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
conserved, whereas the regions between them, postulated to comprise flexible
linkers, are more variable (Majumdar et al., 1996; Morris et al., 1992; Theill et
al., 1989). An additional region, the β-domain, can be inserted within the TAD as
a consequence of alternative splicing, two splice forms occurring in which the β-
domain is present or absent (Delhase et al., 1995; Morris et al., 1992; Theill et
al., 1992). The two splice variants in mammals have substantially different
biological properties which have been studied extensively (Diamond and
Gutierrez-Hartmann, 1996, 2000; Jonsen et al., 2009; Sánchez-Pacheco et al.,
1998; Sporici et al., 2005), but their physiological roles are not well defined.
Additional splice variants of POU1F1 have been described in sheep (Bastos et
al., 2006), but it is unclear whether these play a specific biological role.

The view that the domains in POU1F1 are strongly conserved is based on a
relatively small number of mammalian and non-mammalian species. The
availability of genomic data from over 100 mammalian species, including most
of the extant taxonomic orders, makes possible a much fuller study of POU1F1
variation in mammals and its evolutionary significance. The availability of
transcriptomic data for a number of species allows evaluation of POU1F1 splice
variation across mammals. Such a study is reported here. Questions addressed
include 1) are the POU1F1 domain sequences strongly conserved across all
mammals? 2) is there evidence for variable rates of evolution as seen for the
target genes of POU1F1, GH and prolactin (Li et al., 2005; Wallis, 1996, 2008;
Wallis et al., 2000)? 3) To what extent are splice variations in the POU1F1 gene
conserved across mammals, especially with regard to the form containing the β-
domain (the β-variant)?
2. Methods

2.1. Sequences

cDNA sequences for POU1F1 from various mammals were obtained by searching the publicly 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
linker and hinge regions and the TAD are rather more variable. The β-domain is very variable, particularly at the C-terminal end (Fig. 2). The sequence of dog POU1F1 shows rather high variation, especially in the TAD and hinge region.

3.2. Rates of Evolution

3.2.1. Complete POU1F1

Analysis of the POU1F1 CDS alignment for subgroup 1 (including β-domain) by the codeml method gave a dN/dS ratio of 0.085, showing that the protein overall is fairly strongly conserved (Table 1). Similar results were obtained for the other subgroups. However, as noted above, some domains appear to be more strongly conserved than others, so this value is an average; individual domains/regions are considered separately below. Codeml analysis also indicated that there was significant variation in dN/dS between species; this was largely due to an increased rate of evolution on the lineage leading to dog, for which branch dN/dS was significantly elevated (0.18; P<0.05, likelihood ratio test).

3.2.2. POU-specific domain and homeodomain

Analysis of the POU-specific domain and homeodomain, separately, using codeml gave very low values for dN/dS (Table 1), confirming the strong conservation deduced from visual inspection and previous reports. There was no evidence for rate variation between species, including dog.

3.2.3. TAD

Analysis of the TAD (excluding β-domain) by codeml gave values for dN/dS (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
13

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
variant lacking exons 3-5 was not detected. The splice variant lacking exon 3 was detected at a similar level in cow, but at a much lower level (<1%) in rat, dog and human.

3.4. Conclusions

Previous work on POU1F1 concluded that the main domains identified within the protein are strongly conserved, while regions between these (hinge and linker regions) are more variable. The present survey of POU1F1 sequences derived for a large number of mammalian species generally confirms this, except for the β-domain, which is very variable.

The variability of the β-domain is reflected in a high dN/dS ratio - 0.91 for the POU1F1 alignment including representatives from all main mammalian groups (Subgroup 1; Table 1). This is close to the ratio expected for neutral evolution (1.0), suggesting that this domain is not subject to functional constraints, and may have no specific function. High dN/dS values (0.52-1.31) for the β-domain were obtained for each of the main mammalian groups examined separately, with no value significantly different from 1.0. Lack of specific function for the β-domain is also supported by the observation that in a number of species stop codons or indels in the β-domain would prevent production of a functional protein product, although the corresponding splice variant is produced (in marmoset anyway) at a level similar to that in other species. As reported previously (Diamond and Gutierrez-Hartmann, 1996) the 5'/N-terminal half of the β-domain does seem to more be strongly conserved, but this reflects 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


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


The 26-amino acid β431 motif of the Pit-1β transcription factor is a dominant and independent repressor domain. Mol. Endocrinol. 23, 1371-1384.


Takagi, M., Kamasaki, H., Yagi, H., Fukuzawa, R., Narumi, S., Hasegawa, T., 2017. A novel heterozygous intronic mutation in POU1F1 is associated with
combined pituitary hormone deficiency. Endocr. J. 64, 229-234.


Wallis, M., 2008. Mammalian genome projects reveal new growth hormone (GH) sequences. Characterization of the GH-encoding genes of armadillo (Dasypus novemcinctus), hedgehog (Erinaceus europaeus), bat (Myotis lucifugus), hyrax (Procavia capensis), shrew (Sorex araneus), ground squirrel (Spermophilus tridecemlineatus), elephant (Loxodonta africana), cat (Felis catus) and opossum (Monodelphis domestica). Gen. Comp. Endocrinol. 155, 271-279.

Legends for Figures

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

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

Fig. 3. Phylogenetic trees for selected POU1F1 coding sequences (subgroup 1) based on dS and dN values. The trees were derived using coding sequences excluding β-domain. Numbers on selected branches are dN/dS ratios; note the accelerated evolution on the branch leading to dog (Clu). The overall dN/dS ratio for this sequence set was 0.055 (Table 1); that for the branch leading to dog was 0.147. Full species names are as in Figure 2B plus Can Colobus angolensis (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 brandii* (Brandt's bat), *Ete Echinops telfairi* (Madagascar

541  hedgehog). Scale bars indicate substitutions/nucleotide site.
Table 1
Rates of evolution (dN/dS) for sequences encoding the domains of POU1F1.

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

* including representatives of main mammalian groups (Group 1; see Fig. 3 for species included in this group)
** including tree shrew and flying lemur
Table 2  
Alternative splicing at the *POU1F1* exon 1-exon 2 junction

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

<sup>a</sup> number of separate experiments for this species in this project

<sup>b</sup> total hits including exon1-exon2 border

<sup>c</sup> mean % β-domain for the number of experiments shown (calculated for each experiment as hits including β-domain as percentage of all hits including exon1-exon2 border)

<sup>d</sup> SEM calculated where number of experiments is 3 or more

<sup>e</sup> data from 4 different *Macaca* species
Fig. 1

[Diagram showing domains and exons in a protein sequence]
### (A) Nucleotide sequences

<table>
<thead>
<tr>
<th>Species</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homo sapiens</td>
<td>GTCCCATCTATTTTGTCTTTGATCCAAACTCCTAAATGTTTGTGCACACATTTCTCGGTGACAACGTTGGGAAACACA</td>
</tr>
<tr>
<td>Monodelphis domestica</td>
<td>opossum</td>
</tr>
<tr>
<td>Ornythorhynchus anatinus</td>
<td>platypus</td>
</tr>
<tr>
<td>Orycteropus afer</td>
<td>bat</td>
</tr>
<tr>
<td>Elephantulus edwardii</td>
<td>shrew</td>
</tr>
<tr>
<td>Chrysochloris asiatica</td>
<td>golden hamster</td>
</tr>
<tr>
<td>Trichechus manatus</td>
<td>manatee</td>
</tr>
<tr>
<td>Aardvark</td>
<td>aardvark</td>
</tr>
<tr>
<td>Tasmanian devil</td>
<td>tasmanian devil</td>
</tr>
<tr>
<td>Platypus</td>
<td>platypus</td>
</tr>
<tr>
<td>Mouse</td>
<td>mouse</td>
</tr>
<tr>
<td>Standing Polychord</td>
<td>standing polychord</td>
</tr>
</tbody>
</table>

### (B) Protein sequences

<table>
<thead>
<tr>
<th>Species</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homo sapiens</td>
<td>Roso sapiens</td>
</tr>
<tr>
<td>Monodelphis domestica</td>
<td>opossum</td>
</tr>
<tr>
<td>Orycteropus afer</td>
<td>bat</td>
</tr>
<tr>
<td>Elephantulus edwardii</td>
<td>shrew</td>
</tr>
<tr>
<td>Chrysochloris asiatica</td>
<td>golden hamster</td>
</tr>
<tr>
<td>Trichechus manatus</td>
<td>manatee</td>
</tr>
<tr>
<td>Aardvark</td>
<td>aardvark</td>
</tr>
<tr>
<td>Tasmanian devil</td>
<td>tasmanian devil</td>
</tr>
<tr>
<td>Platypus</td>
<td>platypus</td>
</tr>
<tr>
<td>Mouse</td>
<td>mouse</td>
</tr>
<tr>
<td>Standing Polychord</td>
<td>standing polychord</td>
</tr>
</tbody>
</table>
Fig. 3