INTERNATIONAL JOURNAL OF
IMMUNOGENETICS

doi: 10.1111/iji.12230

Adaptive evolution of the MHC class III-encoded receptor RAGE in primates and murine rodents

X. Wu*,[1], J. Wu†,[1], C. W. Thompson‡ & Y. Li†

Summary

The receptor for advanced glycation end products is associated with a series of physiological and pathological processes. Here, we studied the evolution of this multiligand receptor in primates and murine rodents. The evolutionary analyses reveal that adaptive selection had contributed to the variation at a number of amino acid sites in both taxa. Further, the major adaptively selected sites of both taxa are located on the extracellular ligand- and intracellular adaptor-binding regions and receptor oligomerization-related surfaces. The co-occurrence of adaptive evolution on the homologue domains suggests that they could play similar roles in these taxa. In terms of advantage fitness, the adaptive changes at these sites could contribute to host defence against the potential challenges towards these interactions and relevant signalling pathways, or the specificity of these essential points.

Introduction

The major histocompatibility complex (MHC) class III gene *AGER* encodes a receptor for advanced glycation end products (RAGE) that is critical for various aspects of physiological and pathological processes (Hofmann *et al.*, 1999; Cohen, 2013; Sorci *et al.*, 2013). It is expressed in a variety of cell types, including endothelial cells, monocytes/macrophages and T-lymphocytes (Brett *et al.*, 1993; Chen *et al.*, 2004; Moser *et al.*, 2007). In the context of immune and inflammatory responses, RAGE can sense endogenous ligands and subsequently induce cellular activation through intracellular signal pathways, such as mitogen-activated protein kinases, Rho family small GTPases Rac/Cdc42, Janus kinase–signal transducer and activator of transcription, and nuclear factor-κB (NF-κB) (Lander *et al.*, 1997; Huttunen *et al.*, 1999; Taguchi *et al.*, 2000; Huttunen & Rauvala, 2004).

Structurally, RAGE is a member of the immunoglobulin superfamily and exhibits similar key structural features in mammals (Neeper *et al.*, 1992; Schmidt *et al.*, 1994; Xu *et al.*, 2013; Sessa *et al.*, 2014). The extracellular part of RAGE contains a V-type immunoglobulin domain followed by two C-type immunoglobulin domains (C1 and C2). The structures of the V domain (Xie *et al.*, 2008; Xue *et al.*, 2011; Sirois *et al.*, 2013), V-C1 domains (Koch *et al.*, 2010; Park *et al.*, 2010; Sirois *et al.*, 2013; Xu *et al.*, 2013; Yatime & Andersen, 2013), C1–C2 domains (Sirois *et al.*, 2013) and V-C1–C2 domains (Yatime & Andersen, 2013) have been provided. These Ig-like domains consist of β-strands, which are connected to form two β-sheets. For V domain, strands A, C, C′, F and G form one β-sheet, and strands B, D and E form the second sheet. For C1 domain, strands A, B, D and E are embed in one sheet, and strands C, C′, F and G are embed in the second sheet, whereas two additional strands, A′ and G′, belong to neither of two sheets (Koch *et al.*, 2010). For putative structure of C2 domain, strands A, B and E form one sheet, and strands A′, C, F, G and G′ form the second sheet (Yatime & Andersen, 2013). With disulphide bond, the β-sheets folded into a β-sandwich structure (Bork *et al.*, 1994; Xue *et al.*, 2011). The domains V and C1 form a rigid structural unit, and residues at the interdomain interface contact through several interdomain hydrogen bonds and hydrophobic interactions (Koch *et al.*, 2010; Park *et al.*, 2010). In contrast, the C2 domain connects to V-C1 through a flexible linker (Park *et al.*, 2010; Xu *et al.*, 2013). The other domain includes an N-terminal signal sequence to target RAGE to the cell membrane. Following the extracellular domains are a single transmembrane spanning segment and a 46 amino acid-long cytoplasmic tail (ctRAGE). The solution structure of ctRAGE contains an unusual

*Department of Nephrology, The Second Affiliated Hospital of Nanjing Medical University, Nanjing, China, †College of Animal Science and Technology, Sichuan Agricultural University, Yaan, China and ‡Department of Ecology and Evolutionary Biology and Museum of Zoology, University of Michigan, Ann Arbor, MI, USA

Received 6 October 2014; revised 25 June 2015; accepted 19 July 2015

Correspondence: Dr. Yan Li, College of Animal Science and Technology, Sichuan Agricultural University, Yaan, China. Tel: +86 835 2886080; Fax: +86 835 2886080; E-mail: liyan@sicau.edu.cn

[1]X. Wu and J. Wu contributed equally to this work and should be considered co-first authors.

hooklike α-turn (residues 2–15) and an unstructured C-terminal tail (residues 16–42) (Rai *et al.*, 2012). The cytoplasmic domain shares little homologous features with any known phosphotyrosine, phosphothreonine, or phosphoserine motifs, but the domain is essential for RAGE-mediated signal transduction and subsequent function (Hofmann *et al.*, 1999; Kislinger *et al.*, 1999; Yatime & Andersen, 2013). In addition to the cell-bound form, RAGE also presents circulating secretory types, which are composed of only the extracellular domains.

Receptor for advanced glycation end products is notable for recognizing a diverse set of endogenous ligands, which do not share any obvious structural characteristics. The ligands initially identified were advanced glycation end products (AGEs), which form in normal physiological processes or in response to cellular stress caused by various pathologies (Brownlee *et al.*, 1984; Schmidt *et al.*, 2000). In addition to AGEs, RAGE can recognize nonsecreted cytosolic molecules, including several members of proinflammatory cytokine S100 family (Donato *et al.*, 2013) and high-mobility group protein box-1 (HMGB1) (Lotze & Tracey, 2005). The nuclear protein HMGB1 can be released by necrotic cells and subsequently evokes the immune response to infection, injury and inflammation (Taguchi *et al.*, 2000). The S100 family represents a subgroup of EF-hand calcium-binding proteins (Heizmann *et al.*, 2002). The components of S100 proteins are released from activated inflammatory cells and characteristically accumulate at sites of chronic inflammation (Zimmer *et al.*, 1995; Schafer & Heizmann, 1996). The release of these S100 family proteins into the extracellular space contributes importantly to the inflammatory response through interaction with receptors, including RAGE (Hofmann *et al.*, 1999; Donato *et al.*, 2013). Besides protein ligands, RAGE is a receptor for nucleic acid and consequently mounts inflammatory responses to infection and/or tissue damage (Sirois *et al.*, 2013; Xu *et al.*, 2013). As RAGE has been identified to bind to a variety of individual ligands, it is regarded as a pattern recognition receptor (PRR; Chavakis *et al.*, 2003; Yan *et al.*, 2008; Kalea *et al.*, 2009). Interestingly, the ctRAGE can also recruit different adaptor proteins, such as the extracellular signal-regulated kinase Erk1, diaphanous-1 or Toll/interleukin-1 receptor domain containing adaptor protein (Ishihara *et al.*, 2003; Hudson *et al.*, 2008; Sakaguchi *et al.*, 2011; Rai *et al.*, 2012), subsequently serves a crucial role in signal transduction.

The apparently wide spectrum of the extracellular ligands and intracellular adaptors makes the recognition strategies of RAGE highly sophisticated and interesting in understanding its roles in physiological and pathological processes. It has been reported that the interaction between RAGE-ligand occurs mainly within the V domain (Kislinger *et al.*, 1999; Taguchi *et al.*, 2000; Leclerc *et al.*, 2007; Ostendorp *et al.*, 2007; Xie *et al.*, 2007, 2008). However, recent studies on RAGE structure and function demonstrated that the C1 and C2 domains also are involved in binding ligands. RAGE can recognize S100A12, a member of S100 family, mediated only by the C1 domain (Xie *et al.*, 2007). The VC1 structural unit might act in binding N_{ε}-carboxy-methyllysine, one of the major AGE structures, and Ca^{2+}-S100B with increased affinity than that of the isolated V domain (Dattilo *et al.*, 2007; Xue *et al.*, 2011). The C1 and C2 domains can cooperate to bind to S100A6 despite the presence of the flexible linker between them (Leclerc *et al.*, 2007). In terms of intracellular signal propagation, some uncertainty remains over the structural nature and mechanism of ctRAGE in recognizing the intracellular adaptors because of the disordered structure and absence of known homologous motifs used for signal transduction (Rai *et al.*, 2012).

In addition to their divergence regarding ligand recognition, the extracellular domains of RAGE contribute to the oligomerization, which is a prerequisite for its signal transduction (Dattilo *et al.*, 2007; Ostendorp *et al.*, 2007; Xie *et al.*, 2007, 2008; Koch *et al.*, 2010; Zong *et al.*, 2010; Rai *et al.*, 2012; Xu *et al.*, 2013; Yatime & Andersen, 2013). As already reported, V domain has an important role in RAGE oligomerization. During the process of oligomerization, several nonoverlapped functional surfaces of the V domain mediate the oligomer formation and heparan sulphate can serve as an additional stabilizing factor to sustain the assemblage (Xie *et al.*, 2008; Xu *et al.*, 2013). Moreover, it has been identified that C1, C2 and/or the transmembrane regions also might be responsible for the oligomerization to transduce their signal (Xie *et al.*, 2007; Wei *et al.*, 2012; Xu *et al.*, 2013; Yatime & Andersen, 2013). Considering the critical and complex roles of RAGE in physiological and pathological processes, it is of interest to understand the atomic level structural basis and the involvement strategy of functional domains by which RAGE binds these individual extracellular ligands and intracellular adaptors and undertakes oligomerization.

Despite the progress in elucidating the structure, function and clinical relevance of RAGE, the evolution of this PRR remains, to some extent, deficient (Sessa *et al.*, 2014). The overall organization of the MHC class III, in which gene *AGER* is located, appears to be highly conserved in human and mouse, and all the genes identified in human are shared with mouse (Trowsdale, 1995; Kumanovics *et al.*, 2003). Also, the primary structures of RAGE are well conserved in mammals. For example, RAGE of mouse exhibits considerable amino acid sequence similarity (78%) with its human homologue. Moreover, RAGE generally presents a conserved picture characterized by its tertiary structural and functional roles in human and mouse (Xu *et al.*, 2013). However, little is known about whether RAGE has been influenced by factors, such as microbes in evolutionary arms race, or endogenous changes. Previous studies have suggested that

pathogens have evolved a range of anti-immune molecular methods including blockage of signalling pathways to counter both innate and adaptive immune mechanisms (Finlay & McFadden, 2006). If such items occurred on RAGE, variations in the receptor for countering such challenges could provide insights into the features relevant to its structure and function.

An important step towards our understanding structural and functional innovations in host immune-related genes includes the description of genetic variation. In this study, we analysed the patterns of nucleotide variation for this receptor in primates and murine rodents. Our results suggest that adaptive selection led to rapid evolution of RAGE in both taxa. The majority of adaptively selected amino acid sites were implicated in two types of functions: (i) the recognition of extracellular ligand and intracellular adaptor and (ii) receptor oligomerization.

Materials and methods

Samples

To investigate the evolutionary patterns and processes of *AGER*, nucleotide sequences of nine primate species were obtained from GenBank. For murines, *AGER* sequences were collected from 11 species. The sequences of *Mus musculus* and *Rattus norvegicus* were obtained from GenBank. Sequence data for the other murine taxa were generated in this study. The coding regions of *AGER* sequences from these samples were defined based on their GenBank entries of *Homo sapiens* and *M. musculus*. Full details of our taxonomic sampling are in Table S1 (Supporting Information).

DNA sequencing and alignment

The complete *AGER* sequences of nine murine samples were amplified from genomic DNA. A number of primers have been designed which are capable of amplifying the entire coding region from different species. Primer sequences are available upon request. The PCR amplification products were purified and cloned using the pMD 19-T Vector kit (TaKaRa). The inserts were sequenced using the vector primers. The coding regions of *AGER* sequences from these samples were defined based on their GenBank entries of *M. musculus* and *R. norvegicus*. The sequences of primate and murine taxa were aligned independently using ClustalW (Thompson *et al.*, 1997) and checked manually.

Evolutionary analyses

To provide comprehensive information of adaptive selection within gene *AGER* during primate and/or murine evolution, we performed evolutionary analyses for each taxon at different levels. First, we analysed the general substitution patterns of complete coding region. Because the four domains (V, C1, C2 and ctRAGE) perform distinct functions in physiological and pathological processes (see above), we also assayed the evolvement of these domains between pairs of species (Table S2 in Supporting Information). For each data set, we calculated the ratios of the number of nonsynonymous substitutions per nonsynonymous site (d_N) and the number of synonymous substitutions per synonymous site (d_S) using Pamilo–Bianchi–Li's method (Li, 1997). Deviations of d_N/d_S from neutrality were analysed with the *Z*-test and Fisher's exact test (Zhang *et al.*, 1997; Nei & Kumar, 2000). A d_N/d_S ratio that exceeds one significantly indicates positive selection, while a $d_N/d_S < 1$ reflects negative selection.

After examining adaptive selection at the entire coding region and functional domain levels, we then detected the individual sites that have been influenced by positive selection covering the complete coding region for both primates and murines with CODEML in the PAML package (Yang, 1997, 2007). Program CODEML adopts codon-based models (M) for studying the evolutionary mechanism of amino acid sites by comparing synonymous and nonsynonymous substitution rates (Goldman & Yang, 1994). For each taxon, we calculated likelihood values with three pairs of models (M0 versus M2, M7 versus M8 and M8a versus M8). Null models M0, M7 and M8a assumed that all codons evolved neutrally or under purifying selection (d_N/d_S values ≤ 1); alternative models M2 and M8 allowed a proportion of sites to be under positive selection with $d_N/d_S > 1$. For each pair comparison, a likelihood ratio test (LRT) was used to estimate the difference between them (Nielsen & Yang, 1998; Swanson *et al.*, 2003). Under model M2 and M8, a Bayes empirical Bayes (BEB) procedure was used to identify the amino acid site influenced by positive selection (Yang *et al.*, 2005).

Results and discussion

In primates and murine rodents, adaptive selection could have effects on the evolution of RAGE

The molecular signature of adaptive selection can be identified through comparative sequence analysis among species. To detect the signature of adaptive selection, we analysed the sequence variations of *AGER* in the primates and murine rodents. For primate sequences, none of the ratio comparisons exceeded one significantly. However, the ratios of *Pan troglodytes* versus *Gorilla gorilla* from Fisher's exact test ($P = 0.066$) were marginally significant. Further, the d_N/d_S ratios of *H. sapiens* to two *Pan* species, and *P. paniscus* to *G. gorilla* exceeded one, deviating from neutral expectation. These results suggest that *RAGE* potentially experienced positive selection in primates. For the V domain, we observed that none of the d_N/d_S ratios were larger than one for the possible comparisons. However, a different pattern was found for C1, C2 and ctRAGE coding regions. Twenty-four, three and

thirty pair-wise comparisons have $d_N/d_S > 1$ for these regions, respectively, although the differences are not significant. Our results suggest that these domains might be subject to positive selection, which is consistent with the observations of the entire coding region.

For murine *AGER*, we evaluated selection using the same strategy implemented for primates. It is of note that the d_N for *R. norvegicus* versus *R. rattus* was larger than their d_S, although the difference was not statistically significant. For the four functional domains, a similar pattern was observed in these domains with respect to that of primate sequences. None of the d_N/d_S ratios were larger than one for the possible comparisons for V domain. In contrast, for C1 and C2 domains, there were two (*R. exulans* versus *R. fuscipes* and *R. norvegicus* versus *R. rattus*) and three (*R. rattus* versus *R. fuscipes*, *R. norvegicus*, and *R. tanezumi*) comparisons that exceeded one, although the values were not significantly different. For the ctRAGE domain, the d_N/d_S ratios between *M. caroli* and *M. musculus* slightly exceeded one. These observations infer that RAGE could be influenced by positive selection in murine lineages similar to that of primates.

For the comparisons with $d_N/d_S > 1$, especially that of murine *AGER* data set, the marginally significant and nonsignificant results were probably due to the fact that a small number of amino acid sites are involved in the adaptive evolution. While most amino acids in a protein are under functional constraints, adaptive variation typically occurs at just a few sites (Li, 1997). However, it is also feasible that the selectively neutral mutations might account for such an evolutionary pattern. In this situation, it is valuable to examine the signs of adaptive changes at the individual amino acid sites in RAGE during primate and murine evolution.

Positive selection possibly contributed to the amino acid changes of RAGE

Using a codon-based test in a maximum likelihood framework (Yang, 1997, 2007), we detected the implication of adaptive changes at the individual amino acid sites of RAGE in both taxa (see Materials and methods). For the primate sequences, the LRT results showed that in each comparison, the model that allowed codons to evolve adaptively ($d_N/d_S > 1$) fits the data significantly better than the models for neutrality or under purifying selection ($d_N/d_S \leq 1$) (Table 1). Eight sites had d_N/d_S ratios exhibiting a departure from neutrality in favour of adaptive selection ($P > 0.80$) (Fig. S1 in Supporting Information). Three sites (189, 201 and 221) are located in the C1 domain; site 270 is located in the C2 domain. Another group of sites (369, 376, 385 and 400) are located in the ctRAGE domain.

For the murine data set, the LRT of M2 versus M0 showed that the model for adaptive selection (M2) fits the data significantly better than the model for neutrality or under purifying selection (M0) (Table 1 and Fig. S2 in Supporting Information), suggesting that a proportion of sites could be under positive selection. The BEB analysis indicated that four sites have been subject to positive selection (Table 1). One adaptively changed site (8) lies in the signal sequence. This sequence is removed resulting in the release of the mature RAGE protein. The other three sites (104, 171 and 394) are located in the V, C1 and ctRAGE domains, respectively. These results suggest that adaptive selection had played an important role in shaping the variation at numerous sites on multiple RAGE domains in both the evolution of primates and murine rodents.

Amino acid sites under adaptive selection are located on the extracellular ligand- and intracellular adaptor-binding region or contact surfaces relevant to receptor oligomerization

To further understand the structural and functional significance of evolutionary changes at the adaptively selected amino acid sites, we located them on the three-dimensional structures of mature human and mouse RAGEs (PDB ID: 4lp5, 2LMB and 4im8)

Table 1. Likelihood values and sites inferred to be under positive selection for gene *AGER* in primates and murine rodents

Taxon	Model	l[a]	$-2\Delta L$[b]	Sites under selection identified by PAML
Primates	M0: one-ratio	−2454.98	17.50** (M2 versus M0)	None
	M2: selection	−2446.23		189 (at $P \geq 0.90$)
	M7: β	−2449.55	6.64* (M8 versus M7)	Not allowed
	M8a: β and fixed ω	−2449.54	6.62* (M8 versus M8a)	Not allowed
	M8: β and ω	−2446.23		189 (at $P \geq 0.95$)
				201, 221, 270, 369, 376, 385, 400 (at $0.90 > P \geq 0.80$)
Murine rodents	M0: one-ratio	−3011.57	23.68** (M2 versus M0)	None
	M2: selection	−2099.73		106 (at $P \geq 0.80$)
	M7: β	−3001.52	3.98 (M8 versus M7)	Not allowed
	M8a: β and fixed ω	−3000.10	1.14 (M8 versus M8a)	Not allowed
	M8: β and ω	−2999.53		8, 106, 173, 397 (at $P \geq 0.80$)

[a] l, The log likelihood value under different codon-based model.

[b] $-2\Delta L$, twice the log likelihood difference between pair of models. See Materials and methods for details on LRT.

Single and double asterisks correspond to $P < 0.05$ or $P < 0.01$, respectively (one-tailed t-test). The numbers of amino acid sites represent their position in the primate and murine sequence alignments.

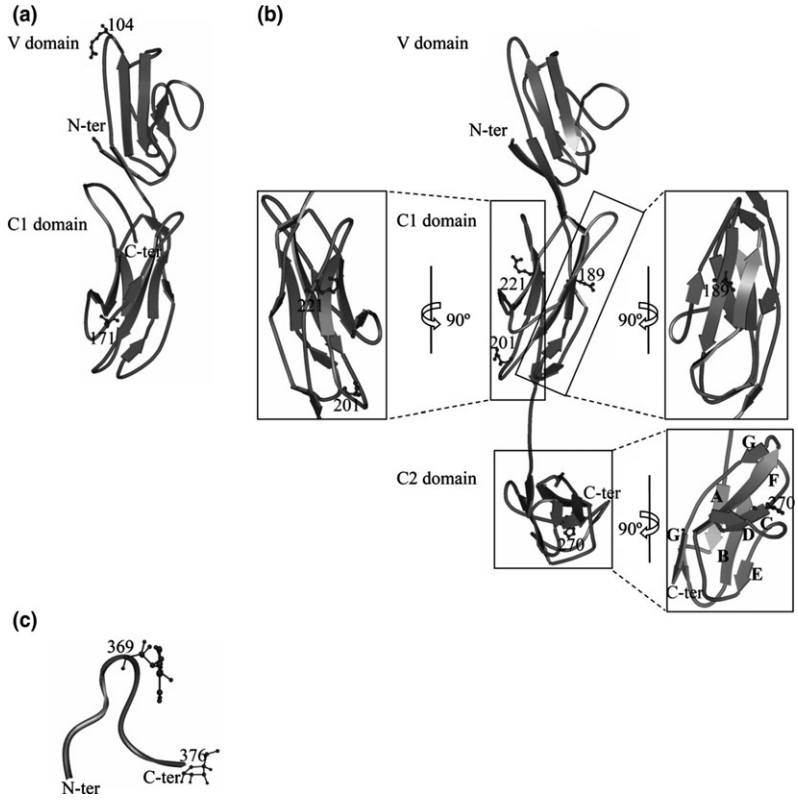

Figure 1. Stick model showing the locations of adaptively selected amino acid sites in mouse receptor for advanced glycation end products (RAGE) VC1 domains (a) (Xu *et al.*, 2013), human RAGE VC1C2 domains (b) (Yatime & Andersen, 2013) and human α-turn of ctRAGE (c) (Rai *et al.*, 2012). In (b), the β-sheets containing sites under selection are shown in the squares. In (c), the binding patch formed by sites 365–367 corresponding to 4–6 of ctRAGE is indicated by yellow colour. The residues of selected amino acids are shown on the three-dimensional structures in purple. The numbers of amino acid sites represent their position in the human and mouse RAGE sequences.

(Fig. 1) (Rai *et al.*, 2012; Xu *et al.*, 2013; Yatime & Andersen, 2013). In the V domain, the adaptively selected site 106 of murine sequences alignment (correspond to site 104 of mouse and site 105 of human RAGE) is located on a positively charged patch (Figs 1a and S3 in Supporting Information). In the human RAGE, the residue at this and nearby sites contributes to the binding of ligand Ca^{2+}-S100B (Koch *et al.*, 2010). This patch also appears to make contact with AGE-BSA and DNA (Matsumoto *et al.*, 2008; Xu *et al.*, 2013). In addition, this structural unit is responsible for binding of RAGE to heparan sulphate, which can stabilize the organization of the oligomers (Xu *et al.*, 2013). Site-directed mutagenesis studies suggested that the nearby residues (sites 104 and 107) of the site 105 in the human homologue (Fig. S3 in Supporting Information) play a critical role in RAGE-heparan sulphate oligomeric complexes (Xu *et al.*, 2013). Consistently to the previous functional perspective, the presence of adaptive evolution at this site can be a reflection of advantage(s) in contacting with the endogenous ligands.

On the C1 domain, three adaptively selected sites (189, 201 and 221) of primate RAGE and one site (173) of murine alignment data set (correspond to site 171 of mouse and site 172 of human RAGE) fall into disparate functional regions (Figs 1a,b and S3 in Supporting Information). Although the sites 173 and 189 are in different taxon, they both are located on the C1–C1 contact interface that could mediate the oligomerization of RAGE (Xie *et al.*, 2007; Koch *et al.*, 2010), in which the residue at site 189 can form a bridge with site 162′ in adjacent molecules to stabilize the oligomers (Koch *et al.*, 2010). The site 221 in primates is distributed in a positively charged patch on the edge of the β-sandwich (Fig. 1b). This patch is juxtaposed to the electropositive surface of V domain, forming a highly electrostatic platform (Koch *et al.*, 2010; Park *et al.*, 2010; Yatime & Andersen, 2013). An essay on the crystal structure of RAGE suggested that the sites within the packed platforms of RAGE dimer interact with the acidic residues in C2 domain of another symmetry-related RAGE to form a multimeric complex (Yatime & Andersen, 2013). In addition, the structural significance of the electropositive platform fits well to the negative feature of the diverse RAGE ligands (Koch *et al.*, 2010). Therefore, the variation at site 221 might contribute to recognizing specificity in adhesion molecules.

Site 201 is another adaptively selected site in primate RAGE, lying in the loop (residues 195–205)

between strands E and F of C1 domain (Koch *et al.*, 2010). Compared with β-sheet and loop topology of C1-set Ig molecules (Bork *et al.*, 1994), this loop is elongated by residues 198–201 in both primate and murine RAGEs (Koch *et al.*, 2010). The residues at site 198 of this loop also interact with C1–C2 linker through a hydrogen bond to stabilize the C terminus of C1 (Koch *et al.*, 2010). Taken together, the elongated length of E-F loop and the adaptive changes at site 201 suggest that this loop and its interaction with linker could be important for the function of RAGE.

For C2 domain, previous studies have showed that several sites (245, 337, 274 or 312) can mediate a multimeric receptor complex with a platform formed by V and C1 domains (Koch *et al.*, 2010; Park *et al.*, 2010; Yatime & Andersen, 2013). However, to date, there has been no further annotation to describe the structural features of this domain in ligand recognition or signalling transduction. The selected site 270 appears to lie in strand C of this domain (Yatime & Andersen, 2013). It is worth noting that this site and three above-mentioned sites (245, 274 and 312) are colocated on a cleft mainly comprised of strands A′, G′, G, F and C (Yatime & Andersen, 2013). Sites 270 and 245 are at the bottom of the cleft, and sites 274 and 312 are on the side wall (Fig. 1b). The combined results of our evolution analysis and previous structure essay suggest that the cleft could act as a structural and functional unit in the interaction with the basic platform of V and C1 domains.

For RAGE cytoplasmic domain, we observed that positively selected sites fall in two domains. One group of selected sites in the primate data set (369 and 376 correspond to 8 and 15 of ctRAGE, respectively) are located on the α-turn (residues 2–15, PDB ID: 2LMB) (Rai *et al.*, 2012). The hook-like shape α-turn consists of a hydrophobic patch interacting with its intracellular partner mammalian diaphanous-1 (mDia1) (Hudson *et al.*, 2008; Rai *et al.*, 2012). The selected site 369 closes to this binding patch formed by site 365–367 (Fig. 1c). Thus, the location of selected site inferred its functional significance in binding intracellular adaptors. The remaining portion of the gene in primates (385 and 400 correspond to 24 and 39 of ctRAGE, respectively) lies in C terminus, as does site 397 of murine rodent sequence (corresponding to mouse site 394 and human site 35) (Fig. S3 in Supporting Information). The dynamics and disorder of structure could not provide further clues for predicting the function of this domain. The presence of adaptive selection at the sites within this region, in both primate and murine sequences, led us to propose that these sites could have been responsible for the signalling cascades that are activated based on differential interactions. Alternatively, the changes might have constituted a functional adaptation to overcome microbial insults that block the signalling pathways.

Our analyses suggest that adaptive selection has driven the evolution at a number of RAGE sites in both primates and murine rodents. More interestingly, although none of individual sites were influenced in both taxa, we observed that the majority of these sites are located on the extracellular ligand- and intracellular adaptor-binding regions and oligomerization-related surfaces. It is becoming increasingly clear that, despite the numerous insults used by microbial pathogens, several general anti-immune mechanisms underlying these strategies are shared between these diverse pathogens (Finlay & McFadden, 2006; Elde & Malik, 2009). In these scenarios, the similar distribution pattern of these adaptively selected sites would most likely represent shared strategies of the RAGEs to counter these insulting mechanisms. In terms of physiological function, not only would the shared diversity strategies of RAGEs be in favour of their interactions with a large variety of ligands and adaptors (see above), but these general strategies would also maintain their critical functions in the signalling cascades (Lander *et al.*, 1997; Huttunen *et al.*, 1999; Taguchi *et al.*, 2000; Huttunen & Rauvala, 2004). It should be noted that the existence of the differences between two taxa, such as the site 106 of murine and site 270 of primate RAGE, suggests a different mode of diversity. They may have been due to methodological differences of the receptors to overcome different environmental and/or endogenous challenges. Certainly, further studies are needed to validate the functional implication(s) of these domains and adaptively selected sites on the exact nature of RAGE.

In conclusion, our analyses provide evidence that adaptive selection has driven the evolution of RAGE in primates and murine rodents. The evolutionary significance of this immune receptor could shed more insights into its multiligand recognition, oligomerization and strategies for countering a variety of challenges.

Acknowledgements

We thank Annie Orth for providing the *Mus pahari* sample. We also thank University of Michigan Museum of Zoology for providing the *Maxomys surifer* and *Maxomys whiteheadi* samples. We are indebted to Vernon Tintinger for providing the *Rattus exulans*, *R. fuscipes*, *R. rattus* and *R. tanezumi* DNA samples. We are grateful to Yu-Teh Kirk Lin for providing the *Mus caroli* samples. Research was supported by Specialized Research Fund for the Doctoral Program of Higher Education to Y.L. (Grant No. 20115103120007).

References

Bork, P., Holm, L. & Sander, C. (1994) The immunoglobulin fold. Structural classification, sequence patterns and common core. *Journal of Molecular Biology*, **242**, 309.

Brett, J., Schmidt, A.M., Yan, S.D., Zou, Y.S., Weidman, E., Pinsky, D. *et al.* (1993) Survey of the distribution of a newly characterized receptor for advanced glycation end products in tissues. *American Journal of Pathology*, **143**, 1699.

Brownlee, M., Vlassara, H. & Cerami, A. (1984) Nonenzymatic glycosylation and the pathogenesis of diabetic complications. *Annals of Internal Medicine*, **101**, 527.

Chavakis, T., Bierhaus, A., Al-Fakhri, N., Schneider, D., Witte, S., Linn, T. *et al.* (2003) The pattern recognition receptor (rage) is a counterreceptor for leukocyte integrins: a novel pathway for inflammatory cell recruitment. *Journal of Experimental Medicine*, **198**, 1507.

Chen, Y., Yan, S.S., Colgan, J., Zhang, H.P., Luban, J., Schmidt, A.M., Stern, D. & Herold, K.C. (2004) Blockade of late stages of autoimmune diabetes by inhibition of the receptor for advanced glycation end products. *Journal of Immunology*, **173**, 1399.

Cohen, M.M. Jr (2013) Perspectives on rage signaling and its role in cardiovascular disease. *American Journal of Medical Genetics. Part A*, **161A**, 2750.

Dattilo, B.M., Fritz, G., Leclerc, E., Kooi, C.W., Heizmann, C.W. & Chazin, W.J. (2007) The extracellular region of the receptor for advanced glycation end products is composed of two independent structural units. *Biochemistry*, **46**, 6957.

Donato, R., Cannon, B.R., Sorci, G., Riuzzi, F., Hsu, K., Weber, D.J. & Geczy, C.L. (2013) Functions of s100 proteins. *Current Molecular Medicine*, **13**, 24.

Elde, N.C. & Malik, H.S. (2009) The evolutionary conundrum of pathogen mimicry. *Nature Reviews Microbiology*, **7**, 787.

Finlay, B.B. & McFadden, G. (2006) Anti-immunology: evasion of the host immune system by bacterial and viral pathogens. *Cell*, **124**, 767.

Goldman, N. & Yang, Z. (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Molecular Biology and Evolution*, **11**, 725.

Heizmann, C.W., Fritz, G. & Schafer, B.W. (2002) S100 proteins: structure, functions and pathology. *Frontiers in Bioscience*, **7**, d1356.

Hofmann, M.A., Drury, S., Fu, C., Qu, W., Taguchi, A., Lu, Y. *et al.* (1999) Rage mediates a novel proinflammatory axis: a central cell surface receptor for s100/calgranulin polypeptides. *Cell*, **97**, 889.

Hudson, B.I., Kalea, A.Z., Del Mar Arriero, M., Harja, E., Boulanger, E., D'Agati, V. & Schmidt, A.M. (2008) Interaction of the rage cytoplasmic domain with diaphanous-1 is required for ligand-stimulated cellular migration through activation of Rac1 and Cdc42. *Journal of Biological Chemistry*, **283**, 34457.

Huttunen, H.J. & Rauvala, H. (2004) Amphoterin as an extracellular regulator of cell motility: from discovery to disease. *Journal of Internal Medicine*, **255**, 351.

Huttunen, H.J., Fages, C. & Rauvala, H. (1999) Receptor for advanced glycation end products (rage)-mediated neurite outgrowth and activation of nf-kappab require the cytoplasmic domain of the receptor but different downstream signaling pathways. *Journal of Biological Chemistry*, **274**, 19919.

Ishihara, K., Tsutsumi, K., Kawane, S., Nakajima, M. & Kasaoka, T. (2003) The receptor for advanced glycation end-products (rage) directly binds to erk by a d-domain-like docking site. *FEBS Letters*, **550**, 107.

Kalea, A.Z., Schmidt, A.M. & Hudson, B.I. (2009) Rage: a novel biological and genetic marker for vascular disease. *Clinical Science (London)*, **116**, 621.

Kislinger, T., Fu, C., Huber, B., Qu, W., Taguchi, A., Du Yan, S. *et al.* (1999) N(epsilon)-(carboxymethyl)lysine adducts of proteins are ligands for receptor for advanced glycation end products that activate cell signaling pathways and modulate gene expression. *Journal of Biological Chemistry*, **274**, 31740.

Koch, M., Chitayat, S., Dattilo, B.M., Schiefner, A., Diez, J., Chazin, W.J. & Fritz, G. (2010) Structural basis for ligand recognition and activation of rage. *Structure*, **18**, 1342.

Kumanovics, A., Takada, T. & Lindahl, K.F. (2003) Genomic organization of the mammalian mhc. *Annual Review of Immunology*, **21**, 629.

Lander, H.M., Tauras, J.M., Ogiste, J.S., Hori, O., Moss, R.A. & Schmidt, A.M. (1997) Activation of the receptor for advanced glycation end products triggers a p21(ras)-dependent mitogen-activated protein kinase pathway regulated by oxidant stress. *Journal of Biological Chemistry*, **272**, 17810.

Leclerc, E., Fritz, G., Weibel, M., Heizmann, C.W. & Galichet, A. (2007) S100b and s100a6 differentially modulate cell survival by interacting with distinct rage (receptor for advanced glycation end products) immunoglobulin domains. *Journal of Biological Chemistry*, **282**, 31317.

Li, W.-H. (1997) *Molecular Evolution*. Sinauer Associates, Sunderland, MA.

Lotze, M.T. & Tracey, K.J. (2005) High-mobility group box 1 protein (hmgb1): nuclear weapon in the immune arsenal. *Nature Reviews Immunology*, **5**, 331.

Matsumoto, S., Yoshida, T., Murata, H., Harada, S., Fujita, N., Nakamura, S. *et al.* (2008) Solution structure of the variable-type domain of the receptor for advanced glycation end products: new insight into age-rage interaction. *Biochemistry*, **47**, 12299.

Moser, B., Desai, D.D., Downie, M.P., Chen, Y., Yan, S.F., Herold, K., Schmidt, A.M. & Clynes, R. (2007) Receptor for advanced glycation end products expression on t cells contributes to antigen-specific cellular expansion in vivo. *Journal of Immunology*, **179**, 8051.

Neeper, M., Schmidt, A.M., Brett, J., Yan, S.D., Wang, F., Pan, Y.C., Elliston, K., Stern, D. & Shaw, A. (1992) Cloning and expression of a cell surface receptor for advanced glycosylation end products of proteins. *Journal of Biological Chemistry*, **267**, 14998.

Nei, M. & Kumar, S. (2000) *Molecular Evolution and Phylogenetics*. Oxford University Press, Oxford.

Nielsen, R. & Yang, Z. (1998) Likelihood models for detecting positively selected amino acid sites and applications to the hiv-1 envelope gene. *Genetics*, **148**, 929.

Ostendorp, T., Leclerc, E., Galichet, A., Koch, M., Demling, N., Weigle, B., Heizmann, C.W., Kroneck, P.M. & Fritz, G. (2007) Structural and functional insights into rage activation by multimeric s100b. *EMBO Journal*, **26**, 3868.

Park, H., Adsit, F.G. & Boyington, J.C. (2010) The 1.5 a crystal structure of human receptor for advanced glycation endproducts (rage) ectodomains reveals unique features determining ligand binding. *Journal of Biological Chemistry*, **285**, 40762.

Rai, V., Maldonado, A.Y., Burz, D.S., Reverdatto, S., Yan, S.F., Schmidt, A.M. & Shekhtman, A. (2012) Signal transduction in receptor for advanced glycation end products (rage): solution structure of c-terminal rage (ctrage) and its binding to mdia1. *Journal of Biological Chemistry*, **287**, 5133.

Sakaguchi, M., Murata, H., Yamamoto, K., Ono, T., Sakaguchi, Y., Motoyama, A., Hibino, T., Kataoka, K. & Huh, N.H. (2011) Tirap, an adaptor protein for tlr2/4, transduces a signal from rage phosphorylated upon ligand binding. *PLoS ONE*, **6**, e23132.

Schafer, B.W. & Heizmann, C.W. (1996) The s100 family of ef-hand calcium-binding proteins: functions and pathology. *Trends in Biochemical Sciences*, **21**, 134.

Schmidt, A.M., Hofmann, M., Taguchi, A., Yan, S.D. & Stern, D.M. (2000) Rage: a multiligand receptor contributing to the cellular response in diabetic vasculopathy and inflammation. *Seminars in Thrombosis and Hemostasis*, **26**, 485.

Sessa, L., Gatti, E., Zeni, F., Antonelli, A., Catucci, A., Koch, M., Pompilio, G., Fritz, G., Raucci, A. & Bianchi, M.E. (2014) The receptor for advanced glycation end-products (rage) is only present in mammals, and belongs to a family of cell adhesion molecules (CAMs). *PLoS ONE*, **9**, e86903.

Sirois, C.M., Jin, T., Miller, A.L., Bertheloot, D., Nakamura, H., Horvath, G.L. *et al.* (2013) Rage is a nucleic acid receptor that promotes inflammatory responses to DNA. *Journal of Experimental Medicine*, **210**, 2447.
Sorci, G., Riuzzi, F., Giambanco, I. & Donato, R. (2013) Rage in tissue homeostasis, repair and regeneration. *Biochimica et Biophysica Acta*, **1833**, 101.
Swanson, W.J., Nielsen, R. & Yang, Q. (2003) Pervasive adaptive evolution in mammalian fertilization proteins. *Molecular Biology and Evolution*, **20**, 18.
Taguchi, A., Blood, D.C., del Toro, G., Canet, A., Lee, D.C., Qu, W. *et al.* (2000) Blockade of rage-amphoterin signalling suppresses tumour growth and metastases. *Nature*, **405**, 354.
Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. & Higgins, D.G. (1997) The clustal_x windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, **25**, 4876.
Trowsdale, J. (1995) "Both man & bird & beast": comparative organization of MHC genes. *Immunogenetics*, **41**, 1.
Wei, W., Lampe, L., Park, S., Vangara, B.S., Waldo, G.S., Cabantous, S., Subaran, S.S., Yang, D., Lakatta, E.G. & Lin, L. (2012) Disulfide bonds within the c2 domain of rage play key roles in its dimerization and biogenesis. *PLoS ONE*, **7**, e50736.
Xie, J., Burz, D.S., He, W., Bronstein, I.B., Lednev, I. & Shekhtman, A. (2007) Hexameric calgranulin c (s100a12) binds to the receptor for advanced glycated end products (rage) using symmetric hydrophobic target-binding patches. *Journal of Biological Chemistry*, **282**, 4218.
Xie, J., Reverdatto, S., Frolov, A., Hoffmann, R., Burz, D.S. & Shekhtman, A. (2008) Structural basis for pattern recognition by the receptor for advanced glycation end products (rage). *Journal of Biological Chemistry*, **283**, 27255.
Xu, D., Young, J.H., Krahn, J.M., Song, D., Corbett, K.D., Chazin, W.J., Pedersen, L.C. & Esko, J.D. (2013) Stable rage-heparan sulfate complexes are essential for signal transduction. *ACS Chemical Biology*, **8**, 1611.
Xue, J., Rai, V., Singer, D., Chabierski, S., Xie, J., Reverdatto, S., Burz, D.S., Schmidt, A.M., Hoffmann, R. & Shekhtman, A. (2011) Advanced glycation end product recognition by the receptor for ages. *Structure*, **19**, 722.
Yan, S.F., Ramasamy, R. & Schmidt, A.M. (2008) Mechanisms of disease: advanced glycation end-products and their receptor in inflammation and diabetes complications. *Nature Clinical Practice Endocrinology & Metabolism*, **4**, 285.
Yang, Z. (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *Computer Applications in the Biosciences*, **13**, 555.
Yang, Z. (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, **24**, 1586.
Yang, Z., Wong, W.S. & Nielsen, R. (2005) Bayes empirical Bayes inference of amino acid sites under positive selection. *Molecular Biology and Evolution*, **22**, 1107.
Yatime, L. & Andersen, G.R. (2013) Structural insights into the oligomerization mode of the human receptor for advanced glycation end-products. *FEBS Journal*, **280**, 6556.
Zhang, J., Kumar, S. & Nei, M. (1997) Small-sample tests of episodic adaptive evolution: a case study of primate lysozymes. *Molecular Biology and Evolution*, **14**, 1335.
Zimmer, D.B., Cornwall, E.H., Landar, A. & Song, W. (1995) The s100 protein family: history, function, and expression. *Brain Research Bulletin*, **37**, 417.
Zong, H., Madden, A., Ward, M., Mooney, M.H., Elliott, C.T. & Stitt, A.W. (2010) Homodimerization is essential for the receptor for advanced glycation end products (rage)-mediated signal transduction. *Journal of Biological Chemistry*, **285**, 23137.

Supporting Information

Additional supporting information may be found in the online version of this article:

Table S1 Information regarding the samples and sequences included in this study.
Table S2 The range of functional domains of RAGE from human and mouse. (Neeper *et al.*, 1992; Schmidt *et al.*, 1994; Xu *et al.*, 2013).
Figure S1 Alignment of the amino acid sequences of primate RAGEs. The asterisks denote the site under positive selection, as indicated by the ML test.
Figure S2 Alignment of the amino acid sequences of murine rodent RAGEs. The asterisks denote the site under positive selection, as indicated by the ML test.
Figure S3 Alignment of the amino acid sequences of human and mouse.