Coevolution of insulin-like growth factors, insulin and their receptors and binding proteins in new world monkeys

Article (Accepted Version)

Coevolution of Insulin-Like Growth Factors, Insulin and Their Receptors and Binding Proteins in New World Monkeys

Michael Wallis

Biochemistry Department, School of Life Sciences, University of Sussex, Falmer, Brighton BN1 9QG. UK

Proofs and Correspondence to: M.Wallis, address as above;
Tel: 01273 472552; email: M.Wallis@sussex.ac.uk

Key words: IGF-I; IGF1R; insulin receptor; IGFBP; New World primates; molecular evolution;

Abstract

Previous work has shown that the evolution of both insulin-like growth factor 1 (IGF1) and insulin shows an episode of accelerated change on the branch leading to New World monkeys (NWM). Here the possibility that this is accompanied by a corresponding episode of accelerated evolution of IGF1 receptor (IGF1R), insulin receptor (IR) and/or IGF binding proteins (IGFBPs) was investigated. Analysis of receptor sequences from a range of primates and some non-primate mammals showed that accelerated evolution did indeed occur on this branch in the case of IGF1R and IR, but not for the similar insulin receptor-related receptor (IRRR) which does not bind insulin or IGF1. Marked accelerated evolution on this branch was also seen for some IGFBPs, but not the mannose 6-phosphate/IGF2 receptor or epidermal growth factor receptor. The rate of evolution slowed before divergence of the lineages leading to the NWM for which sequences are available (*Callithrix* and *Saimiri*). For the IGF1R and IR the accelerated evolution was most marked for the extracellular domains (ectodomains). Application of the branch-sites method showed dN/dS ratios significantly greater than 1.0 for both receptor ectodomains and for IGFBP1, and allowed identification of residues likely to have been subject to selection. These residues were concentrated in the N-terminal half of the IGF1R ectodomain but the C-terminal half of the IR ectodomain, which could have implications for the formation of hybrid receptors. Overall the results suggest that adaptive coevolution of IGF1, insulin and their receptors and some IGFBPs occurred during the evolution of NWM. For the most part, the residues that change on this branch could not be associated with specific functional aspects (ligand binding, receptor dimerization, glycosylation) and the physiological significance of this coevolution remains to be established.

1. Introduction

In primates the sequences of IGFs and insulin are strongly conserved, except on the lineage leading to New World monkeys (NWM), where for both insulin and IGF1 an episode of rapid evolution occurred [1-4]. As a consequence the sequence of each hormone in NWM differs markedly from that in other primates. A corresponding burst of change is not seen for IGF2, but a single substitution may have functional significance [4]. For insulin the rapid evolution has led to a change in its receptor-binding and biological activity in NWM, though the full significance of this is not clear [1]. These observations suggest that on the lineage leading to NWM there was a coevolution of the insulin and IGF1 genes, perhaps associated with a change in the relative roles of these molecules in regulation of metabolism and growth [4]. The question then arises as to whether associated changes also occurred in the receptors for these hormones, and the binding proteins with which IGFs associate.

The insulin receptor (IR) and IGF1 receptor (IGF1R) are large, 2-chain, membrane-spanning glycoproteins, possessing intrinsic tyrosine kinase activity and widely distributed in target tissues [5-7]. They are structurally similar (about 56% overall sequence identity) [5], and also similar to the insulin receptor-related receptor (IRRR) [8], which does not bind insulin or IGF and may act as an alkali sensor regulating bicarbonate excess [9] and/or function in male sexual differentiation [10]. The genes for IR, IGF1R and IRRP in mammals are large and comprise 21-23 exons. In each case a single precursor is cleaved to give the two chains (α and β), which remain attached via cystine bridging and exist in the plasma membrane as cystine-linked dimers. In many mammals the IR gene is subject to alternative splicing, with exon 11 (36 nucleotides) being included (IR-B) or excluded (IR-A). The proportions of the two forms vary according to tissue of expression and developmental stage [11,12]. IR-B binds primarily

insulin, IR-A both insulin and IGF2 while IGF1R has high affinity for IGF1 and IGF2 and much lower affinity for insulin [12,13].

Structural studies on the IR and IGF1R [14-17] have revealed a similar overall domain organization. The ectodomain of the receptor forms an inverted V, with three fibronectin-III-like (FnIII) domains on the C-terminal arm of the V and leucine rich 1 (LR1), cystine rich (CR) and leucine rich 2 (LR2) domains on the N-terminal arm. In the dimer the FnIII domains of one monomer are closely associated with LR1, CR and LR2 of the other. Hybrid dimers (one IR monomer and one IGF1R monomer) can form, and may be of physiological significance [18,19]. The dimeric insulin receptor ectodomain has two equivalent ligand binding sites, both involving contributions from each monomer [20,21] and showing negative cooperativity. Insulin and IGFs each contain two receptor-binding regions. One of these (site 1), binds to a site on the receptor including LR1 from one monomer and αCT (at the C-terminus of the receptor α chain) on the second. For insulin, binding site 1 was recently defined in detail [22]. Site 2 on insulin/IGF appears to bind to a less well-defined site around the FnIII-1/FnIII-2 interface on the second monomer; for IR, alanine scanning data have partially identified this site [23], but there are no equivalent data for IGF1R.

IGFBPs comprise a family of six main proteins with high affinity for IGF1 and IGF2, but not for insulin. They are produced in many locations and are found in the circulation and in many tissues. They modulate the actions of IGFs in various ways, increasing activity by extending half life and promoting tissue location, but decreasing activity by blocking receptor binding. They also have IGF-independent activities [24,25]. Each IGFBP comprises three domains, the N- and C-domains binding to separate sites on IGF, while the linker- (L-) domain provides sites for proteolytic cleavage and tissue location. The structures of N- and C-domains of

various IGFBPs have been determined, and these and mutational studies have defined the IGF binding sites [26,27].

The recent availability of extensive genomic and transcriptomic data potentially increases substantially the sequence data available for insulin, IGFs, their receptors and IGFBPs. Recently the genomic sequence of a NWM, the marmoset (*Callithrix jacchus*) has been described in detail [28] providing evidence for rapid evolution of several proteins involved in growth regulation, including IGF1R and IGFBP2; this was linked to the reduction in size seen in the callitrichine NWM (marmosets and tamarins) [28,29]. Here sequences of the IGF1R, IR, IRRR and IGFBPs of primates and several non-primate outgroups have been derived from available databases, and examined in order to investigate further which of these proteins showed an episode of rapid evolution on the lineage leading to NWM, alongside the accelerated evolution noted previously for their cognate ligands [4].

2. Methods

2.1. Data

Sequences of IGF1R, IR, IRRR, IGFBPs, epidermal growth factor receptor (EGFR) and cation independent mannose 6-phosphate/IGF2 receptor (M6P/IGF2R) for a wide range of primates and some non-primate mammals were obtained from genomic and transcriptomic databases, as detailed in Supplementary Table S1 (IGF1R, IR, IRRR), Supplementary Table S2 (IGFBPs) and Supplementary Table S3 (EGFR, M6P/IGF2R). Sequences encoding signal peptides were not available for all species, so analyses were confined to mature proteins. Likewise, for some species exon 11 of IR could not be identified, and analyses were carried out on the exon 11 minus (IR-A) form of the receptor.

2.2. Sequence Analysis

Alignments of nucleotide (coding) and protein sequences were obtained using clustalw [30], with manual adjustment where appropriate. Phylogenetic analysis to determine branch lengths was carried out using the codeml programme in paml [31] using amino acid sequences and a defined tree based on conventional understanding of primate phylogeny [32,33]. To assess the variability of sequence evolution the ratio between nonsynonymous and synonymous nucleotide substitutions (dN/dS) was used. Synonymous substitutions do not affect the protein sequence, and are therefore not subject to the selective constraints maintaining protein structure in evolution [34] . The dN/dS ratio therefore gives an indication of the rate of protein evolution relative to the underlying 'neutral' rate. The significance of elevated dN/dS ratios on the branch to NWM was tested using the likelihood ratio test, comparing model 2 (two dN/dS ratios) with model 0 (one dN/dS ratio) [31].

Whether variable rates reflected adaptive evolution was investigated using alignments of coding sequences and the branch-site method, Model A of the codeml programme [31,35]; whether dN/dS was significantly greater than 1.0 was tested using the likelihood ratio test, and a null model in which dN/dS was fixed at 1.0.

2.3. Structural models

Substitutions on the branch to NWM were mapped onto 3-dimensional models of the IR (pdb entries 2DTG and 3LOH respectively) [16,36] and IGFBP1 and 4 (pdb entries 1DSQ and 2DSR) [27] using PyMol (The PyMOL Molecular Graphics System, Version 1.3, Schrödinger, LLC).

3. Results and discussion

3.1. Episodic evolution of IGF1 and insulin receptors in mammals.

Sequences of primate IGF1Rs, IRs and IRRRs were derived from available sequence databases (Supplementary Table S1). These were aligned using clustalw. Alignments of mature proteins (Supplementary Figs. S1-S3) were subjected to phylogenetic analysis, using the codeml programme and a defined tree. Fig. 1 shows trees constructed using the protein sequences, with branch lengths based on numbers of substitutions. It is clear that the rate of evolution increased markedly on the branch leading to NWM for both IGF1R and IR but not for IRRR, although overall IRRR is less conserved than IR or IGF1R. This suggests that the accelerated evolution seen previously on this branch for insulin and IGF1 ([4] and confirmed by trees included in Fig. 1) was accompanied by a corresponding acceleration in the rate for the receptors of these molecules, but not for IRRR which does not bind either IGFs or insulin. The accelerated evolution of IGF1R seen here accords with previous reports [28,29].

Also showing a fairly high rate of evolution in the trees for IR, IGF1R and IRRR in Fig. 1 are the branches to rodents, *Cavia* and *Mus*. This may reflect the relatively rapid molecular evolution seen generally in rodents [37]. Notably, for IR, the rate for *Cavia* is only slightly higher than that for *Mus*, despite the very high rate seen for *Cavia* (and other hystricognath rodents) for insulin itself, but not IGF1. For IGF1R the branch leading to rabbit (*Oryctolagus*) also shows a relatively rapid evolution.

The sequences of IR and IGF1R can be divided into 2 major regions, with clearly distinct functions - extracellular domain (ecd; ectodomain; ligand binding) and intracellular domain (icd; signal transduction). Is the accelerated evolution seen on the branch to NWM spread across both of these, or associated mainly with one part of the molecule? It is clear from Fig. 2 that the accelerated evolution is associated primarily with the ectodomain (~14-fold

increase in rate), though a smaller rate increase (2-3-fold) is also seen for the intracellular domain. The relatively high evolutionary rate (long branch lengths) for rodents is again seen for the trees in Fig. 2. The branch for *Cavia* is only modestly longer than that for *Mus* in IR ectodomain, but notably in every other tree the *Mus* branch is longer than the *Cavia* branch, suggesting that there may have been an increase in evolutionary rate for the *Cavia* branch in IR ectodomain, though this is modest compared with the increased rate for insulin itself. The increased evolutionary rate for *Oryctolagus* IGF1R noted above is seen to be most marked for the intracellular domain (Fig. 2).

3.2. Significance of the accelerated evolution of IGF1R and IR

An episode of rapid evolution could be due to relaxed purifying selection, presumably reflecting loss of function, or to selection/adaptive change, reflecting a gain or change in function. Rather little information is available about the functioning of the IR or IGF1R or their ligands in NWM. A NWM (owl monkey, *Aotus trivirgatus*) insulin had lowered potency (about 20% that of human insulin) for binding to human IR [1]; binding to IR from NWM was not investigated. However, several considerations support the idea that the accelerated evolution of the genes for insulin, IGF1 and their receptors on the branch leading to NWM is adaptive in nature. (1) In the receptors the change is associated primarily with the extracellular domain; if an increased evolutionary rate resulted from loss of function one might expect it to be more evenly distributed across the whole molecule. (2) Following the episode of rapid evolution, the rate appears to decrease markedly. Thus, the divergence time for Platyrrhini/Catarrhini (NWM/Old World monkeys and apes) is ~45 million years ago (MYA) and for *Saimiri*/*Callithrix* ~19 MYA [38-41]. In Fig. 1, data for IRRR accord with these timings, with the average branch length following the Callithrix/Saimiri split being slightly shorter than that preceding it. However, for IR and IGF1R the branch preceding the

Callithrix/*Saimiri* split is 10-16 times longer than the branches to *Callithrix*/*Saimiri*. This would accord with a change in function during the period of accelerated evolution, followed by a return to stringent purifying selection associated with retention of the changed function. (3) Both insulin and IGF1 are crucial physiological regulators. While it is conceivable that loss of function of one of them could occur during the course of evolution, loss of function of both on a single lineage would be an extraordinary coincidence.

A statistical test of the significance of an episode of accelerated evolution can be performed by studying the ratio of nonsynonymous to synonymous substitutions (dN/dS). The former affect protein sequence and are normally subject to stringent purifying selection, while the latter have no effect on protein sequence and are thought to be largely adaptively neutral and little affected by purifying selection [42]. As a consequence, for most genes, most of the time, dN/dS is << 1.0. If a gene loses function and purifying selection no longer applies, dN/dS is expected to approach 1.0. For a gene showing adaptive evolution (positive selection) dN/dS is expected to increase. If dN/dS significantly exceeds 1.0 positive selection is clearly established, though failure to reach this statistical level does not prove that the change is not adaptive.

dN/dS ratios were determined for branches on the phylogenetic trees for IGF1R, IR and IRRR constructed from alignments of coding sequences for the mature proteins. Overall dN/dS was low, as expected, confirming stringent purifying selection. dN/dS increased significantly on the branch to NWM for IR (9 fold) and IGF1R (9 fold) but not for IRRR (Table 1). The effect for IR and IGF1R was more marked if just the ectodomain was considered (13 and 16-fold increase in dN/dS respectively), but the values of dN/dS were not greater than 1.0, so clear

evidence for positive selection is not provided. Increases for the intracellular domain, were more modest, 3.7 and 2.6 fold for IR and IGF1R respectively, but still significant.

The branch-sites method in codeml [35] provides a discriminating test for positive selection at specific sites on a specified branch on a phylogenetic tree. Applying this method to the trees for IR and IGF1R ectodomain coding sequences identified in each case a considerable number of sites (55 and 60 respectively) on the branch leading to NWM that were potentially subject to positive selection (Supplementary Table S4). The overall dN/dS values for these sites were significantly greater than 1.0 for the ectodomains of both IR (dN/dS = 2.42; 2 x Δlog likelihood (lnL) = 3.89; $p < 0.05$) and IGF1R (dN/dS= 3.35; 2 x ΔlnL = 8.68; $p<0.01$). The result provides clear-cut evidence for positive selection on the branch leading to NWM for both IR and IGF1R.

3.3. Distribution of substitutions in the ectodomains of IGF1R and IR

The branch-sites method identified 55-60 sites in the ectodomains of IR and IGF1R (Supplementary Table S4) that are likely to have evolved by positive selection on the branch leading to NWM. These sites corresponded closely with those identified as changing on this branch by visual inspection of sequences. In order to investigate the potential role of these sites in the functioning of the receptors, their positions within the 3D structure of the extracellular domains were examined.

The sites changing on the branch to NWM were mapped onto the 3D structure of the insulin receptor ectodomain (Fig. 3); in the case of IGF1R equivalent sites in the IR were mapped, based on sequence alignment of IR and IGF1R. In each case changes are seen in all 6 domains of the ectodomain, but there is a marked difference between IR and IGF1R for these

changing sites. The number of changes is similar for the IR (55 residues) and IGF1R (60 residues), but for IR the majority of changing sites (62%) fall in the C-terminal half of the receptor ectodomain (Fn-III-1, Fn-III-2 and Fn-III-3) while for IGF1R the majority (67%) are in the N-terminal half (LR1,CR, LR2). The difference is statistically significant (Fisher's Exact Test, $P<0.005$). For the IR, the Fn-III-2 domain (including the region of the insert domain around the cleavage point, which is mostly not included in the crystal structure) showed the most changes, while for IGF1R it was the CR domain (Fig. 3).

This difference between the regions of the IR and IGF1R showing most changes on the lineage to NWM could be of significance for the formation of heterodimers between IR and IGF1R, since here the regions showing most changes (N-terminal half of extracellular domain for IGF1R, C-terminal half for IR) will be apposed in the dimer, as will the corresponding regions showing fewest changes.

3.4. Functional aspects of substitutions in the ectodomains of IGF1R and IR

Ligand-binding sites. The sites on IGF1R and IR involved in binding ligand have been investigated by site-directed mutagenesis [23,43,44], computer modelling [45,46] and, in the case of IR, crystallography of insulin-receptor complexes [22]. Two main sites have been identified, involving both chains and corresponding to two binding sites on the ligands. In IR site 1 involves mainly LR1 and αCT, while site 2 involves domains Fn-III-1 and Fn-III-2 and possibly LR2. Site 1 is little changed on the branch to NWM, and none of the key residues identified as involved in insulin binding appears to be altered on this branch. However, residue 715 in αCT, immediately following F714 which appears to be crucial for insulin binding, does change (Val in human, Ile in NWM); conversion of this residue to Ala leads to lowered insulin binding [47]. In the case of site 2 several residues in domains Fn-III-1 and

Fn-III-2 are changed on the branch to NWM, and one of these, residue 591 (Asp in human, Ala in NWM) has been shown by mutation to Ala to be important for insulin binding [23].

For IGF1R, again domains L1 and αCT have been shown by alanine scanning to be involved in site 1 binding [43], but also CR which appears to play a part in binding of IGF1 but not IGF2 [48]. For L1 and αCT there are few changes on the branch leading to NWM and none of these coincides with a residue identified as involved in IGF binding. For CR many residues change on the branch to NWM (see above). None of these has been implicated in IGF1 binding by alanine scanning studies, though, in the homology model for IGF-receptor complex [46] several were implicated in the ligand/receptor interface. Site 2 for IGF1R has not been identified experimentally, but the homology model places it on Fn-III-1 and Fn-III-2, as for insulin. A number of residues changing on the branch to NWM are located in this region, two of which are implicated in ligand-receptor interaction in the homology model [46].

Dimerization. A set of residues that is clearly important for effective functioning of IR and IGF1R is that involved in dimerization. The IR ectodomain structure 3LOH.pdb [36] was used to investigate this, identifying those residues in one monomer that contact (are within 4Å of) the second monomer. 69 such residues were identified, although this may not be a complete list since some regions of the ectodomain are not defined in this structure. Only two of these were residues that change on the branch to NWM, suggesting that the interface between monomers does not change much during this episode of rapid evolution. An equivalent structure is not available for the IGF1R, but identifying sites potentially involved in dimerization on the basis of alignment of IR and IGF1R, again showed that only two of these sites changed during the period of accelerated evolution on the branch leading to NWM.

It is notable that when IR and IGF1R are compared these 69 interface sites are poorly conserved. Only 32 (46%) are identical, compared with 59% overall identity for the two receptors, suggesting that there may be substantial differences between the modes of dimerization of IR and IGF1R. This could be important in regulating formation of cross-hybrids between IR and IGF1R, though it is known that such hybrids do form, and may be of physiological importance [13,18,19].

Glycosylation. Glycosylation plays an important role in the structure and function of IR and IGF1R. About 17 Asn residues are N-glycosylated in the IR ectodomain [49], and a similar number are potentially glycosylated in the IGF1R ectodomain. In IR the crucial motif for the N-glycosylation site (Asn-X-Ser/Thr) is retained on the branch to NWM in every case; at one site Thr is replaced by Ser, but this should not affect glycosylation. For IGF1R, at one site Asn is replaced by Asp on the branch to NWM, which would prevent glycosylation; again Thr replaces Ser at one site. There are also 6 sites that are O-glycosylated in human IR, all near the N-terminus of the β-chain [50]; these do not appear to be essential for receptor assembly, ligand binding or receptor autophosphorylation. One of these sites changed (Thr -> Met) on the branch to NWM, while for a second a substitution is seen within NWM. The equivalent region in IGF1R appears to be less extensively O-glycosylated, but the extent of this is not fully established.

Conformational changes. Conformational changes follow binding of ligands to IR and IGF1R, and are important in transduction of signal and in negative cooperativity. Details of the structural changes involved are poorly understood. The C-terminal segment of the α chain of IR, αCT, appears to play an important role in signal transduction as well as insulin binding [47]. The region of αCT involved in insulin binding is mostly unchanged on the branch

leading to NWM, except for the most C-terminal residue (Val in human, Ile in NWM), but there are several changes in the sequence immediately following this, preceding the cleavage site. Whether these are involved in signal transduction is unclear - they are not defined in the crystallographic structure. The structural basis of negative cooperativity is not well defined, and it is not possible to establish whether residues involved in this include those that changed along the branch to NWM.

Naturally-occurring mutants. Over 100 human IR mutants associated with insulin resistance are listed in the Human Gene Mutation Database [51]. Of about 60 sites in the ectodomain where missense mutations or single amino acid deletions have been described (all of which have significant clinical consequences), only one coincides with a site that changes on the branch to NWM, a mutation at position 879, in the FnIII-3 domain (mutation Ala -> Gly, causing non-insulin dependent diabetes; change Ala -> Thr on branch to NWM). Fewer human IGF1R mutations have been reported [51]. Of about 15 sites where missense mutations or single amino acid deletions have been described, most are associated with growth retardation, but 4 are associated with longevity [52]. One of the latter sites coincides with a site that changes on the branch to NWM.

3.5. IGFBPs

Coding sequences for the six main IGFBPs were aligned and translated. Protein sequence alignments for mature proteins were used to construct phylogenetic trees using the codeml programme and a defined tree. Results are shown in Fig. 4. For IGFBP1, 2, 4 and 5 the branch to NWM is clearly elongated, indicating rapid evolution. In many cases rapid evolution is also seen for some of the non-primate outgroups. In all cases dN/dS is increased on the branch to NWM (Table 2), and for IGFBP1 dN/dS is significantly greater than 1.0. In

every case the mean dN/dS value was lower for terminal branches to individual NWM (*Callithrix* and *Saimiri*) than on the branch preceding divergence of these species, indicating that after the burst of rapid evolution on the branch leading to NWM, the rate fell. Analysis by the branch-sites method [35] showed values of dN/dS exceeding 1.0 on this branch for all IGFBPs, but for only IGFBP1 was this statistically significant.

The accelerated evolution on the branch to NWM is most marked for IGFBP1 and IGFBP4 and is clear from sequence alignments (Fig. 5; equivalent alignments for IGFs 2,3,5 and 6 are shown in Supplementary Fig. S4). Notably for both IGFBP1 and IGFBP4, the NWM sequences are more different from the human sequence than are those of the phylogenetically much more distant prosimian *Otolemur* (bushbaby) or non-primate species. Indicated in Fig. 5 is the division of each protein into 3 domains (in addition to signal peptide), N-domain, L-domain and C-domain. For IGFBP1 the NWM L-domain is very divergent, and the presence of deletions makes accurate alignment difficult. The rapid evolution of the protein on the branch to NWM was not due to misalignment in this region, because analysis of just N- and C-domains for IGFBP1 by the codeml branch-sites method confirmed dN/dS > 1.0 (Table 2).

For each IGFBP, residues changing on the branch to NWM were identified by the branch-sites method (Table 2). These residues were distributed across the 3 domains, with no marked predominance in any one. A few of the changing sites were identified as involved in IGF1 binding sites on N- and C-domain, but again without any marked preponderance. Residues that were markedly changed in NWM IGFBP1 compared with the human protein included five Ser residues subject to phosphorylation, which alters the affinity for IGF [53,54] (only one of which is conserved in NWM IGFBP1), and the Arg-Gly-Asp motif near the C-

terminus which is required for binding of integrin α5β1 and for effects on cell migration [55] and which is conserved in other mammals, but not lower vertebrates [56].

3.6. Specificity of the accelerated evolution in NWM

The markedly accelerated evolution seen on the lineage leading to NWM for insulin, IGF1, IR, IGF1R and several IGFBPs shows specificity in that for the two receptors it is largely confined to the ectodomains and shows a different pattern for IR and IGF1R. Furthermore, it is not seen for the related IRRR which does not bind insulin or IGFs. Analysis of the marmoset genome [28] revealed 37 genes, including *IGF1R,* showing positive selection on the lineage leading to marmoset (before and after the *Saimiri/Callithrix* split), suggesting that rapid evolution on this lineage is not common. To assess further the specificity of the accelerated evolution seen here a number of additional genes are considered.

Growth hormone (GH), prolactin and their receptors. GH and prolactin are involved in regulation of growth and lactation. They show a pattern of episodic evolution in mammals, with in each case markedly accelerated evolution in primates, but *preceding* the NWM/OWM-apes split [57]. The ecds of the GH and prolactin receptors show a corresponding episode of rapid change [57,58]. These hormones thus resemble insulin, IGF1 and their receptors in showing episodic evolution, but the pattern of change is quite distinct.

Epidermal growth factor (EGF) and its receptor (EGFR). The EGFR is a tyrosine kinase receptor, like IR and IGF1R [59]. EGF shows accelerated evolution/positive selection on

the lineage to marmoset [28]. Phylogenetic analysis shows no corresponding episode of rapid change in EGFR, either for the whole receptor or the ecd (Supplementary Fig. S5).

Cation-independent mannose 6-phosphate/IGF2 receptor (M6P/IGF2R). The M6P/IGF2R binds IGF2 as well as mannose 6-phosphate-containing glycoproteins, probably targeting the growth factor for degradation. It has a very large ecd, comprising 15 domains of which 2 (domains 11 and 13) are involved in IGF2 binding [60,61]. Phylogenetic analysis reveals no marked acceleration of evolution on the lineage leading to NWM for the whole receptor or for the IGF2-binding domains (Supplementary Fig. S6).

4. Conclusions

Analysis of the marmoset genome [28,29] showed an acceleration in the rate of evolution of IGF1R and IGFBP2, and it was suggested that this was related to the reduced size seen in callitrichine NWM. The present study confirms this acceleration, but establishes that it largely stopped before divergence of *Callithrix* and *Saimiri*, i.e. *before* the size reduction associated with callitrichine evolution. This episode of accelerated evolution applied specifically to IGF1, insulin, IGF1R, IR and several IGFBPs. Various factors suggest that adaptive evolution was involved, and in the case of the receptors and IGFBP1 this is confirmed by the branch-sites analysis. Whether the periods of accelerated evolution coincided exactly cannot be established on the basis of the available evidence, but in each case the increased rate of evolution started after divergence of NWM from OWM/apes, and was largely completed before divergence of *Saimiri* and *Callithrix*. It would clearly be a remarkable coincidence if all these molecules showed a burst of rapid evolution independently on this branch. It seems likely therefore that coevolution of these structurally-

and functionally-related molecules occurred, associated with a change in their roles in regulating cellular metabolism and growth. Further information about the functions of insulin and IGFs in NWM is needed in order to be able to assess the physiological significance of this change, and such assessment seems desirable in view of the importance of NWM, especially the marmoset, as a model for human biology and disease [62].

As far as can be ascertained the changes that occurred on the lineage leading to NWM mostly did not specifically involve alteration of features most clearly involved in biological function, including residues directly involved in receptor-ligand interactions, sites where mutation is known to cause significant loss of function, or side chains involved in dimerization. However, the structural basis of some functional aspects, including for receptors the second binding site and negative cooperativity, is poorly understood, and significant changes here cannot be excluded. Effects on the formation of heterodimers between IR and IGF1R are possible, given the different distributions of changes in their extracellular domains.

The changes occurring on the lineage leading to NWM are in marked contrast to those seen in the other group showing accelerated evolution of insulin. In the hystricognath rodents, including guinea pig, insulin sequence is markedly variable and different from that in other mammals - much more so than is the case for NWM [63-65]. However, here the sequences of IGF1 and IGF2 are strongly conserved [66]. There is a small increase in the rate of evolution of IR (Figs 1 and 2), but much less than that seen for NWM, and no acceleration of the rate for IGF1R. It has been suggested [67] that the changes in hystricognath insulin were a consequence of lack of environmental zinc, normally a component of insulin crystals and secretory granules. This explanation would accord with the observation that the changes in

this group are mainly confined to insulin, but is unlikely to apply in the case of NWM, given the marked differences in the pattern of variation seen.

References

[1] S. Seino, D.F. Steiner, G.I. Bell, Sequence of a New World primate insulin having low biological potency and immunoreactivity, Proc. Nat. Acad. Sci. USA 84 (1987) 7423–7427.

[2] J-H. Yu, J. Eng, R.S. Yalow, Isolation and amino acid sequences of squirrel monkey (*Saimiri sciurea*) insulin and glucagon, Proc. Nat. Acad. Sci. USA 87 (1990) 9766–9768.

[3] M. Wallis, Episodic evolution of protein hormones in mammals, J. Mol. Evol. 53 (2001) 10–18.

[4] M. Wallis, New insulin-like growth factor (IGF)-precursor sequences from mammalian genomes: the molecular evolution of IGFs and associated peptides in primates. Growth Hormone & IGF Res. 19 (2009) 12–23.

[5] Y. Ebina, L. Ellis, K. Jarnagin, M. Edery, L. Graf, E. Clauser, J. Ou, F. Masiarz, Y.W. Kan, I.D. Goldfine, R.A. Roth, W.J. Rutter, The human insulin receptor cDNA: the structural basis for hormone-activated transmembrane signalling, Cell 40 (1985) 747-758.

[6] A. Ullrich, J.R. Bell, E.Y. Chen, R. Herrera, L.M. Petruzzelli, T.J. Dull, A. Gray, L. Coussens, Y.-C. Liao, M. Tsubokawa, A. Mason, P.H. Seeburg, C. Grunfeld, O.M. Rosen, J.

Ramachandran, Human insulin receptor and its relationship to the tyrosine kinase family of oncogenes, Nature 313 (1985) 756-761.

[7] A. Ullrich, A. Gray, A.W. Tam, T. Yang-Feng, M. Tsubokawa, C. Collins, W. Henzel, T. Le Bon, S. Kathuria, E. Chen, S. Jacobs, U. Francke, J. Ramachandran, Y. Fujita-Yamaguchi, Insulin-like growth factor I receptor primary structure: comparison with insulin receptor suggests structural determinants that define functional specificity, EMBO J. 5 (1986) 2503-2512.

[8] P. Shier, V.M. Watt, Primary structure of a putative receptor for a ligand of the insulin family, J. Biol. Chem. 264 (1989) 14605-14608.

[9] I.E. Deyev, F. Sohet, K.P. Vassilenko, O.V. Serova, N.V. Popova, S.A. Zozulya, E.B. Burova, P. Houillier, D.I. Rzhevsky, A.A. Berchatova, A.N. Murashev, A.O. Chugunov, R.G. Efremov, N.N. Nikol'sky, E. Bertelli, D. Eladari, A.G.Petrenko, Insulin receptor related receptor as an extracellular alkali sensor, Cell Metabolism 13 (2011) 679-689.

[10] S. Nef, S. Verma-Kurvari, J. Merenmies, J-D. Vassalli, A. Efstratiadis, D. Accili, L.F. Parada, Testis determination requires insulin receptor family function in mice. Nature 426 (2003) 291–295.

[11] D.E. Moller, A. Yokota, J.F. Caro, J.S. Flier, Tissue-specific expression of two alternatively spliced insulin receptor mRNAs in man. Mol. Endocrinol. 3 (1989) 1263-1269.

[12] A. Belfiore, F. Frasca, G. Pandini, L. Sciacca, R. Vigneri, Insulin receptor isoforms and insulin receptor/insulin-like growth factor receptor hybrids in physiology and disease, Endocr. Rev. 30 (2009) 586-623.

[13] F. Frasca, G. Pandini, P. Scalia, L. Sciacca, R. Mineo, A. Costantino, I.D. Goldfine, A. Belfiore, R. Vigneri, Insulin receptor isoform A, a newly recognized high-affinity insulin-like growth factor II receptor in fetal and cancer cells, Mol. Cell. Biol. 19 (1999) 3278-3288.

[14] T.P.J. Garrett, N.M. McKern, M. Lou, M.J. Frenkel, J.D. Bentley, G.O. Lovrecz, T.C. Elleman, L.J. Cosgrove, C.W. Ward, Crystal structure of the first three domains of the type-1 insulin-like growth factor receptor, Nature 394 (1998) 395-399.

[15] M. Lou, T.P.J. Garrett, N.M. McKern, P.A. Hoyne, V.C. Epa, J.D. Bentley, G.O. Lovrecz, L.J. Cosgrove, M.J. Frenkel, C.W. Ward, The first three domains of the insulin receptor differ structurally from the insulin-like growth factor 1 receptor in the regions governing ligand specificity, Proc. Nat. Acad. Sci. USA 103 (2006) 12429–12434.

[16] N.M. McKern, M.C. Lawrence, V.A. Streltsov, M.Z. Lou, T.E. Adams, G.O. Lovrecz, T.C. Elleman, K.M. Richards, J.D. Bentley, P.A. Pilling, P.A. Hoyne, K.A. Cartledge, T.M. Pham, J.L. Lewis, S.E. Sankovich,.V. Stoichevska, E. Da Silva, C.P. Robinson, M.J. Frenkel, L.G. Sparrow, R.T. Fernley, V.C. Epa, C.W. Ward, Structure of the insulin receptor ectodomain reveals a folded-over conformation, Nature 443 (2006) 218–221.

[17] A.E. Whitten, B.J. Smith, J.G. Menting, M.B. Margetts, N.M. McKern, G.O. Lovrecz,

T.E. Adams, K. Richards, J.D. Bentley, J. Trewhella, C.W. Ward, M.C. Lawrence, Solution structure of ectodomains of the insulin receptor family: the ectodomain of the type 1 insulin-like growth factor receptor displays asymmetry of ligand binding accompanied by limited conformational change, J. Mol. Biol. 394 (2009) 878-892.

[18] M.A. Soos, K. Siddle, Immunological relationships between receptors for insulin and insulin-like growth factor I. Evidence for structural heterogeneity of insulin-like growth factor I receptors involving hybrids with insulin receptors, Biochem. J. 263 (1989) 553-563.

[19] S. Benyoucef, K.H. Surinya, D. Hadaschik, K. Siddle, Characterization of insulin/IGF hybrid receptors: contributions of the insulin receptor L2 and Fn1 domains and the alternatively spliced exon 11 sequence to ligand binding and receptor activation, Biochem. J. 403 (2007) 603–613.

[20] P. De Meyts, The insulin receptor: a prototype for dimeric, allosteric membrane receptors? Trends Biochem. Sci. 33 (2008) 376-384.

[21] C.W. Ward, J.G. Menting, M.C. Lawrence, The insulin receptor changes conformation in unforeseen ways on ligand binding: Sharpening the picture of insulin receptor activation, Bioessays 35 (2013) 945–954.

[22] J.G. Menting, J. Whittaker, M.B. Margetts, L.J. Whittaker, G.K.W. Kong, B.J. Smith, C.J. Watson, L. Žáková, E. Kletvíková, J. Jiráček, S.J. Chan, D.F. Steiner, G.G. Dodson, A.M. Brzozowski, M.A. Weiss, C.W. Ward, M.C. Lawrence, How insulin engages its

primary binding site on the insulin receptor, Nature 493 (2013) 241-245.

[23] L. Whittaker, C. Hao, W. Fu, J. Whittaker, High-affinity insulin binding: insulin interacts with two receptor ligand binding sites, Biochemistry 47 (2008) 12900-12909.

[24] R.C. Baxter, IGF binding proteins in cancer: mechanistic and clinical insights, Nature Reviews, Cancer 14 (2014) 329-341.

[25] C. Duan, Q. Xu, Roles of insulin-like growth factor (IGF) binding proteins in regulating IGF actions, Gen. Comp. Endocrinol. 142 (2005) 44-52.

[26] D.R. Clemmons, Use of mutagenesis to probe IGF-binding protein structure/function relationships, Endocrine Revs. 22 (2001) 800-817.

[27] T. Sitar, G.M. Popowicz, I. Siwanowicz, R. Huber, T.A. Holak, Structural basis for the inhibition of insulin-like growth factors by insulin-like growth factor-binding proteins, Proc. Nat. Acad. Sci. USA 111 (2006) 13028-13033.

[28] K.C. Worley and 110 other authors, The common marmoset genome provides insight into primate biology and evolution, Nature Genetics 46 (2014) 850-857.

[29] R.A. Harris , S.D. Tardif, T. Vinar, D.E. Wildman, J.N. Rutherford, J. Rogers, K.C. Worley, K.M. Aagaard, Evolutionary genetics and implications of small size and twinning in callitrichine primates, Proc. Nat. Acad. Sci. USA 111 (2014) 1467-1472.

[30] D.G. Higgins, P.M. Sharp, CLUSTAL: a package for performing multiple sequence alignment on a microcomputer, Gene 73 (1988) 237–244.

[31] Z. Yang, PAML 4: phylogenetic analysis by maximum likelihood, Mol. Biol. Evol. 24 (2007) 1586–1591.

[32] W.J. Murphy, P.A. Pevzner, S.J. O'Brien, Mammalian phylogenomics comes of age, Trends Gen. 20 (2004) 631-639.

[33] D.E. Wildman, N.M. Jameson, J.C. Opazo, S.V. Yi, A fully resolved genus level phylogeny of neotropical primates (Platyrrhini), Mol. Phylogen. Evo. 53 (2009) 694-702.

[34] T. Gojobori, E.N. Moriyama, M. Kimura, Statistical methods for estimating sequence divergence, Methods Enzymol. 183 (1990) 531–550.

[35] J. Zhang, R. Nielsen, Z. Yang, Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level, Mol. Biol. Evol. 22 (2005) 2472–2479.

[36] B.J. Smith, K. Huang, G. Kong, S.J. Chan, S. Nakagawa, J.G. Menting, S. Hu, J. Whittaker, D.F. Steiner, P.G. Katsoyannis, C.W. Ward, M.A. Weiss, M.C. Lawrence, Structural resolution of a tandem hormone-binding element in the insulin receptor and its implications for design of peptide agonists, Proc. Nat. Acad. Sci. USA 107 (2010) 6771–6776.

[37] X. Gu, W-H. Li, Higher rates of amino acid substitution in rodents than in humans, Mol. Phylogen. Evo. 1 (1992) 211-214.

[38] M.E. Steiper, N.M. Young, Primate molecular divergence dates, Mol. Phylogen. Evo. 41 (2006) 384-394.

[39] H.J. Chatterjee, S.Y.W. Ho, I. Barnes, C. Groves, Estimating the phylogeny and divergence times of primates using a supermatrix approach, BMC Evol. Biol. 9 (2009) article 259.

[40] R.D.Wilkinson, M.E. Steiper, C. Soligo, R.D. Martin, Z.H. Yang, S. Tavaré, Dating primate divergences through an integrated analysis of palaeontological and molecular data. Syst. Biol. 60 (2011) 16-31.

[41] P. Perelman, W.E. Johnson, C. Roos, H.N. Seuánez, J.E. Horvath, M.A.M. Moreira, B. Kessing, J. Pontius, M. Roelke, R. Rumpler, M.P.C. Schneider, A, Silva, S.J. O'Brien, J. Pecon-Slattery, A molecular phylogeny of living primates, PLoS Genetics 7 (2011) article e1001342.

[42] Z. Yang, J.P. Bielawski, Statistical methods for detecting molecular adaptation, Trends Ecol. Evol. 15 (2000) 496-503.

[43] J. Whittaker, A.V. Groth, D.C. Mynarcik, L. Pluzek, V.L. Gadsbøll, L.J. Whittaker, Alanine scanning mutagenesis of a Type 1 insulin-like growth factor receptor ligand binding site, J. Biol. Chem. 276 (2001) 43980-43986.

[44] J. Whittaker, L. Whittaker, Characterization of the functional insulin binding epitopes of the full-length insulin receptor, J. Biol. Chem. 280 (2005) 20932–20936.

[45] V.C. Epa, C.W. Ward, Model for the complex between the insulin-like growth factor I and its receptor: towards designing antagonists for the IGF-1 receptor, Protein Engineering, Design & Selection 19 (2006) 377-384.

[46] H. Vashisth, C.F. Abrams, All-atom structural models for complexes of insulin-like growth factors IGF1 and IGF2 with their cognate receptor, J. Mol. Biol. 400 (2010) 645-658.

[47] J. Whittaker, L.J. Whittaker, C.T. Roberts, N.B. Phillips, F. Ismail-Beigi, M.C. Lawrence, M.A. Weiss, α-Helical element at the hormone-binding surface of the insulin receptor functions as a signaling element to activate its tyrosine kinase, Proc. Nat. Acad. Sci. USA 109 (2012) 11166–11171.

[48] H. Sørensen, L. Whittaker, J. Hinrichsen, A. Groth, J. Whittaker, Mapping of the insulin-like growth factor II binding site of the Type I insulin-like growth factor receptor by alanine scanning mutagenesis, FEBS Lett. 565 (2004) 19-22.

[49] L.G. Sparrow, M.C. Lawrence, J.J. Gorman, P.M. Strike, C.P. Robinson, N.M. McKern, C.W. Ward, N-linked glycans of the human insulin receptor and their distribution over the crystal structure, Proteins: Struct. Funct. Bioinformatics 71 (2008) 426–439.

[50] L.G. Sparrow, J.J. Gorman, P.M. Strike, C.P. Robinson, N.M. McKern, V.C. Epa, C.W. Ward, The location and characterisation of the O-linked glycans of the human insulin receptor, Proteins: Struct. Funct. Bioinformatics 66 (2007) 261–265.

[51] P.D. Stenson, M. Mort, E.V. Ball, K. Shaw, A.D. Phillips, D.N. Cooper, The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine, Hum. Genet. 133 (2014) 1-9.

[52] Y. Suh, G. Atzmon, M-O. Cho, D. Hwang, B. Liu, D.J. Leahy, N. Barzilai, P. Cohen, Functionally significant insulin-like growth factor I receptor mutations in centenarians, Proc. Nat. Acad. Sci. USA 105 (2008) 3438-3442.

[53] J.I. Jones, W.H. Busby, G. Wright, C.E. Smith, N.M. Kimack, D.R. Clemmons, Identification of the sites of phosphorylation in insulin-like growth factor binding protein-1. Regulation of its affinity by phosphorylation of serine 101, J. Biol. Chem. 268, (1993) 1125-1131.

[54] L. Dolcini , A. Sala, M. Campagnoli, S. Labò, M. Valli, L. Visai, L. Minchiotti, H.L. Monaco, M. Galliano, Identification of the amniotic fluid insulin-like growth factor binding protein-1 phosphorylation sites and propensity to proteolysis of the isoforms, FEBS J. 276 (2009) 6033-6045

[55] J.I. Jones, A. Gockerman, W.H. Busby, G. Wright, D.R. Clemmons, Insulin-like growth factor binding protein 1 stimulates cell migration and binds to the α5β1 integrin by

means of its Arg-Gly-Asp sequence, Proc. Nat. Acad. Sci. USA 90 (1993) 10553–10557.

[56] D. Ocampo Daza, G. Sundström, C.A. Bergqvist, C. Duan, D. Larhammar, Evolution of the insulin-like growth factor binding protein (IGFBP) family, Endocrinology 152 (2011) 2278-2289.

[57] I.A. Forsyth, M. Wallis, Growth hormone and prolactin-molecular and functional evolution. J. Mammary Gland Biol. Neoplasia 7 (2002) 291-312.

[58] Y. Li, M. Wallis, Y.-P. Zhang, Episodic evolution of prolactin receptor gene in mammals: coevolution with its ligand, J. Mol. Endocrinol. 35 (2005) 411-419.

[59] A. Ullrich, L. Coussens, J.S. Hayflick, T.J. Dull, A. Gray, A.W. Tam, J. Lee, Y. Yarden, T.A. Libermann, J. Schlessinger, J. Downward, E.L.V. Mayes, N. Whittle, M.D. Waterfield, P.H. Seeburg, Human epidermal growth factor receptor cDNA sequence and aberrant expression of the amplified gene in A431 epidermoid carcinoma cells, Nature 309 (1984) 418-425

[60] D.O. Morgan, J.C. Edman, D.N. Standring, V.A. Fried, M.C. Smith, R.A. Roth, W.J. Rutter, Insulin-like growth factor II receptor as a multifunctional binding protein, Nature 329 (1987) 301-307.

[61] P. Lobel, N.M. Dahms, S. Kornfeld, Cloning and sequence analysis of the cation-independent mannose 6-phosphate receptor, J. Biol. Chem. 263 (1988) 2563-2570.

[62] B.A. 't Hart, D.H. Abbott, K. Nakamura, E. Fuchs, The marmoset monkey: a multi-purpose preclinical and translational model of human biology and disease, Drug Discovery Today 17 (2012) 1160-1165.

[63] J.J. Beintema, R.N. Campagne, Molecular evolution of rodent insulins, Mol. Biol. Evol. 4 (1987) 10–18.

[64] J.M. Conlon, Evolution of the insulin molecule: insights into structure-activity and phylogenetic relationships, Peptides 22 (2001) 1183-1193.

[65] J.C. Opazo, R.E. Palma, F. Melo, E.P. Lessa, Adaptive evolution of the insulin gene in caviomorph rodents, Mol. Biol. Evol. 22 (2005) 1290–1298.

[66] G.I. Bell, M.M. Stempien, N.M. Fong, S. Seino, Sequence of a cDNA encoding guinea pig IGF-I, Nucl. Acids Res. 18 (1990) 4275.

[67] T.L. Blundell, S.P. Wood, Is the evolution of insulin Darwinian or due to selectively neutral mutation? Nature 257 (1975) 197–203.

Tables

Table 1. dN/dS values for phylogenetic trees - IR, IGF1R and IRRR

	dN/dS					
	IR		IGF1R		IRRR	
	all branches except to NWM	branch to NWM	all branches except to NWM	branch to NWM	all branches except to NWM	branch to NWM
complete CDS	0.032	0.287[a]	0.031	0.294[a]	0.089	0.140[d]
ecd	0.030	0.399[a]	0.026	0.410[a]	0.066	0.079[d]
icd	0.024	0.089[b]	0.032	0.082[c]	0.125	0.140[d]

The significance of elevated dN/dS values on the branch to NWM was tested using the likelihood ratio test, comparing model 2 in codeml (2 dN/dS ratio) with model 0 (one dN/dS ratio) [31,42].
a $2x\Delta lnL$ =120.2-135.6; $P<<0.001$
b $2x\Delta lnL$ =6.94 $P<0.01$
c $2x\Delta lnL$ =4.32 $P<0.05$
d $2x\Delta lnL$ =0.04-1.72 $P>0.1$

Table 2. dN/dS values for phylogenetic trees - IGFBPs

	dN/dS				
	all branches except to NWM	branch below NWM	branches to *Callithrix* and *Saimiri*(mean)	branch below NWM from branch-sites analysis	Number of sites under selection
IGFBP1	0.202	2.95***	0.476	6.82**	84
IGFBP1 N&C domain	0.128	2.53***	0.523	8.33**	56
IGFBP2	0.107	0.422***	0.061	2.90	20
IGFBP3	0.088	0.449***	0.036	2.00	8
IGFBP4	0.042	0.647***	0.020	4.45	21
IGFBP5	0.027	0.415***	0.018	1.76	14
IGFBP6	0.250	0.672*	0.016	2.11	11

The significance of elevated dN/dS values on the branch to NWM was tested using the likelihood ratio test, comparing model 2 in codeml (2 dN/dS ratio) with model 0 (one dN/dS ratio) [31,42].
* $2x\Delta lnL$ =4.68; $P<0.05$; *** $2x\Delta lnL$ =14.6-70.8; $P<0.001$ (comparison with all branches except to NWM)
** $2x\Delta lnL$ =17.8-20.6; $P<0.001$ (comparison with 1.0 in branch-sites method)

Figure legends

Fig. 1. Phylogenetic trees for IR, IGF1R, IRRR, insulin and IGF1. Trees were based on alignments of protein sequences, and constructed using codeml (with a defined tree) to determine branch lengths. The branch to NWM is shown as a heavy line. The scale bars show mean substitutions per site. Species included are: *Homo* (man), *Pan* (chimpanzee), *Gorilla* (gorilla), *Pongo* (orangutan), *Macaca* (rhesus macaque), *Papio* (baboon), *Callithrix* (marmoset), *Saimiri* (squirrel monkey), *Otolemur* (bushbaby), *Tupaia* (tree shrew), *Oryctolagus* (rabbit), *Mus* (mouse), *Cavia* (guinea pig), *Cani*s (dog). Full species names are given in Supplementary Table S1.

Fig. 2. Phylogenetic trees for extracellular (ecd) and intracellular (icd) domains of IR, IGF1R and IRRR. Trees were based on alignments of protein sequences, and constructed using codeml (with a defined tree), to determine branch lengths. The branch to NWM is shown as a heavy line. The scale bars show mean substitutions per site. Species included are indicated in Fig. 1 legend.

Fig. 3. Distribution of substitutions on the ectodomains of IR and IGF1R. Structures were produced using Pymol, and the 3D structure of the IR ectodomain ([16], pdb structure 2DTG, single monomer, Fab structures excluded). Residues shown in space-fill mode are those which change on the lineage to NWM. Note that the insertion domain (residues 656-754, including αCT), is not shown in the crystal structure. LR1 is largely located behind CR. Numbers following domain names are the numbers of changes on the branch to NWM for that domain.

Fig 4. Phylogenetic trees for IGFBP1-6. Trees were based on alignments of mature protein sequences, and constructed using codeml (with a defined tree) to determine branch lengths. The branch to NWM is shown as a heavy line. Scale bars indicate mean substitutions per site. Species included are indicated in Fig. 1 legend, plus *Sus* (pig), used where *Canis* sequence not available.

Fig. 5. Sequence alignments for IGFBP1 and IGFBP4. The human sequence is shown in full, identical residues are shown by . and gaps by -. Number of differences from human are shown at the end of of each sequence. The N- and C-domains are shaded grey. Species included are indicated in Fig. 1 legend.

Fig. 1.

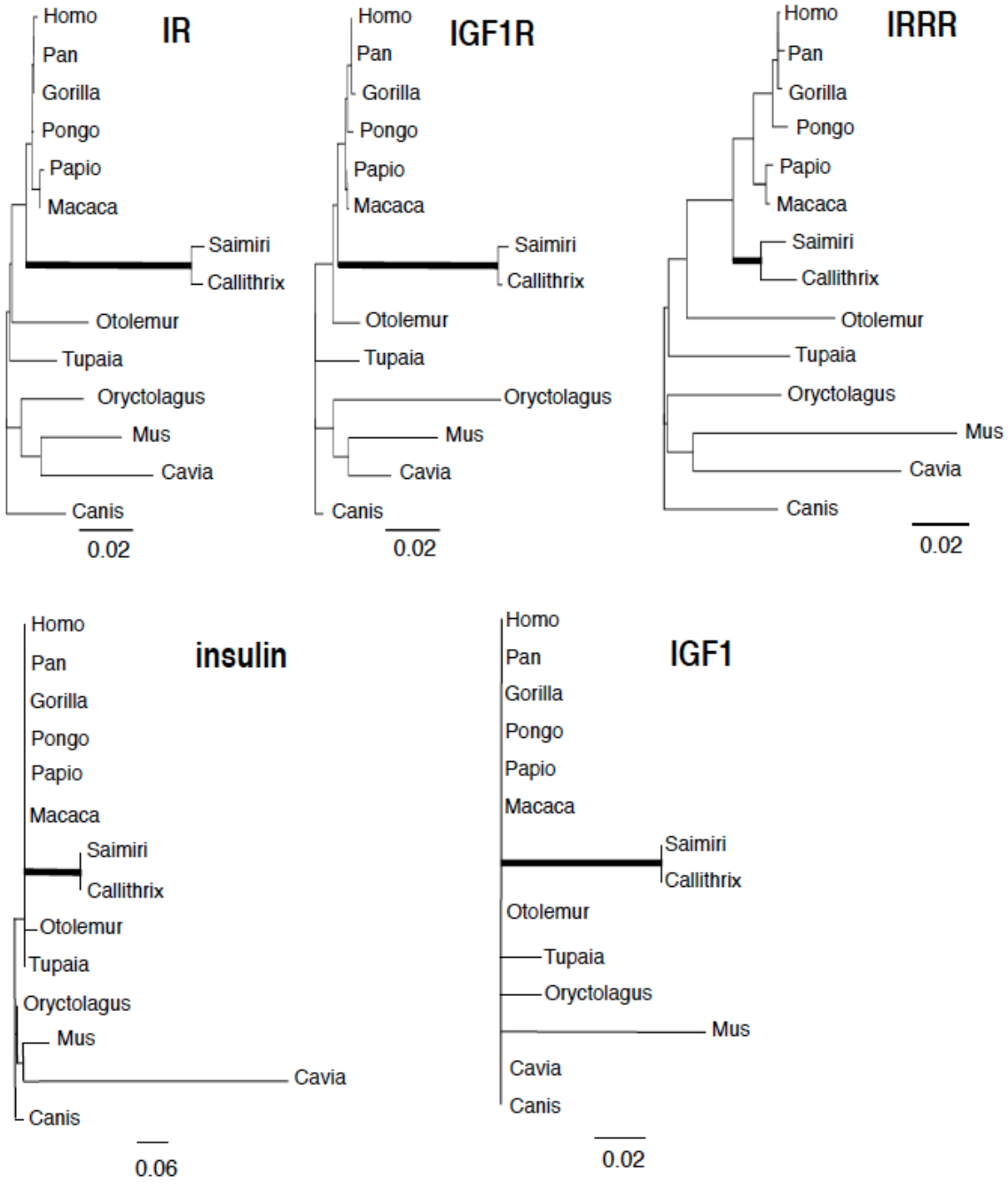

IR
Homo
Pan
Gorilla
Pongo
Papio
Macaca
Saimiri
Callithrix
Otolemur
Tupaia
Oryctolagus
Mus
Cavia
Canis
0.02
IGF1R
Homo
Pan
Gorilla
Pongo
Papio
Macaca
Saimiri
Callithrix
Otolemur
Tupaia
Oryctolagus
Mus
Cavia
Canis
0.02
IRRR
Homo
Pan
Gorilla
Pongo
Papio
Macaca
Saimiri
Callithrix
Otolemur
Tupaia
Oryctolagus
Mus
Cavia
Canis
0.02
insulin
Homo
Pan
Gorilla
Pongo
Papio
Macaca
Saimiri
Callithrix
Otolemur
Tupaia
Oryctolagus
Mus
Cavia
Canis
0.06
IGF1
Homo
Pan
Gorilla
Pongo
Papio
Macaca
Saimiri
Callithrix
Otolemur
Tupaia
Oryctolagus
Mus
Cavia
Canis
0.02

Fig. 2.

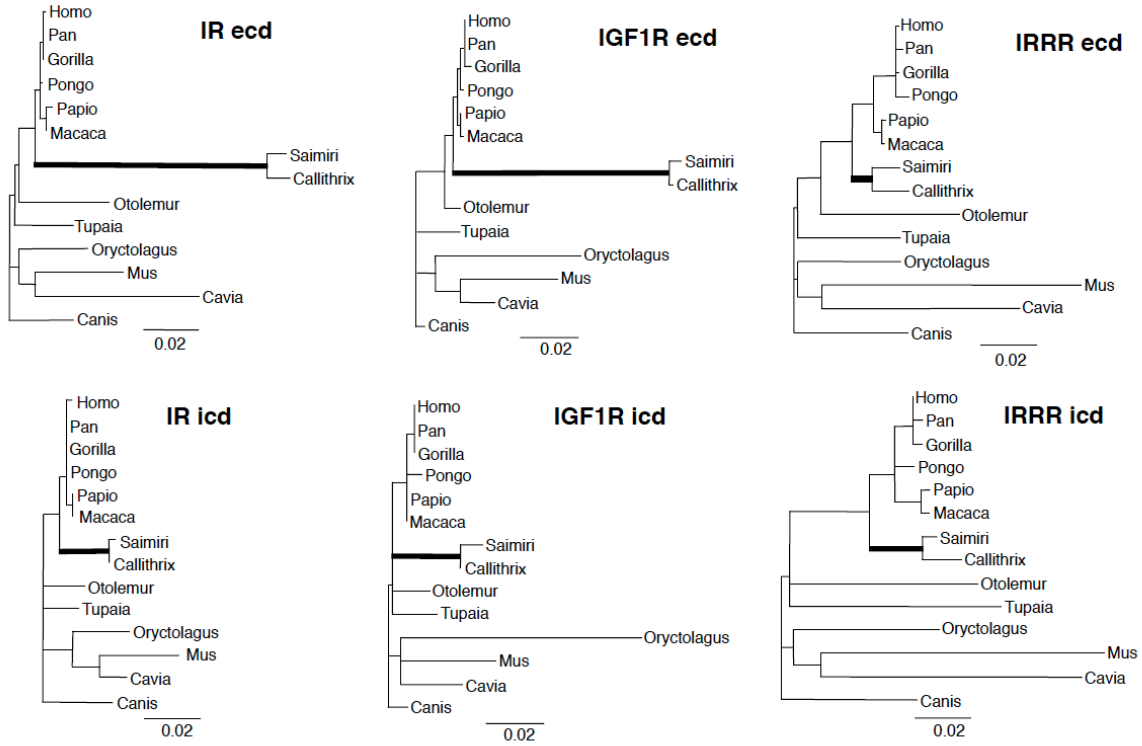

IR ecd
Homo
Pan
Gorilla
Pongo
Papio
Macaca
Saimiri
Callithrix
Otolemur
Tupaia
Oryctolagus
Mus
Cavia
Canis
0.02
IGF1R ecd
Homo
Pan
Gorilla
Pongo
Papio
Macaca
Saimiri
Callithrix
Otolemur
Tupaia
Oryctolagus
Mus
Cavia
Canis
0.02
IRRR ecd
Homo
Pan
Gorilla
Pongo
Papio
Macaca
Saimiri
Callithrix
Otolemur
Tupaia
Oryctolagus
Mus
Cavia
Canis
0.02
IR icd
Homo
Pan
Gorilla
Pongo
Papio
Macaca
Saimiri
Callithrix
Otolemur
Tupaia
Oryctolagus
Mus
Cavia
Canis
0.02
IGF1R icd
Homo
Pan
Gorilla
Pongo
Papio
Macaca
Saimiri
Callithrix
Otolemur
Tupaia
Oryctolagus
Mus
Cavia
Canis
0.02
IRRR icd
Homo
Pan
Gorilla
Pongo
Papio
Macaca
Saimiri
Callithrix
Otolemur
Tupaia
Oryctolagus
Mus
Cavia
Canis
0.02

Fig. 3.

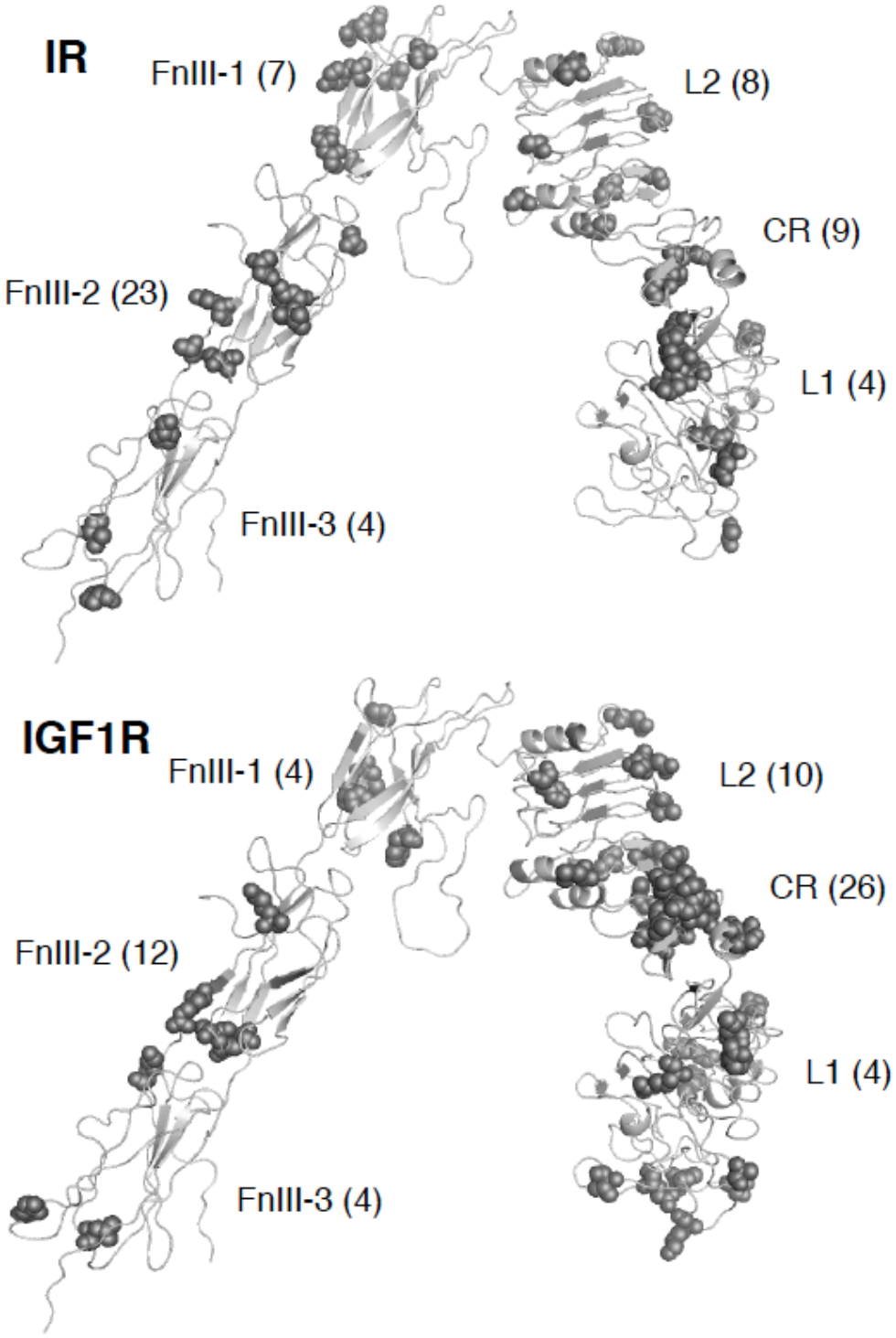

IR
FnIII-1 (7)
L2 (8)
CR (9)
FnIII-2 (23)
L1 (4)
FnIII-3 (4)
IGF1R
FnIII-1 (4)
L2 (10)
CR (26)
FnIII-2 (12)
L1 (4)
FnIII-3 (4)

Fig. 4.

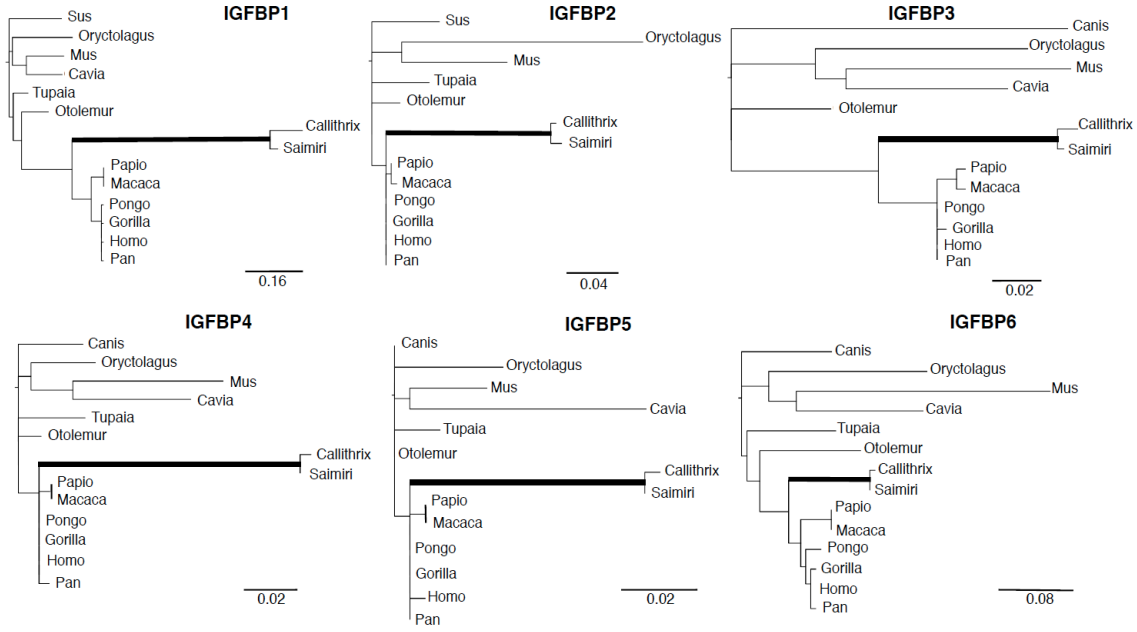
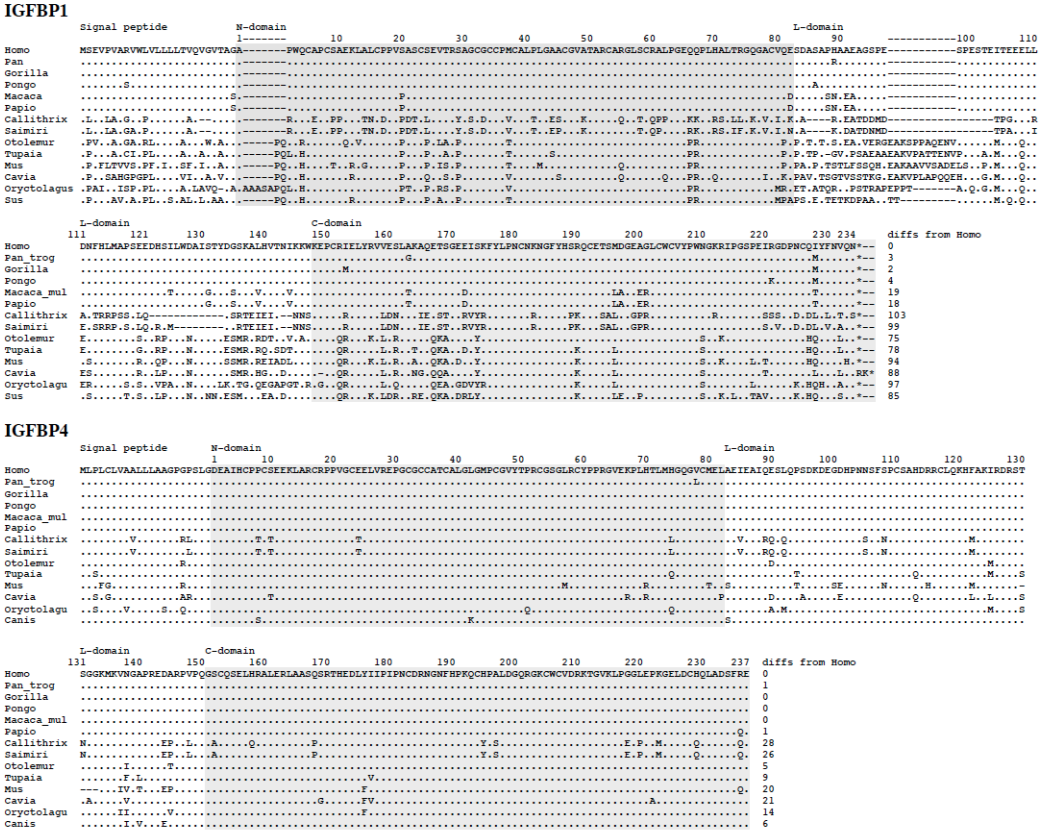

IGFBP1
Sus
Oryctolagus
Mus
Cavia
Tupaia
Otolemur
Callithrix
Saimiri
Papio
Macaca
Pongo
Gorilla
Homo
Pan
0.16
IGFBP2
Sus
Oryctolagus
Mus
Tupaia
Otolemur
Callithrix
Saimiri
Papio
Macaca
Pongo
Gorilla
Homo
Pan
0.04
IGFBP3
Canis
Oryctolagus
Mus
Cavia
Otolemur
Callithrix
Saimiri
Papio
Macaca
Pongo
Gorilla
Homo
Pan
0.02
IGFBP4
Canis
Oryctolagus
Mus
Cavia
Tupaia
Otolemur
Callithrix
Saimiri
Papio
Macaca
Pongo
Gorilla
Homo
Pan
0.02
IGFBP5
Canis
Oryctolagus
Mus
Cavia
Tupaia
Otolemur
Callithrix
Saimiri
Papio
Macaca
Pongo
Gorilla
Homo
Pan
0.02
IGFBP6
Canis
Oryctolagus
Mus
Cavia
Tupaia
Otolemur
Callithrix
Saimiri
Papio
Macaca
Pongo
Gorilla
Homo
Pan
0.08

Fig. 5.


IGFBP1
Signal peptide
N-domain
L-domain
Homo
Pan
Gorilla
Pongo
Macaca
Papio
Callithrix
Saimiri
Otolemur
Tupaia
Mus
Cavia
Oryctolagus
Sus
C-domain
diffs from Homo
IGFBP4
Signal peptide
N-domain
L-domain
C-domain
diffs from Homo