

# The molecular evolution of sperm zonadhesin

HOLGER HERLYN\* and HANS ZISCHLER

*Institute of Anthropology, Johannes Gutenberg University, Mainz, Germany*

**ABSTRACT** Based on pioneering work of Hardy and Garbers, zonadhesin has become one of the best studied sperm ligands in boreoeutherian mammals, both from a biochemical and evolutionary perspective. Zonadhesin is a mosaic-type protein that localizes to the apical head of spermatozoa. In pig, cattle, rabbit and primates, zonadhesin precursor essentially consists of two or three MAM (meprin/A5 antigen/mu receptor tyrosine phosphatase) domains, one mucin-like domain, one incomplete and four complete D domains (homologous to vWFD). Mouse zonadhesin is distinguished from this general pattern by 20 extra partial D3 domains. While concerted evolution drives the divergence of the mucin-like domain in the ortholog comparison, MAM and D domains mainly diverge under the influence of drift and positive selection, both in the paralog and ortholog comparison. As can be seen particularly well within a putative binding region in the most C-terminal MAM domain, positive selection not only causes amino acid exchanges, but also promotes changes in the pattern of predicted posttranslational modification. Moving window and correlation analyses of sequence evolution and sexual body dimorphism further suggest that sexual selection, especially sperm competition, drives zonadhesin divergence. However, considering its *zona pellucida* avidity, female cryptic choice might as well contribute to zonadhesin evolution. Despite the general tendency for divergence of zonadhesin, conservation by negative selection dominates the evolution of most codon sites. In accordance, the distribution of EGF (epidermal growth factor)-like motifs, DP-doublets, single cysteines and CGLC motifs suggests a wide conservation of processing, folding and oligomerization of zonadhesin in pig, rabbit and primates.

**KEY WORDS:** *selection, mucin, von Willebrand factor, tandem repeat, sperm binding, zona pellucida*

## Introduction

One of the key steps in the fertilization cascade is the recognition, binding, and penetration of the egg zona pellucida (ZP) by the mature spermatozoon. From an evolutionary perspective the genes and proteins involved are of particular interest as they determine species-specific gamete recognition, which highlights their role in the evolution and maintenance of biological diversity. From the medical point of view the proteins implicated in ZP interaction are highly attractive candidates for the development of contraceptive vaccines (Naz *et al.*, 2000). However, molecular studies on sperm-ZP interaction are generally hampered by the comparatively small fraction of mature spermatozoa in ejaculates (Hardy and Garbers 1995). A great leap forward in this field is represented by the work of Hardy and Garbers (1994) who concentrated proteins from large quantities of pig (*Sus scrofa*) sperm membranes through an affinity matrix made of porcine ZP. One of the isolated proteins that turned out to bind species-specifically to ZP (Hardy and Garbers 1994) was termed

zonadhesin (ZAN), thus referring to its ZP binding avidity (Hardy and Garbers 1995). Based on sequences from cloned cDNA fragments, Hardy and Garbers (1995) characterized pig ZAN precursor as a novel mosaic-type protein of 2,447 amino acids (aa) containing five tandem repeats homologous to von Willebrand factor (vWF) D. The same study provided first time evidence based on Northern blotting and *in situ* hybridisation that pig ZAN is specifically expressed in testis, especially in haploid spermatids (Hardy and Garbers 1995). In the following years, Hardy and

---

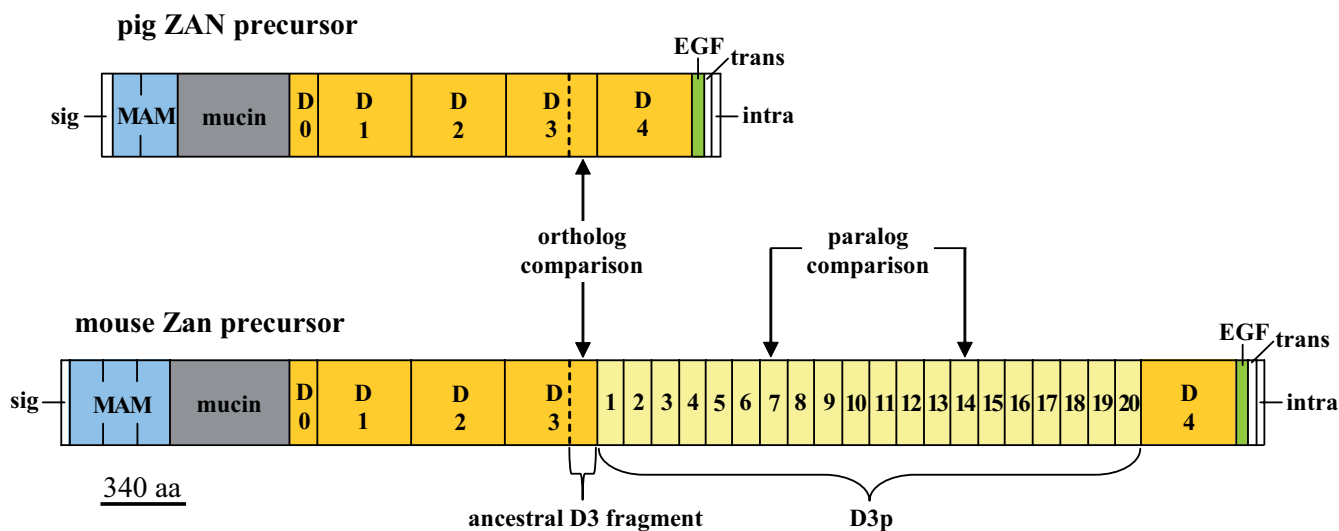
*Abbreviations used in this paper:* D0-4, D domains 1-4; D3p1-20, mouse partial D3 domains 1-20; dn, nonsynonymous substitution rate; ds, synonymous substitution rate; EGF, epidermal growth factor; intra, intracellular segment; MAM, meprin/A5 antigen/mu receptor tyrosine phosphatase domain; ML, maximum likelihood; MRCA, most recent common ancestor; ORF, open reading frame; pnc, conservative nonsynonymous substitution rate; pnr, radical nonsynonymous substitution rate; sig, signal peptide; trans, transmembrane segment; ZAN; zonadhesin; Zan, mouse zonadhesin; ZP, zona pellucida; vWF, von Willebrand factor.

\*Address correspondence to: Holger Herlyn. Institute of Anthropology, Johannes Gutenberg University, Mainz, Germany. Fax: +49-6131-3923799.  
e-mail: herlyn@uni-mainz.de

Published online: 4 July 2008

0214-6282/2008/\$35.00

© UBC Press  
Printed in Spain



**Fig. 1. Schematic depiction of the domain architecture of pig (ZAN) and mouse zonadhesin (Zan) precursor.** Pig ZAN is a mosaic-type protein comprising (from N- to C-terminus) signal peptide (sig), two meprin/A5 antigen/mu receptor tyrosine phosphatase domains (MAM), mucin-like domains (mucin), one incomplete and four complete D domains (D0-4), EGF-like domain (EGF), transmembrane segment (trans), and a basic segment (intra). Mouse Zan contains one additional MAM and 20 partial D3 domains (D3p1-20), derived from the C-terminal 120 amino acids of D3 (ancestral D3 fragment). Only one out of a total of 190 possible pairwise sequence comparisons among D3p1-20 has been depicted. The C-termini of pig and mouse D3 were compared within ortholog analysis (see legend Fig. 7 for all species included). The domain architecture was redrawn from Gao and Garbers (1998).

colleagues used immunoprecipitation and other methods to unravel the activation of pig ZAN by processing, glycosylation and oligomerization (Hickox *et al.*, 2001; Bi *et al.*, 2003). Beyond this, Hardy and collaborators were first to document that transcripts of human ZAN undergo alternative splicing into six variants (see AY046055 at NCBI; see also Gasper and Swanson 2006). The extensive genetic and biochemical characterization of ZAN/ZAN presented by Hardy and colleagues provided an invaluable basis for the follow-up studies of the last years. The present review aims at giving an account of the structure, localization, activation, function and, last but not least, evolution of ZAN/ZAN, as it has been achieved since the pioneering work of Hardy and Garbers in 1994.

### ZAN - gene structure and testis specific expression in Boreoeutheria

Due to the ongoing advances in chromosome and genome sequencing we know more and more about the gene structure of ZAN in diverse species. The 48 exons of human ZAN, for instance, localize to chromosome 7q22 (Glöckner *et al.*, 2001; see also Gao *et al.*, 1997). They span ~64 kb and contain an open reading frame of 2,811 aa (see transcript ENST00000349350, release 40). To give another example, puffer fish (*Takifugu rubripes*) Zan contains 47 exons that code for a protein of 2,525 aa (Hunt *et al.*, 2005). The two examples reflect the currently known phylogenetic range of ZAN, thus illustrating its considerable phylogenetic age of at least 450 million years (see Gilligan *et al.*, 2002 for the human - puffer fish split).

Northern blotting and *in situ* hybridization suggest that pig ZAN is expressed only in testis (Hardy and Garbers 1995; see also expression profile under NCBI accession Ssc.14486). Expression profiles from NCBI confirm a testis specific transcription of

human ZAN (UniGene accession Hs.307004). Mouse Zan, on the other hand, might be expressed exclusively in testis (Gao and Garbers 1998) or in diverse organs such as testis, brain, kidney, and eye (Mm.7984). Irrespective of possible additional organs of expression in species such as mouse, all the investigated mammalian species exhibit ZAN expression in the testis. This is different from the situation in zebrafish (*Danio rerio*) where expression is restricted to the gut according to RT-PCR and Northern blotting (Hunt *et al.*, 2005). As puffer fish Zan is also expressed in the gut, Hunt *et al.*, (2005) speculate that the expression of ZAN in testis and its implication in fertilization represents a younger evolutionary novelty of Mammalia and Boreoeutheria, respectively.

### ZAN is a mosaic protein containing MAM, mucin and vWF related D domains

The open reading frame (ORF) of pig ZAN codes for a signal peptide, two MAM (meprin/A5 antigen/mu receptor tyrosine phosphatase) domains, one mucin-like domain, one partial and four complete D (homologous to vWFD) domains and a C-terminus composed of EGF (epidermal growth factor)-like domain, transmembrane segment and a short basic tail (Fig. 1; Hardy and Garbers 1995). Genomic and cDNA data have now confirmed the characteristic succession of MAM, mucin-like, and D domains for several other boreoeutherian mammals including European rabbit (*Oryctolagus cuniculus*), dog (*Canis familiaris*), cattle (*Bos taurus*), and diverse primates (for cDNA data, see Lea *et al.*, 2001; Herlyn and Zischler 2005a,b; Gasper and Swanson 2006; for genomic data, see e.g. ENSEMBL).

ZAN was apparently prone to domain expansion and loss, with the consequence that the number of subunits differs between species. Mouse Zan, for instance, comprises three MAM do-

mains, in contrast to two as reported for pig and other species (Hardy and Garbers 1995; Gao and Garbers 1998). The mucin-like domain of pig and rabbit ZAN may serve as another example: whilst 53 hepta-repeats make up the mucin-like domain of pig ZAN, 20 hepta-repeats have been counted in rabbit ZAN (Hardy and Garbers 1995; Lea *et al.*, 2001). In line with the varying numbers of MAM domains and small aa repeats, the number of D domains differs between orthologs. While rabbit and primates exhibit the «pig-like» pattern of one partial and four complete D domains (Fig. 1; Hardy and Garbers 1995; Lea *et al.*, 2001; Herlyn and Zischler 2005b), mouse Zan is distinguished by 20 extra partial D domains derived from the N-terminus of domain D3 (D3p1-20 in Fig. 1; Gao and Garbers 1998).

Despite the interspecies differences in the number of subunits, a general pattern emerges from the available data. The extra repeats described for mouse Zan (Gao and Garbers, 1998), for example, probably represent an evolutionary novelty of house mouse or a larger rodent taxon (Herlyn and Zischler 2006). Moreover, though the number of small aa repeats differs between species, there is only one mucin-like domain regardless of the investigated species (for a summary, see, Hunt *et al.*, 2005). On the other hand, the question whether two or three MAM domains were present in the most recent common ancestor (MRCA) of Boreoeutheria has to remain unsettled. Consequently, we postulate two or three MAM domains, one mucin-like domain, one partial and four complete D domains for ZAN of the Boreoeutheria MRCA (see also Hunt *et al.*, 2005). This pattern appears to have been maintained within primates, including humans (Herlyn and Zischler 2005a,b; Gao and Garbers 2006).

### ZAN localizes to the acrosome in spermatids and spermatozoa of Boreoeutheria

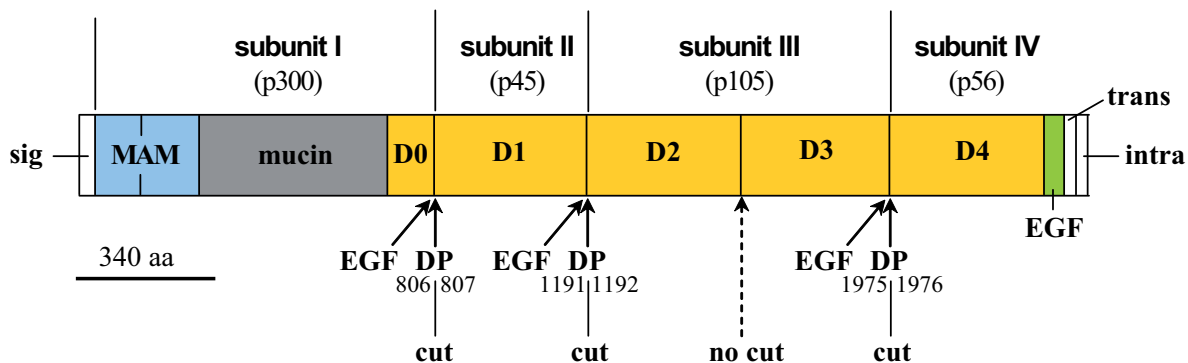
Immunofluorescence and immunoelectron microscopy of spermatozoa from different boreoeutherian species such as horse, pig, mouse, rat, rabbit, cattle, mole, hamster, and humans repeatedly localized zonadhesin at the apical head of spermatozoa (Gao and Garbers 1998; Hickox *et al.*, 2001; Lea *et al.*, 2001; Breazeale *et al.*, 2002; Bi *et al.*, 2003; Olson *et al.*, 2004). Additionally, ZAN might be expressed in Sertoli-cells as inferred from

immunofluorescence data of rabbit testis (Lea *et al.*, 2001). The latter finding gave rise to the speculation that ZAN, in particular MAM and mucin-like domains, could mediate spermatid-Sertoli cell adhesion during spermatogenesis. Processing of MAM and mucin-like domains may then release the maturing spermatids from Sertoli cells (Lea *et al.*, 2001; see also Gao and Garbers 1998). However, recent ultrastructural investigations of pig and hamster spermatids and spermatozoa have revealed that ZAN does not localize to the cell surface, which makes its involvement in spermatid-Sertoli cell adhesion quite unlikely (Bi *et al.*, 2003; Olson *et al.*, 2004). Instead, immunoelectron microscopy demonstrated a dynamic redistribution of ZAN during spermatogenesis and epididymal maturation of sperm, from the inner to the outer acrosomal membrane and from the outer acrosomal membrane to the acrosomal matrix (Bi *et al.*, 2003; Olson *et al.*, 2004). Irrespective of the specific substructure, ZAN apparently localizes to the acrosome so that a post-acrosomal reaction function appears more likely than an involvement in spermatid-Sertoli cell adhesion, at least in hamster and pig.

### ZAN D (and MAM?) domains function in ZP adhesion

The domains constituting the major part of ZAN (MAM, mucin-like and D domains) account for the binding activity of a plethora of proteins (reviewed in Gao and Garbers 1998). Direct evidence for the binding partner of ZAN came from its aforementioned isolation through a ZP affinity matrix. Subsequent Western blotting and partial aa sequencing of the isolated peptides revealed that the ZP avidity of oligomers containing D1-3 is particularly strong (Hardy and Garbers 1994, 1995; see also Hickox *et al.*, 2001). Meanwhile, ZP binding activity has also been demonstrated for a rabbit ZAN D4 construct (Lea *et al.*, 2001) so that ZP apparently represents the common binding partner of ZAN D1-4 domains.

Though we are less sure about the binding partner(s) of ZAN MAM, mucin-like, and D0 domains, indirect evidence suggests an involvement of at least ZAN MAM domains in ZP binding (present contribution). Using the phage peptide display technique, Naz *et al.*, (2000) identified nine ZP binding peptides with the consensus GHRGRRVGLGGGGRIGG (Consensus17 in Naz *et al.*, 2000).



**Fig. 2. Schematic depiction of the processing pattern of pig zonadhesin (ZAN) precursor.** Pig ZAN precursor is hydrolysed at D<sup>806</sup>-p<sup>807</sup>, D<sup>1191</sup>-p<sup>1192</sup> and D<sup>1975</sup>-p<sup>1976</sup>. The processing produces four subunits (subunits I-IV) of different molecular weight (p300, p45, p105, and p56). Note that each hydrolysed bond is preceded by an epidermal growth factor (EGF)-like domain at the C-terminus of the upstream D domain. No cleavage takes place in the absence of EGF-like motif and DP doublet (see subunit III). For additional abbreviations see legend Figure 1. The domain architecture was redrawn from Gao and Garbers (1998). For the original description of pig ZAN processing, see Hickox *et al.*, (2001) and Bi *et al.* (2003).

We aligned (BioEdit; Hall 1999) this consensus to the 2,811 aa of human ZAN transcript ENST00000349350 (release40) and found that it coincides nearly completely with a putative binding region of 30 aa at the N-terminus of the most C-terminal ZAN MAM domain (see below; Herlyn and Zischler 2005a). The apparent differences between the sequences might be caused by the fact that only part of the peptides used for the inference of Consensus17 is orthologous to the putative MAM domain binding region. Thus, it appears possible that not only ZAN D domains but also ZAN MAM domains function in ZP binding. Such a binding might - in analogy to the binding of lysin to the vitelline envelope in marine gastropods (Swanson and Vacquier 2002) - create a hole in the ZP through which the spermatozoon passes to reach the egg cell membrane.

### Activation of ZAN by processing, glycosylation and oligomerization

Using immunoprecipitation, aa sequencing and other techniques, Hickox *et al.*, (2001) and Bi *et al.*, (2003) were able to demonstrate that the pig ZAN precursor is hydrolysed in the testis at D<sup>806</sup>-P<sup>807</sup>, D<sup>1191</sup>-P<sup>1192</sup> and D<sup>1975</sup>-P<sup>1976</sup> (Fig. 2). The processing products were termed p300, p105, p45, and p56 according to their molecular weights (Bi *et al.*, 2003). Not surprisingly, the molecular weights of processing products vary between taxa due to the presence/absence of posttranslational modifications such as glycosylations. For simplicity, we here use abstract terminology that is not related to weight but aims at facilitating comparisons between species (Fig. 2; Herlyn and Zischler 2005b). Subunit I (p300 in pig) spans MAM-D0 domains and the first seven aa of D1 domain. Subunit II (p45 in pig, p43 in rabbit) consists of the nearly complete D1 domain and the first seven aa of D2 domain. Subunit III (p105 in pig and horse, p97 in rabbit) comprises most of D2 domain, the complete D3 domain and the first seven aa of D4 domain. Finally, the C-terminus of ZAN precursor is termed subunit IV (p56 in pig, p60 in horse, p58 in rabbit) (for molecular weights, see Lea *et al.*, 2001, Bi *et al.*, 2003, and Breazeale *et al.*, 2004).

Whether hydrolysis of ZAN precursor into subunits I-IV takes place autocatalytically or endoproteolytically is not yet clear (Lea *et al.*, 2001; Bi *et al.*, 2003). In case of an endoproteolytic cleavage, ZAN processing might be enhanced by a dibasic endoprotease (Lea *et al.*, 2001) as was documented for vWF proprotein (Wise *et al.*, 1990). Gao and Garbers (1998) proposed that the binding of such an endoprotease to ZAN precursor could be facilitated by the EGF-like domain downstream of domain D4. In line with this, Herlyn and Zischler (2005b) described that each of the hydrolysed DP bonds is preceded by an EGF-like motif in the upstream D domain (separated by about 70 aa in pig ZAN). Provided that both EGF-like motif and DP doublet are required for proper cleavage, they postulated a pig-like processing of ZAN precursor for cattle and primates (Table 1; see Herlyn and Zischler 2005b). In line with Zan's specific domain architecture

TABLE 1

### PREDICTED PROCESSING OF PRECURSOR ZONADHESIN ACCORDING TO PRESENCE (+) AND ABSENCE (-) OF EGF-LIKE MOTIFS AND DP-DOUBLETS IN DOMAINS D0-D4

Investigated species (coding DNA)	Domains and motifs								Pig-like processing (predicted)		
	D0		D1		D2		D3			D4	
	EGF	DP	EGF	DP	EGF	DP	EGF	DP		EGF	DP
<i>Homo sapiens</i>	+	+	+	+	-	-	+	+			yes
<i>Papio hamadryas</i>	+	+	+	+	+	-	+	+			yes
<i>Macaca mulatta</i>	+	+	+	+	+	-	+	+			yes
<i>Macaca fascicularis</i>	+	+	+	+	+	-	+	+			yes
<i>Saimiri sciureus</i>	+	+	+	+	-	-	+	+			yes
<i>Saguinus oedipus</i>	+	+	+	+	-	-	+	+			yes
<i>Callithrix jacchus</i>	+	+	+	+	-	-	+	+			yes
<i>Microcebus murinus</i>	+	+	+	+	-	-	+	+			yes
<i>Oryctolagus cuniculus</i>	-	+	+	+	+	-	+	-			no*
<i>Mus musculus</i>	+	-	+	+	-	-	-	+			no
<i>Bos taurus</i>	+	+	+	+	-	-	+	+			yes
<i>Sus scrofa</i>	+	+	+	+	-	-	+	+			-

EGF, EGF-like motif; DP, DP-doublet. N-terminal DP-bonds represent putative cleavage sites, particularly when the preceding D-domain contains a C-terminal EGF-like motif (C.C.[2][GP][FYW].[6]C). *S. scrofa* (below) was taken as reference (see Fig. 2; Hickox *et al.*, 2001; Bi *et al.*, 2003). Note: The asterisk marks a conflict between prediction and empirical data in rabbit ZAN.

(Gao and Garbers 1998; Fig. 1), the motif pattern suggests another mode of processing of mouse Zan (Table 1). The presence/absence of motifs might, thus, represent a good predictor for the processing of ZAN precursor. On the other hand, there is experimental evidence for an analog processing of pig and rabbit ZAN (Hardy and Garbers 1995; Lea *et al.*, 2001) that cannot be brought in line with the expectations from motif distribution (Table 1).

Processing into subunits does not represent the only mechanism contributing to the activation of ZAN. Subunits I-III of pig and rabbit ZAN, for instance, undergo strong glycosylation as indicated by considerable mobility shifts after enzymatic deglycosylation (Lea *et al.*, 2001; Bi *et al.*, 2003). Apart from processing and glycosylation, a third mechanism contributes to the activation of ZAN, i.e. the formation of dimers. The dimers of subunits II and III (p105/p45 in pig, p97/p43 in rabbit) are particularly relevant for ZAN activation as they exhibit increased ZP binding avidity and, moreover, represent the basic unit of higher molecular oligomers. Since dimers and oligomers decompose under reducing conditions, their formation apparently results from disulfide bonding (Hardy and Garbers 1994, 1995; Lea *et al.*, 2001; Hickox *et al.*, 2001; Bi *et al.*, 2003).

The oligomerization of vWF is mediated by interchain disulfide bonding of the cysteine residues in CGLC motifs (Mayadas and Wagner 1992). Remarkably, the CGLC motif also occurs in ZAN D1 and D2 of pig, rabbit, and primates. The consensus CG[nonpolar]C is even conserved across ZAN D1-3 of pig, cattle, rabbit, mouse, and primates (Hardy and Garbers 1995; Gao and Garbers 1998; Lea *et al.*, 2001; for physico-chemical properties,

<sup>1</sup> It is generally assumed that synonymous nucleotide substitutions evolve neutrally. Consequently, dn/ds is 1 when neither synonymous nor nonsynonymous substitutions are functionally effective (neutral evolution). On the other hand, dn/ds values < 1 point to an overall reduction of individual fitness by amino acid exchanges (negative, or purifying selection). Finally, dn/ds values > 1 suggest that amino acid changes increase the fitness of an individual. In the latter case, evolutionary biologists speak of positive (Darwinian) selection. Irrespective of the terminology, dn/ds values > 1 indicate adaptive, fast evolution.



These two examples illustrate that concerted evolution has also promoted ZAN divergence between species, at least in case of the mucin-like domain (for mouse and human consensus, see Gao and Garbers 1998 and Gasper and Swanson 2006).

### Positively selected sites accumulate across motifs and a putative binding region

One interesting outcome of recent sequence analyses of *ZAN* cDNA fragments encoding MAM and D domains is that candidate sites of positive selection accumulate across non-conserved myristylation, phosphorylation and glycosylation motifs (Fig. 3B; Herlyn and Zischler 2005a,b). A similar association of motifs and positively selected sites has been reported e.g. for certain regulatory proteins of HIV-1 (De Oliveira *et al.*, 2004) and can be taken as a hint whereafter positive selection promotes changes in posttranslational modification of ZAN (Herlyn and Zischler 2005a,b).

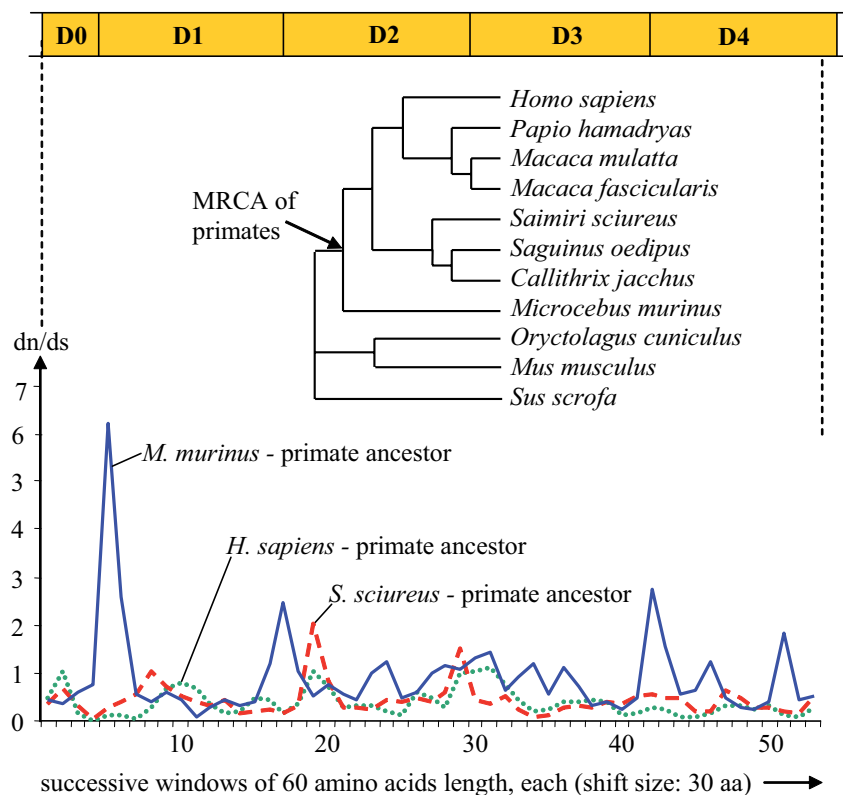
The association of positively selected sites and predicted motifs is particularly striking within a 30 aa fragment at the N-terminus of the most C-terminal ZAN MAM domain (see Fig. 3B; Herlyn and

Zischler 2005a). Similar accumulations of positively selected sites are known from binding regions of diverse proteins such as human leukocyte antigen and major histocompatibility complex I (Hughes *et al.*, 1990; Suzuki and Gojobori 1999). Remarkably, a likewise pattern has even been reported for a putative binding region of bindin, the sperm acrosomal protein of sea urchins that bonds the sperm to the egg (Biermann 1998; see paper by Zigler this issue), so that the respective ZAN fragment might also represent a binding region. This conclusion is supported by accessibility predictions whereafter the respective 30 aa fragment is exposed in mature ZAN (Herlyn and Zischler 2005a). On the other hand, the three dimensional structure and binding properties of ZAN have still to be examined experimentally to validate the putative adhesive function of the 30 aa fragment.

### Sexual selection promotes interspecies divergence of ZAN

As outlined above, positive selection (protein adaptive evolution) plays a contributory role in the evolution of *ZAN*, particularly in regions coding for MAM and D domains. However, the term «positive selection» simply means that selection favors aa exchanges that are in turn assumed to be advantageous for survival of the individual. «Positive selection» does not say whether the preferential propagation of an allele from one generation to the next is due to enhanced viability («ecological selection») or increased fertility («sexual selection») of the carriers of specific alleles. Molecular genetic studies point to an important impact of sexual selection on the evolution of sperm-egg interacting proteins in marine invertebrates, which are free-spawners such as abalones and sea urchins. It is proposed that closely related species rapidly change the primary structure of the sperm-egg interacting proteins, thus reinforcing reproductive barriers even under sympatric conditions (Swanson and Vacquier 2002). However in the case of animals with internal fertilization, the evolution of sperm proteins might also be under the influence of alternative mating strategies. Primates are particularly suited for studies addressing this question as their behavior is particularly well documented.

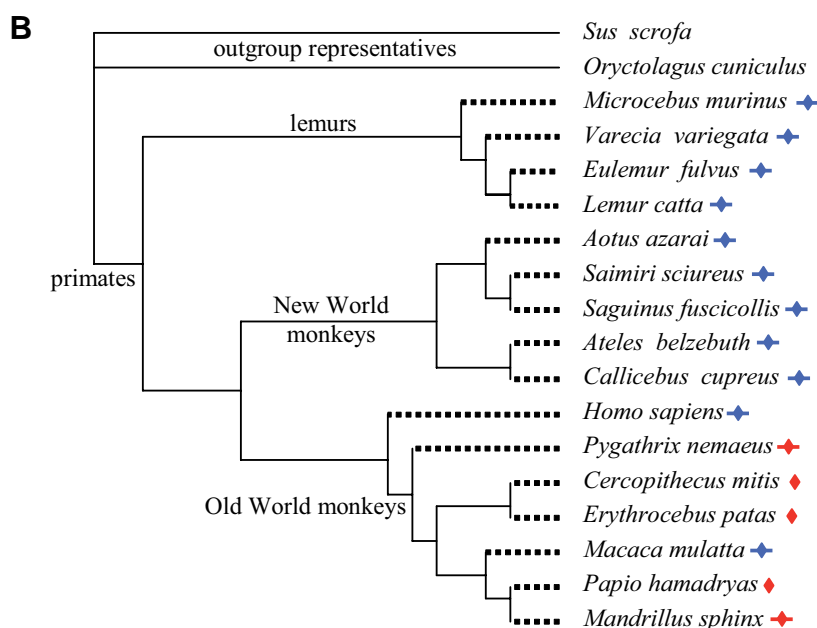
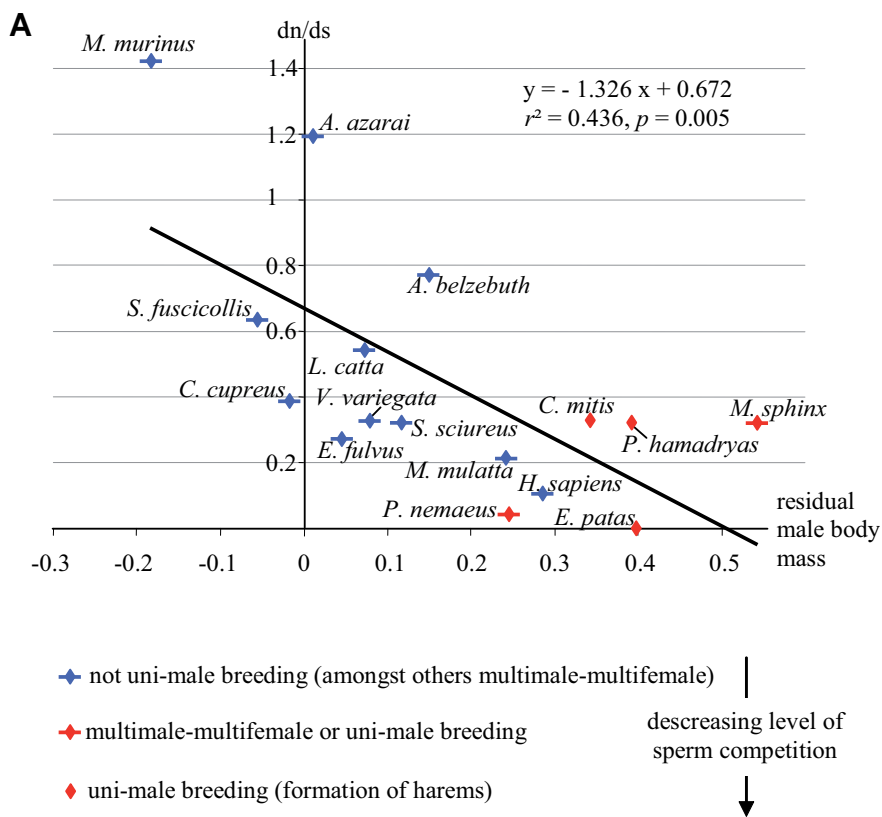
The first evidence for a correlation between sequence evolution of *ZAN* and sexual selection came from moving window analysis of dn/ds along cDNA fragments encoding ZAN domains D0-4. This showed that dn/ds peaks were highest in pairwise sequence comparisons including the gray mouse lemur (*Microcebus murinus*), the presumably most promiscuous species included in the primate sample (Fig. 4; Herlyn Zischler 2005a, for the mating system of *M. murinus*, see, e.g., Fietz 1999). In a recent study, we further investigated how sexual selection influences *ZAN* sequence evolution. The dataset under scrutiny contained concatenated MAM and D domain encoding fragments from 16 primate species, including lemurs, New World monkeys, Old World monkeys, and humans (Herlyn and Zischler



**Fig. 4. Moving window analysis of cDNAs coding for zonadhesin domains D0-4 (~ 4.9 kb).** dn/ds was deduced for 60 amino acid (aa) windows (shift size: 30 aa) from pairwise comparisons between a primate ortholog (for the species, see tree) and a hypothetical sequence reconstructed for the most recent common ancestor (MRCA) of primates. For simplicity, only three curves (out of eight) are shown in the graph. Note: dn/ds peaks are highest when comparing the ancestral sequence with the cDNA of the presumably most promiscuous primate species included, the gray mouse lemur *M. murinus* (for mating systems, see legend Fig. 5). The PAML package (Yang 1997) was used for sequence reconstruction. The primate phylogeny was adopted from Smith and Cheverud (2002). Moving window analysis was carried out with CRANN (Creevey and McInerney 2002). The graph was adapted from Herlyn and Zischler (2005b).

2007). We used dn/ds estimates for the terminal branches of a commonly accepted phylogeny (Fig. 5B; for the phylogeny, see e.g. Smith and Cheverud 2002) as a measure of ZAN evolution in single species. Sexual dimorphism in body weight was taken as an easily available measure of sperm competition (for a discussion of this point, see Herlyn and Zischler 2007). Plotting the species-specific dn/ds values against residual male body weights revealed a significant negative correlation between sequence evolution of the concatenated ZAN fragments and body weight dimorphism in primates (Fig. 5A; Herlyn and Zischler 2007). As expected, the ZAN fragments turned out to evolve faster in sexually less dimorphic species where estrous females mate with multiple partners (e.g., *M. murinus*; Fietz 1999) than in more dimorphic - and often living in harems - species where males are more successful at monopolizing estrous females (e.g. red guenon, *Erythrocebus patas*; Fig. 5A). Therefore, correlation study results suggest that sexual selection, especially competition amongst sperm («sperm competition»), contributes to sequence divergence of ZAN MAM and D domains. However, since ZAN is a sperm ligand, it is rather likely that not only sperm competition but also selection by the female binding partner («cryptic female choice») contributes to the evolution of the MAM and D domain encoding fragments analyzed.

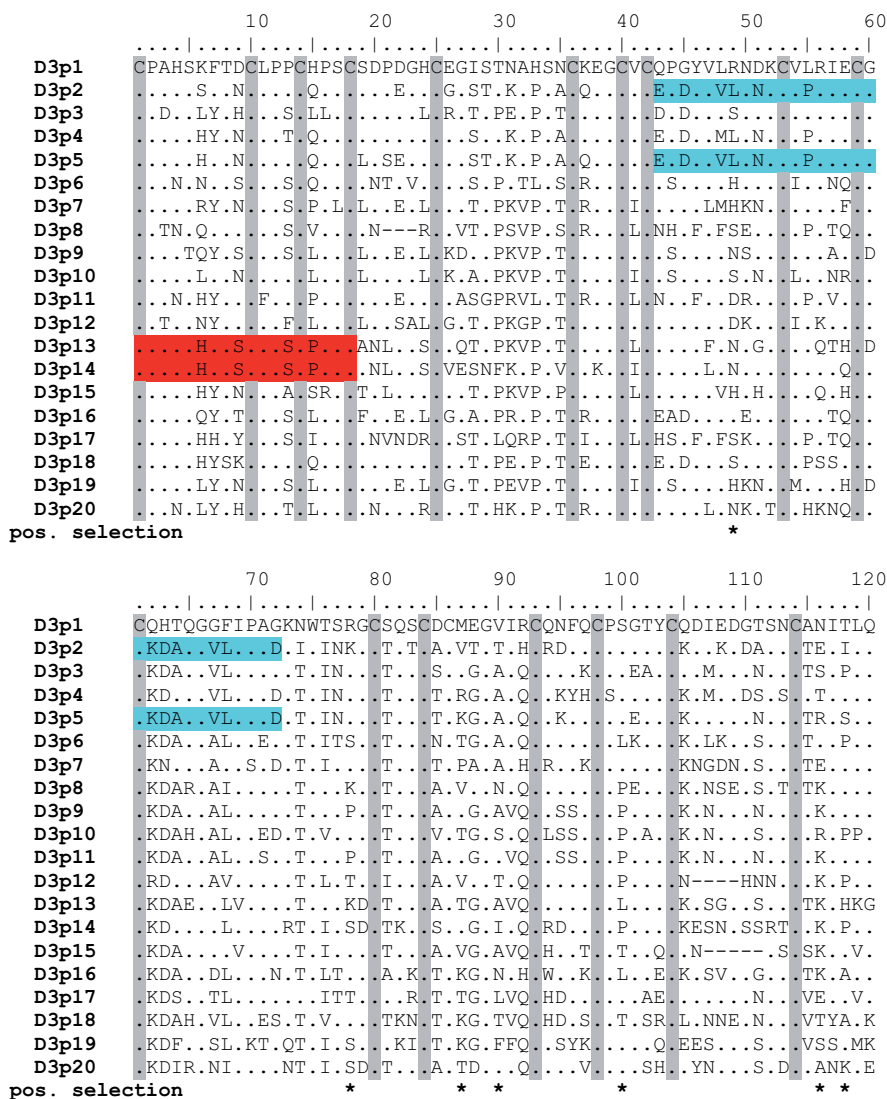
Since ZAN D1-4 and possibly also ZAN MAM domains bind ZP (see above), it is quite probable that the fragments used for the correlation study presented here coevolve in adaptation to a constantly changing ZP receptor (see Turner and Hoekstra, 2008). The question that remained unanswered is the actual sequence(s) that might function as ZP receptors of ZAN MAM and D domains. Based on a sample comprising sequences from five Boreoeutheria, Swanson *et al.*, (2001) found signatures of positive selection for single codon sites of two ZP components, i.e. ZP2 and ZP3 (Swanson *et al.*, 2001). In case of ZP3, the candidate sites of positive selection even turned out to be concentrated within distinct fragments as to be expected for binding regions. ZAN MAM domains, D domains, and



**Fig. 5. Zonadhesin evolves more slowly in primate species with less sperm competition. (A)** Significant negative correlation between evolution (dn/ds) of MAM and D domain encoding zonadhesin fragments and body weight dimorphism (male residuals) in primates.

The curve suggests a lower level of sperm competition in sexually more dimorphic primate species. Apparently, males of these species are more successful at monopolizing estrous females. Note: as expected, uni-male mating («harem living») species cluster down-right in the graph. The dn/ds values used for regression are PAML (Yang 1997) estimates for the terminal branches (dotted lines) of the phylogeny depicted in (B). The alignment analyzed comprised three concatenated fragments of altogether ~ 555 bp length. For primate phylogeny see e.g. Smith and Cheverud (2002). Data on body weight and mating system were taken from literature (for citations see Herlyn and Zischler 2007). The graph was adapted from Herlyn and Zischler (2007). In contrast to the previously published version, *M. sphinx* has not been classified as «uni-male breeding» in accordance with a personal communication from Alan Dixon.

**Fig. 6. Amino acid alignment of mouse partial D3 repeats (D3p1-20).** Dots indicate identity with the reference D3p1. Gray shading highlights conserved cysteines (C). Partial gene conversion between D3p2 and D3p5 (turquoise) and D3p13 and D3p14 (red) has been detected by GENECONV (Sawyer 1989). Asterisks below indicate candidate sites of positive selection (pos. selection;  $p_{(dn/ds>1)} > 0.99$ ) as identified by a branch-site model implemented in PAML (Yang 1997). See Herlyn and Zischler (2006) for the tree used for sequence analysis.



species compared, and the substitution model chosen. So far however, published dn/ds estimates consistently indicate moderate negative selection for ZAN or ZAN fragments, respectively (see Swanson *et al.*, 2003; Herlyn and Zischler 2005a,b; Gasper and Swanson 2006). In other words, the vast majority of codon sites are under negative selection, i.e. selection against fitness-reducing aa changes. Specifically, cysteines are under negative selection in both paralogs and between species comparisons. In particular, the CG[nonpolar]C motif is conserved among pig, cattle, rabbit, mouse, and primates as well as between ZAN D1, D2, and D3 of single species (see above). Likewise, single cysteine residues are widely conserved between species (see vertical shading in Fig. 3B; e.g. Herlyn and Zischler 2005a) and between duplicate domains of single species (see vertical shading in Fig. 6; e.g. Hardy and Garbers 1995, Gao and Garbers 1998). These data reflect the well established phenomenon that disulfide bonds and, thus, folding and oligomerization are negatively selected. As in other proteins, the protein backbone of ZAN apparently needs to meet certain requirements regarding shape and physico-chemical properties in order to retain functionality. It can thus be summarized that the pattern of ZAN/evolution between species includes: between species divergence through positive selection, drift, and concerted evolution, despite a strong tendency for conservation of the protein backbone.

the fragments used for correlation study might thus coevolve in adaptation to one or more rapidly changing ZP fragments.

**Conservation of protein backbone and oligomerization by negative selection**

Depending on the model used, an estimated proportion of 87-99% of the codon sites coding for ZAN D domains show signatures of negative selection or neutral evolution (Herlyn and Zischler 2005b). Likewise, 88-97% of codon sites encoding MAM domains fall into the dn/ds interval (0,1) (Herlyn and Zischler 2005a). As can be seen from the estimates, these values are not absolute and vary depending on the fragment examined, the

**Evolution in the paralog comparison: mouse partial D3 repeats**

Some of the evolutionary mechanisms shaping ZAN (positive selection, negative selection, gene conversion) are condensed within the 20 partial D3 repeats of mouse (D3p1-20; Fig. 6). As mentioned above, these partial repeats are derived from the 120

<sup>2</sup> Nonsynonymous nucleotide substitutions can be distinguished according to their effect on physico-chemical properties. Whilst a conservative nucleotide substitution does not affect the physico-chemical properties (e.g. from polar to polar), a radical substitution does (e.g., from polar to nonpolar). Like synonymous substitutions, conservative nonsynonymous substitutions are expected to evolve (nearly) neutrally. pnr/pnc can then be used as a measure of sequence evolution, with pnr being the rate of radical substitutions to radical sites and pnc being the rate of conservative substitutions to conservative sites. In analogy to dn/ds, pnr/pnc values > 1, = 1, and < 1 are taken as evidence of positive selection, neutral evolution, and negative selection, respectively.

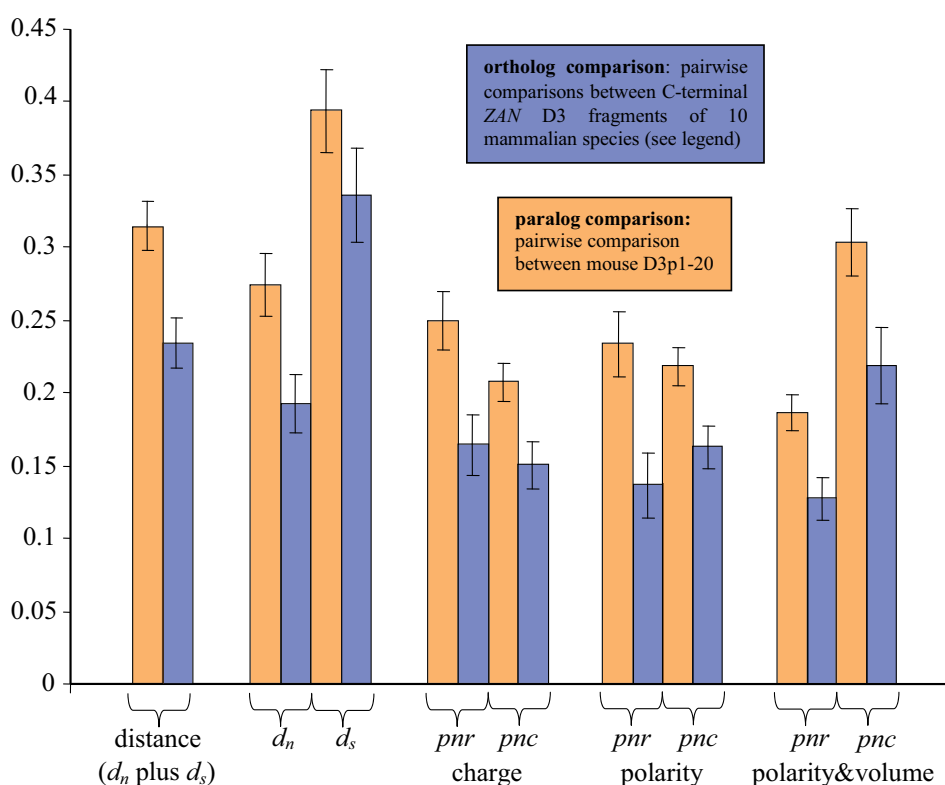
C-terminal aa of domain D3 (Gao and Garbers 1998). Considering that mouse D3p1-20 are phylogenetically younger than domains D0-4 (Herlyn and Zischler 2006), pairwise distances should be higher in the paralog (D3p1-20) than in the ortholog comparison (corresponding C-termini of mouse D3, pig D3, human D3 etc.). Surprisingly, the pairwise distances turned out to be higher in the paralog-ortholog comparison irrespective of the distance measure used ( $d_n$ ,  $d_s$ ,  $pnc$ ,  $pnc^2$ ) (Fig. 7). As there was no evidence for saturation in either the paralog or ortholog dataset (see legend Fig. 7), the authors concluded that the paralogs must have diverged faster than the orthologs. ML based analysis revealed stronger evidence of positive selection among paralogs than orthologs, which could explain the higher distances observed in the paralog-ortholog comparison (Fig. 6; Herlyn and Zischler 2006).

Beyond this, mouse partial D3 repeats can serve as an extreme example for the conservation of cysteines in the comparison of domain duplicates. For instance, all 18 cysteines are fully conserved when comparing the sequences amongst each other (Fig. 6; Gao and Garbers 1998; Herlyn and Zischler 2006). Homogenization by partial gene conversion might finally account for the sequence identity of the N-termini of mouse D3p13 and D3p14 as well as for the central 30 aa of D3p2 and D3p5 (Fig. 6, Herlyn and Zischler 2006). Nevertheless, divergence by positive selection and drift so far outbalanced trends for conservation by negative selection and homogenization by concerted evolution. This seems to be a general principle in the evolution of D0-4 domains (Herlyn and Zischler 2006) and might also apply to MAM domains.

## Conclusion

Based on pioneering work of Hardy and Garbers, ZAN has come to represent one of the best investigated mammalian sperm ligands from both a biochemical and evolutionary perspective. ZAN is a mosaic-type protein that localizes to the apical head of spermatozoa. While rabbit, pig, cattle and primate ZAN precursors essentially consist of two or three MAM domains, one mucin-like domain, one incomplete and four complete D domains, mouse ZAN differs from the common architecture by an extra set of 20 partial D3 repeats.

Sequence comparison between species and within species consistently indicates that positive selection has contributed to the sequence evolution of ZAN. A putative binding region of the most C-terminal MAM domain is characterized by a particular



**Fig. 7. Higher mean distances in the paralog-ortholog comparison.** The ortholog comparison was based on the homolog C-terminal D3 fragments from *Mus musculus*, *Oryctolagus cuniculus*, *Macaca fascicularis*, *Papio hamadryas*, *Homo sapiens*, *Saguinus oedipus*, *Callithrix jacchus*, *Saimiri sciureus*, *Microcebus murinus*, and *Sus scrofa*. cDNAs coding for mouse partial D3 repeats (mouse D3p1-20) were used for paralog comparison. Vertical lines represent standard errors. Given the younger phylogenetic age of mouse D3p1-20 compared to the orthologs (see text), the rate of sequence evolution must have been higher in paralogs than in orthologs. Pairwise sequence comparisons were carried out using MEGA (Kumar et al., 2004) and SCR3 (Hughes et al., 1990). Note: Results from DAMBE (Xia and Xie 2001) analysis ruled out that saturation led to an underestimation of the distance estimates in the ortholog comparison.  $d_n$ , rate of nonsynonymous substitutions;  $d_s$ , rate of synonymous substitutions;  $pnc$ , rate of conservative nonsynonymous substitutions;  $pnr$ , rate of radical nonsynonymous substitutions. Note:  $pnc$  and  $pnr$  were deduced using amino acid classifications by charge, polarity, and polarity+volume. The figure was adapted from Herlyn and Zischler (2006).

accumulation of positively selected codon sites. Furthermore, between species comparisons of MAM and D domain encoding cDNAs revealed an enrichment of candidate sites of positive selection across non-conserved motifs. While positive selection and drift are the driving forces behind the divergence of MAM and D domains, divergence of the mucin-like domains is mainly driven by concerted evolution of the basic units, small repeats of mostly seven aa length. Notwithstanding a general tendency for divergence, most codon sites are negatively selected probably because of processing and folding constraints.

A negative correlation between ZAN evolution and sexual dimorphism in body weight suggests that sexual selection, in particular sperm competition, could represent the driving force behind the divergence of single MAM and D domain encoding sites. Hence, the level of sperm competition is lower in uni-male breeding and sexually more dimorphic species because males are more successful at maintaining exclusive breeding access to their mates. Given that D domains and maybe also MAM domains

bind ZP, it seems reasonable to assume that cryptic female choice by one or more constantly changing ZP fragments also contributes to sequence evolution of ZAN.

#### Acknowledgements

We gratefully acknowledge the financial support of the German Primate Center Göttingen in an early phase of the project. Additional financial support was provided by the German Research Foundation (He 3487/1-1). Our studies would not have been possible without the material generously provided by the members of the Departments of Primate Genetics (Dr. C. Roos), Reproductive Biology (Prof. K. Hodges; Dr. A. Schrod), Neurobiology (Prof. U. Jürgens), and Veterinary Medicine and Primate Husbandry (Prof. F.-J. Kaup) at the German Primate Center Göttingen. We are moreover indebted to Dr. M. Perret (Laboratoire d'Ecologie, Brunoy/France) and Prof. B. Brenig (Veterinary Medicine, University of Göttingen) for contributing further material. Last but not least, we thank C. Schwiegl (Max Planck Institute for Biophysical Chemistry, Göttingen) for technical support when we launched the zonadhesin project in 2001.

#### References

- ANISIMOVA, M., BIELAWSKI, J. P. and YANG, Z. (2001) Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Mol Biol Evol* 18: 1585–92.
- BI, M., HICKOX, J. R., WINFREY, V. P., OLSON, G. E., HARDY, D. M. (2003) Processing, localization and binding activity of zonadhesin suggest a function in sperm adhesion to the zona pellucida during exocytosis of the acrosome. *Biochem J* 375: 477–88.
- BIERMANN, C. H. (1998) The molecular evolution of sperm bindin in six species of sea urchins (Echinoidea: Strongylocentrotidae). *Mol Biol Evol* 15: 1761–71.
- BREAZEALE, K. R., BRADY, H. A., BI, M., UPPULURI, S., THOMPSON, L. D., BRUEMMER, J. E., HARDY, D. M. (2002) Biochemical properties and localization of zonadhesin in equine spermatozoa. *Theriogenology* 58: 359–62.
- CREEVEY, C. and MCINERNEY, J. O. (2002) An algorithm for detecting directional and non-directional positive selection, neutrality and negative selection in protein coding DNA sequences. *Gene* 300: 43–51.
- DE OLIVEIRA, T., SALEMI, M., GORDON, M., VANDAMME, A. M., VAN RENSBURG, 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: 1047–58.
- FIETZ, J. (1999) Mating system of *Microcebus murinus*. *Am J Primatol* 48: 127–133.
- GAO, Z., HARUMI, T. and GARBERS, D. L. (1997) Chromosome localization of the mouse zonadhesin gene and the human zonadhesin gene (ZAN). *Genomics* 41: 119–22.
- GAO, Z. and 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: 3415–21.
- GASPER, J. and SWANSON, W. J. (2006) Molecular population genetics of the gene encoding the human fertilization protein zonadhesin reveals rapid adaptive evolution. *Am J Hum Genet* 79: 820–30.
- GILLIGAN, P., BRENNER, S. and VENKATESH, B. (2002) Fugu and human sequence comparison identifies novel human genes and conserved non-coding sequences. *Gene* 294: 35–44.
- GLÖCKNER, G., SCHERER, S., SCHATTEVOY, R., BORIGHT, A., WEBER, J., TSUI, L. C., ROSENTHAL, A. (1998) Large-scale sequencing of two regions in human chromosome 7q22: analysis of 650 kb of genomic sequence around the EPO and CUTL1 loci reveals 17 genes. *Genome Res* 8: 1060–73.
- HALL, T. A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* 41: 95–8.
- HARDY, D. M. and GARBERS, D. L. (1994) Species-specific binding of sperm proteins to the extracellular matrix (zona pellucida) of the egg. *J Biol Chem* 269: 19000–4.
- HARDY, D. M. and 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: 26025–28.
- HERLYN, H. and ZISCHLER, H. (2005a) Sequence evolution, processing, and posttranslational modification of zonadhesin D domains in primates, as inferred from cDNA data. *Gene* 362: 85–97.
- HERLYN, H. and ZISCHLER, H. (2005b) Identification of a positively evolving putative binding region with increased variability in posttranslational motifs in zonadhesin MAM domain 2. *Mol Phylogenet Evol* 37: 62–72.
- HERLYN, H. and ZISCHLER, H. (2006) Tandem repetitive D domains of the sperm ligand zonadhesin evolve faster in the paralogue than in the orthologue comparison. *J Mol Evol* 63: 602–11.
- HERLYN, H. and ZISCHLER, H. (2007) Sequence evolution of the sperm ligand zonadhesin correlates negatively with body weight dimorphism in primates. *Evolution Int J Org Evolution* 61: 289–98.
- HICKOX, J. R., BI, M. and HARDY, D. M. (2001) Heterogeneous processing and zona pellucida binding activity of pig zonadhesin. *J Biol Chem* 276: 41502–9.
- HUGHES, A. L., OTA, T. and NEI, M. (1990) Positive Darwinian selection promotes charge profile diversity in the antigen-binding cleft of class I major-histocompatibility-complex molecules. *Mol Biol Evol* 7: 515–24.
- HUNT, P. N. D., WILSON, M. D., VON SCHALBURG, K. R., DAVIDSON, W. S., KOOP, B. F. (2005) Expression and genomic organization of zonadhesin-like genes in three species of fish give insight into the evolutionary history of a mosaic protein. *BMC Genomics* 6: 165.
- KOSAKOVSKY POND, S. L., FROST, S. D. and MUSE, S. V. (2005) HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 21: 676–9.
- KUMAR, S., TAMURA, K. and NEI, M. (2004) MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform* 5: 150–63.
- LEA, I. A., SIVASHANMUGAM, P. and O'RAND, M. G. (2001) Zonadhesin: characterization, localization, and zona pellucida binding. *Biol Reprod* 65: 1691–700.
- MAYADAS, T. N. and WAGNER, D. D. (1992) Vicinal cysteines in the prosequence play a role in von Willebrand factor multimer assembly. *Proc Natl Acad Sci USA* 89: 3531–5.
- NAZ, R. K., ZHU, X. and KADAM, A. L. (2000) Identification of human sperm peptide sequence involved in egg binding for immunocontraception. *Biol Reprod* 62: 318–24.
- OLSON, G. E., WINFREY, V. P., BI, M., HARDY, D. M., NAGDAS, S. K. (2004) Zonadhesin assembly into the hamster sperm acrosomal matrix occurs by distinct targeting strategies during spermiogenesis and maturation in the epididymis. *Biol Reprod* 71: 1128–34.
- ROST, B. (1996) PredictProtein. *Methods Enzymol* 266: 525–39.
- SAWYER, S. (1989) Statistical tests for detecting gene conversion. *Mol Biol Evol* 6: 526–38.
- SMITH, R. J. and CHEVERUD, J. M. (2002) Scaling of sexual dimorphism in body mass: a phylogenetic analysis of Rensch's Rule in Primates. *Int J Primatol* 23: 1095–135.
- SUZUKI, Y. and GOJOBORI, T. (1999) A method for detecting positive selection at single amino acid sites. *Mol Biol Evol* 16: 1315–28.
- SWANSON, W. J., NIELSEN, R. and YANG, Q. (2003) Pervasive adaptive evolution in mammalian fertilization proteins. *Mol Biol Evol* 20: 18–20.
- SWANSON, W. J. and 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. (2001) Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. *Proc Natl Acad Sci USA* 98: 2509–14.
- TURNER, L. M. and HOEKSTRA, H. E. (2008) Causes and consequences of the evolution of reproductive proteins. *Int. J. Dev. Biol.* 52: 769–780.
- WISE, R. J., BARR, P. J., WONG, P. A., KIEFER, M. C., BRAKE, A. J., KAUFMAN, R. J. (1990) Expression of a human proprotein processing enzyme: correct cleavage of the von Willebrand factor precursor at a paired basic amino acid site. *Proc Natl Acad Sci USA* 87: 9378–82.
- XIA, X. and XIE, Z. (2001) DAMBE: data analysis in molecular biology and evolution. *J Hered* 92: 371–3.
- YANG, Z. (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* 13: 555–6.
- ZHANG, J. (2000) Rates of conservative and radical nonsynonymous nucleotide substitutions in mammalian nuclear genes. *J Mol Evol* 50: 56–68.

## Related, previously published *Int. J. Dev. Biol.* articles

See our recent Special Issue ***Developmental Biology in Poland*** edited by Tarkowski, Maleszewski and Kloc at:  
<http://www.ijdb.ehu.es/web/contents.php?vol=52&issue=2-3>

See our recent Special Issue ***Ear Development*** edited by Fernando Giraldez and Bernd Fritsch at:  
<http://www.ijdb.ehu.es/web/contents.php?vol=51&issue=6-7>

### **Follicular cell differentiation in polytrophic ovaries of a moth midge, *Tinearia alternata***

Marta Mazurkiewicz and Janusz Kubrakiewicz  
*Int. J. Dev. Biol.* (2008) 52: 267-278

### **Peter Holland, homeobox genes and the developmental basis of animal diversity**

Sebastian M. Shimeld  
*Int. J. Dev. Biol.* (2008) 52: 3-7

### **Genetic control of gamete quality in the mouse - a tribute to Halina Krzanowska**

Jozefa Styrna  
*Int. J. Dev. Biol.* (2008) 52: 195-199

### **Molecular aspects of avian oogenesis and fertilisation**

Bozenna Olszanska and Urszula Stepinska  
*Int. J. Dev. Biol.* (2008) 52: 187-194

### **Key apoptosis regulating proteins are down-regulated during postnatal tissue development**

Shane D. Madden, Maryanne Donovan and Thomas G. Cotter  
*Int. J. Dev. Biol.* (2007) 51: 415-424

### **Molecular evolution of the vertebrate mechanosensory cell and ear**

Bernd Fritsch, Kirk W. Beisel, Sarah Pauley and Garrett Soukup  
*Int. J. Dev. Biol.* (2007) 51: 663-678

### **The molecular biology of ear development - "Twenty years are nothing"**

Fernando Giraldez and Bernd Fritsch  
*Int. J. Dev. Biol.* (2007) 51: 429-438

### **Gametophyte interaction and sexual reproduction: how plants make a zygote**

Leonor C. Boavida, Ana Maria Vieira, Jörg D. Becker and José A. Feijó  
*Int. J. Dev. Biol.* (2005) 49: 615-632

2006 ISI **\*\*Impact Factor = 3.577\*\***

