Characterisation of the GH gene cluster in a new-world monkey, the marmoset (Callithrix jacchus)

O C Wallis and M Wallis
Biochemistry Laboratory, School of Biological Sciences, University of Sussex, Brighton BN1 9QG, UK

Abstract

In most mammals pituitary GH is encoded by a single gene with no close relatives. However, in man the GH gene has been shown to be one of a cluster of five closely related genes, four of which are expressed in the placenta. Rhesus monkey also expresses at least five closely related GH-like genes, although the genomic organisation of these has not been fully reported. Here we describe the cloning and characterisation of GH-like genes in a new-world monkey, the marmoset (Callithrix jacchus). This species possesses a cluster of eight GH-like ‘genes’. The gene at the 5′ end of this cluster encodes pituitary GH and is similar to that encoding human GH. Five of the eight marmoset ‘genes’ are probably pseudogenes, since they include mutations which would prevent normal expression, including stop codons and small insertions/deletions that would change the reading frame. In one case a large part of a gene is deleted, and in another a large insertion is introduced into an exon. The remaining two marmoset genes are potentially expressible, as proteins with sequences substantially different (at 25–30% of all residues) from that of marmoset GH itself; whether and in which tissue(s) such expression actually occurs is not yet known. None of the marmoset genes is clearly equivalent to any of the human GH-like genes expressed in the placenta, and this and phylogenetic analysis suggest that the duplications that gave rise to the marmoset GH gene cluster occurred independently of those that gave rise to the corresponding cluster in man. Although it includes more ‘genes’, the marmoset cluster extends over a shorter region of chromosomal DNA (about 35 kb) than does the human GH gene cluster (about 50 kb).

Introduction

Pituitary growth hormone (GH) is expressed in the pituitary gland of all mammals as a single-chain protein of about 190 amino acid residues. The gene for GH comprises five exons and four introns. In most mammals there is just a single gene for GH, but in some artiodactyls there are two GH-like genes (Valinsky et al. 1990, Wallis et al. 1998), and in man there is a cluster of five GH-like genes (Chen et al. 1989).

The human GH gene cluster extends over 50–60 kb of DNA and has been studied in detail. The gene for pituitary GH lies at the 5′ end of the cluster (Chen et al. 1989). The remaining four genes are expressed in the placenta, and include two genes (hCS-A and hCS-B) coding for placental lactogen (choriomammosomatotrophin), one (hCS-L) for an alternatively spliced placental lactogen variant expressed at very low levels, and one (hGH-V) for placental GH variant, which is thought to take over the function of pituitary GH during pregnancy (Alsat et al. 1997). The cluster clearly arose by tandem duplications of the GH gene. The sequence of human GH differs markedly from that of non-primate GHS, indicating that a burst of rapid evolution occurred during the evolution of primate GHS (Wallis 1994, 1996). The sequence of each of the placental GH-like genes is much more similar to that of human GH than the latter is to any non-primate GH sequence, suggesting that the duplications that gave rise to the human GH gene cluster occurred independently of those that gave rise to the corresponding cluster in man. Although it includes more ‘genes’, the marmoset cluster extends over a shorter region of chromosomal DNA (about 35 kb) than does the human GH gene cluster (about 50 kb).
between the two placental lactogen genes, and suggesting that sequence similarities may not always give a clear guide to the events involved in evolution of this cluster.

The organisation of GH genes in primates other than man is less well defined. The rhesus monkey *(Macaca mulatta)* has been shown to express four GH-like genes in the placenta in addition to pituitary GH (Golos *et al.* 1993), but how the genes for these are organised at the genomic level is not clear; the four genes expressed in the placenta may not correspond exactly to those seen in the human. The rhesus GH sequence is very similar to that of man. In at least some lower primates, including the slow loris (Wallis *et al.* 2001) and bushbaby (Adkins *et al.* 2001), there appears to be only one gene encoding a GH-like protein, as in non-primates, and the sequence of this GH is very similar to that of pig and other non-primate GHs, indicating that the burst of rapid change that occurred during the evolution of primate GH occurred after the divergence of lines leading to prosimians and simians. In another lower primate, the tarsier, there may be several GH-like genes (Liu *et al.* 2001), but these have not yet been fully characterised. In new-world monkeys the sequence of GH is similar to that of human GH (Liu *et al.* 2001, Wallis *et al.* 2001), establishing that the episode of rapid change occurred before divergence of lines leading to new-world and old-world monkeys. Our previous studies (Wallis *et al.* 2001) indicated that new-world monkeys possess a cluster of GH-like genes, but the detailed nature of this cluster was not elucidated.

Here we describe studies that have established that a new-world monkey, the marmoset (*Callithrix jacchus*), possesses a cluster of eight GH-like genes. One of these, at the 5′ end of the cluster, encodes pituitary GH. The function of the remainder is not yet clear, but it is likely that several are functionless pseudogenes.

**Materials and methods**

**Amplification of marmoset GH-like genes**

Genomic DNA from male marmoset (*Callithrix jacchus*) was a kind gift from Dr David Hunt (Institute of Ophthalmology, University of London). Oligonucleotide primers for PCR were obtained from MWG-Biotech (Milton Keynes, Bucks, UK) and Sigma-Genosys (Pampisford, Cambs, UK). Regions of the marmoset GH-like gene cluster were amplified using PCR (Saiki 1990) and the primer pairs listed in Table 1. PCR reactions were carried out using *Pfu Turbo* DNA polymerase or Herculase Enhanced DNA polymerase (Stratagene, La Jolla, CA, USA). Reactions contained 50 pmol of each primer, 100–200 ng genomic DNA, 10 µl 10 × *Pfu Turbo* reaction buffer or 10 × *Herculase* reaction buffer (as provided by Stratagene for the enzyme used), 0·2 mM of each dNTP and 2·5 units *Pfu Turbo* DNA polymerase or 5 units Herculase Enhanced DNA polymerase, adjusted to a final volume of 100 µl with H₂O. PCR was carried out for 31 cycles using a denaturation temperature of 94 °C, an annealing temperature of 60–65 °C and an extension temperature of 72 °C. The times for the stages of the cycles were based on those recommended by Stratagene for the enzyme used except that the extension time was 1·5 min/kb of DNA to be amplified. The reaction was stopped after a final extension time of 10 min at 72 °C. Sizes of PCR products were analysed by subjecting samples to 1% agarose gel electrophoresis followed by ethidium bromide staining.

Amplification of the marmoset GH gene was achieved using PCR and primer set 1 (Table 1) as described previously (Wallis *et al.* 2001). This primer set also amplified several other genes. Other primers (sets 2–8) were used to amplify other GH-like genes within the GH cluster, and regions between genes were amplified using primers designed to run from the 3′ end of one gene to the 5′ end of another (sets 9–15). Primers for amplification of regions at the 5′ (sets 16 and 17) and 3′ (sets 18–20) ends of the gene cluster were based on the (by that stage) known sequence from the marmoset cluster, and published sequence of the human GH gene cluster (Chen *et al.* 1989).

**Cloning and sequencing of the marmoset GH-like genes**

PCR products were cloned into the phagemid PCR-Script Amp(+) vector, according to the instructions supplied with the pCR-Script Amp SK(+) cloning kit (Stratagene), and transformed into ultracompetent *E. coli* cells supplied with the kit. Initially double-stranded DNA carrying PCR products was subjected to sequencing, using an ABI 343A automatic sequencer and dye-primer (Amersham, Little Chalfont, Bucks, UK).
and dye-terminator (Perkin Elmer, Warrington, Cheshire, UK) kits. Later sequencing was carried out by the sequencing services provided by Genescreen Ltd (New Milton, Hants, UK), SeqLab (Gottingen, Germany) or MWG-Biotech (Ebersberg, Germany).

Sequence analysis

Sequences were aligned using the CLUSTALw program (Higgins & Sharp 1988). Sequence alignments can be found at the following website: http://www.biols.sussex.ac.uk/Home/Mike_Wallis/GHAlign/. Assembly of sequences to give the overall ordering of the marmoset GH gene cluster was carried out by identifying overlapping sequences (Fig. 1). Phylogenetic analysis was carried out using the program MacClade 3 (Maddison & Maddison 1992) and parsimony, likelihood and distance (neighbour-joining; Saitou & Nei (1987)) methods in the PAUP 4 package (Swofford 1998).

Table 1 Primer pairs used for PCR of marmoset GH gene cluster. Primer pairs 1–8 were used to clone ‘genes’, primer pairs 9–15 were used to clone intergene regions, primer pairs 16–20 were used to clone regions at the extreme 5’ and 3’ of the cluster (see Fig. 1). Note that in several cases the same oligonucleotide appears two or more times in the table, paired with a different partner oligonucleotide.

<table>
<thead>
<tr>
<th>Primer pair</th>
<th>5’ primer</th>
<th>3’ primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>TGGCTATTCCTGACATCTCCTTCCGCC</td>
<td>CCACCCCATATAATTAGAGAAGGACAC</td>
</tr>
<tr>
<td>2</td>
<td>TGGCTATTCCTGACATCTCCTTCCGCC</td>
<td>CCATCCCATATAATTAGAGAAGGTCAC</td>
</tr>
<tr>
<td>3</td>
<td>TGGCTATTCCTGACATCTCCTTCCGCC</td>
<td>AAGAAATGCCTTCAGTCATCTCTGG</td>
</tr>
<tr>
<td>4</td>
<td>TGGCTATTCCTGACATCTCCTTCCCTC</td>
<td>CACACTCCATAATTATTTAGAGAAGGACAC</td>
</tr>
<tr>
<td>5</td>
<td>TGGCTATTCCTGACATCTCCTTCCCTC</td>
<td>CACCCCCATATAATTACAGAAGGACAC</td>
</tr>
<tr>
<td>6</td>
<td>TGGCTATTCCTGACATCTCCTTCCCTC</td>
<td>CATCCCATATAATTAGAGAAGGACAC</td>
</tr>
<tr>
<td>7</td>
<td>TCTCGTGTCCCAGGAGAGAAGATA</td>
<td>CACCCCCATATAATTAGAGAAGGACAC</td>
</tr>
<tr>
<td>8</td>
<td>TGGCTATTCCTGACATCCTTCCCTC</td>
<td>GGATGTCGTTGTTGAGACGCCTACG</td>
</tr>
<tr>
<td>9</td>
<td>CAGCTTCGAGGGGCTGTGCCCT</td>
<td>AAGAAATGCCTTCAGTCATCTCTGG</td>
</tr>
<tr>
<td>10</td>
<td>GGTCGAGACATCTCCTTGAGATGGTC</td>
<td>CTCGACGCAATTCGATAGTAGGAG</td>
</tr>
<tr>
<td>11</td>
<td>CAGCTTCGAGGGGCTGTGCCCT</td>
<td>CTCGACGCAATTCGATAGTAGGAG</td>
</tr>
<tr>
<td>12</td>
<td>GGAGCCTTTGACAGGGGCTGTGCC</td>
<td>CCAGCTTGGGAGGCGATAGGCC</td>
</tr>
<tr>
<td>13</td>
<td>AGAAGAAAGGAACTCGAGAAAATCC</td>
<td>GCCAGGAGAGGAGAGAGAGAGAG</td>
</tr>
<tr>
<td>14</td>
<td>ACAGCGCTTCGAGGACAGACACTG</td>
<td>ACCACTGACGTTGCTGTACTATGG</td>
</tr>
<tr>
<td>15</td>
<td>GGCTATTCCTGACATCTCCTTCCGCC</td>
<td>CTGAGCCATTCGCTAGTAGTAGGAG</td>
</tr>
<tr>
<td>16</td>
<td>GCCCAAGGGAGTGAGTGAGGATGTC</td>
<td>GGAGCAGACATGTCAGAAAATCC</td>
</tr>
<tr>
<td>17</td>
<td>GCCCAAGGGAGTGAGTGAGGATGTC</td>
<td>CAATGTCACATTCCGCCACCTG</td>
</tr>
<tr>
<td>18</td>
<td>TGGCTATTCCTGACATCTCCTTCCGCC</td>
<td>CAATGTCACATTCCGCCACCTG</td>
</tr>
<tr>
<td>19</td>
<td>TGGCTATTCCTGACATCTCCTTCCGCC</td>
<td>GGAGCAGACATGTCAGAAAATCC</td>
</tr>
<tr>
<td>20</td>
<td>TGGCTATTCCTGACATCTCCTTCCGCC</td>
<td>AAGAAATGCCTTCAGGACCTACG</td>
</tr>
</tbody>
</table>

Results and discussion

The organisation of GH-like genes in the marmoset

Eight unique GH-like gene sequences were identified using the PCR approach and sequence analysis as described. The sequences of all of these ‘genes’ have been determined in full, using at least two independently derived clones in each case. These eight gene sequences were assembled into a unique cluster by cloning inter-gene regions and matching sequences to give overlaps (Fig. 1). Sequences at the 3’ and 5’ ends of each intergene region were determined, again based on at least two independently derived clones in each case. Intergenic lengths (Fig. 1) were based on sizes of PCR products. Sequences of the eight marmoset GH-like genes are available from the EMBL/Genbank/DDBJ database (Accession numbers AJ297563; AJ489807-AJ489813).

The gene at the 5’ end of the cluster corresponds to the gene for marmoset pituitary GH, as
described previously (Wallis et al. 2001). The sequences of the seven ‘genes’ downstream of this all differ considerably from the marmoset or human GH genes. Five contain mutations which would prevent ‘normal’ gene expression; these are probably pseudogenes and have been designated as such (ψgene) in Fig. 1. Thus, ψgene 2 (Fig. 1) contains an insertion of two bases in exon 3 which would disrupt the reading frame and introduce a new stop codon downstream. ψGene 3 includes a stop codon in the signal peptide which would be expected to prevent normal translation of this gene and a large insertion in exon 3 corresponding to the reverse sequence of parts of intron 4 and exon 5. ψGene 4 contains a very large deletion removing exons 2, 3 and 4 and associated introns; expression of a GH-like protein would be impossible. The signal peptide of ψgene 7 contains a deletion of two codons, and a substitution in the last codon of the signal peptide (Ala to Asp) which may prevent its normal processing; furthermore, two deletions of single bases in exon 2 would lead to frame shifts disrupting the sequence of most of the mature protein, eventually leading to premature termination. ψGene 8 has a single base deletion in exon 4 which would again lead to a reading frame shift and disruption of the final third of the protein sequence.

Sequences at the 5’ and 3’ ends of the gene cluster were amplified and cloned using 5’ and 3’ primers.
(respectively) primers based on unique sequence at the ends of the human GH gene cluster (i.e. sequences outside the duplicated parts of that cluster). These would appear to establish the limits of the marmoset cluster. The presence of other GH-like genes outside the cluster in marmoset cannot be excluded, but failure to find any amplified genes that cannot be reconciled with the organisation shown in Fig. 1 argues against such additional genes.

**Protein sequences**

Figure 2 shows an alignment of the derived amino acid sequences of marmoset GH-like genes 1, 5 and 6, and of sequences of various other GHs and related proteins. The sequence of human GH is shown in full. For other sequences, – indicates a residue identical to that in human GH, ● indicates a gap. Abbreviations: hGH, human GH; rh-mGH, rhesus monkey GH; marmGH, marmoset GH (gene 1); hGH-v, human placental GH; hCS-A and hCS-B, human chorionic somatomammotrophins (placental lactogens) A and B; hCS-L, hCS-like protein; rh-mGHv, rhesus monkey GH variant (placental); rh-mCS1, 2 and 3, rhesus monkey chorionic somatomammotrophins (placental lactogens) 1, 2 and 3; marmGene5 and 6, marmoset GH-like proteins 5 and 6 (see Fig. 1); s1GH, slow lorises GH; pigGH, pig GH.

**Figure 2** Alignment of amino acid sequences derived from sequences of marmoset GH-like genes 1, 5 and 6, and of sequences of various other GHs and related proteins. The sequence of human GH is shown in full. For other sequences, – indicates a residue identical to that in human GH, ● indicates a gap. Abbreviations: hGH, human GH; rh-mGH, rhesus monkey GH; marmGH, marmoset GH (gene 1); hGH-v, human placental GH; hCS-A and hCS-B, human chorionic somatomammotrophins (placental lactogens) A and B; hCS-L, hCS-like protein; rh-mGHv, rhesus monkey GH variant (placental); rh-mCS1, 2 and 3, rhesus monkey chorionic somatomammotrophins (placental lactogens) 1, 2 and 3; marmGene5 and 6, marmoset GH-like proteins 5 and 6 (see Fig. 1); s1GH, slow lorises GH; pigGH, pig GH.
of human placental GH-like genes. Whether these marmoset genes are actually expressed has not yet been established. The potential significance of the amino acid differences that they show compared with human or marmoset GH can be evaluated in the light of the model of the 3-dimensional structure of the human GH-receptor complex described by De Vos et al. (1992). Overall when sequences of proteins derived from gene 5 or 6 are compared with marmoset GH, 26–27% of all residues differ. Twenty-one residues can be identified that contribute to the hydrophobic core of the GH molecule, which are presumably of particular importance in maintaining the overall structure/fold. Of these only three (14%) differ when the potential product of either gene 5 or gene 6 is compared with marmoset GH, and these substitutions are conservative. On the other hand, 47 residues can be identified as being close to (within 5 Å of) residues in binding sites on the receptor; of these 16–18 (34–38%) differ between gene 5 and 6 products and marmoset GH. Among these is residue 172 (numbering as in Fig. 2; Asp in human, rhesus monkey and marmoset GHs, His in non-primate and prosimian GHs), which is thought to play a major part in determining the species specificity of higher primate GHs (Behncken et al. 1997). Interestingly residue 172 in marmoset GH-like genes 5 (Asn) and 6 (Ser) differs from that in both higher primate and non-primate GHs. Thus, core residues likely to contribute to the overall structure/fold of the protein are quite strongly conserved whereas those that are probably involved in determining receptor-binding affinity and specificity are very variable. These results suggest that the gene 5 and gene 6 products are probably functional proteins, but that their interactions with receptors are likely to differ markedly from those of GH itself.

Regulatory elements

Analysis of sequences upstream of the ATG start site of the human and other GH genes has revealed various regulatory sequences, and many of these have been identified upstream of the marmoset GH gene (Wallis et al. 2001), including a TATA box, distal and proximal Pit-1 elements (Theill & Karin 1993, Krawczak 1999), a cAMP response element (CRE) (Eberhardt et al. 1996) and a negative response element (NRE3) which probably represents a binding site for transcription factor YY1 (Park & Roe 1996). Analysis of the marmoset GH-like genes reveals that some of these regulatory sequences are strongly conserved in all cases, even in the pseudogenes, whereas others are not (Fig. 3). Thus the TATA box is retained in all cases except gene 2, where a deletion removes it completely. In several cases (including the two potentially expressible genes, 5 and 6) the proximal Pit-1 element is substantially altered, making it unlikely that its normal function is retained, but the distal Pit-1 site is strongly conserved as is the NRE. The CRE element is very variable.

Slater et al. (1985) identified a glucocorticoid regulatory element in intron 1 of the human GH gene. The sequence of this is not conserved in the marmoset GH gene, or any of the genes in the marmoset GH-like gene cluster. On the other hand, several of the putative negative thyroid hormone responsive elements (TREs) that have been identified in the 3’ untranslated region of the human GH gene (Zhang et al. 1992) appear to be conserved in the marmoset GH gene and to a lesser extent in the GH-like genes. Two TREs have also been identified in the promoter region of human GH and placental lactogen genes (Leidig et al. 1992). These both bind thyroid hormone receptor, but only induce increased transcription of the placental lactogen gene, not the GH gene. The proximal TRE occupies positions −128 to −104 (numbering as Fig. 3). Interestingly in this region the sequence of the eight marmoset GH-like genes is more similar to that of the human placental lactogen genes than the human GH gene, and two positions (−113 and −111) identified as crucial for TRE responsiveness (Leidig et al. 1992) are occupied by G and A respectively in all the marmoset GH-like genes, as in human placental lactogen genes but not the human GH gene. On the other hand, there is a deletion of four bases at the 5′ end of this putative TRE in the marmoset GH gene (but not genes 5 and 6), which would probably prevent its functioning. Thus it may well be that a positive TRE is found in marmoset GH-like genes 5 and 6, but not in the GH gene, a situation functionally similar to that seen for the human gene cluster.

Evolution of GH gene clusters

Phylogenetic analysis using the neighbour-joining method in PAUP (Swofford 1998) of the sequences
corresponding to coding regions or equivalent of the marmoset GH-like genes and pseudogenes and GH-like genes from other primates and pig gave the phylogenetic tree shown in Fig. 4. This tree suggests that the eight GH-related 'genes' in marmoset are more closely related to each other than to any of the GH-related genes in rhesus monkey, man, lower primates or pig (marmoset afii9852 gene 4 was excluded from the data set used for Fig. 4, because of its very large deletion, but separate analysis using just sequences equivalent to that retained in γgene 4 confirmed that this too is more closely related to the other marmoset genes than that of any other group). Similar results were obtained using alternative phylogenetic analysis methods (parsimony or maximum likelihood); the clustering of all of the marmoset GH-like gene sequences, to the exclusion of all others, was strongly supported (over 90% of bootstrap replicates) in each case. The implication of this is that the gene duplications that gave rise to the GH-like gene clusters in marmoset and man occurred independently, after divergence of the lines leading to new-world monkeys and old-world monkeys/apes. On the other hand clear similarity between at least some of the GH-like genes expressed in rhesus monkey and human placenta (Fig. 4, and Wallis (1996)) suggests that this is not the case for these two species – at least one of the gene duplications that gave rise to the placental lactogens...
(choriomammosomatotrophins) must have occurred before divergence of lines leading to man and old-world monkeys. The separate origins of the marmoset and human/rhesus monkey GH-like gene clusters is also in agreement with the lack of similarity at the protein level, as referred to above, and with the much closer spacing of genes within the marmoset cluster, with about 3.5 kb between genes in the former, but 7–14 kb between genes in the latter (Fig. 5).

For the most part, differences between the eight marmoset GH-like genes involve substitution of one nucleotide by another or insertion or deletion of a small number of nucleotides. There are two exceptions. Gene 4 includes a large deletion which removes a substantial part of the gene, including exons 2, 3 and 4, introns 2 and 3, and parts of introns 1 and 4. Nevertheless the remaining sequence up- and downstream of this deletion is very similar to that of other genes in the cluster. Gene 3 on the other hand includes an insertion of about 270 nucleotides in exon 3, corresponding to a duplication of parts of intron 4 and exon 5, with the sequence reversed. In neither of these cases is there evidence of a duplicated sequence flanking the insert or deletion, and the underlying mechanisms are not clear. Similarly, the reasons for the rapid, apparently independent expansion of GH-like gene clusters in lineages leading to the marmoset and man are not clear, although it is possible that it is associated with the period of rapid evolution of GH in primates. It has been suggested that this episode of rapid change was due to the process of ‘function switching’, with the hormone carrying out two functions of varying relative importance (Wallis 1997). Duplication(s) giving rise to two or more copies of the GH genes,
followed by divergent evolution, could then allow efficient, separate regulation of these two functions. The observation that the rate of evolution apparently falls after the gene duplication accords with such a mechanism.

Acknowledgements

We thank Dr David Hunt for the gift of marmoset genomic DNA.

References

Adkins RM, Nekrutenko A & Li W-H 2001 Bushbaby growth hormone is much more similar to nonprimate growth hormones than to rhesus monkey and human growth hormones. Molecular Biology and Evolution 18 35–60.


Valinsky A, Shani M & Gootwine E 1990 Restriction fragment length polymorphism in sheep at the growth hormone locus is the result of variation in gene number. Animal Biotechnology 1 133–144.


Wallis M 1997 Function switching as a basis for bursts of rapid change during the evolution of pituitary growth hormone. Journal of Molecular Evolution 44 348–350.


Received 25 January 2002
Accepted 29 April 2002