General and Comparative Endocrinology 260 (2018) 100–106

Contents lists available at ScienceDirect

General and Comparative Endocrinology

journal homepage: www.elsevier.com/locate/ygcen

Research paper

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

ARTICLE INFO

Article history:
Received 18 November 2017
Revised 5 January 2018
Accepted 9 January 2018
Available online 12 January 2018

Keywords:
Pit-1
POU1F1
β-domain
Molecular evolution
Evolutionary rates

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-Puig 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.

E-mail address: m.wallis@sussex.ac.uk

https://doi.org/10.1016/j.ygcen.2018.01.005
0016-6480/

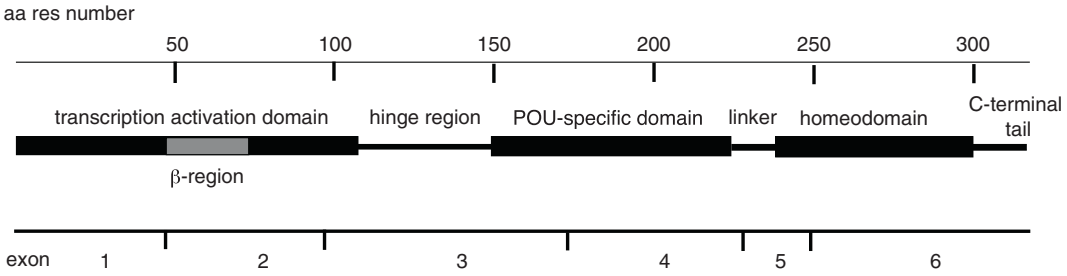

Fig. 1. Overall structure of POU1F1. The domains of the protein are indicated by alternating thick and thin lines. The β-domain is shown in grey. Numbers above indicate amino acid residue numbers within the protein. Numbers below indicate the distribution of the 6 exons of the POU1F1 gene; 5′ utr and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

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., 2005)? 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 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 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,