–130

Zimmerman EC, Jones CM, Fet V, Hogan BLM, Magnuson MA (1991) Nucleotide sequence of mouse SCIP cDNA, a POU-domain transcription factor. Nucleic Acids Res 19:956