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...
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 additionally 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).

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 EST00000349350, 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 boreoeuthern 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-
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 spermatozoon (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 GHRGRVGLGGGGRIGG (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 D806-P807, D1191-P1192 and D1975-P1976. 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 D806, D807, D1191, D1192 and D1975 (p176) (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 D4. In line with this, Herlyn and Zischler (2005b) described that the binding of such an endoprotease to ZAN precursor could be facilitated by the EGF-like domain downstream of D4. In line with this, Herlyn and Zischler (2005b) described that the binding of such an endoprotease to ZAN precursor could be facilitated by the EGF-like domain downstream of D4. 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 (Gao and Garbers 1998; Fig. 1), the motif pattern suggests another mode of processing of mouse Zan (Table 1). The absence/presence 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 C[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, 1 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.

**TABLE 1**

<table>
<thead>
<tr>
<th>Investigated species (coding DNA)</th>
<th>D0</th>
<th>D1</th>
<th>D2</th>
<th>D3</th>
<th>D4</th>
<th>Pig-like processing (predicted)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homo sapiens</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>yes</td>
</tr>
<tr>
<td>Papio hamadryas</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>yes</td>
</tr>
<tr>
<td>Macaca mulatta</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>yes</td>
</tr>
<tr>
<td>Macaca fascicularis</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>yes</td>
</tr>
<tr>
<td>Saimiri sciureus</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>yes</td>
</tr>
<tr>
<td>Saginus oedipus</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>yes</td>
</tr>
<tr>
<td>Calithece cucalus</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>yes</td>
</tr>
<tr>
<td>Microcebus murinus</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>yes</td>
</tr>
<tr>
<td>Oryctolagus cuniculus</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>yes</td>
</tr>
<tr>
<td>Mus musculus</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>yes</td>
</tr>
<tr>
<td>Bos taurus</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>yes</td>
</tr>
<tr>
<td>Sus scrofa</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>yes</td>
</tr>
</tbody>
</table>

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* 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.
see Zhang 2000) so that di- and oligomerization through disulfide bonding may represent a general principle of ZAN activation (Herlyn and Zischler 2005b).

**ZAN divergence by positive selection, drift and concerted evolution**

Positive selection, i.e. selection for aa changes, contributes to the rapid evolution of many genes involved in mating behavior, fertilization, spermatogenesis, and sex determination (reviewed in Swanson and Vacquier 2002). Thus, it was not surprising that positive selection (protein adaptive evolution) was also detected for single codon sites of ZAN, soon after sufficient GenBank data were available for maximum likelihood (ML) analysis of the ratio of nonsynonymous substitutions/nonsynonymous sites to synonymous substitutions/synonymous sites (dn/ds) (Swanson et al., 2003). A drawback of this first analysis was that the ML methods used were not robust with the small sample comprising a total of four sequences from human, pig, house mouse, and European rabbit (Anisimova et al., 2001; see Gasper and Swanson 2006). Nevertheless, subsequent ML analyses of sequences from more species (particularly primates) confirmed the occurrence of positively selected codon sites across ZAN (see asterisks in Fig. 3B; Herlyn and Zischler 2005a,b; Gasper and Swanson 2006). Beyond this, positive selection has also been deduced from human population data (Gasper and Swanson 2006).

Positive selection is not the only force driving divergence of ZAN and ZAN. Though more apparent on the nucleotide level, drift may also have played a contributory role in ZAN divergence. Sequence analysis, moreover, suggests that unequal crossing-over, intragenic gene conversion and/or slippage contributed to ZAN evolution. In case of the mucin-like domain, the homogenizing effect of these mechanisms is so strong that its basic units, small aa repeats of mostly seven aa lengths, were subjected to concerted evolution. Due to the homogenizing effect of concerted evolution, the 53 hepta-peptides shaping the mucin-like domain of pig ZAN, can be summarized by the consensus aa sequence PTE(K/R)(P/T)T(V/I) (Hardy and Garbers 1995). The 20 mucin-like hepta-repeats of rabbit ZAN, as another example, share the consensus sequence (P/T)TVP(P/T)E(P/E) (Lea et al., 2001).

---

**Fig. 3. Identification of a putative binding region in the most C-terminal zonadhesin (ZAN) MAM domain.**

(A) Amino acid (aa) alignment of human ZAN and a 17 aa consensus (Consensus17) inferred from ZP binding human peptides. Only the relevant part of the altogether 2,611 aa that make up transcript ENST00000349350 (release 40) is shown. Consensus17 is based on nine ZP binding peptides identified by Naz (for more details, see Herlyn and Zischler 2005a). Consensus motifs: G[^EDRKHPFYW][^P][STAGCN][^P] N-myristylation; N[^P][ST][^P] N-glycosylation; SG.G glycosaminoglycan attachment; [ST][RK] protein kinase C phosphorylation; [ST][2]DE casein kinase II phosphorylation.

(B) solvent accessibility

Consumer17

---

**putative binding region**

---

**positive selection**

---

**Fig. 3.** Identification of a putative binding region in the most C-terminal zonadhesin (ZAN) MAM domain. (A) Amino acid (aa) alignment of human ZAN and a 17 aa consensus (Consensus17) inferred from ZP binding human peptides. Only the relevant part of the altogether 2,611 aa that make up transcript ENST00000349350 (release 40) is shown. Consensus17 is based on nine ZP binding peptides identified by Naz et al., (2000). Note: Consensus17 almost completely aligns to the boxed binding region predicted by Herlyn and Zischler (2005a). Differences between the sequences might result from the circumstance that not every peptide used for the inference of Consensus17 is ortholog to the putative MAM domain binding region. The sequences were aligned using BioEdit (Hall 1999). Consensus17 was taken from Naz et al., (2000). (B) Amino acid alignment of the N-terminus of the most C-terminal zonadhesin (ZAN) MAM domain (MAM3 in mouse). Only the human reference (H. sapiens) is shown in detail. Dots indicate identity with the reference; dashes represent gaps. Gray background highlights conserved cysteines. Solvent accessibility prediction refers to the human reference (e = exposed, space character = intermediary, b = buried). Asterisks highlight candidate sites for positive selection (shaded asterisks: $P_{\text{dev}}(\text{aa}) \geq 0.95$). The coloring marks posttranslational modifications (red: single motifs; turquoise: overlapping motifs). Note: exposed aa, motifs, and candidate sites of positive selection accumulate within the boxed putative binding region of about 30 aa. Motif search and solvent accessibility prediction was performed using the PredictProtein server (Rost 1996). Candidate sites of positive selection were inferred from cDNA data running a beta model implemented in PAML (Yang 1997) and HyPhy (Kosakovský Pond et al., 2005) and the phylogeny shown in Fig. 4 (for more details, see Herlyn and Zischler 2005a). Consensus motifs: G[^EDRKHPFYW][2][STAGCN][^P] N-myristylation; N[^P][ST][^P] N-glycosylation; SG.G glycosaminoglycan attachment; [ST][RK] protein kinase C phosphorylation; [ST][2]DE casein kinase II phosphorylation.
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 consensuses, 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 2005a).
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 primates. Apparently, males of these species are more successful at monopolizing estrous females. Note: as expected, uni-male mating («harem living») species cluster downward-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 Dixson.

\[ y = -1.326x + 0.672 \]
\[ r^2 = 0.436, p = 0.005 \]
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 codons 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 features of negative selection or neutral evolution (Herlyn and Zischler 2005a,b). Some of the evolutionary mechanisms shaping ZAN (positive selection, drift, and concerted evolution) despite a strong tendency for conservation of species divergence through positive selection, drift, and concerted evolution (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.

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 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.

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 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.

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 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.

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 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.
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 (dn, ds, pnc, pnc2) (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 been represented as 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 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 3497/1-1). 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. Schrodt), 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. Schwiegk (Max Planck Institute for Biophysical Chemistry, Göttingen) for technical support when we launched the zonadhesin project in 2001.

References


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

See our recent Special Issue *Developmental Biology in Poland* edited by Tarkowski, Maleszewski and Kloc at:

See our recent Special Issue *Ear Development* edited by Fernando Giraldez and Bernd Fritzsch at:

**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

**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 Fritzsch, 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 Fritzsch
*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