Molecular evolution of GH in primates: characterisation of the GH genes from slow loris and marmoset defines an episode of rapid evolutionary change

O C Wallis, Y-P Zhang[1] and M Wallis

Biochemistry Laboratory, School of Biological Sciences and Centre for the Study of Evolution,
 University of Sussex, Brighton BN1 9QG, UK

[1]Laboratory of Cellular and Molecular Evolution, Kunming Institute of Zoology,
 Chinese Academy of Sciences, Kunming, China

(Requests for offprints should be addressed to O C Wallis, Biochemistry Laboratory, School of Biological
 Sciences, University of Sussex, Brighton BN1 9QG, UK; Email: C.Wallis@sussex.ac.uk)

ABSTRACT

Pituitary growth hormone (GH), like several other protein hormones, shows an unusual episodic pattern of molecular evolution in which sustained bursts of rapid change are imposed on long periods of very slow evolution (near-stasis). A marked period of rapid change occurred in the evolution of GH in primates or a primate ancestor, and gave rise to the species specificity that is characteristic of human GH. We have defined more precisely the position of this burst by cloning and sequencing the GH genes for a prosimian, the slow loris (*Nycticebus pygmaeus*) and a New World monkey, marmoset (*Callithrix jacchus*). Slow loris GH is very similar in sequence to pig GH, demonstrating that the period of rapid change occurred during primate evolution, after the separation of lines leading to prosimians and higher primates. The putative marmoset GH is similar in sequence to human GH, demonstrating that the accelerated evolution occurred before divergence of New World monkeys and Old World monkeys/apes. The burst of change was confined largely to coding sequence for mature GH, and is not marked in other components of the gene sequence including signal peptide, 5′ upstream region and introns. A number of factors support the idea that this episode of rapid change was due to positive adaptive selection. Thus (1) there is no apparent loss of function of GH in man compared with non-primates, (2) after the episode of rapid change the rate of evolution fell towards the slow basal level that is seen for most mammalian GHs, (3) the accelerated rate of substitution for the exons of the GH gene significantly exceeds that for introns, and (4) the amino acids contributing to the hydrophobic core of GH are strongly conserved when higher primate and other GH sequences are compared, and for coding sequences other than that coding for hydrophobic core residues the rate of substitution for non-synonymous sites (K_A) is significantly greater than that for synonymous sites (K_S). In slow loris, as in most non-primate mammals, there is no evidence for duplication of the GH gene, but in marmoset, as in rhesus monkey and man, the putative GH gene is one of a cluster of closely related genes.

Journal of Molecular Endocrinology (2001) **26**, 249–258

INTRODUCTION

Pituitary growth hormone (GH) shows marked species specificity (Nicoll *et al*. 1986, Wallis 1989). In particular GH from non-primate mammals is not active in man – with the consequence that, until the advent of recombinant DNA-derived material, human hypopituitary dwarfism was treated with GH extracted from human pituitaries. The biological specificity is paralleled by marked differences in amino acid sequence. This does not reflect simply a high rate of evolution, but an unusual episodic pattern of molecular evolution in which prolonged periods during which sequence is strongly conserved (near-stasis; for example, pig and dog GHs are identical) are interrupted by occasional bursts of rapid change (Wallis 1981, 1994, 1996, Ohta 1993). For evolution of GH in mammals two such bursts of rapid change have been

identified, during the evolution of artiodactyls (prior to radiation of the ruminants) and early in primate evolution (or in an ancestor of primates). For GH evolution in mammals the periods of near-stasis predominate, occupying about 90% of the sampled evolutionary time, but most observed change in amino acid sequence (about 80%) occurred during the rapid bursts (Wallis 1994). Episodic evolution of this type appears to have occurred in a number of other protein hormones and cytokines (Wallis 2001), including prolactin (Wallis 1981, 2000, Curlewis *et al.* 1998), insulin (Blundell & Wood 1975, Beintema & Campagne 1987) and interleukin-2 (Zelus *et al.* 2000, Zhang & Nei 2000) and the mechanisms underlying it are therefore of general interest.

The burst of rapid change of GH in the lineage leading to higher primates was particularly marked, with substitutions at about 35% of all amino acid residues. Previous studies have established that this burst occurred before the separation of lineages leading to man and Old World monkeys (sequences of GHs of man and rhesus monkey are very similar) and after the divergence of most orders of placental mammals (Wallis 1994), but lack of information about GH in lower primates, or in mammalian orders most closely related to primates, has prevented its more precise location. The period of rapid change also occurred before the gene duplications that gave rise to a cluster of GH-like genes in higher primates, most of which are expressed in the placenta (since all of the proteins expressed by the human GH gene cluster are much more similar to human GH than they are to non-primate GHs) (Chen *et al.* 1989, Wallis 1996). There is no evidence for duplication of the GH gene in non-primate mammals, with the exception of a very recent duplication in caprine ruminants (Valinsky *et al.* 1990, Wallis *et al.* 1998). In order to define the episode of rapid change in primate GH evolution more precisely we have cloned and sequenced the GH genes from a prosimian, the slow loris (*Nycticebus pygmaeus*), and a New World monkey, the marmoset (*Callithrix jacchus*). This has enabled the burst of rapid change to be located on the primate lineage, after divergence of the slow loris and related prosimians from the line leading to higher primates, but before divergence of the New World monkeys and the Old World monkeys/apes.

MATERIALS AND METHODS

Polymerase chain reaction (PCR) and Southern blotting

Slow loris

Genomic DNA was prepared from liver of a female slow loris (*Nycticebus pygmaeus*) obtained in the Yunnan province of China. Oligonucleotide primers used for PCR were obtained from MWG-BIOTECH (Milton Keynes, UK). Three primer pairs were successfully used in amplifying the slow loris GH gene; their sequences were as follows. Primer pair 1: TGCTCTGCCTGCCCTGGACT CAG, sense primer (exon 2); CGAAGCGGC GACACTTCATGACC, antisense primer (exon 5). Primer pair 2: GAACAGGATGAGTGGGAG GAGGTTC, sense primer (5′ upstream); GTG CAGGTCCTTCTTGAAGCAAGAG, antisense primer (exon 5). Primer pair 3: CATGCCTTTG TCCAGCCTGTTTGCC, sense primer (exon 2); GACACCTAGTCAGACAAAATGATGCAAC, antisense primer (3′ untranslated region (utr)).

Portions of the slow loris GH gene were amplified by PCR (Saiki 1990) using these primer pairs (50 pmol each primer), 100 ng genomic DNA, 2·5 units *PfuTurbo* DNA polymerase (Stratagene, La Jolla, CA, USA), 0·2 mM of each dNTP and 10 µl of $10 \times Pfu$ reaction buffer (200 mM Tris–HCl, pH 8·8, 100 mM KCl, 100 mM $(NH_4)_2SO_4$, 20 mM $MgSO_4$, 1% Triton X-100 and 1 mg/ml nuclease-free bovine serum albumin) adjusted to a final reaction volume of 100 µl with H_2O. An initial long (5 min) denaturation step at 94 °C was followed by 3 min at 60 °C, 5 min at 72 °C and then 30 cycles of 1 min at 94 °C, 1 min at 60 °C and 5 min at 72 °C. The reaction was stopped after a final extension time of 10 min at 72 °C. The size and purity of the PCR products was estimated by subjecting samples to 1% agarose gel electrophoresis followed by ethidium bromide staining. The identity of the PCR products was confirmed by Southern blot analysis (Southern 1975), using an α^{32}P-labelled oGH cDNA probe (Warwick *et al.* 1989) or human GH cDNA probe (made from a plasmid kindly provided by Dr J Martial, University of Liege, Belgium; Martial *et al.* 1979).

Marmoset

Genomic DNA from male marmoset (*Callithrix jacchus* was a kind gift from Dr David Hunt) (Institute of Opthalmology, University of London). The marmoset GH gene was amplified by PCR using the following primers (primer pair 4): TGGCTATCCTGACATCCTTTCCCGC, sense primer (5′ upstream); CCACCCCATAATATTA GAGAAGGACAC, antisense primer (3′ downstream). These primers were based on the known human GH gene sequence, and on sequence information obtained in our laboratory for GH-like genes from the New World monkey *Cebus albifrons* (OC Wallis & M Wallis, unpublished observations). Use of these and related primers revealed the presence of a considerable number of GH-like

genes. Of these one (1–1) was identified as the putative marmoset GH gene on the basis of its resemblance to the human GH gene, and its presence (like the human gene) at the extreme 5′ end of the gene cluster. Procedures used to amplify this marmoset GH gene were the same as those used for the slow loris gene (see above).

Cloning and sequencing of the slow loris and marmoset GH genes

PCR products were cloned into the phagemid pCR-Script Amp SK(+) vector according to the instructions of the pCR-Script Amp SK(+) cloning kit manufacturers (Stratagene), and transformed into ultracompetent *E. coli* cells supplied with the kit. Double-stranded phagemid DNA carrying PCR products was subjected to sequencing, using an ABI 343A automatic sequencer and dye-primer (Amersham, Little Chalfont, Buckinghamshire, UK) and dye-terminator (Perkin Elmer, Warrington, UK) cycle sequencing reaction kits. The slow loris and marmoset GH gene sequences have been deposited in the EMBL/Genbank/DDBJ database (Accession Nos. AJ297562 and AJ297563 respectively).

Sequence analysis

Sequences were aligned using the CLUSTALW program (Higgins & Sharp 1988) with gap introduction and extension penalties of 10 and 5 respectively, followed by manual adjustment, with no increase in number of gaps. Sequence alignments can be found at the following website: http://www.biols.sussex.ac.uk/Home/Mike_Wallis/GHAlign/. For analysis of rates of GH evolution a phylogenetic tree for mammalian evolution was based on that of Kumar & Hedges (1998) which was derived from a large amount of molecular evidence. Trees based solely on GH sequences deviate markedly from conventional expectations for mammalian evolution, presumably because of the very variable rates of evolution (Felsenstein 1978). GH gene sequence alignments were used to determine branch lengths on this defined tree, using the parsimony-based programme MacClade 3 (Maddison & Maddison 1992) or distance-method programmes (dnadist or protdist and FITCH) in the PHYLIP package (Felsenstein 1993). The neighbor-joining method (Saitou & Nei 1987, Neighbor in the PHYLIP package) was also used, although its use was limited because of inability to accept defined trees. Non-synonymous (K_A) and synonymous (K_S) substitution rates in coding sequence were determined using the method of Nei & Gojobori (1986),

with correction for transition/transversion ratio as given by Zhang *et al*. (1998) and Zhang & Nei (2000); matrices of these were used as input for the FITCH programme. Significance of differences between substitution rates was assessed using Student's *t*-test as described by Graur & Li (1999).

RESULTS AND DISCUSSION

Amplification and cloning of GH genes from slow loris and marmoset

A fragment of the slow loris GH gene (~ 1 kb) was amplified using PCR and primer pair 1. This hybridised with both human and (more strongly) ovine cDNA probes and was cloned and shown to have a GH-like sequence. Clones covering the entire coding region of the slow loris GH gene and extending into 5′ upstream and 3′ downstream sequence were obtained using primer sets 2 and 3 where in each case one primer was based on sequence determined from the clone obtained using primer set 1, and the other was based on known sequence in the 5′ upstream region (pig GH gene; primer set 2) or 3′ downstream region (human GH gene; primer set 3). PCR products of appropriate size were obtained in each case, cloned and shown to possess GH gene-related sequence.

Marmoset GH-related genes were PCR amplified using primer pair 4. These produced a product of expected size (1·9 kb). When this was cloned it was shown to correspond to several different sequences, all similar to the human GH gene, indicating that as in human and rhesus monkey (Chen *et al*. 1989, Golos *et al*. 1993) a family of GH-like genes occurs in the marmoset. At least 12 GH-like genes have now been identified in this species, and amplification and cloning of intergene sequences indicates that these are organised into a gene cluster. One of the cloned genes (1–1) has been tentatively identified as marmoset pituitary GH, on the grounds that (1) like human GH it is at the 5′ end of the gene cluster, (2) the protein that it encodes is more similar to human GH than that encoded by any of the other genes identified and (3) it shows the 5′ regulatory sequences and the five exons that are characteristic of the human GH gene. This gene has been characterised in detail and is described here. For each of the other GH-like genes identified in marmoset, the encoded protein sequence is much more similar to the human GH sequence than those of non-primate or slow loris GHs.

The clones corresponding to the slow loris and marmoset GH genes were subjected to DNA sequencing. Both DNA strands were sequenced, all ambiguities were resolved, and sequencing of
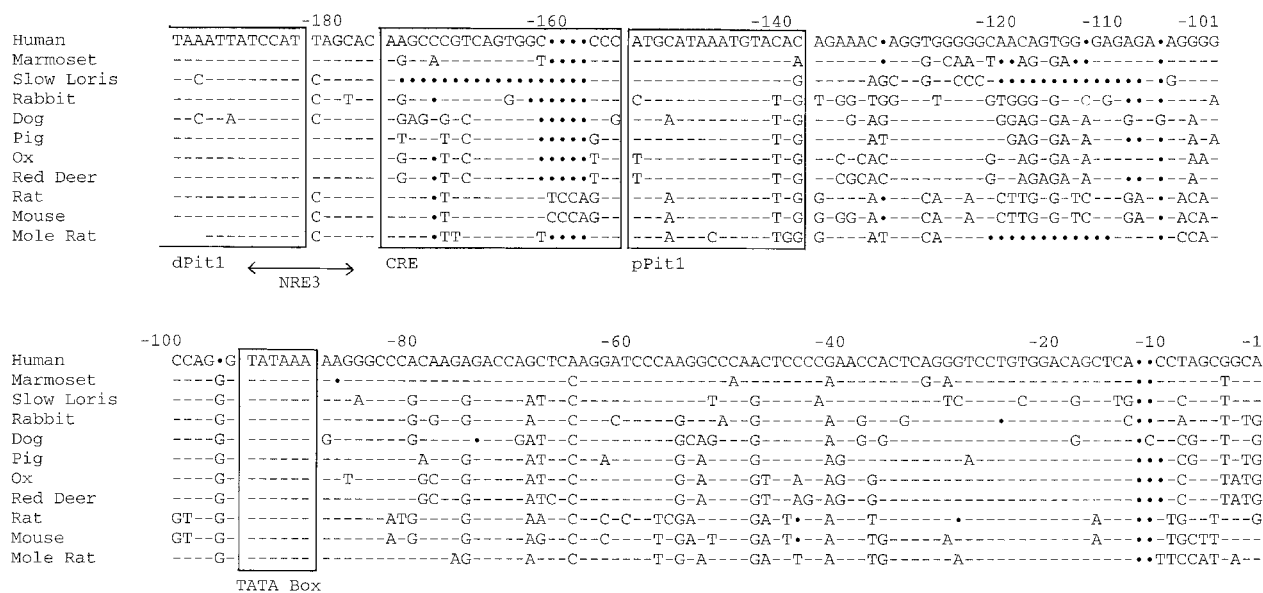
```
                  -180              -160                 -140                -120     -110    -101
Human      TAAATTATCCAT TAGCAC AAGCCCGTCAGTGGC••••CCC ATGCATAAATGTACAC AGAAAC•AGGTGGGGGCAACAGTGG•GAGAGA•AGGGG
Marmoset   ------------ ------ -G--A---------T•••---- ---------------A ------•---G-CAA-T••AG-GA••------•-----
Slow Loris --C--------- C----- •••••••••••••••••••••-- -------------G -----AGC--G--CCC-••••••••••••--G----
Rabbit     ------------ C--T-- -G--•-------G-••••••-- C------------T-G T-GG-TGG---T----GTGGG-G--C-G-••-•----A
Dog        --C--A------ C----- -GAG-G-C------•••••--G ---A---------T-G ---G-AG----------GGAG-GA-A---G--G--A--
Pig        ------------ ------ -T---T-C------•••••--- -----AT---------T-G -----GAG-GA-A---•••---A-A
Ox         ------------ ------ -G--•T-C------•••••T-- T------------T-G --C-CAC------G--AG-GA-A-------AA-
Red Deer   ------------ ------ -G--•T-C------•••••T-- T------------T-G --CGCAC---------G--AGAGA-A---•••--A--
Rat        ------------ C----- ----•T---------TCCAG-- ---A---------T-G G----A•---CA-A--CTTG-G-TC---GA--•ACA-
Mouse      ------------ C----- ----•T---------CCCAG-- ---A---------T-G G-GG-A•---CA--A--CTTG-G-TC---GA--•ACA-
Mole Rat   --------- C----- ----•TT-------T•••--- ---A---C------TGG G----AT---CA----•••••••••••••-----•CCA-
           dPit1    ←         CRE                    pPit1
                     NRE3
```

```
            -100               -80               -60              -40          -20      -10      -1
Human      CCAG•G TATAAA AAGGGCCCACAAGAGACCAGCTCAAGGATCCCAAGGCCCAACTCCCCGAACCACTCAGGGTCCTGTGGACAGCTCA-••CCTAGCGGCA
Marmoset   ---G- ------ -•-------------------------C-------------A--------A--------G-A------------------•------T---
Slow Loris ---G- ------ ---A----G----G-----AT--C-------------T---G-----A-----------TC----C---G---TG•--C---T---
Rabbit     ---G- ------ --------G-G---G-----A---C----C-----G---A--G------A-G---G---------•-----------C•••-A---T-TG
Dog        ---G- ------ G-------G-----•---GAT---------GCAG---G-------A-G-G-----------G-----C--CG--T--G
Pig        ---G- ------ ---------A---G----AT--C--A------G-A----G------AG----------A-----------•••-CG--T-TG
Ox         ---G- ------ --T------GC-G-----AT--C---------G-A----GT--A--AG--G--------------------•••-C---TATG
Red Deer   ---G- ------ ---------GC-G-----ATC-C---------G-A----GT--AG-AG--G---------------------•••-C---TATG
Rat        GT--G- ------ ------ATG----G-----AA--C--C-C--TCGA-----GA-T•--A---T-------•-----------A---•-TG--T---G
Mouse      GT--G- ------ ------A-G----G-----AG--C--C----T-GA-T--GA-T•--A--TG-----A------------A---•-TGCTT----
Mole Rat   ---G- ------ --------AG-----A---C-------T-G-A----GA--T--A---TG------A-------------•••TTCCAT-A--
           TATA Box
```

FIGURE 1.   Alignment of 5′ upstream sequences for GH genes from primates and selected non-primates. Positions of regulatory elements are shown, as discussed in the text (dPit1 and pPit1, distal and proximal Pit-1 elements; CRE, cyclic AMP response element; NRE3, negative regulatory element). The sequence for the human GH gene is shown in full and other sequences are compared with this; – represents identity to the human GH sequence, • represents a deletion.

second, independent clones revealed no discrepancies. Unique sequences were established for the GH genes from these two species and are available from the EMBL/Genbank/DDBJ database (Accession No. AJ297562 (slow loris) and AJ297563 (marmoset)). For slow loris no evidence was seen for multiple GH genes of the sort seen in human (Chen *et al.* 1989), rhesus monkey (Golos *et al.* 1993) or New World monkeys (discussed here) or for allelic variation of the kind noted in the mole rat GH gene (Lioupis *et al.* 1999).

The nucleotide sequences determined for the slow loris (1718 bp; PCR primers excluded) and marmoset (1881 bp; PCR primers excluded) GH genes have an overall organisation similar to that of GH genes in other placental mammals (e.g. ox – Woychik *et al.* 1982, red deer – Lioupis *et al.* 1997, pig – Vize & Wells 1987, rabbit – Wallis & Wallis 1995, human – Chen *et al.* 1989), and the following interpretation has been based on these. In each case the coding sequence is split into five exons by four moderately sized introns, the positions of which are identical to those in other mammalian GH genes. The sequences include 159 bp (slow loris) or 352 bp (marmoset) of sequence 5′ of the start codon, and 96 bp (slow loris) or 117 bp (marmoset) 3′ of the TAG stop codon (including a potential polyadenylation signal, AATAAA). The exon–intron junctions (identified by comparison with other

mammalian GH gene sequences) show boundaries conforming to the GT–AG rule (Breathnach & Chambon 1981). The sizes of the introns in slow loris and marmoset are similar to those of other mammalian GH genes, though intron 3 is rather variable in length (92 bp in human and marmoset genes, 163 bp in slow loris and 197 bp in pig). The repetitive sequence element found in the second intron of the rat GH gene (Barta *et al.* 1981) is not seen in the slow loris or marmoset sequences.

Regulatory elements

Various putative regulatory sequences reported in the 5′ sequences of other GH genes are seen in the slow loris and marmoset GH genes, including a TATA box (Fig. 1). A negative regulatory element (NRE3) is conserved in most mammals, including slow loris and marmoset, and probably represents a binding site for transcription factor YY1 (Park & Roe 1996). Two putative binding sites for the Pit-1 transcription factor (Theill & Karin 1993) are seen in corresponding positions for other mammalian GH genes (Krawczak *et al.* 1999), the distal one overlapping the NRE3. It is notable that the distal and proximal sites in slow loris are much closer than in all other mammalian species studied, reflecting a deletion of about 14 nucleotides in the slow loris

gene, which removes the site of a cyclic AMP response element (CRE) found in the human GH gene (Eberhardt *et al*. 1996). This region is not deleted in other mammalian GH genes but, with the exception of marmoset and rabbit (Wallis & Wallis 1995), sequence differences (mutations in the crucial CGTCA motif) make it unlikely that it functions as a CRE.

A glucocorticoid response element, present in the first intron of the human GH gene (Slater *et al*. 1985), does not appear to be present in either the marmoset or the slow loris gene. A number of putative negative thyroid hormone responsive elements (TREs) have been identified in the 3′ untranslated region (utr) of the human GH gene (Zhang *et al*. 1992); several of these appear to be present in the marmoset gene, but not in the slow loris gene. The differences between these various regulatory elements in human, marmoset and slow loris GH genes suggest that the physiological regulation of GH in these primate species may show significant differences.

Slow loris and marmoset GHs

The derived amino acid sequences for marmoset and slow loris preGHs comprise a signal peptide (26 or 27 residues) and a mature GH sequence (190 residues). The slow loris signal peptide sequence is longer by one residue than that for other mammalian preGHs (except for ruminants). The slow loris and marmoset signal peptides differ from each other and those of other eutherian preGHs at 27–46% of all residues. The signal peptide sequences are thus generally less conserved than those of the mature GHs.

An alignment showing the derived amino acid sequences for slow loris and marmoset GHs compared with sequences of various other mammalian GHs is shown in Fig. 2. The sequence of slow loris GH is similar to that of pig GH (differing at 4 residues); the latter has been established previously as being very similar to the ancestral sequence for eutherian GHs (Wallis 1994). Like this ancestral sequence the slow loris GH sequence differs markedly (at 63 residues) from that of human GH. The putative marmoset GH sequence is similar to that of human GH (differing at 17 residues). It is possible to assess the potential significance of the sequence differences observed in the light of the 3-D models available for hormone-receptor complexes for human (crystal structure – De Vos *et al*. 1992) and ovine (homology model – Wallis *et al*. 1998) GHs. Slow loris GH differs from pig GH at four amino acid positions. None of these substitutions is shared with human GH. All are

conservative, none is in the hydrophobic core, but one (Leu[104]) involves a site that interacts with the receptor (in binding site 2) in other species. This could have an effect on binding efficacy. Marmoset GH differs from human GH at 17 residues. Three of these (Ala[62], Ser[63] and Lys[64] in marmoset) are located in receptor binding site I (de Vos *et al*. 1992); the changes are all conservative in nature, but mutagenesis of these residues has revealed some effect on binding to the receptor (Cunningham & Wells 1989), so a significant effect cannot be ruled out. The other residues that differ between marmoset and human GHs are not located close to either binding site.

Of the many differences between non-primate and human GHs, that at position 169 (His in pig and other non-primates, Asp at the equivalent position, 171, in man) appears to be most important in determining species specificity (Behncken *et al*. 1997). Slow loris GH has His at this position (Fig. 2) and would therefore be expected to be inactive in man and higher primates. Marmoset GH has Asp[171], which would probably confer activity in man.

Molecular evolution of GH in primates

The demonstration (Fig. 2) that the sequence of slow loris GH is very similar to that of pig GH establishes that the episode of rapid change in primate GH evolution occurred after divergence of slow loris and related prosimians from the line leading to higher primates. The similarity between marmoset GH and human GH seen in Fig. 2 indicates that the rapid change was essentially completed by the time of divergence of lineages leading to New World monkeys and Old World monkeys/apes. The burst of rapid change must therefore have occurred over a shorter period of evolutionary time than has been previously established, as illustrated in the phylogenetic tree shown in Fig. 3. The topology and divergence times for the tree illustrated are mostly based on those of Kumar & Hedges (1998), with branch lengths determined using the protdist and FITCH programmes of the PHYLIP package. Use of alternative trees (including that in which all orders of eutherian mammals diverged at the same point in evolutionary time – Wallis 1994), or use of the MacClade programme to derive branch lengths, gave very similar patterns for GH evolution. Use of the neighbor-joining method of Saitou & Nei (1987) also gave a tree showing a period of rapid evolution of GH in the period preceding divergence of New and Old World monkeys, although inability to include a user-defined tree with this method meant that the tree

```
                      10        20        30        40        50        60        70        80        90       100
Pig             FPAMPLSSLFANAVLRAQH LHQLAADTYKEFERAYIPEG QRYS-IQNAQAAFCFSETIP APTGKDEAQQRSDVELLRFS LLLIQSWLGPVQFLSRVFTN
Alpaca          ------------------  ---------------T-----  ----•------------  -------------------  --------------------
Ox              A----S--G---------  --------F-----T-----  ----•----T-V---------  -----N----K--L----I-  ----------L---------
Sheep/goat      A----S--G---------  --------F-----T-----  ----•----T-V---------  -----N----K--L----I-  ----------L---------
Red deer        ----S--G----------  --------F-----T-----  ----•----T-V---------  -----N----K--L----I-  ----------L---------
Dog             ------------------  -------------------  ----•------------  -------------------  --------------------
Cat             ------------------  -------------------  ----•------------  -------------------  --------------------
American mink   ------------------  ---------D---------  ----•------------  -------------M-----  --------------------
Horse           ------------------  -------------------  ----•------------  -------------M-----  ------------L-------
Rabbit          ------------------  -------------------  ----•------------  -------------M-----  ----------------A---
Slow loris      ------------------  -------------------  ----•------------  -------------M-----  ------------L-------
Man             --TI---R--D--M---HR -----F---Q---E----KE -K--FL--P-TSL----S-- T-SNRE-T--K-NL----I- --------E-----RS--A-
Monkey(rhesus)  --TI---R--D--M---HR -----F---Q---E----KE -K--FL--P-TSL----S-- T-SNRE-T--K-NL----I- --------E-----RS--A-
Marmoset        --TI---R-LD--M---HR -----F---Q---E----KE -K--FL--P-TSL----S-- T-AS-K-T--K-NL----M- --------FE-----RS--A-
Elephant        ------------------  -------------------  ----•------------  -------------------  --------------------
Rat             ------------------  -------------------  ----•------------  -----E-----T-M-----  ----------------I---
Mouse           ----------S-------  -------------------  ----•------------  -----E-----T-M-----  ----------------I---
Hamster         ------------------  -------------------  ----•----T-------  -----E-------M-----  ----------------I---
Mole rat        -------N----------  -------------------  ----•------------  -----E-------M-----  --------------------
Guinea pig      ----------G-------  -------------T-----  ----•-H-T-T--------  ---D-E----------H--  --------------------

Possum          ------------------  ----V--------T----A --H-•--ST-T---------  -------------------  --------S-----------
```

```
                     110       120       130       140       150       160       170       180       190
Pig             SLVFGTSD-RVYEKLKDLEE GIQALMRELEDGSPRAGQIL KQTYDKFDTNLRSDDALLKN YGLLSCFKKDLHKAETYLRV MKCRRFVESSCAF
Alpaca          --------•----------  -------------------  R------------------  -------------------  -------------     2
Ox              --------•----------  --L--------T-------  ----------M--------  -------R-----T------  ------G-A----    18
Sheep/goat      --------•----------  --L-------VT-------  ----------M--------  -------R-----T------  ------G-A----    19
Red deer        --------•----------  --L--------T-------  ----------M--------  -------R-----T------  ------G-A----    18
Dog             --------•----------  -------------------  -------------------  -------------------  -------------     0
Cat             --------•----------  ---------------G----  -------------------  -------------------  -------------     1
American mink   --------•----------  ---------------P--  -------------------  -------------------  -------------     3
Horse           --------•-------R----  -------------------  -------------------  -------------------  -------------     3
Rabbit          T-------•----------  -------------V--L-  ----------G--------  -------------------  ----------V-     7
Slow loris      ---L----•----------  -------------V----  -------------------  -------------------  -------------     4
Man             ---Y-A--SN--DL------  ---T--GR-------T---F ----S-----SHN------  ----Y--R--MD-V--F--I VQ--•S--G--G-    62
Monkey(rhesus)  ---Y---YSD--DL------  --T--GR-----S-T---F ----S-----SHNN-----  ----Y--R--MD-I--F--I VQ--•S--G--G-    64
Marmoset        --LY-V--SD---Y------  ---T--GR-------T-E-F M---R---V-SQNN-----  ----Y--R--MD-V--F--I VQ--•S--G--G-    67
Elephant        --------•----------  -------------P--V-  -------------M-----  -------------------  -------------     3
Rat             --M-----•----------  ------Q--------I----  --------A-M--------  -------------------  ------A------    10
Mouse           --M-----•----------  ------Q--------V----  --------A-M--------  -------------------  -------------    10
Hamster         --M-----•----------  ------Q--------V----  ----------M--------  -------------------  -------------     8
Mole Rat        --------•--F-------  -------------L----L-  ----------M--------  -------------------  -------------     7
Guinea pig      --------•----------  -------------T------  ------------N-----  -------R----RT------  -------------    13

Possum          --------•-------R----  ------Q------S-G-LV- -T-----------E-----  -------------------  -------------    16
```

FIGURE 2.   Alignment of mammalian GH sequences. The sequence of pig GH is shown in full (this is thought to be very similar to the ancestral GH sequence for placental mammals – Wallis 1994). Other sequences are compared with this; – indicates identity, • indicates gap. Numbers to the right of the alignment indicate the number of differences from the pig sequence. Note the similarity of slow loris GH to the sequences of pig and other non-primate GHs, and of marmoset GH to human GH. The sources of sequences used for the analysis are cited in Wallis (1994, 1996), except for the marsupial possum (Saunders *et al*. 1998), mole rat (Lioupis *et al*. 1999), guinea pig (Adkins *et al*. 2000, Odorico *et al*. Genbank accession number AF233853) and slow loris and marmoset (this report; accession numbers AJ297562 and AJ297563).

obtained deviated substantially from conventionally accepted mammalian phylogeny.

The episode of rapid change seen for GH evolution in primates appears to be specific to the coding sequence for the mature protein hormone; the pattern of evolution seen for other components of the GH gene is different. Thus, when sequences of the signal peptide, 5′-untranslated region or introns are analysed the burst of rapid evolution seen for the protein hormone is absent (Fig. 4). When synonymous substitutions in the coding region (i.e. substitutions that do not affect the amino acid sequence) are examined, some increase in the corresponding substitution rate is apparent; this may reflect the observation that a high rate for non-synonymous substitutions is often accompanied by a high rate for synonymous substitutions (Graur 1985, Mouchiroud *et al*. 1995). That the burst of change is specific to the protein-coding component of the gene indicates that its cause relates to the protein, resulting from either adaptive change in response to selection, or loss of selective constraints due to loss of function. A number of factors militate against loss of function and consequent lowered purifying selection. Thus, there is no evidence for loss of growth-promoting action
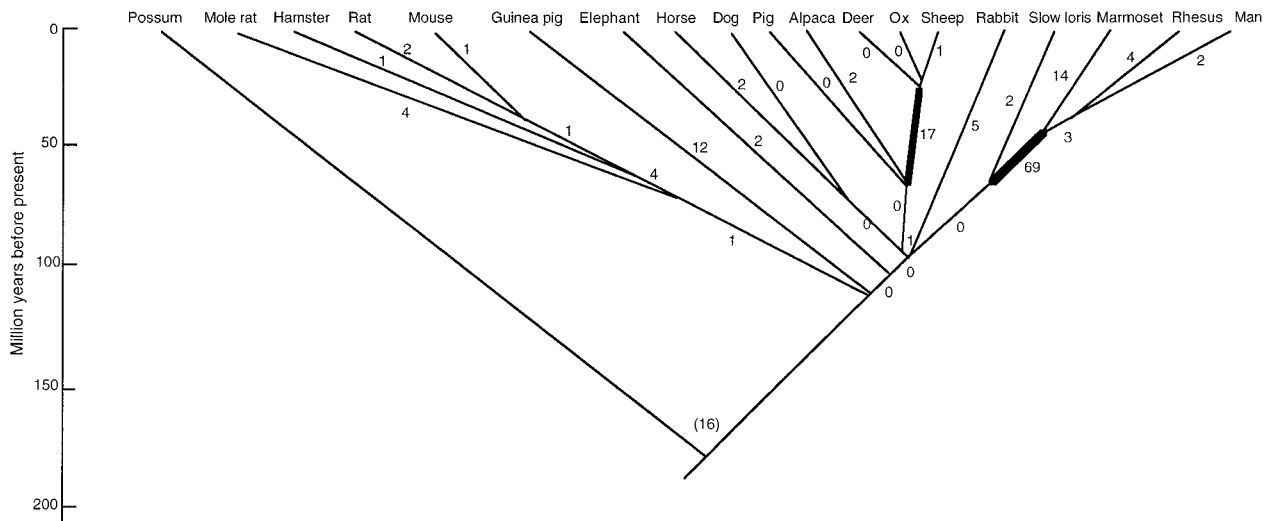
FIGURE 3.   Phylogenetic tree for mammalian GHs. The tree was constructed using the programmes protdist and FITCH in the PHYLIP package, with a user-defined tree. The vertical axis shows time (million years before present; MYBP). Divergence times were based on those of Kumar & Hedges (1998), except for divergence times for separation of lines leading to slow loris and higher primates (60 MYBP) and New and Old World monkeys (40 MYBP) which were from Szalay & Delson (1979). Numbers of substitutions are indicated along the branches of the tree. Thick lines indicate periods of rapid molecular evolution for GH. Note that the tree has been simplified by omitting data from some species of Cetartiodactyla and Carnivora.

in human GH, though there are small changes in the metabolic actions of the hormone and acquisition of species specificity. It is also notable that the episode of rapid change is followed by return to a period of slow evolution; it is difficult to see how loss of selective constraint accompanied by 'neutral' substitution of about 35% of all amino acid residues could result in a molecule which was subsequently sufficiently well adapted to its original (or a new) function to allow its almost complete preservation by purifying selection.

A more formal demonstration that the burst of rapid change in primate GH evolution reflects a period of positive adaptive change would be established if the rate of evolution for non-synonymous substitutions (K_A) significantly exceeded the rate for those parts of the gene thought to be subject to little or no purifying selection. The ratio of K_A to the rate for synonymous substitutions (K_S) has frequently been used for this purpose (Endo *et al.* 1996, Messier & Stewart 1997). The K_A/K_S ratio for GH evolution increases from 0·015 (for slowly evolving regions of the tree shown in Fig. 3; Wallis 1996) to 1·26 for the burst in primate evolution. This is a substantial increase, but not significantly greater than 1·0 (i.e. K_A is not significantly greater than K_S). Limiting factors include the low starting point and the apparent increase in K_S itself during the burst (Fig. 4). An alternative approach is to compare the rate of

evolution for the coding region with the rate for introns (defined as K_{INTR}). The K_A value for the sequence coding for the mature GH protein and K_{INTR} for introns 1, 2 and 4 (intron 3 is very variable in length, and therefore difficult to align properly) are shown in Fig. 4c and e, expressed as substitutions/site/year $\times 10^9$. For the periods before, during and after the burst of rapid change in primate GH evolution the value of K_{INTR} shows little change. K_A increases as expected during the burst, to a value that is significantly greater ($P < 0·05$) than K_{INTR}, providing further evidence that positive selection has driven the rapid change seen for GH evolution in primates.

The availability of the 3-D structure of human GH bound to two receptor molecules (de Vos *et al.* 1992) allows the residues within the hormone to be allocated to specific functional categories, and it is of interest to assess whether the pattern of episodic evolution identified for the whole hormone applies equally to these categories. Twenty-nine residues have been identified as comprising the hydrophobic core of GH. These are quite strongly conserved, with only three differences between human and slow loris GHs (10%); interestingly all three of these differences occur in the C-terminal half of helix 4. Forty-eight residues occur in or close to (within 5 Å of) receptor binding site 1 or 2. Twenty-two (46%) of these differ when human and slow loris GH sequences are compared. The episode of rapid
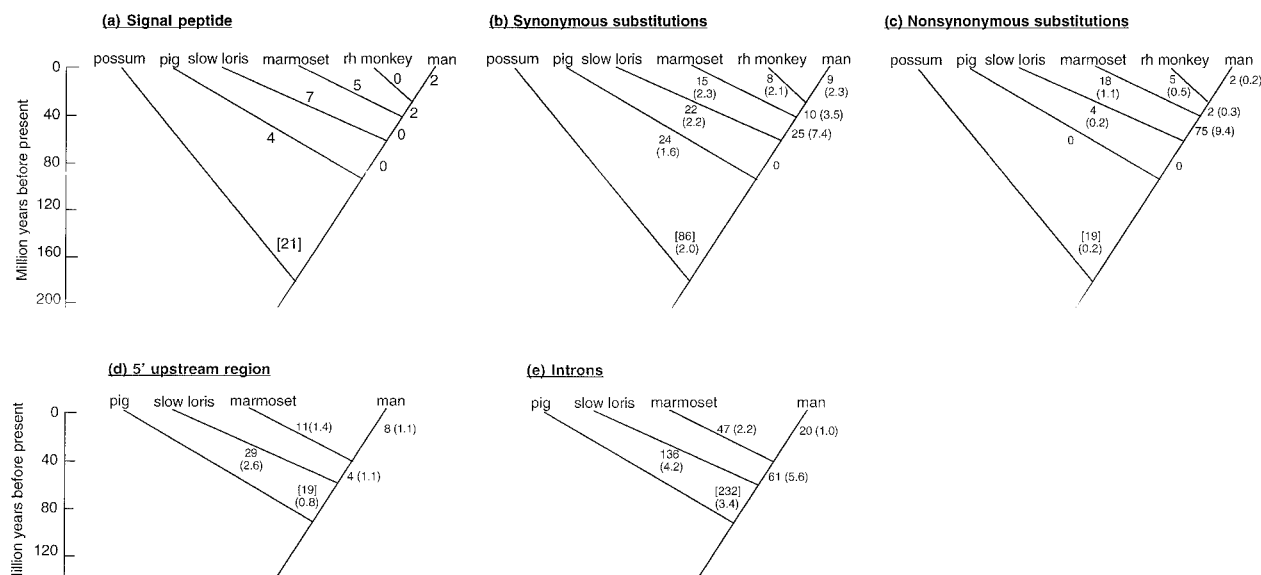
FIGURE 4.   Phylogenetic trees for primate GHs, showing changes undergone by various components of the GH gene. The trees shown are equivalent to the primate branch of the tree shown in Fig. 3, with pig and marsupial (possum) as outgroups. Numbers on branches are amino acid substitutions (a) or nucleotide substitutions (b–e), calculated using the protdist (a) or dnadist (d and e) programmes or the Nei & Gojobori (1986) method (b and c) to produce a matrix for the FITCH programme of the PHYLIP package. Numbers in parentheses (b–e) are substitution rates (substitutions/nucleotide site/year $\times 10^9$). The marsupial is not included in (d) or (e) because only a cDNA sequence is available.

acceleration identified in Figs 3 and 4 is thus particularly marked for residues associated with receptor binding, but much less obvious for residues associated with the hydrophobic core. For the former, K_A/K_S increases to 1·37, but again K_A is not significantly greater than K_S. K_A for residues associated with receptor binding sites (0·267) is significantly greater than that for residues associated with the hydrophobic core (0·038) ($P<0·05$). For residues other than those contributing to the hydrophobic core K_A/K_S is 1·77 and K_A is significantly greater than K_S ($P<0·05$), providing strong evidence for positive selection as the basis of the episode of rapid evolution. It is notable that there is also some evidence for an episode of rapid change in the GH receptor during primate evolution (Ohta 1993), and that a relatively high proportion of the substitutions accepted during evolution of GH in primates is radical in nature (Zhang 2000), though the substitution rate for radical changes is lower than that for conservative ones.

Thus, the burst of rapid evolution is largely specific to the coding sequence of mature GH, probably reflecting alterations in the function of the protein, leading to adaptive changes. The nature of these changes is not clear. Major differences in the physiological function of GH in primates and non-primates are not apparent, though human GH shows greater lactogenic (prolactin-like) activity than GHs from non-primate mammals and some differences in receptor-binding properties (Cadman & Wallis 1981, Amit *et al.* 1992), as well as the marked species specificity that has been referred to. During primate evolution duplication of the GH gene led to the appearance of a gene cluster and family of proteins, most of which are expressed in the placenta (Chen *et al.* 1989). Similarity between the genes in this cluster indicates that the gene duplications occurred *after* the episode of rapid change in primate evolution (Wallis 1996). A mechanism to explain the burst of rapid change during primate GH evolution, 'function switching', has been proposed (Wallis 1997) in which the hormone assumes a second function in addition to its main role in growth promotion, and fluctuations in the importance of this second role lead to many adaptive substitutions in the protein with rather little overall change in function. The second role could involve the metabolic actions of GH or interaction with lactogenic receptors. The pressure for rapid change would be relieved if gene duplication(s) gave rise to additional copies of the GH gene which could adopt the second function, leaving the original gene to be involved solely in regulation of growth. This would accord with the

situation in primate GH, where return of the rapid evolutionary change to a basal level occurred at about the same time as gene duplication.

ACKNOWLEDGEMENTS

NOTE ADDED IN PROOF

Recently Adkins *et al*. (2001) have reported the characterisation of the GH gene from another prosimian, the bushbaby (*Galago senegalensis*). This is similar to the gene for slow loris GH: sequences of bushbaby and slow loris GHs differ at only 1 amino acid residue.

REFERENCES

Adkins RM, Vandeberg J & Li W-H 2000 Molecular evolution of growth hormone and receptor in the guinea-pig, a mammal unresponsive to growth hormone. *Gene* **246** 357–363.

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** 55–60.

Amit T, Hochberg Z, Waters MJ & Barkey RJ 1992 Growth hormone- and prolactin-binding proteins in mammalian serum. *Endocrinology* **131** 1793–1803.

Barta A, Richards RI, Baxter JD & Shine J 1981 Primary structure and evolution of rat growth hormone gene. *PNAS* **78** 4867–4871.

Behncken SN, Rowlinson SW, Rowland JE, Conway-Campbell BL, Monks TA & Waters MJ 1997 Aspartate 171 is the major primate-specific determinant of human growth hormone. Engineering porcine growth hormone to activate the human receptor. *Journal of Biological Chemistry* **272** 27077–27083.

Beintema JJ & Campagne RN 1987 Molecular evolution of rodent insulins. *Molecular Biology and Evolution* **4** 10–18.

Blundell TL & Wood SP 1975 Is the evolution of insulin Darwinian or due to selectively neutral mutation? *Nature* **257** 197–203.

Breathnach R & Chambon P 1981 Organization and expression of eucaryotic split genes coding for proteins. *Annual Review of Biochemistry* **50** 349–383.

Cadman HF & Wallis M 1981 An investigation of the sites that bind human somatotropin (growth hormone) in the liver of the pregnant rabbit. *Biochemical Journal* **198** 605–614.

Chen EY, Liao Y-C, Smith DH, Barrera-Saldana HA, Gelinas RE & Seeburg PH 1989 The human growth hormone locus: nucleotide sequence, biology, and evolution. *Genomics* **4** 479–497.

Cunningham BC & Wells JA 1989 High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. *Science* **244** 1081–1085.

Curlewis JD, Saunders MC, Kuang J, Harrison GA & Cooper DW 1998 Cloning and sequence analysis of a pituitary prolactin cDNA from the brushtail possum (*Trichosurus vulpecula*). *General and Comparative Endocrinology* **111** 61–67.

De Vos AM, Ultsch M & Kossiakoff AA 1992 Human growth hormone and extracellular domain of its receptor: crystal structure of the complex. *Science* **255** 306–312.

Eberhardt NJ, Jiang S-W, Shepard AR, Arnold AM & Trujillo MA 1996 Hormonal and cell-specific regulation of the human growth hormone and chorionic somatomammotropin genes. *Progress in Nucleic Acid Research and Molecular Biology* **54** 127–163.

Endo T, Ikeo K & Gojobori T 1996 Large-scale search for genes on which positive selection may operate. *Molecular Biology and Evolution* **13** 685–690.

Felsenstein J 1978 Cases in which parsimony or compatability methods will be positively misleading. *Systematic Zoology* **27** 401–410.

Felsenstein J 1993 PHYLIP (Phylogeny Inference Package) version 3·5c. Distributed by the author. Department of Genetics, University of Washington, Seattle, USA.

Golos TG, Durning M, Fisher JM & Fowler PD 1993 Cloning of four growth hormone/chorionic somatomammotropin-related complementary deoxyribonucleic acids differentially expressed during pregnancy in the rhesus monkey placenta. *Endocrinology* **133** 1744–1752.

Graur D 1985 Amino acid composition and the evolutionary rates of protein-coding genes. *Journal of Molecular Evolution* **22** 53–62.

Graur D & Li W-H 1999 *Fundamentals of Molecular Evolution,* edn 2. Sunderland MA, USA: Sinauer.

Higgins DG & Sharp PM 1988 CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene* **73** 237–244.

Krawczak M, Chuzhanova NA & Cooper DN 1999 Evolution of the proximal promoter region of the mammalian growth hormone gene. *Gene* **237** 143–151.

Kumar S & Hedges SB 1998 A molecular timescale for vertebrate evolution. *Nature* **392** 917–920.

Lioupis A, Wallis OC & Wallis M 1997 Cloning and characterisation of the gene encoding red deer (*Cervus elaphus*) growth hormone: implications for the molecular evolution of growth hormone in artiodactyls. *Journal of Molecular Endocrinology* **19** 259–266.

Lioupis A, Nevo E & Wallis M 1999 Cloning and characterisation of the gene encoding mole rat (*Spalax ehrenbergi*) growth hormone. *Journal of Molecular Endocrinology* **22** 29–36.

Maddison WP & Maddison DR 1992 *MacClade*: *Analysis of Phylogeny and Character Evolution,* Version 3·0. Sunderland, MA, USA: Sinauer.

Martial JA, Hallewell RA, Baxter JD & Goodman HM 1979 Human growth hormone complementary DNA cloning and expression in bacteria. *Science* **205** 602–607.

Messier W & Stewart C-B 1997 Episodic adaptive evolution of primate lysozymes. *Nature* **385** 151–154.

Mouchiroud D, Gautier C & Bernardi G 1995 Frequencies of synonymous substitutions in mammals are gene-specific and correlated with frequencies of nonsynonymous substitutions. *Journal of Molecular Evolution* **40** 107–113.

Nei M & Gojobori T 1986 Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution* **3** 418–426.

Nicoll CS, Mayer GL & Russell SM 1986 Structural features of prolactins and growth hormones that can be related to their biological properties. *Endocrine Reviews* **7** 169–203.

Ohta T 1993 Pattern of nucleotide substitutions in growth hormone-prolactin gene family: a paradigm for evolution by gene duplication. *Genetics* **134** 1271–1276.

Park K-Y & Roe J-H 1996 Identification of a negative regulatory site in the upstream region of bovine growth hormone gene. *Biochemical and Biophysical Research Communications* **219** 354–358.

Saiki RK 1990 Amplification of genomic DNA. In *PCR Protocols*, pp 13–19. Eds MA Innis, DH Gelfand, JJ Sninsky & TJ White. San Diego, USA: Academic Press.

Saitou N & Nei M 1987 The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4** 406–425.

Saunders MC, Deakin J, Harrison GA & Curlewis JD 1998 cDNA cloning of growth hormone from the brushtail possum (*Trichosurus vulpecula*). *General and Comparative Endocrinology* **111** 68–75.

Slater EP, Rabenau O, Karin M, Baxter JD & Beato M 1985 Glucocorticoid receptor binding and activation of a hetero-logous promoter by dexamethasone by the first intron of the human growth hormone gene. *Molecular Cell Biology* **5** 2984–2992.

Southern EM 1975 Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology* **98** 503–517.

Szalay FS & Delson E 1979 *Evolutionary History of the Primates*. New York: Academic Press.

Theill LE & Karin M 1993 Transcriptional control of GH expression and anterior pituitary development. *Endocrine Reviews* **14** 670–689.

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** 135–144.

Vize PD & Wells JRE 1987 Isolation and characterization of the porcine growth hormone gene. *Gene* **55** 339–344.

Wallis M 1981 The molecular evolution of pituitary growth hormone, prolactin and placental lactogen: a protein family showing variable rates of evolution. *Journal of Molecular Evolution* **17** 10–18.

Wallis M 1989 Species specificity and structure-function relationships of growth hormone. In *Biotechnology in Growth Regulation*, pp 3–14. Eds RB Heap, CG Prosser & GE Lamming. London: Butterworths.

Wallis M 1994 Variable evolutionary rates in the molecular evolution of mammalian growth hormones. *Journal of Molecular Evolution* **38** 619–627.

Wallis M 1996 The molecular evolution of vertebrate growth hormones: a pattern of near-stasis interrupted by sustained bursts of rapid change. *Journal of Molecular Evolution* **43** 93–100.

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.

Wallis M 2000 Episodic evolution of protein hormones: molecular evolution of pituitary prolactin. *Journal of Molecular Evolution* **50** 465–473.

Wallis M 2001 Episodic evolution of protein hormones in mammals. *Journal of Molecular Evolution* (In press).

Wallis M, Lioupis A & Wallis OC 1998 Duplicate growth hormone genes in sheep and goat. *Journal of Molecular Endocrinology* **21** 1–5.

Wallis OC & Wallis M 1995 Cloning and characterisation of the rabbit growth hormone-encoding gene. *Gene* **163** 253–256.

Warwick JM, Wallis OC & Wallis M 1989 Cloning, sequence and expression in *Escherichia coli* of cDNA for ovine pregrowth hormone. *Biochimica et Biophysica Acta* **1008** 247–250.

Woychik RP, Camper SA, Lyons RH, Horowitz S, Goodwin EC & Rottman FM 1982 Cloning and nucleotide sequencing of the bovine growth hormone gene. *Nucleic Acids Research* **10** 7197–7210.

Zelus D, Robinson-Rechavi M, Delacre M, Auriault C & Laudet V 2000 Fast evolution of interleukin-2 in mammals and positive selection in ruminants. *Journal of Molecular Evolution* **51** 234–244.

Zhang J 2000 Rates of conservative and radical nonsynonymous nucleotide substitutions in mammalian nuclear genes. *Journal of Molecular Evolution* **50** 56–68.

Zhang J & Nei M 2000 Positive selection in the evolution of mammalian interleukin-2 genes. *Molecular Biology and Evolution* **17** 1413–1416.

Zhang J, Rosenberg HF & Nei M 1998 Positive Darwinian selection after gene duplication in primate ribonuclease genes. *PNAS* **95** 3708–3713.

Zhang W, Brooks RL, Silversides DW, West BL, Leidig F, Baxter JD & Eberhardt NL 1992 Negative thyroid hormone control of human growth hormone gene expression is mediated by 3′-untranslated/3′-flanking DNA. *Journal of Biological Chemistry* **267** 15056–15063.