Evolution of the POU1F1 transcription factor in mammals: rapid change of the alternatively-spliced β-domain

Article (Accepted Version)

Evolution of the POU1F1 transcription factor in mammals: rapid change of the alternatively-spliced β-domain

Michael Wallis

Biochemistry and Biomedicine Group, School of Life Sciences, University of Sussex, Brighton, BN1 9QG. UK

Corresponding author address as above; email: m.wallis@sussex.ac.uk

Short title: Evolution of POU1F1 in mammals

Keywords: Pit-1, POU1F1, β-domain, molecular evolution, evolutionary rates

Declaration of Interest: None

Abstract

The POU1F1 (Pit-1) transcription factor is important in regulating expression of growth hormone, prolactin and TSH β-subunit, and controlling development of the anterior pituitary cells in which these hormones are produced. POU1F1 is a conserved protein comprising three main domains, an N-terminal transcription activation domain (TAD), a POU-specific domain and a C-terminal homeodomain. Within the TAD, a β-domain can be inserted by alternative splicing, giving an extended 'β-variant' with altered properties. Here sequence data from over 100 species were used to assess the variability of POU1F1 in mammals. This showed that the POU-specific domain and homeodomain are very strongly conserved, and that the TAD is somewhat less conserved, as are linker and hinge regions between these main domains. On the other hand, the β-domain is very variable, apparently evolving at a rate not significantly different from that expected for unconstrained, neutral evolution. In several species stop and/or frameshift mutations within the β-domain would prevent expression of the β-variant as a functional protein. In most species expression of the β-variant is low (<5% of total *POU1F1* expression). The rate of evolution of POU1F1 in mammals shows little variation, though the lineage leading to dog does show an episode of accelerated change. This comparative genomics study suggests that in most mammalian species POU1F1 variants produced by alternative splicing may have little physiological significance.

1. Introduction

The transcription factor POU1F1 (Pit-1, GHF-1) plays a crucial part in regulating the development of the anterior pituitary gland and the expression of specific pituitary hormones. Mutations in the *POU1F1* gene can lead to failure of the development of cells expressing growth hormone (GH), prolactin and TSH in mice and humans (Andersen and Rosenfeld, 2001; Cohen and Radovick 2002; Kelberman et al., 2009; Li et al., 1990; Radovick et al., 1992). The expression of GH, prolactin and the β-subunit of TSH is regulated by POU1F1, and promoters for genes encoding these hormones, *POU1F1* itself, and various associated proteins, contain binding sites for POU1F1 (Baumeister et al., 2000; Chen et al., 1990; Ellestad and Porter, 2013; Featherstone et al., 2012; Fox et al., 1990; Herman et al., 2012; Nowakowski and Maurer, 1994; Scully et al., 2000). In the adult, POU1F1 is expressed at high levels in somatotropes, lactotropes and thyrotropes. It is expressed in most other cell types at very low levels if at all, though significant expression has been reported in human placenta, hemopoietic and lymphoid tissues, and mammary gland (Bamberger et al., 1995; Delhase et al., 1993; Gil-Puig et al., 2005). Expression levels in breast tumours, and tumour-derived cell lines are often higher than those in normal breast tissue, and appear to be associated with enhanced proliferation and metastasis (Gil-Pig et al., 2005; Ben-Batalla et al., 2010).

POU1F1 is a member of the POU family of transcription factors, and like other members of the family has a multi-domain structure, with an N-terminal transcription activation (TAD) domain, a POU-specific domain and a C-terminal homeodomain (Theill et al., 1989) (Fig. 1). These domains are strongly

63 conserved, whereas the regions between them, postulated to comprise flexible
64 linkers, are more variable (Majumdar et al., 1996; Morris et al., 1992; Theill et
65 al., 1989). An additional region, the β-domain, can be inserted within the TAD as
66 a consequence of alternative splicing, two splice forms occurring in which the β-
67 domain is present or absent (Delhase et al., 1995; Morris et al., 1992; Theill et
68 al., 1992). The two splice variants in mammals have substantially different
69 biological properties which have been studied extensively (Diamond and
70 Gutierrez-Hartmann, 1996, 2000; Jonsen et al., 2009; Sánchez-Pacheco et al.,
71 1998; Sporici et al., 2005), but their physiological roles are not well defined.
72 Additional splice variants of POU1F1 have been described in sheep (Bastos et
73 al., 2006), but it is unclear whether these play a specific biological role.
74
75 The view that the domains in POU1F1 are strongly conserved is based on a
76 relatively small number of mammalian and non-mammalian species. The
77 availability of genomic data from over 100 mammalian species, including most
78 of the extant taxonomic orders, makes possible a much fuller study of POU1F1
79 variation in mammals and its evolutionary significance. The availability of
80 transcriptomic data for a number of species allows evaluation of POU1F1 splice
81 variation across mammals. Such a study is reported here. Questions addressed
82 include 1) are the POU1F1 domain sequences strongly conserved across all
83 mammals? 2) is there evidence for variable rates of evolution as seen for the
84 target genes of POU1F1, GH and prolactin (Li et al., 2005; Wallis, 1996, 2008;
85 Wallis et al., 2000)? 3) To what extent are splice variations in the POU1F1 gene
86 conserved across mammals, especially with regard to the form containing the β-
87 domain (the β-variant)?

2. Methods

2.1. Sequences

cDNA sequences for *POU1F1* from various mammals were obtained by searching the publically available ncbi nucleotide database using BLAST (Altschul et al., 1990) with human POU1F1 β-variant cDNA as Query. In all cases they were checked against appropriate wgs or sra databases (https://trace.ncbi.nlm.nih.gov) using BLAST. Additional sequences were obtained by searching sra databases using BLAST and sequences from related species. Sequences were aligned in Mesquite (Maddison and Maddison, 2016) and translated to protein sequences. Sources for all the sequences used and full CDS and protein alignments are given in Supplementary Table 1 and Supplementary Figs. 1 and 2. Domains within sequences were assigned on the basis of Fig. 1.

2.2. Sequence analysis - evolutionary rates

To analyse evolutionary rates of different regions within the *POU1F1* CDS sequences, the codeml programme in the paml package (Phylogenetic Analysis by Maximum Likelihood; Yang, 2007) was used to determine the ratio (dN/dS) of nonsynonymous substitutions (which alter amino acid sequence) to synonymous substitutions (which do not). For most coding sequences dN/dS is low, reflecting maintenance of functional sequence by purifying selection. For a sequence with little or no specific function dN/dS approaches 1.0, the neutral rate of evolution. If dN/dS is significantly greater than 1.0, the sequence is undergoing rapid adaptive evolution by natural selection, though a value lower

than 1.0 does not necessarily rule out adaptive evolution.

Alignments of CDS sequences corresponding to all or subregions of the *POU1F1* mRNA were analysed using the codeml method (Yang, 2007), using a defined phylogenetic tree. Significance of differences between dN/dS ratios was tested using the likelihood ratio test (Yang, 2007).

2.3. Splicing patterns

Splicing patterns for the *POU1F1* gene were determined by analysing transcriptomes available for various species through the sra database (https://trace.ncbi.nlm.nih.gov/Traces/sra). In each case, *POU1F1*-related sequences were identified using BLAST with the appropriate CDS as query, and analysed to identify hits overlapping splice junctions.

3. Results and Discussion

3.1. POU1F1 Sequences

Complete *POU1F1* coding sequences were derived for a total of 113 mammalian species. Analysing all these sequences together using codeml took an excessively long time, and they were therefore divided into subgroups: (1) subgroup 1 including representatives from each of the main mammalian groups (38 spp), (2) primates, tree shrew and flying lemur (32 spp), (3) rodents and lagomorphs (19 spp), (4) Laurasiatheria (48 spp), (5) Xenarthra, Afrotheria, Marsupialia and Prototheria (14 spp). Individual species included in each of these groups (plus outgroups) are indicated in the sequence alignments given in Supplementary Figs. 1 and 2.

In no species was there clear evidence for more than one *POU1F1* gene. However, in several cases there was evidence of polymorphism, and in some of these it is conceivable that this could reflect the presence of two very similar (duplicate) genes rather than polymorphisms. In all such cases intra-specific variation was less than between-species variation (based on comparison with closely related species), so the analysis would not be affected.

Alignment of POU1F1 sequences was straightforward, with only a few insertions or deletions (indels) required. Visual assessment of alignments (Supplementary Fig. 2) indicated that the POU-specific and homeodomain domains are very strongly conserved, as suggested previously on the basis of comparison of a few species (Majumdar et al., 1996; Morris et al., 1992; Theill et al., 1989), and that

157 linker and hinge regions and the TAD are rather more variable. The β-domain is
158 very variable, particularly at the C-terminal end (Fig. 2). The sequence of dog
159 POU1F1 shows rather high variation, especially in the TAD and hinge region.
160
161 *3.2. Rates of Evolution*
162 *3.2.1. Complete POU1F1*
163 Analysis of the *POU1F1* CDS alignment for subgroup 1 (including β-domain) by
164 the codeml method gave a dN/dS ratio of 0.085, showing that the protein overall
165 is fairly strongly conserved (Table 1). Similar results were obtained for the other
166 subgroups. However, as noted above, some domains appear to be more strongly
167 conserved than others, so this value is an average; individual domains/regions are
168 considered separately below. Codeml analysis also indicated that there was
169 significant variation in dN/dS between species; this was largely due to an
170 increased rate of evolution on the lineage leading to dog, for which branch dN/dS
171 was significantly elevated (0.18; $P<0.05$, likelihood ratio test).
172
173 *3.2.2. POU-specific domain and homeodomain*
174 Analysis of the POU-specific domain and homeodomain, separately, using
175 codeml gave very low values for dN/dS (Table 1), confirming the strong
176 conservation deduced from visual inspection and previous reports. There was no
177 evidence for rate variation between species, including dog.
178
179 *3.2.3. TAD*
180 Analysis of the TAD (excluding β-domain) by codeml gave values for dN/dS
181 (0.084 for subgroup 1) similar to that obtained for POU1F1 overall (Table 1),

indicating that this domain is fairly strongly conserved, but less so than the homeodomain or POU-specific domain. Again, there was no evidence for rate variation between species; dN/dS was elevated on the branch leading to dog, but not significantly.

3.2.4. Hinge and linker regions

The hinge region between TAD and POU-specific domain is rather more variable than either of these, with dN/dS 0.146. Similarly, the short linker region (dN/dS 0.052) is more variable than its flanking POU-specific and homeodomains (Table 1). Nevertheless, both these sequences are quite strongly conserved. Neither shows evidence for rate variation between species.

3.2.5. C-terminal tail

The short C-terminal tail is rather variable. Indels in some species, and truncation in marsupials make detailed analysis difficult.

3.2.6. β-domain

Analysis of the β-domain by codeml (alignment for subgroup 1) gave a high value for dN/dS of 0.91, and similarly high values were obtained with alignments for other subsets of sequences (Table 1). In no case was the value significantly different from 1.0. This corresponds to the ratio expected for a sequence evolving by neutral evolution, unconstrained by the purifying selection imposed by functional constraints. This suggests, but does not prove, lack of function for this specific protein sequence - elevated evolutionary rate could also be due to positive selection (with dN/dS not necessarily exceeding 1.00), although in this

case one might expect to see rate variation between groups or species, which is not apparent.

However, lack of function of the β-domain (and presumably therefore the β-variant of POU1F1) is also indicated by the presence in some species of mutations in this domain which would prevent expression of the intact protein (Fig. 2). Thus in a prosimian (*Daubentonia madagascariensis*; aye aye) and an afrotherian (*Elephantulus edwardii*; elephant shrew) a stop codon in the sequence encoding the β-domain would prevent translation of the following sequence (including POU specific domain and homeodomain). In the New World monkey marmoset (three species, *Callithrix jacchus, C. kuhlii* and *C. geoffroyi*) there are two separate deletions in the β-domain, of two and one nucleotides respectively; between these the reading frame is changed, with introduction of a stop codon. In the related New World monkey tamarin (*Saguinas midas*) just one of these deletions occurs, changing the reading frame of the rest of the protein. In pangolin (*Manolis pentadactyla*) insertion of two nucleotides into the β-domain sequence would again change the reading frame for the rest of the protein.

As has been noted previously (Diamond & Gutierrez-Hartmann 1996) the N-terminal half of the β-domain is more conserved than the C-terminal half. However, examination of the CDS alignment shows that this applies to the non-coding nucleotide sequence as well as the protein sequence; the high dN/dS value for this region is due to low dS as well as high dN. Exceptions to high conservation of this region are guinea pig, elephant shrew and tenrec. The C-terminal end of the β-domain corresponds to the 14-residue insert found in the

Pit-1T variant, specific to thyrotropes (Haugen et al., 1993) and here too a high dN/dS value suggests lack of specific function. The deletions noted above in the β-domain for marmoset and tamarin fall in this region, and would be expected to prevent expression of a functional Pit-1T, but the stop codons in the β-domain af aye-aye and elephant shrew fall upstream of this region.

In the marsupial and monotreme species for which data are available, substitutions at the 3' end of the β-domain-encoding sequence alter the ..AG required for this sequence to be spliced out. However, a potential alternative splice site is introduced 3 nucleotides into exon 2. Analysis of available transcriptomic data for opposum (*Monodelphis domestica*; low expression of POU1F1 seen in transcriptomes from various tissues and whole newborn, but not available for isolated pituitary) indicates that this is used in most cases (the β-domain is retained in only one of 12 instances identified). For other marsupials and for monotremes the available transcriptomic data give no useful information on this aspect.

3.2.7. Variation of evolutionary rate between groups and species

Overall, although there is clear evidence for variation in evolutionary rates between different regions of the POU1F1 sequence, there is rather little evidence for rate variation between groups and species (Table 1). The rate (dN/dS) for Xenarthra is relatively low (Table 1), though sequences for only two species (armadillo and sloth) are available for this Eutherian superorder.

A species for which the rate of evolution is relatively high, as noted above, is the

dog. Sequences for wolf and domestic dog breeds were identical. This was studied further by examining the sequences of a number of species closely related to dog (family Canidae; fox, *Vulpes* and dhole, *Cuon*). The data for these additional species were incomplete, but did show POU1F1 sequences similar to that of dog, indicating that accelerated POU1F1 evolution occurred on the lineage leading to Canidae (given that sequences of other Caniformia - bear, panda and ferret - were conserved) (Supplementary Fig. 2). The phylogenetic trees shown in Fig. 3, based on dN and dS values, illustrate this. GH and prolactin, expression of which is controlled by POU1F1, show a markedly episodic pattern of evolution, but interestingly for these proteins the lineage leading to Canidae does not show accelerated evolution (Li et al., 2005; Wallis 1996, 2008; Wallis et al., 2000). The increased rate of POU1F1 evolution on the branch leading to dog, was confined to the TAD, hinge region and N-terminal part of the POU-specific domain (encoded by exons 1-3), but whether it was due to adaptive change or loss of function could not be determined.

3.3. Alternative splicing of the POU1F1 gene

The *POU1F1* gene product is subject to alternative splicing, giving a number of variant forms of the protein, some of which have already been discussed. The availability of transcriptomic data for a number of mammalian species enables the extent and nature of such alternative splicing to be examined.

Alternative splicing of *POU1F1* at the exon 1/exon 2 splice site (giving inclusion or exclusion of the β-domain) was assessed for those species for which transcriptomic databases were available for pituitary tissue or cells. Results are

shown in Table 2. Expression levels for the β-variant were low compared with the variant in which the β-domain is excluded - lower than 5% in all species examined except rat (12.2%) and sooty mangabey (*Cercocebus*; 6.5%). The level was particularly low (0.38%) in dog. Notably the level in marmoset (1.9%) was comparable with that in several other species, despite the fact that production of functional β-variant protein in marmoset is not possible, owing to a stop codon (see above). The expression level seen for β-domain in rat agrees closely with that originally reported by Theill et al. (1992) and Morris et al. (1992), who found a ratio of 1:7 for the variants including or excluding the β-domain. However, the results found here suggest that the rat may be exceptional, with most other species examined showing much lower expression levels of the β-variant.

In many species, additional variants caused by alternative splicing at the exon 1/exon 2 junction were observed. In most cases their incidence was less than 1% that of the main variant. Exceptions were a variant in which the splice site was 6 nucleotides into exon 2 (potentially producing a variant two amino acids shorter than normal; incidence 1-2% that of the normal variant in several species including human, cow and sheep) and a variant in which exon 2 is excluded (very rare except in the naked mole rat, *Heterocephalus*, where its incidence is about 25% that of the normal variant).

Bastos et al. (2006) reported *POU1F1* splice variants in sheep in which exon 3, or exons 3-5 were lacking. Analysis of sheep pituitary transcriptomes revealed the presence of the former, at about 6% relative to the normal variant, but the

307 variant lacking exons 3-5 was not detected. The splice variant lacking exon 3 was
308 detected at a similar level in cow, but at a much lower level (<1%) in rat, dog and
309 human.
310
311 *3.4. Conclusions*
312 Previous work on POU1F1 concluded that the main domains identified within the
313 protein are strongly conserved, while regions between these (hinge and linker
314 regions) are more variable. The present survey of POU1F1 sequences derived for
315 a large number of mammalian species generally confirms this, except for the β-
316 domain, which is very variable.
317
318 The variability of the β-domain is reflected in a high dN/dS ratio - 0.91 for the
319 POU1F1 alignment including representatives from all main mammalian groups
320 (Subgroup 1; Table 1). This is close to the ratio expected for neutral evolution
321 (1.0), suggesting that this domain is not subject to functional constraints, and
322 may have no specific function. High dN/dS values (0.52-1.31) for the β-domain
323 were obtained for each of the main mammalian groups examined separately, with
324 no value significantly different from 1.0. Lack of specific function for the β-
325 domain is also supported by the observation that in a number of species stop
326 codons or indels in the β-domain would prevent production of a functional
327 protein product, although the corresponding splice variant is produced (in
328 marmoset anyway) at a level similar to that in other species. As reported
329 previously (Diamond and Gutierrez-Hartmann, 1996) the 5'/N-terminal half of
330 the β-domain does seem to more be strongly conserved, but this reflects
331 conservation at the DNA/RNA level, both synonymous and nonsynonymous

substitutions, so the high dN/dS ratio is maintained.

Also of interest in the light of these results is a recent report of a human patient with combined pituitary hormone deficiency resulting from a *POU1F1* mutation in which the shorter ("normal") splice variant is missing, but the β-variant is retained, suggesting that the latter cannot substitute functionally for the former (Takagi et al., 2017)

Although the β-splice variant of POU1F1 clearly does have different biological properties from the shorter normal variant, the above observations suggest that its physiological significance is limited; in a few species a functional protein cannot be produced, and in others the very high variability of the β-domain suggests lack of specific function. It is also notable that in most species for which data is available, transcriptional databases indicate that the incidence of β-domain inclusion is low, comparable with retention of introns. Overall the results obtained here are consistent with the idea that for *POU1F1*, in most species, there is a single main transcript with variants produced by alternative splicing being of little biological significance. A similar situation may apply for many other genes where alternative splicing has been described (Tress et al., 2017).

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

Andersen, B., Rosenfeld, M.G., 2001. POU domain factors in the neuroendocrine system: lessons from developmental biology provide insights into human disease. Endocr. Rev. 22, 2-35.

Bamberger, A.-M., Bamberger, C.M., Pu, L.-P., Puy, L.A., Loh, Y.P., Asa, S.L., 1995. Expression of *pit-1* messenger ribonucleic acid and protein in the human placenta. J. Clin. Endocrinol. Metab. 80, 2021–2026.

Bastos, E., Ávila, S., Cravador, A., Renaville, R., Guedes-Pinto, H., Castrillo, J.L., 2006. Identification and characterization of four splicing variants of ovine *POU1F1* gene. Gene 382, 12-19.

Baumeister, H., Wegner, M., Richter, D., Meyerhof, W., 2000. Dual regulation of somatostatin receptor subtype 1 gene expression by Pit-1 in anterior pituitary GH3 cells. Mol. Endocrinol. 14, 255-271.

Ben-Batalla, I., Seoane, S., Garcia-Caballero, T., Gallego, R., Macia, M., Gonzalez, L.O., Vizoso, F., Perez-Fernandez, R., 2010. Deregulation of the Pit-1 transcription factor in human breast cancer cells promotes tumor growth and metastasis. J. Clin. Inv. 120, 4289-4302.

Chen, R., Ingraham, H.A., Treacy, M.N., Albert, V.R., Wilson, L., Rosenfeld, M.G., 1990. Autoregulation of *pit-1* gene expression mediated by two *cis*-active promoter elements. Nature 346. 583-586.

Cohen, L.E., Radovick, S., 2002. Molecular basis of combined pituitary hormone deficiences. Endocr. Rev. 23, 431-442.

Delhase, M., Vergani, P., Malur, A., Hooghe-Peters, E.L., Hooghe, R.J. 1993. The transcription factor Pit-1/GHF-1 is expressed in hemopoietic and lymphoid tissues. Eur. J. Immunol. 23, 951–955.

Delhase, M., Vila, V., Hooghe-Peters, E.L., Castrillo, J.L., 1995. A novel pituitary transcription factor is produced by alternative splicing of the human *GHF-1/PIT-1* gene. Gene 155, 273-275.

Diamond, S.E., Gutierrez-Hartmann, A., 1996. A 26-amino acid insertion domain defines a functional transcription switch motif in Pit-1β. J. Biol. Chem. 271, 28925-28932.

Diamond, S.E., Gutierrez-Hartmann, A., 2000. The Pit-1β domain dictates active repression and alteration of histone acetylation of the proximal prolactin promoter. J. Biol. Chem. 275, 30977-30986.

Ellestad, L.E., Porter, T.E., 2013. *Ras-dva* is a novel Pit-1- and glucocorticoid-

406 regulated gene in the embryonic anterior pituitary gland. Endocrinology 154,
407 308-319.
408
409 Featherstone, K., White, M.R.H., Davis, J.R.E., 2012. The prolactin gene: a
410 paradigm of tissue-specific gene regulation with complex temporal transcription
411 dynamics. J. Neuroendocrinol. 24, 977-990.
412
413 Fox, S.R., Jing, M.T.C., Casanova, J., Ye, Z.-S., Stanley, F., Samuels, H.H.,
414 1990. The homeodomain protein, Pit-1/GHF-1, is capable of binding to and
415 activating cell-specific elements of both the growth hormone and prolactin gene
416 promoters. Mol. Endocrinol. 4, 1069-1080.
417
418 Gil-Puig, C., Seoane, S., Blanco, M., Macia, M., Garcia-Caballero, T., Segura,
419 C., Perez-Fernandez, R., 2005. Pit-1 is expressed in normal and tumorous
420 human breast and regulates GH secretion and cell proliferation. Eur. J.
421 Endocrinol. 153, 335–344.
422
423 Haugen, B.R., Wood, W.M., Gordon, D.F., Ridgway, E.C., 1993. A thyrotrope-
424 specific variant of Pit-1 transactivates the thyrotropin β Promoter. J. Biol. Chem.
425 268, 20818-20824.
426
427 Herman, J.-P., Julien, N., Guillen, S., Enjalbert, A., Pellegrini, I., Franc, J.-L.,
428 2012. Research resource: A genome-wide study identifies potential new target
429 genes for POU1F1. Mol. Endocrinol. 26, 1455-1463.
430

431 Jonsen, M.D., Duval, D.L., Gutierrez-Hartmann, A. 2009. The 26-amino acid β-
432 motif of the Pit-1β transcription factor is a dominant and independent repressor
433 domain. Mol. Endocrinol. 23, 1371-1384.
434
435 Kelberman, D., Rizzoti, K., Lovell-Badge, R., Robinson, I.C.A.F., Dattani, M.T.,
436 2009. Genetic regulation of pituitary gland development in human and mouse.
437 Endocr. Rev. 30, 790-829.
438
439 Li, S., Crenshaw, E.B., Rawson, E.J., Simmons, D.M., Swanson, L.W.,
440 Rosenfeld, M.G., 1990. Dwarf locus mutants lacking three pituitary cell types
441 result from mutations in the POU-domain gene *pit-1*. Nature 347, 528-533.
442
443 Li, Y., Wallis, M., Zhang, Y.-P., 2005. Episodic evolution of prolactin receptor
444 gene in mammals: coevolution with its ligand. J. Mol. Endocrinol. 35, 411-419.
445
446 Maddison, W.P., Maddison, D.R., 2016. Mesquite: a modular system for
447 evolutionary analysis. Version 3.10. http://mesquiteproject.org.
448
449 Majumdar, S., Irwin, D.M., Elsholtz, H.P., 1996, Selective constraints on the
450 activation domain of transcription factor Pit-1. Proc. Natl. Acad. Sci. U.S.A. 93,
451 10256-10261.
452
453 Morris, A.E., Kloss, B., McChesney, R.E., Bancroft, C., Chasin, A., 1992. An
454 alternatively spliced Pit-1 isoform altered in its ability to trans-activate. Nucl.
455 Acids Res. 20, 1355-1361.

Nowakowski, B.E., Maurer, R.A., 1994. Multiple Pit-1-binding sites facilitate estrogen responsiveness of the prolactin gene. Mol. Endocrinol. 8, 1742-1749.

Radovick, S., Nations, M., Du, Y., Berg, L.A., Weintraub, B.D., Wondisford, F.E., 1992. A mutation in the POU-homeodomain of Pit-1 responsible for combined pituitary hormone deficiency. Science 257, 1115-1118.

Sánchez-Pacheco, A., Peña, P., Palomino, T., Güell, A., Castrillo, J.L., Aranda, A., 1998. The transcription factor GHF-1, but not the splice variant GHF-2, cooperates with thyroid hormone and retinoic acid receptors to stimulate rat growth hormone gene expression. FEBS Lett. 422, 103-107.

Scully, K.M., Jacobson, E.M., Jepsen, K., Lunyak, V., Viadiu, H., Carrière, C., Rose, D.W., Hooshmand, F., Aggarwal, A.K., Rosenfeld, M.G., 2000. Allosteric effects of Pit-1 DNA sites on long-term repression in cell type specification. Science 290, 1127-1131.

Sporici, R,A,, Hodskins, J.S., Locasto, D.M., Meszaros, L.B., Ferry, A.L., Weidner, A.M., Rinehart, C.A., Bailey, J.C., Mains, I.M., Diamond, S.E., 2005. Repression of the prolactin promoter: a functional consequence of the heterodimerization between Pit-1 and Pit-1β. J. Mol. Endocrinol. 35, 317-331.

Takagi, M., Kamasaki, H., Yagi, H., Fukuzawa, R., Narumi, S., Hasegawa, T., 2017. A novel heterozygous intronic mutation in *POU1F1* is associated with

481 combined pituitary hormone deficiency. Endocr. J. 64, 229-234.
482
483 Theill, L.E., Castrillo, J.-L., Wu, D., Karin, M., 1989. Dissection of functional
484 domains of the pituitary-specific transcription factor GHF-1. Nature 342, 945-
485 948.
486
487 Theill, L.E., Hattori, K., Lazzaro, D., Castrillo, J.-L., Karin, M., 1992.
488 Differential splicing of the GHF1 primary transcript gives rise to two
489 functionally distinct homeodomain proteins. EMBO J. 11, 2261-2269.
490
491 Tress, M.L., Abascal, F., Valencia, A., 2017. Alternative splicing may not be the
492 key to proteome complexity. Trends Biochem. Sci. 42, 98-110.
493
494 Wallis, M., 1996. The molecular evolution of vertebrate growth hormones: a
495 pattern of near-stasis interrupted by sustained bursts of rapid change. J. Mol.
496 Evo. 43, 93-100.
497
498 Wallis, M., 2008. Mammalian genome projects reveal new growth hormone
499 (GH) sequences. Characterization of the GH-encoding genes of armadillo
500 (*Dasypus novemcinctus*), hedgehog (*Erinaceus europaeus*), bat (*Myotis*
501 *lucifugus*), hyrax (*Procavia capensis*), shrew (*Sorex araneus*), ground squirrel
502 (*Spermophilus tridecemlineatus*), elephant (*Loxodonta africana*), cat (*Felis*
503 *catus*) and opossum (*Monodelphis domestica*). Gen. Comp. Endocrinol. 155,
504 271-279.
505

506 Wallis, O.C., Mac-Kwashie, A.O., Makri, G., Wallis, M., 2005. Molecular
507 evolution of prolactin in primates. J. Mol. Evo. 60, 606-614.
508
509 Yang, Z., 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol.
510 Biol. Evo. 24. 1586–1591.
511
512

513 **Legends for Figures**

514

515 **Fig. 1.** Overall structure of POU1F1. The domains of the protein are indicated
516 by alternating thick and thin lines. The β-domain is shown in grey. Numbers
517 above indicate amino acid residue numbers within the protein. Numbers below
518 indicate the distribution of the 6 exons of the POU1F1 gene; 5' utr and 3' utr
519 extensions of exons 1 and 6 are not included.

520

521 **Fig. 2.** Sequence alignment of selected β-domains. A) Nucleotide sequences. B)
522 Derived amino acid sequences. The full sequence for human β-domain is shown
523 on the top line; for sequences of other species . indicates identity to human.
524 Positions where indels lead to changes in reading frame are shaded light grey
525 (Cja, Smi, Mpe); for Mpe an insertion of GC occurs in the position indicated.
526 Locations of stop codons are indicated by dark grey shading. Two additional
527 marmoset species (*Callithrix kuhlii, C. geoffroyi*) had identical sequence to Cja.
528 Full species names and common names are given to the right of the amino acid
529 sequences.

530

531 **Fig. 3.** Phylogenetic trees for selected POU1F1 coding sequences (subgroup 1)
532 based on dS and dN values. The trees were derived using coding sequences
533 excluding β-domain**.** Numbers on selected branches are dN/dS ratios; note the
534 accelerated evolution on the branch leading to dog (Clu). The overall dN/dS ratio
535 for this sequence set was 0.055 (Table 1); that for the branch leading to dog was
536 0.147. Full species names are as in Figure 2B plus Can *Colobus angolensis*
537 (colobus monkey), Bac *Balaenoptera acutorostrata* (minke whale), Bbu

538 *Bubalus bubalis* (water buffalo), Uma *Ursus maritimus* (polar bear), Lwe
539 *Leptonychotes weddellii* (Weddell seal), Pva *Pteropus vampyrus* (large flying
540 fox), Mbr *Myotis brandtii* (Brandt's bat), Ete *Echinops telfairi* (Madagascar
541 hedgehog). Scale bars indicate substitutions/nucleotide site.
542
543
544

Table 1

Rates of evolution (dN/dS) for sequences encoding the domains of POU1F1.

Domain	Subgroup 1*	Primates**	Glires	Laurasiatheria	Afrotheria
full CDS including β domain	0.085	0.085	0.083	0.079	0.077
full CDS excluding β domain	0.055	0.048	0.055	0.051	0.038
TAD	0.084	0.078	0.139	0.081	0.032
β-domain	0.907	1.255	1.308	1.011	0.789
hinge region	0.146	0.111	0.106	0.113	0.211
POU specific	0.008	0.007	0.003	0.003	0.000
Linker region	0.052	0.052	0.057	0.078	0.124
Homeodomain	0.017	0.023	0.023	0.014	0.005

* including representatives of main mammalian groups (Group 1; see Fig. 3 for species included in this group)

** including tree shrew and flying lemur

572 **Table 2**
573 Alternative splicing at the *POU1F1* exon 1-exon 2 junction
574

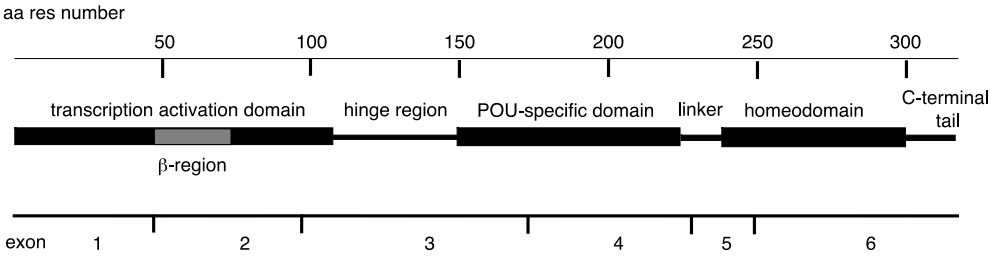
Species	Project	tissue	expts[a]	n[b]	% β-domain[c]	SEM[d]
Primates						
Homo sapiens	SRP035346	pituitary adenoma	9	3786	2.5	0.63
Pan troglodytes	SRP051959	pituitary	1	860	2.8	
Macaca spp[e]	SRP051959, SRP048677	pituitary	6	5744	3.9	0.51
Cercocebus atys	SRP051959	pituitary	1	345	6.5	
Papio anubis	SRP051959	pituitary	1	965	2.7	
Chlorocebus sabaeus	SRP033127	pituitary	5	4592	3.0	0.11
Callithrix jacchus	SRP051959	pituitary	1	375	1.9	
Rodentia						
Rattus norvegicus	SRP017586, SRP075804	pituitary cells GH4C1 cells	3	1272	12.2	1.45
Heterocephalus glaber	SRP061363	pituitary	1	173	1.2	
Cetartiodactyla						
Bos taurus	SRP070150	pituitary	10	1584	2.3	0.48
Bos taurus	SRP052656	pituitary	5	866	4.2	0.10
Ovis aries	ERP005642	pituitary	11	2501	4.3	0.67
Capra hircus	SRP069238	pituitary	4	146	3.9	2.22
Carnivora						
Canis lupus	SRP055477	pituitary	4	733	0.4	0.23
Ailuropoda melanocota	SRP063482	pituitary	1	247	1.2	

575
576 a) number of separate experiments for this species in this project
577 b) total hits including exon1-exon2 border
578 c) mean % β-domain for the number of experiments shown (calculated for each experiment as hits
579 including β-domain as percentage of all hits including exon1-exon2 border)
580 d) SEM calculated where number of experiments is 3 or more
581 e) data from 4 different *Macaca* species
582

583

584

585 Fig.1



586

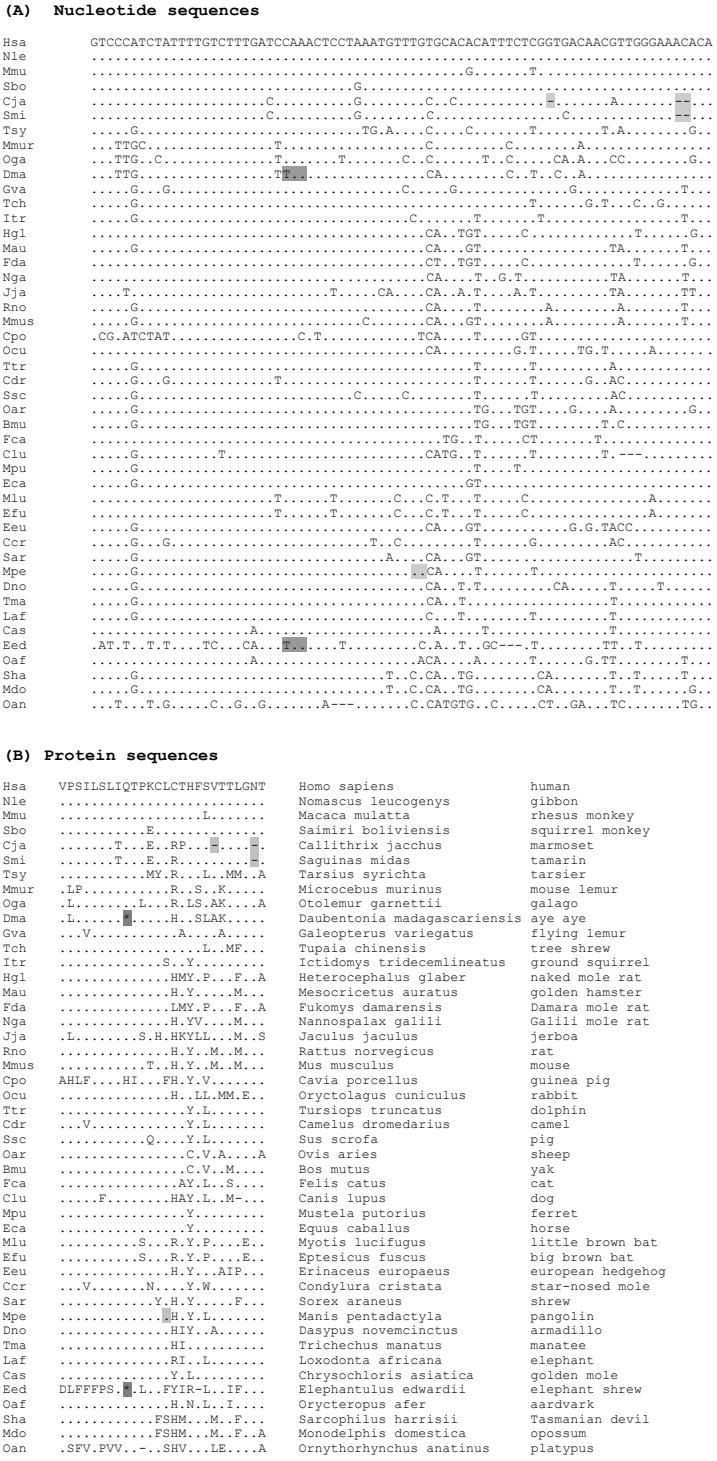
587

588

589

590

591 Fig.2



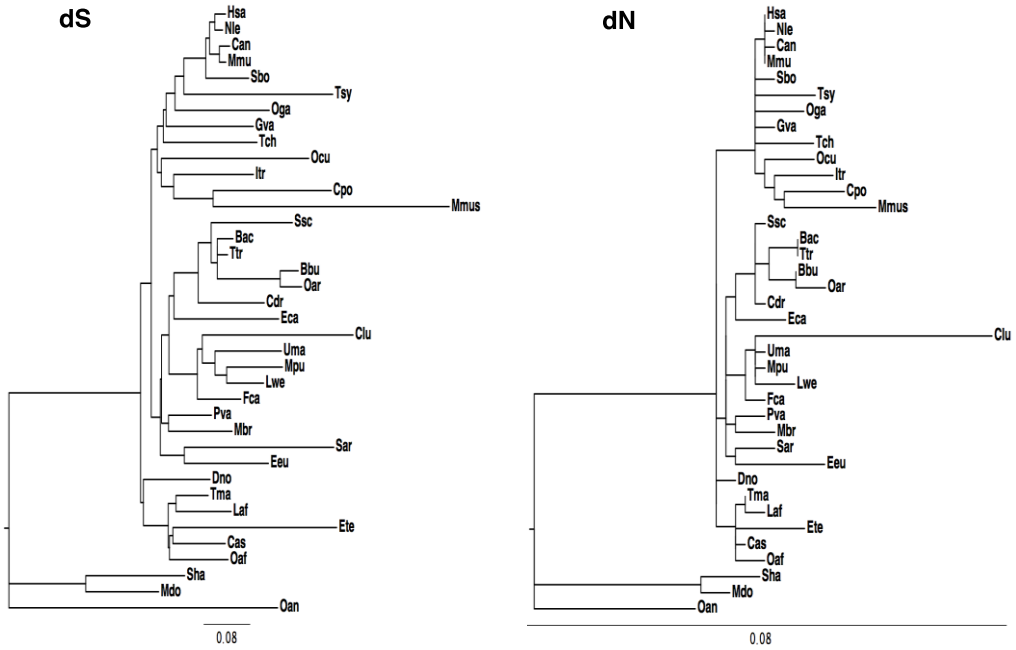
592

593

594

595 Fig.3

596



597