

Identification of a positively evolving putative binding region with increased variability in posttranslational motifs in zonadhesin MAM domain 2

Holger Herlyn*, Hans Zischler

Institute of Anthropology, University of Mainz, Colonel-Kleinmann-Weg 2 (SB II) D-55099, Germany

Received 12 October 2004; revised 25 February 2005

Available online 31 May 2005

Abstract

Positive selection has been shown to be pervasive in sex-related proteins of many metazoan taxa. However, we are only beginning to understand molecular evolutionary processes on the lineage to humans. To elucidate the evolution of proteins involved in human reproduction, we studied the sequence evolution of MAM domains of the sperm-ligand zonadhesin in respect to single amino acid sites, solvent accessibility, and posttranslational modification. GenBank-data were supplemented by new cDNA-sequences of a representative non-human primate panel. Solvent accessibility predictions identified a probably exposed fragment of 30 amino acids belonging to MAM domain 2 (i.e., MAM domain 3 in mouse). The fragment is characterized by significantly increased rate of positively selected amino acid sites and exhibits high variability in predicted posttranslational modification, and, thus, might represent a binding region in the mature protein. At the same time, there is a significant coincidence of positively selected amino acid sites and non-conserved posttranslational motifs. We conclude that the binding specificity of zonadhesin MAM domains, especially of the presumed epitope, is achieved by positive selection at the level of single amino acid sites and posttranslational modifications, respectively.

© 2005 Elsevier Inc. All rights reserved.

Keywords: Positive selection; Posttranslational modification; Zonadhesin; MAM domains; N-myristylation; N-glycosylation; Phosphorylation; Sperm-egg interaction

1. Introduction

Positive or adaptive evolution has been shown to be common in genes involved in mating behaviour, fertilization, spermatogenesis, and sex determination of many taxa, using as measure the ratio of non-synonymous to synonymous nucleotide substitution rate ($d_n/d_s = \omega$) > 1. Most work has been carried out on marine sea urchins and molluscs (reviewed in Swanson and Vacquier, 2002). For molluscs of the genus *Haliotis* Linnaeus, 1758 (abalones), detailed studies have shown

that the sperm-ligand lysin evolves in adaptation to a constantly changing egg-receptor (Swanson et al., 2001a,b; Swanson and Vacquier, 1998). Positive evolution has also been reported in sex-related proteins of mammals (e.g., Civetta, 2003; Dorus et al., 2004; Rooney and Zhang, 1999; Swanson et al., 2001a,b, 2003; Torgerson et al., 2002; Torgerson and Singh, 2003; Wyckoff et al., 2000). However, none of the previous studies included a representative primate panel comprising Lemuroidea, Platyrrhini, Cercopithecidae, and Hominoidea, which might have important biomedical implications on conception and contraception in humans. Beyond this, a denser sampling also minimizes the negative effect of saturation which generally impairs analyses through multiple substitutions.

* Corresponding author. Fax: +49 6131 3923799.
E-mail address: herlyn@uni-mainz.de (H. Herlyn).

Zonadhesin was first isolated from pig (*Sus scrofa* Linnaeus, 1758) based on its species-specific zona pellucida binding activity (Hardy and Garbers, 1994). In pig, human (*Homo sapiens* Linnaeus, 1758), and European rabbit (*Oryctolagus cuniculus* Linnaeus, 1758) zonadhesin essentially consists of two MAM (mepirin/A5 antigen/mu receptor tyrosine phosphatase) domains, a mucin-like multiple tandem repeat, and five so called D domains that are homologous to von Willebrand D (Gao and Garbers, 1998; Hardy and Garbers, 1995; Lea et al., 2001). Zonadhesin of the house mouse (*Mus musculus* Linnaeus, 1758) is distinguished from this general structure by three MAM domains and 20 additional incomplete D domains (Gao and Garbers, 1998). It was originally assumed that MAM domains and mucin-like repeat prevent non-specific interactions between spermatozoa and the female reproductive tract (Gao and Garbers, 1998). However, an involvement of MAM domains and mucin-like repeat in the testis-internal binding of Sertoli-cells and germ cells has also been discussed (Lea et al., 2001). Whether at the level of binding to the female reproductive tract or at the level of germ cell–Sertoli cell interaction, the great relevance of both steps for successful reproduction (see e.g., Syed and Hecht, 2002; Töpfer-Petersen, 1999) points to the general importance of the specific binding of zonadhesin MAM domains.

Zonadhesin has recently been suggested to be under positive selection (Swanson et al., 2003). However, this analysis was carried out on the basis of orthologues of only four distantly related species (pig, European rabbit, house mouse, and human). Here, we elucidate the evolution of zonadhesin MAM domains 1 and 2 including seven new sequences from a representative non-human primate panel, comprising one Lemuroidea species (gray mouse lemur), three Platyrrhini (common squirrel monkey, cotton-top tamarin, and white-tufted-ear marmoset), and three Cercopithecidae (hamadryas baboon, crab-eating macaque, and rhesus monkey; for scientific names see Section 2). Based on the present dataset, we provide significant evidence that the zonadhesin domains 1 and 2 comprise positively selected sites. In particular, a previously undescribed rapidly evolving 30 amino acid (aa) fragment of MAM domain 2 is identified (MAM domain 3 in mouse), which might represent a binding-loop within the mature protein. Variability of posttranslational motifs, especially of N-myristylation motifs, is greatest across the presumptive epitope. Moreover, there is a significant overlap of positively selected codon sites and variable posttranslational motifs. Thus, specificity in the binding of MAM domains seems to be achieved by positive evolution at the level of single codon sites and posttranslational modifications, respectively.

2. Materials and methods

2.1. RNA isolation, cDNA synthesis, amplification, cloning, and sequencing

Testes were taken during diagnostic necropsies from one Lemuroidea species (gray mouse lemur, *Microcebus murinus*, Miller, 1777), three Platyrrhini (common squirrel monkey, *Saimiri sciureus* Linnaeus, 1758; cotton-top tamarin, *Saguinus oedipus* Linnaeus, 1758; and white-tufted-ear marmoset, *Callithrix jacchus* Linnaeus, 1758), and three Cercopithecidae (hamadryas baboon, *Papio hamadryas* Linnaeus, 1758; crab-eating macaque, *Macaca fascicularis* Raffles, 1821; and rhesus monkey, *Macaca mulatta* Zimmermann, 1780), and stored at -80°C . Total RNA was isolated using TRI reagent (Molecular Research Center). First-strand cDNA synthesis was carried out using the SuperScript II and ThermoScript RT (both Gibco-BRL) protocols and random primers. Subsequently, two overlapping fragments coding for zonadhesin domains MAM 1 and 2 were amplified by wax-mediated hotstart PCR. Each reaction (30 μl) contained 0.5 U *Taq* polymerase (AmpliTaq, Perkin Elmer), 15 pmol dNTPs, 10 mM Tris–HCl (pH 8.4), 50 mM KCl, 2 mM MgCl_2 , 0.1% Triton X-100, 1.2 mg/ml bovine serum albumin, and 10 pmol of each primer. For primer sequences and corresponding PCR programs see Table 1. Most PCRs were carried out with primer pairs PCR1for/rev and PCR2for/rev. As primer pair PCR2for/rev did not work out for *S. sciureus*, primer pair PCR3for/rev was used to amplify the 3' end of MAM domain 2 in this species. Primer pairs PCR1for/rev and PCR2for/rev, as well as primer PCR3rev were designed from fragments conserved between pig and human zonadhesin from GenBank (U40024 and AY046055). Primer PCR3for was inferred subsequently from primate sequences. PCR products were separated by agarose gel electrophoresis (1% SeaKem, Biozym) and UV-visualized by ethidium-bromide staining. Separated fragments were isolated, ligated into pGEM T-vector (Promega), and electroporated into *Escherichia coli* (TOP 10, Invitrogen). Positive clones were determined by blue-white screening in combination with vector-PCR, and sequenced on both strands on a Li-Cor system, using a Thermo Sequenase Cycle Sequencing Kit (Amersham). For sequences of vector- and sequencing primers and programs used see Table 1. Consensus-sequences were deduced from forward and reverse sequences of at least two clones each. Concatenation products coding for the complete MAM domain 1 and the nearly complete MAM domain 2 are available from GenBank (AY428846 for *C. jacchus*, AY428848 for *M. fascicularis*, AY428850 for *M. murinus*, AY428852 for *M. mulatta*, AY428854 for *P. hamadryas*, AY428856 for *S. sciureus*, and AY428858 for *S. oedipus*).

Table 1

PCR-primers (PCR1–3for/rev), vector-PCR primers (vec-for/rev), and sequencing primers (seq-for/rev) used in the present study

Primer	Sequence (5'–3')	Fragment length (bp)	PCR programs
PCR1for	ATGCTCCTGGACCCCAAG	~760	2 min, 94 °C; 40× (40 s, 94 °C; 40 s, 59 °C, 50 s, 72 °C); 5 min, 72 °C
PCR1rev	AA(AT)AT(AC)AGCTGCATGGGCTGTTG		
PCR2for	CTTGAG(AG)CTGAC(AG)AGTTCTCC	~410	2 min, 94 °C; 40× (40 s, 94 °C; 40 s, 51 °C; 35 s, 72 °C); 5 min, 72 °C
PCR2rev	CAG(AT)GGTTTCAGAAGG(GT)C		
PCR3for	TTCGATACCACATGTATG	~250	2 min, 94 °C; 40× (40 s, 94 °C; 40 s, 50 °C; 15 s, 72 °C); 5 min, 72 °C
PCR3rev	CT(GC)GACA(AG)GTCCCA(GT)GATTG		
vec-for	GTTTTCCAGTCACGAC	+~250 ^a	5 min, 94 °C; 25× (40 s, 94 °C; 40 s, 52 °C; 1 min, 72 °C); 5 min, 72 °C
vec-rev	GGATAACAATTCACACAGG		
seq-for	AGGGTTTTCCAGTCACGACGTT	+~260 ^a	2 min, 95 °C; 30× (30 s, 95 °C, 40 s, 61 °C, 45 s, 72 °C); for ever, 4 °C
seq-rev	GAGCGGATAACAATTCACACAGG		

^a Fragment lengths of vector-PCR and cycle sequencing = PCR-product + ~250 and ~ 260 bp, respectively.

2.2. Data analyses

2.2.1. Analysis of site-specific sequence evolution

The concatenated sequences encoding MAM domains 1 and 2 were visually aligned with orthologues from pig, house mouse, European rabbit, and human from GenBank (U40024, NM_011741, AF244982, and AF332975), using nucleotide and aa sequences as well as secondary and solvent accessibility predictions from the PredictProtein server (see Section 2.2.2). Based upon an alignment of 828 bp, we performed maximum likelihood analysis of sequence evolution, using PAML package version 13.14. Therefore, we adopted the settings from the default control file in the user's guide (among others: CodonFreq = F3×4; Nielsen and Yang, 1998; Yang, 1997; Yang et al., 2000a,b; see also Anisimova et al., 2001, 2002). An unrooted intree which reflects the widely accepted phylogeny within the sampling was given (Murphy et al., 2001; Smith and Cheverud, 2002). We ran beta null model M7 and beta& ω model M8. While M7 excludes positive selection, M8 allows for a positively selected extra site class with a ratio of non-synonymous to synonymous nucleotide substitution rate = $d_n/d_s = \omega > 1$. To avoid local optima, M8 was run three times with different initial starting values of ω (0.4, 1.0, and 1.4). To check for positive selection across sites, a likelihood ratio test (LRT) was performed (Table 2). Therefore, twice the log-likelihood difference

($2\Delta l$) of M7 and M8 was compared to critical values from a chi-square distribution with the degrees of freedom (df) equal to the difference in the number of parameters between the two models, i.e., $df = 2$. Sites suggested to be under positive selection (given minimum support from posterior probability for positive selection = $p_{(\omega>1)} > 0.5$) were subsequently identified using the naive empirical Bayes approach (NEB) implemented in PAML. We additionally carried out an analogous HyPhy analysis (version 0.99 beta; Kosakovsky Pond et al., 2004). Subsequently, only those sites that were suggested as candidates for positive selection by both PAML M8 and HyPhy M8 were further considered. To assess the possible effect of alignment uncertainties due to indels in the mouse lemur sequence, we performed PAML and HyPhy analysis (M8) excluding the respective orthologue. Furthermore, we ran the free-ratio model (PAML) to infer d_n and d_s tree-lengths for a presumably exposed fragment of 30 codons (see Sections 3 and 4). For visual documentation of rate heterogeneities of d_n and d_s across MAM domains 1 and 2, we performed moving window analysis using CRANN (Creevey and McInerney, 2003). Corresponding to the length of the probably exposed fragment, a window size of 30 codon sites was chosen (shift size: 1 codon site). CRANN provides accumulative d_n and d_s values, i.e., sums of all pairwise distances, for every particular window.

Table 2

Likelihood values and parameter estimates for MAM domains 1 and 2 of zonadhesin

Model	Mean ω	l^a	Estimates of parameters ^a
PAML M7 (beta)	0.358	-3712.639	$p = 0.319, q = 0.571$
PAML M8 (beta& ω)	0.425	-3705.517	$f_0 = 0.884 (f_1 = 0.116)$ $p = 0.640, q = 1.964, \omega = 1.804$
HyPhy M7 (beta)	0.358	-3712.572	$p = 0.316, q = 0.567$
HyPhy M8 (beta& ω)	0.419	-3705.556	$f_0 = 0.966 (f_1 = 0.034)$ $p = 0.666, q = 1.723, \omega = 2.202$

^a f_0 and f_1 , frequencies of ω site classes 0 and 1; l , log likelihood; ω , d_n/d_s rate of non-synonymous to synonymous substitution rates; p and q , beta-shape parameters.

2.2.2. Prediction of solvent accessibility and posttranslational motifs

Solvent accessibility predictions (P3acc) and motif search (PROSITE) were carried out for each orthologue with identical settings using the PredictProtein server (<http://www.embl-heidelberg.de/predictprotein/predict-protein.html>; Rost, 1996). P3acc assesses the orientation of individual aa of a protein by solvent accessibility predictions using a 3 state classification, i.e., buried, intermediary, and exposed. PROSITE identifies protein families and motifs by weighted comparison of DNA sequences with profiled database entries (Hofmann et al., 1999). The evolutionary history of posttranslational motifs was traced back by PAML reconstructions of ancestral sequences. For convenience, the predicted motifs are designated according to their first aa position (e.g., N-glycosylation motif 74, instead of N-glycosylation motif 74–77) in Sections 3 and 4.

2.2.3. Rate comparisons

We used the solvent accessibility predictions as a priori structural information, and analysed the evolution of a probably exposed 30 aa fragment between aa positions 129 and 158 (“foreground”) in comparison to aa positions 1–128 plus 159–276 (“background”). Therefore, one-tailed Fisher’s exact tests were carried out using a web based server (<http://www.physics.csbsju.edu/stats/fisher.form.html>). First, we analysed whether the ratio of positively to negatively selected aa sites was increased across the foreground compared to the background. Second, we assessed whether the foreground was distin-

guished by an accumulation of variable posttranslational motifs (Table 3). As variable we defined those motifs that were not conserved throughout the entire sampling. In detail, we counted the number of aa sites that overlap with variable motifs across the foreground, set them into relation to the number of aa sites that do not overlap with variable motifs across the foreground, and compared the resulting ratio with the respective background ratio. In a third approach, motif predictions were taken as a priori information. Subsequently, Fisher’s exact test was carried out to estimate the possible influence of positive selection at the codon level on changes in posttranslational modification (Table 4). Therefore, the ratio of positively to non-positively selected aa sites (=negatively selected plus neutrally evolving aa sites) across variable motifs was compared to the background ratio. For example, we counted 3 positively selected and 5 non-positively selected aa sites across the two variable N-glycosylation motifs 74 and 142 (see Fig. 2). The resulting foreground ratio (3/5) was then compared to the background ratio of positively to non-positively selected aa sites aside from variable N-glycosylation motifs (=24/244; Table 4).

3. Results

3.1. Analysis of sequence evolution

Based on the present data set, comprising MAM domains 1 and 2 of pig, rabbit, 1 Lemuroidea

Table 3
Fisher’s exact test statistics for the fragment between aa positions 129 and 158

Test	Foreground	Background	Fisher’s exact
For rate independence of positively selected aa sites ($p_{(\omega>1)} > 0.5$)	13/17 ^a	13/233 ^b	$p = 0.000$
For rate independence of positively selected aa sites ($p_{(\omega>1)} > 0.95$)	6/24 ^a	1/245 ^b	$p = 0.000$
For rate independence of variable posttranslational motifs (prediction)	28/2 ^c	57/189 ^d	$p = 0.000$

^a Ratio of positively to non-positively selected aa sites within aa positions 129–158.

^b Ratio of positively to non-positively selected aa sites outside aa positions 129–158.

^c Ratio of aa sites that overlap to aa sites that do not overlap with variable motifs, within aa positions 129–158.

^d Ratio of aa sites that overlap to aa sites that do not overlap with variable motifs, outside aa positions 129–158.

Table 4
Fisher’s exact test statistics for the distribution of positively selected sites^a

Predicted variable motif	Foreground ^b	Background ^c	Fisher’s exact
cAMP- and cGMP-dependent phosphorylation	Not applicable (no variable motifs)		
Protein kinase C phosphorylation	6/26	21/223	$p = 0.074$
Casein kinase II phosphorylation	7/17	20/232	$p = 0.004$
Glycosaminoglycan attachment	1/3	26/246	$p = 0.339$
N-glycosylation	3/5	24/244	$p = 0.034$
N-myristylation	11/36	16/213	$p = 0.002$
All predicted variable motifs together	19/66	8/183	$p = 0.000$

^a Only sites with $p_{(\omega>1)} > 0.5$ from PAML and HyPhy M8 have been considered.

^b Ratio of positively to non-positively selected aa sites across variable motifs.

^c Ratio of positively to non-positively selected aa sites offsite variable motifs.

representative, 3 Platyrrhini, 3 Cercopithecidae, and human, as well as the mouse orthologue (for details see Section 2), PAML analysis provides identical results irrespective of the initial starting ω (0.4, 1.0, and 1.4). Subsequent HyPhy analysis confirms the findings obtained by PAML (Table 2): estimates of beta-shape parameters p and q from beta ω model M8 point to a larger proportion of conserved and a smaller proportion of variable codon sites. Consequently, mean ω values inferred under M8 indicate only moderate conservation across all sites (0.425 and 0.419). On the other hand, lower ω values from the null model M7 (0.358) can be ascribed to the unrealistic restriction that ω is not allowed to exceed 1 (Table 2). In both programs, PAML and HyPhy, model M8 ($l \sim -3706$) has a higher log likelihood than M7 ($l \sim -3713$; Table 2). The LRT comparing M7 and M8 provides significant evidence that MAM domains 1 and 2 are positively selected ($p < 0.001$). PAML M8 and HyPhy M8 suggest a set of 27 (mean $\omega = 1.804$) and 35 (mean $\omega = 2.202$) positively selected aa sites (Table 2). Thereby, HyPhy confirms all the sites pinpointed by PAML. Support obtained from posterior probability for positive selection is significant ($p_{(\omega>1)} \geq 0.95$; PAML and HyPhy) in the case of seven aa sites, i.e., aa sites 131, 137, 143, 147, 151, 156, and 159 (Figs. 2 and 3).

3.2. Distribution of exposed aa sites, positively selected aa sites, and predicted motifs

Taken together the solvent accessibility predictions for all sequences, P3acc identifies one internal fragment of 30 aa length that is most probably exposed. The fragment spans from aa position 129 to aa position 158, and, thus, belongs to MAM domain 2 (i.e., MAM domain 3 in mouse). In contrast, upstream and downstream of the

30 aa fragment the buried character state prevails. Based on this a priori structural information, we elucidate the molecular evolution of the presumably exposed fragment. Positions 129–158 are distinguished by an already visually discernable accumulation of positively selected aa sites. Thirteen out of 27 aa sites with $p_{(\omega>1)} \geq 0.5$ from PAML and HyPhy, and 6 out of the 7 sites with $p_{(\omega>1)} \geq 0.95$ cluster within this 30 aa fragment (Figs. 2, 3). Analysis of the dataset excluding the indel-containing mouse lemur orthologue confirms the accumulation of positively selected sites across positions 129–158. A bias of the results due to alignment uncertainties can therefore be excluded. One-tailed Fisher's exact test supports the visual impression of an increased rate of positively selected aa sites between aa positions 129–158 with high significance ($p = 0.000$). This is the case irrespective of whether only the very likely candidate sites ($p_{(\omega>1)} \geq 0.95$ for positive selection) or all sites with $p_{(\omega>1)} > 0.5$ are taken into account (Table 3). In accordance, moving window analysis (CRANN) provides the maximal accumulative d_n value (=29.922) for the window covering codon positions 129–158. At the same time, d_s is not calculable around the d_n peak due to the absence of synonymous nucleotide exchanges (Fig. 1). Based on an alignment comprising solely codon positions 129 and 158, PAML (free-ratio) estimates d_n and d_s tree-lengths of 2.517 and 1.574, respectively, thus confirming positive selection for the 30 aa fragment.

According to motif scan (PROSITE, per eye), phosphorylation, glycosaminoglycan attachment, N-glycosylation, and N-myristylation contribute to the posttranslational modification of the zonadhesin MAM domains. The single predicted cAMP- and cGMP-dependent protein kinase phosphorylation motif

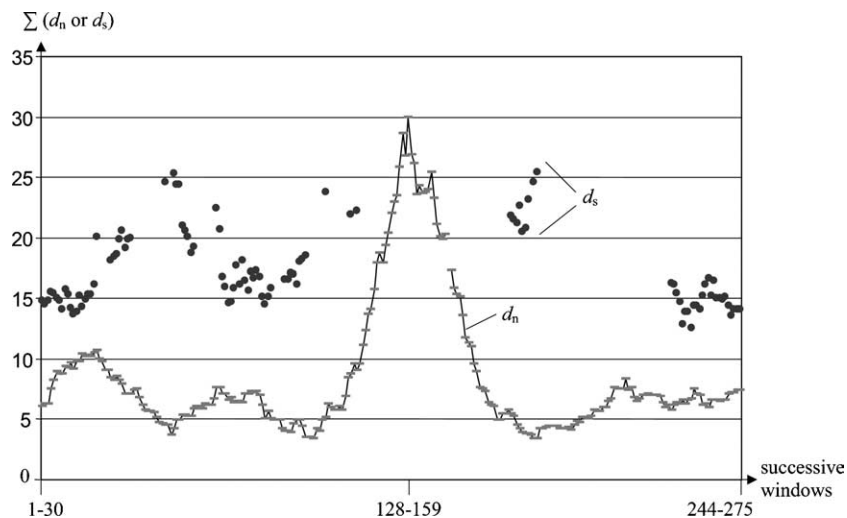


Fig. 1. Results of moving window analysis carried out with CRANN. Abscissa shows the successive windows of 30 codon sites (shift size: 1 codon site). The ordinate indicates accumulative d_n and d_s values, i.e., the respective sums of all pairwise sequence comparisons for each window. d_n peaks around a presumed epitope between aa positions 129 and 158. Around the d_n peak d_s was not calculable due to the absence of synonymous exchanges.

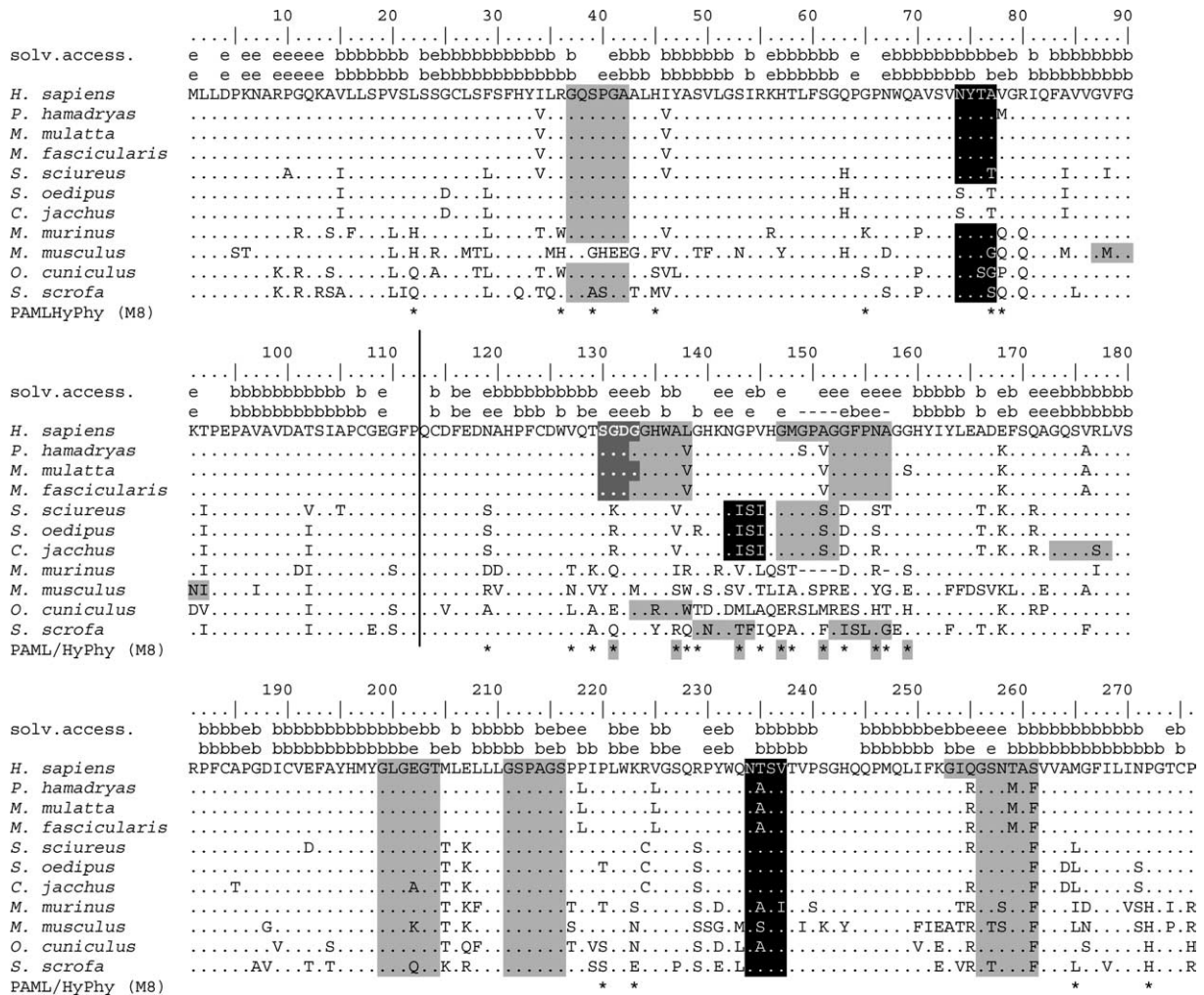


Fig. 2. Distribution of predicted N-myristylation, N-glycosylation and glycosaminoglycan attachment motifs across the near-complete aa alignment of zonadhesin MAM domains 1 and 2. The vertical bar between aa positions 112 and 113 separates MAM domain 1 and 2 (i.e. MAM 2 and MAM 3 in mouse). Most changes are located within a 30 aa spanning fragment of MAM domain 2 (positions 129-158) that presumably belongs to a binding region. Only the reference sequence (*H. sapiens*) is shown in detail. Dots in the alignments indicate congruencies with the *H. sapiens* orthologue, hyphens indicate gaps. Lower case letters above the alignment indicate solvent accessibility predictions (solv. access.; e = exposed, space character = intermediary, b = buried) for single aa sites of *H. sapiens* (upper line) and *M. murinus* (lower line). Amino acid sites suggested to be under positive selection by PAML M8 and HyPhy M8 are highlighted by an asterisk below the alignment. Asterisks with shaded background (*) indicate aa sites with significant support from posterior probability ($p_{(\omega>1)} \geq 0.95$). Consensus sequences: G^[EDRKHPFYW]. {2}[STAGCN]^[P] N-myristylation motif; N^{[P][ST][P]} N-glycosylation motif; SG.G glycosaminoglycan attachment motif.

(motif 55 in Fig. 3) is conserved throughout the entire sampling. In contrast, there are variable (i.e., motifs that are not conserved throughout the entire sampling) protein kinase C phosphorylation, casein kinase II phosphorylation, glycosaminoglycan attachment, N-glycosylation, and N-myristylation motifs (Figs. 2 and 3). At the same time, however, myristylation is known from the N-terminus only, thus questioning the functional relevance of the internal N-myristylation motifs identified here. Irrespectively of these considerations, Fisher's exact test provides significant evidence for an accumulation of variable predicted motifs between aa positions 129 and 158, whether focusing on single motifs ($p = 0.000$ to $p = 0.006$; not shown) or on all motifs together ($p = 0.000$; Table 3).

3.3. Motif evolution

A combined survey of positively selected aa sites and variable motifs shows that positively selected aa sites evoke changes in (predicted) posttranslational modification. Thus, aa replacements along 12 positively selected aa sites are directly "responsible" for changes in the motif patten, i.e., aa sites 127, 129, 131, 137, 139, 143, 147, 151, 153, 156, 220, and 223. Eight out of these 12 aa sites fall again between aa positions 129 and 158, thus emphasizing the special character of this fragment (Figs. 2 and 3). Taken all motifs together, Fisher's exact test suggests an increased rate of positively selected sites across variable motifs compared to the background ratio ($p = 0.000$; Table 4). Based on single motifs, this

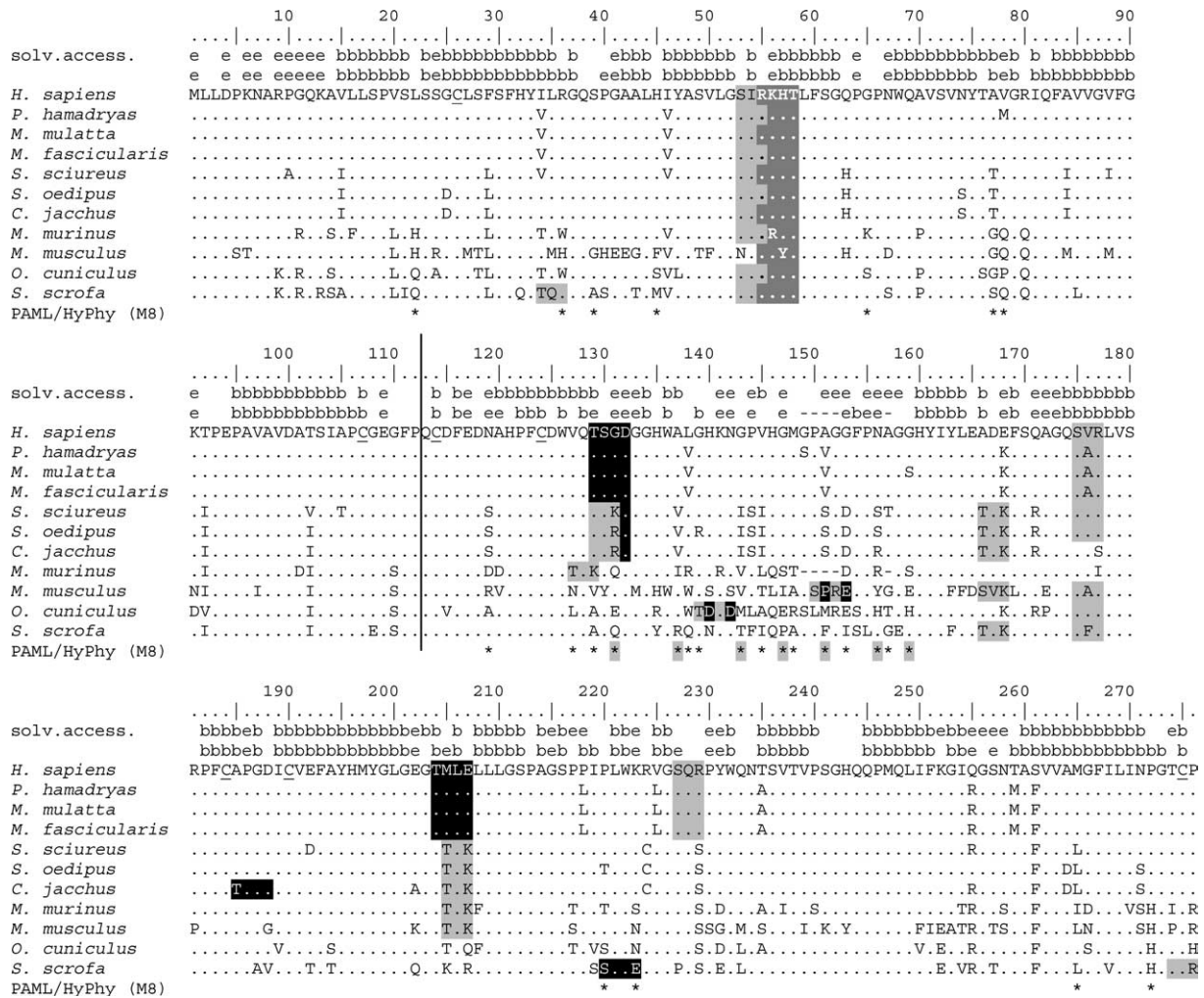


Fig. 3. Distribution of predicted protein kinase C, casein kinase II and cAMP- and cGMP-dependent protein kinase phosphorylation motifs across the near-complete aa alignment of zonadhesin MAM domains 1 and 2. Only the reference sequence (*H. sapiens*) is shown in detail. The striation from aa positions 139 to 142 (ID₁D) and 150 to 153 (SPRE) refers to overlapping protein kinase C and casein kinase II phosphorylation motifs. For further details see legend of Fig. 2. Consensus sequences: [ST].[RK] protein kinase C phosphorylation motif; [ST].[2][DE] casein kinase II phosphorylation motif; [RK]{2}.[ST] cAMP- and cGMP-dependent protein kinase phosphorylation motif.

phenomenon is most obvious in N-myristylation ($p = 0.002$), followed by casein kinase II phosphorylation ($p = 0.004$), and N-glycosylation ($p = 0.034$; Table 4). On the other hand, Fisher's exact test cannot exclude equal rates of positively selected aa sites across background and variable glycosaminoglycan attachment motifs ($p = 0.339$). Neither, Fisher's exact test can rule out equal rates across background and protein kinase C phosphorylation motifs ($p = 0.074$; Table 4). Actually, the coincidence of positively selected codon sites and changes in predicted posttranslational motif pattern might even be underrated by the present approach. For example, the loss of N-myristylation motif 37 in mouse, the loss of N-glycosylation motif 74 in the lineage to *S. oedipus* and *C. jacchus*, and the gain of N-myristylation motifs 87 and 173 in mouse and *C. jacchus*, respectively, are also linked to aa exchanges (Fig. 2). However, synonymous nucleotide exchanges

prevail throughout the sampling at the "responsible" codon positions (41, 74, 91, and 177), and support for positive selection is consequently low (<0.5).

Given PAML reconstructions of ancestral sequences and the probable phylogeny within the sampling (Murphy et al., 2001; Smith and Cheverud, 2002), N-myristylation motifs 37, 199, 211, and 256 as well as N-glycosylation motifs 74 and 234 were already present in the last common ancestor of Primates, Glires (Rodentia + Lagomorpha) and pig, and thus of the Boreoeutheria (see Murphy et al., 2001; Figs. 2 and 4). In analogy, presence of phosphorylation motifs 53, 55, 175, and 205 can be assumed for the boreoeutherian stem species (Figs. 3 and 5). The other posttranslational motifs identified in the present analysis most probably evolved subsequently on the lineages to the recent species, partly in convergence to each other. A twofold independent emergence of N-myristylation motif 133,

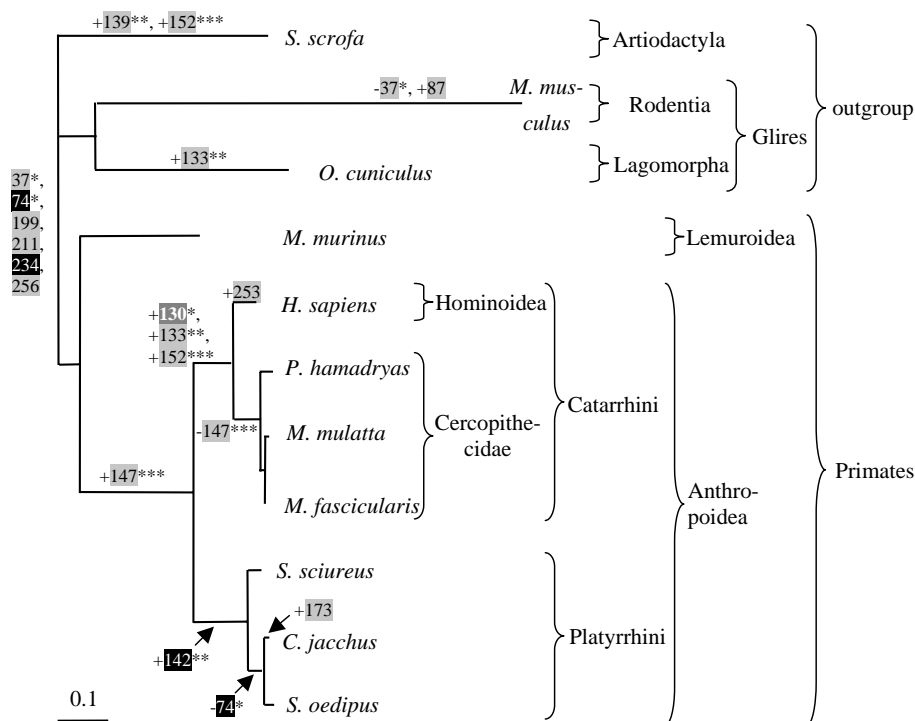


Fig. 4. Variability of predicted **N-myristylation**, **N-glycosylation**, and **glycosaminoglycan attachment** motifs in time plotted on a gene tree of zonadhesin MAM domains 1 and 2. The hypotheses are based on PAML reconstructions of ancestral sequences. Motifs are designated according to their first aa position. Numbers at the base of the tree (37, 74 etc.) refer to motifs that can be assumed for the last common ancestor of all species considered. Positive numbers along branches (+133, +142, +130 etc.) indicate motif gain, negative numbers (–37 etc.) motif loss. The number of asterisks behind each motif refers to the number of positively selected sites that overlap with the respective motif. Branch lengths are given as nucleotide substitutions per codon. Please see Fig. 2 for comparison.

for example, can be assumed for the lineage leading to *O. cuniculus* and Catarrhini (Figs. 2 and 4). Beyond this, the evolution of zonadhesin MAM domains is characterized not only by gaining but also by losing motifs, and thus by changes in general. N-myristylation motif 147, for instance, was probably lost on the lineage leading to Cercopithecidae as suggested by PAML reconstructions of ancestral sequences (Figs. 2 and 4). Taking together all the predicted changes (Figs. 4 and 5), a taxon-specific variability of posttranslational modification can be suggested for zonadhesin MAM domains. Considering the increased rate of positively selected aa sites across variable posttranslational motifs (Table 4), this taxon-specific variability might have been promoted by positive selection on the aa level (see motifs highlighted by an asterisk in Figs. 4 and 5).

4. Discussion

4.1. Sequence evolution and evolution of motifs

Positive selection has been reported previously for zonadhesin in general (see Swanson et al., 2003). Thus, the present evidence for positive evolution of zonadhesin MAM domains fits well into the picture (Table 2). In the

present analysis, 27 aa sites (PAML and HyPhy M8) were suggested for positive selection while previous study identified 6 sites across the homologue fragment (Swanson et al., 2003). Apparently, the comparison of only four distantly related species (pig, rabbit, mouse, and human) impaired the recognition of positively selected sites in the previous approach. The present broadening of the sampling led moreover to higher support for positive selection from posterior probability ($p_{(\omega>1)}$). While $p_{(\omega>1)}$ was <0.90 for all aa sites identified in the previous study (Swanson et al., 2003), a total of 7 sites got significant support from posterior probability ($p_{(\omega>1)} \geq 0.95$; PAML and HyPhy) in the present approach (Figs. 2 and 3). The fact that aa positions 129–158 are distinguished by a probably exposed orientation, enhanced evolution, and increased variability in predicted posttranslational modification has not been described before. The 30 aa sites are moreover characterized by the presence of the sole indels across the entire alignment (see gray mouse lemur sequence in Figs. 2 and 3). Amino acid positions 129–158 might thus form a loop-like structure in the mature protein, where deletions or insertions do not automatically rupture the protein structure. However, these conclusions are drawn from predictions. Future studies have to provide a final proof, whether aa positions 129–158 are really exposed in the mature protein.

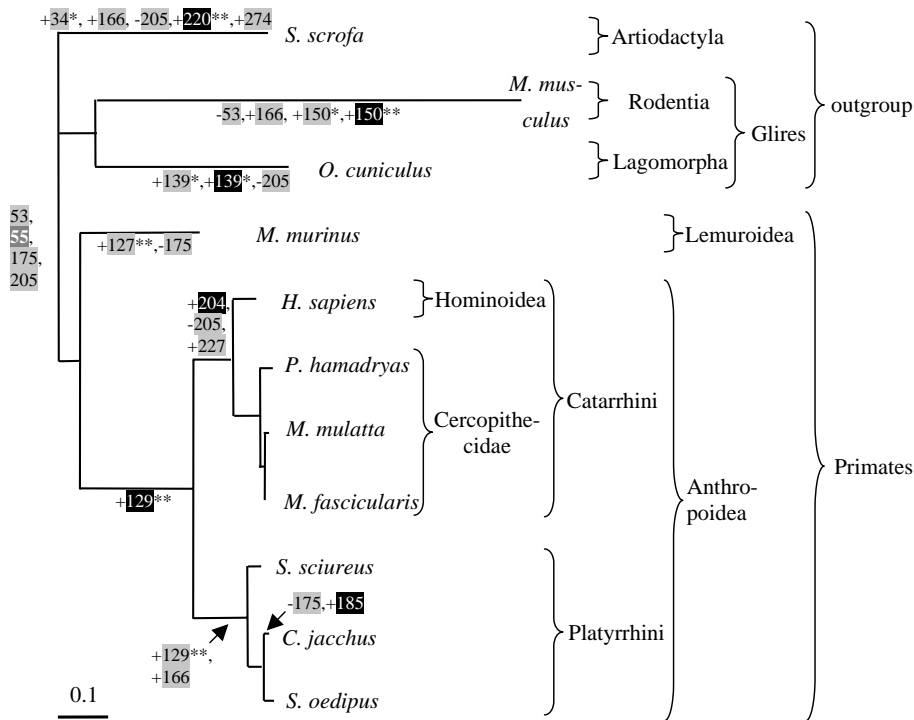


Fig. 5. Variability of predicted protein kinase C, casein kinase II and cAMP- and cGMP-dependent protein kinase motifs in time plotted on a gene tree of zonadhesin MAM domains 1 and 2. The hypotheses are based on PAML reconstructions of ancestral sequences. Labelling of motifs refers to the first aa position. Numbers at the base of the tree (53, 55 etc.) point at motifs that were already realized in the last common ancestor of all species considered. Positive numbers along branches (+34, +220 etc.) refer to the gain of motifs. Negative numbers (–175 etc.) indicate the loss of motifs. Each asterisk represents a positively selected site that overlaps with the preceding motif. Branch lengths refer to nucleotide substitutions per codon. Please compare the data to Fig. 3.

Although it is not clear to what extent predicted motifs represent functional motifs, the present findings suggest that the variability of posttranslational modifications, especially of N-myristylation, casein kinase II phosphorylation, and N-glycosylation, is enhanced by positive selection at the aa levels. On a site-specific level, positive selection has been reported for a variety of sex-related genes. For marine molluscs, for example, it has been shown in detail that the primary structure of the sperm ligand lysin evolves in adaptation to a rapidly changing receptor on the egg surface (Swanson et al., 2001a,b; Swanson and Vacquier, 1998; for a review see Swanson and Vacquier, 2002). Moreover, length polymorphisms such as deletions in case of the mouse lemur sequence (see Fig. 2) have previously been shown to be a consequence of positive selection (Podlaha and Zhang, 2003; see also, e.g., Metz and Palumbi, 1996). To our knowledge, however, the present study yields first evidence that divergence in sperm-ligands might also be promoted by positive selection at the level of posttranslational modification (see Figs. 4 and 5).

4.2. Functional considerations

Accumulations of positively selected sites have previously been reported for, e.g., antigen-binding regions of

proteins like HLA and MHC I (Hughes et al., 1990; Suzuki and Gojobori, 1999) or the disintegrin-like domain of sperm-protein fertilin- β (Civetta, 2003; Torgerson et al., 2002; see also Swanson et al., 2003; Zhu et al., 2000). The enhanced evolution of the 30 aa fragment between positions 129 and 158 might thus represent an adaptation of zonadhesin MAM domains to an unknown binding partner. Posttranslational modifications might affect such a binding. Myristates, for example, are generally known for their role in anchoring proteins to double membranes and other macro-molecules (e.g., Bouamr et al., 2003; Breitenlechner et al., 2004; Loomis et al., 2001; Risinger et al., 1996; Sowadski et al., 1996). However, to gain functionality the internal myristylation motifs predicted here have to become N-terminal first, what might be achieved by posttranslational cleavage. Glycosaminoglycans have also been reported to contribute to the binding of higher molecular structures (e.g., Skidmore et al., 2004). Finally, N-glycosylation and phosphorylation modulate activity in diverse proteins (see, e.g., Bruce et al., 2004; Otto et al., 2004). Given the taxon-specific differences in predicted posttranslational modification (Figs. 2–5), binding specificity of the sperm-ligand zonadhesin might be achieved not only via active aa sites that interact directly with a complementary receptor, but also by posttranslational

modifications that modulate activation, conformation, and binding properties. In particular, aa positions 128–159 seem to be involved in such a modulation of binding properties.

Recently, it has been reported an association between positively selected aa replacements and the emergence of new or alternative myristylation and phosphorylation sites for presumably immune relevant proteins of HIV-1 (De Oliveira et al., 2004). Another study has shown that HIV-1 escapes neutralization antibodies by an adaptively evolving ‘glycan-shield’ (Wei et al., 2003). Similarly, the posttranslational modifications of zonadhesin MAM domains might not only be important for binding but also for evasion of binding. As insemination elicits female immune response (see Johansson et al., 2004), such a masking effect of posttranslational modification could make sense also in the case of sperm ligands. This especially, if one assumes differential posttranslational modification that endows single individuals with a variety of ligands. However, it is not clear whether zonadhesin MAM domains reach the female reproductive tract at all. They might as well be cleaved off from the spermatozoan surface before insemination (Lea et al., 2001). Irrespective of the actual driving force, the observed changes in posttranslational modification of MAM domains probably modify the binding properties of zonadhesin and its decomposition products. Furthermore, positive selection seems to promote changes in the posttranslational modification. As the phenomenon is realized in a broad taxonomic range spanning from viruses to humans, it seems to represent a general biological principle in the evolution of interacting proteins.

5. Conclusions

1. Present analysis (PAML and HyPhy, M8) identifies a set of 27 aa sites under positive selection across zonadhesin MAM domains. Given a 95% threshold, significant support from posterior probability is provided for aa positions 131, 137, 143, 147, 151, 156, and 159. Solvent accessibility prediction suggests an exposed orientation of a 30 aa fragment spanning aa positions 129–158 that belongs to MAM domain 2 (i.e., MAM domain 3 in mouse). Fisher’s exact test indicates an accumulation of positively selected aa sites across this fragment. Moving window analysis (CRANN) and tree-lengths ($d_n = 2.517$, and $d_s = 1.574$; PAML free-ratio) corroborate the accelerated evolution of the 30 aa fragment.

2. Motif search suggests that phosphorylation, glycosaminoglycan attachment, N-glycosylation, and N-myristylation contribute to the posttranslational modification of zonadhesin MAM domains. The 30 aa fragment between aa positions 129 and 158 displays an enhanced variability in predicted posttranslational

motifs. The rate of positively selected sites is significantly enhanced across predicted posttranslational motifs that are not conserved throughout the entire sample. The phenomenon is most apparent in the case of N-myristylation, followed by casein kinase II phosphorylation, and N-glycosylation. We conclude that changes in the predicted pattern of posttranslational modification are promoted by positive selection at the level of single aa sites.

3. The driving force behind the observed pattern of sequence and motif evolution remains to be elucidated. In analogy to, e.g., MHC I, zonadhesin MAM domains might evolve in adaptation to a so far unknown binding partner. Alternatively, the outlined findings might represent adaptations to evade female immune response. Regardless of the actual driving force behind the observed patterns of evolution, a correlation between positively selected aa sites and changes in posttranslational modification seems to represent a general biological principle in the evolution of interacting proteins.

Acknowledgments

We are indebted to Dr. C. Roos and the Departments of Reproductive Biology, Neurobiology, and Veterinary Medicine and Primate Husbandry at the German Primate Center Göttingen, for providing most of the material. Thanks go to Dr. M. Perret (Laboratoire d’Ecologie, Brunoy/France) and Prof. B. Brenig and his group (Veterinary Medicine, University of Göttingen) for the mouse lemur and pig material, respectively. Last but not least, the support of Dr. Sergei Kosakovsky Pond from the Antiviral Research Center at the University of California, San Diego in statistical issues is gratefully acknowledged. This study was supported by the German Research Foundation (HE 3487/1-1).

References

- Anisimova, M., Bielawski, J.P., Yang, Z., 2001. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Mol. Biol. Evol.* 18, 1585–1592.
- Anisimova, M., Bielawski, J.P., Yang, Z., 2002. Accuracy and power of Bayes prediction of amino acid sites under positive selection. *Mol. Biol. Evol.* 19, 950–958.
- Bouamr, F., Scarlata, S., Carter, C., 2003. Role of myristylation in HIV-1 Gag assembly. *Biochemistry* 42, 6408–6417.
- Breitenlechner, C., Engh, R.A., Huber, R., Kinzel, V., Bossemeyer, D., Gassel, M., 2004. The typically disordered N-terminus of PKA can fold as a helix and project the myristylation site into solution. *Biochemistry* 44, 7743–7749.
- Bruce, Y.M., Mikolajczak, S.A., Yoshida, T., Yoshida, R., Kelvin, D.J., Ochi, A., 2004. CD28 T cell costimulatory receptor function is negatively regulated by N-linked carbohydrates. *Biochem. Biophys. Res. Commun.* 317, 60–67.

- Civetta, A., 2003. Positive selection within sperm–egg adhesion domains of fertilin: an ADAM gene with a potential role for fertilization. *Mol. Biol. Evol.* 20, 21–29.
- Creevey, C.J., McInerney, J.O., 2003. CRANN: detecting adaptive evolution in protein-coding DNA sequences. *Bioinformatics* 19, 1726.
- De Oliveira, T., Salemi, M., Gordon, M., Vandamme, A.M., Van Rendsburg, E.J., Engelbrecht, S., Coovadia, H.M., Cassol, S., 2004. Mapping sites of positive selection and amino acid diversification in the HIV genome: An alternative approach to vaccine design? *Genetics* 167 1047–1058.
- Dorus, S., Evans, P.D., Wyckoff, G.J., Choi, S.S., Lahn, B.T., 2004. Rate of molecular evolution of the seminal protein gene SEMG2 correlates with levels of female promiscuity. *Nat. Genet.* 36, 1326–1329.
- Gao, Z., Garbers, D.L., 1998. Species diversity in the structure of zonadhesin, a sperm-specific membrane protein containing multiple cell adhesion molecule-like domains. *J. Biol. Chem.* 273, 3415–3421.
- Hardy, D.M., Garbers, D.L., 1994. Species-specific binding of sperm proteins to the extracellular matrix (zona pellucida) of the egg. *J. Biol. Chem.* 269, 19000–19004.
- Hardy, D.M., Garbers, D.L., 1995. A sperm membrane protein that binds in a species-specific manner to the egg extracellular matrix is homologous to von Willebrand Factor. *J. Biol. Chem.* 270, 26025–26028.
- Hofmann, K., Bucher, P., Falquet, L., Bairoch, A., 1999. The PROSITE database, its status in 1999. *Nucleic Acids Res.* 27, 215–219.
- Hughes, A.L., Ota, T., Nei, M., 1990. Positive Darwinian selection promotes charge profile diversity in the antigen-binding cleft of class I major histocompatibility complex genes in mammals. *Mol. Biol. Evol.* 7, 491–514.
- Johansson, M., Bromfield, J.J., Jasper, M.J., Robertson, S.A., 2004. Semen activates the female immune response during early pregnancy in mice. *Immunology* 112, 290–300.
- Kosakovsky Pond, S.L., Frost, S.D.W., Muse, S.V., 2004. HyPhy: hypothesis testing using phylogenies. *Bioinformatics Advance Access*. Available from: [doi:10.1093/bioinformatics/bti079](https://doi.org/10.1093/bioinformatics/bti079).
- Lea, I.A., Sivashanmugam, P., O'Rand, M.G., 2001. Zonadhesin: characterization, localization, and zona pellucida binding. *Biol. Reprod.* 65, 1691–1700.
- Loomis, J.S., Bowzard, J.B., Courtney, R.J., Wills, J.W., 2001. Intracellular trafficking of the UL11 tegument protein of herpes simplex virus type 1. *J. Virol.* 75, 12019–12209.
- Metz, E.C., Palumbi, S.R., 1996. Positive selection and sequence rearrangements generate extensive polymorphism in the gamete recognition protein binding. *Mol. Biol. Evol.* 13, 397–406.
- Murphy, W.J., Eizirik, E., O'Brien, S.J., Madsen, O., Scally, M., Douady, C.J., Teeling, E., Ryder, O.A., Stanhope, M.J., de Jong, W.W., Springer, M.S., 2001. Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science* 294, 2348–2351.
- Nielsen, R., Yang, Z., 1998. Likelihood models for detecting positively selected amino acid sites and application to the HIV-1 envelope gene. *Genetics* 148, 929–936.
- Otto, V.I., Schlürpf, T., Folkers, G., Cummings, R.F., 2004. Sialylated complex-type N-glycans enhance the signalling activity of soluble intercellular adhesion molecule-1 in mouse. *J. Biol. Chem.* 279, 35201–35209.
- Podlaha, O., Zhang, J., 2003. Positive selection on protein-length in the evolution of a primate sperm ion channel. *Proc. Natl. Acad. Sci. USA* 100, 12241–12246.
- Risinger, M.A., Korsgren, C., Cohen, C.M., 1996. Role of N-myristylation in targeting of band 4.2 (pallidin) in nonerythroid cells. *Exp. Cell Res.* 229, 421–431.
- Rooney, A.P., Zhang, J., 1999. Rapid evolution of a primate sperm protein: relaxation of functional constraint or positive Darwinian selection? *Mol. Biol. Evol.* 16 706–710.
- Rost, B., 1996. PredictProtein. *Methods Enzymol.* 266, 525–539.
- Skidmore, M.A., Patey, S.J., Thanh, N.T., Fernig, D.G., Turnbull, J.E., Xates, E.A., 2004. Attachment of glycosaminoglycan oligosaccharides to thiol-derivatised gold surfaces. *Chem. Commun.* 23, 2700–2701.
- Smith, R.J., Cheverud, J.M., 2002. Scaling of sexual dimorphism in body mass: a phylogenetic analysis of Rensch's Rule in Primates. *Int. J. Primatol.* 23, 1095–1135.
- Sowadski, J.M., Ellis, C.A., Madhusudan, 1996. Detergent binding to unmyristylated protein kinase A—structural implications for the role of myristate. *J. Bioenerg. Biomembr.* 28, 7–12.
- Suzuki, Y., Gojobori, T., 1999. A method for detecting positive selection at single amino acid sites. *Mol. Biol. Evol.* 16, 1315–1328.
- Swanson, W.J., Aquadro, C.F., Vaquier, V.D., 2001a. Polymorphism in abalone fertilization proteins is consistent with the neutral evolution of the egg's receptor for lysin (VERL) and positive Darwinian selection of sperm lysin. *Mol. Biol. Evol.* 18, 376–383.
- Swanson, W.J., Nielsen, R., Yang, Q., 2003. Pervasive adaptive evolution in mammalian fertilization proteins. *Mol. Biol. Evol.* 20, 18–20.
- Swanson, W.J., Vacquier, V.D., 1998. Concerted evolution in an egg receptor for a rapidly evolving sperm protein. *Science* 281, 710–712.
- Swanson, W.J., Vacquier, V.D., 2002. The rapid evolution of reproductive proteins. *Nat. Rev. Genet.* 3, 137–144.
- Swanson, W.J., Yang, Z., Wolfner, M.F., Aquadro, C.F., 2001b. Positive selection drives the evolution of several female reproductive proteins in mammals. *Proc. Natl. Acad. Sci. USA* 98, 2509–2514.
- Syed, V., Hecht, N.B., 2002. Disruption of germ cell–Sertoli cell interactions leads to spermatogenic defects. *Mol. Cell Endocrinol.* 186, 155–157.
- Töpfer-Petersen, E., 1999. Molecules on the sperm's route to fertilization. *J. Exp. Zool.* 285, 259–266.
- Torgerson, D.G., Kulathinal, R.J., Singh, R.S., 2002. Mammalian sperm proteins are rapidly evolving: evidence of positive selection in functionally diverse genes. *Mol. Biol. Evol.* 19, 1973–1980.
- Torgerson, D.G., Singh, R.S., 2003. Sex-linked mammalian sperm proteins evolve faster than autosomal ones. *Mol. Biol. Evol.* 20, 1705–1709.
- Wei, X., Decker, J.M., Wang, S., Hui, H., Kappes, J.C., Wu, X., Salazar-Gonzalez, J.F., Salazar, M.G., Kilby, J.M., Saag, M.S., Komarova, N.L., Nowak, M.A., Hahn, B.H., Kwong, P.D., Shaw, G.M., 2003. Antibody neutralization and escape by HIV-1. *Nature* 422, 212–307.
- Wyckoff, G.J., Wang, W., Wu, C.I., 2000. Rapid evolution of male reproductive genes in the descent of man. *Nature* 403, 304–309.
- Yang, Z., 1997. PAML: A program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. BioSci.* 13, 555–556 (<http://abacus.gene.ucl.ac.uk/software/paml.html>).
- Yang, Z., Nielsen, R., Goldman, N., Krabbe Pedersen, A.-M., 2000a. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155, 431–449.
- Yang, Z., Swanson, W.J., Vacquier, V.D., 2000b. Maximum-likelihood analysis of molecular adaptation in abalone sperm lysin reveals variable selective pressures among lineages and sites. *Mol. Biol. Evol.* 17, 1446–1455.
- Zhu, X., Bansal, N.P., Evans, J.P., 2000. Identification of key functional amino acids of the mouse fertilin beta (ADAM2) disintegrin loop for cell–cell adhesion during fertilization. *J. Biol. Chem.* 275, 7677–7683.