Growth Hormone & IGF Research 25 (2015) 158–167

Contents lists available at ScienceDirect

Growth Hormone & IGF Research

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

Coevolution of insulin-like growth factors, insulin and their receptors and binding proteins in New World Monkeys

Michael Wallis

Biochemistry and Biomedicine Group, School of Life Sciences, University of Sussex, Falmer, Brighton BN1 9QG, UK

ARTICLE INFO

Article history:
Received 16 April 2015
Received in revised form 2 June 2015
Accepted 3 June 2015
Available online 6 June 2015

Keywords:
IGF-I
IGF1R
Insulin receptor
IGFBP
New World Monkeys
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-site 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 IRRR 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

E-mail address: M.Wallis@sussex.ac.uk.

http://dx.doi.org/10.1016/j.ghir.2015.06.004
1096-6374/

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) [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).