Cytogenetic and Genome Research

Original Article

Cytogenet Genome Res
DOI: 10.1159/000509584

Received: March 19, 2020
Accepted: June 8, 2020
Published online: July 18, 2020

Intra-Protein Coevolution Is Increasingly Functional with Greater Proximity to Fertilization

Marcel Kwiatkowski[a] Abdul R. Asif[b] Julia Schumacher[c] Bertram Brenig[d] Hans Zischler[c] Holger Herlyn[c]

[a]Lab for Metabolic Signaling, Institute of Biochemistry, Functional Proteomics and Metabolomics, University of Innsbruck, Innsbruck, Austria; [b]Department of Clinical Chemistry/UMG-Laboratories, University Medical Center Göttingen, Georg August University Göttingen, Göttingen, Germany; [c]Institute of Organismic and Molecular Evolution, Anthropology, University of Mainz, Mainz, Germany; [d]Institute of Veterinary Medicine, Georg August University Göttingen, Göttingen, Germany

Keywords
Evolution · Fertilization · Genome · Protein function · Protein structure

Abstract
Intramolecular coevolution of amino acid sites has repeatedly been studied to improve predictions on protein structure and function. Thereby, the focus was on bacterial proteins with available crystallographic data. However, intramolecular coevolution has not yet been compared between protein sets along a gradient of functional proximity to fertilization. This is especially true for the potential effect of external selective forces on intraprotein coevolution. In this study, we investigated both aspects in equally sized sets of mammalian proteins representing spermatozoa, testis, entire body, and liver. For coevolutionary analyses, we derived the proportion of covarying sites per protein from amino acid alignments of 10 mammalian orthologues each. In confirmation of the validity of our coevolution proxy, we found positive associations with the nonsynonymous or amino acid substitution rate in all protein sets. However, our coevolution proxy negatively correlated with the number of protein interactants (node degree) in male reproductive protein sets alone. In addition, a negative association of our coevolution proxy with protein hydrophobicity was significant in sperm proteins only. Accordingly, the restrictive effect of protein interactants was most pronounced in male reproductive proteins, and the tendency of sperm proteins to form internal structures decreased the more coevolutionary sites they had. Both aspects illustrate that the share of outward and thus functional coevolution increases with greater proximity to fertilization. We found this conclusion confirmed by additional comparisons within sperm proteins. Thus, sperm proteins with high hydrophobicity had the lowest proportions of covarying sites and, according to gene annotations, localized more frequently to internal cellular structures. They should therefore be less exposed to postcopulatory forms of sexual selection. Their counterparts with low hydrophobicity had larger proportions of covarying sites and more often resided at the cell membrane or were secreted. At the cellular level, they are thus closer to externally induced forces of postcopulatory selection which are known for their potential to increase substitution rates. In addition, we show that the intermediary status of the testicular protein set in correlation analyses is probably due to a special combination of reproductive and somatic involvements.

karger@karger.com
www.karger.com/cgr

Karger

Holger Herlyn
Institute of Organismic and Molecular Evolution (iomE), Anthropology
University of Mainz, Anselm-Franz-von-Bentzel-Weg 7
DE–55128 Mainz (Germany)
herlyn@uni-mainz.de

Numerous male reproductive proteins show elevated rates of sequence evolution [e.g., Torgerson et al., 2002; Clark and Swanson, 2005; Haerty et al., 2007]. The probable driving force behind is commonly seen in the various forms of postcopulatory sexual selection, including sperm competition [e.g., Lüke et al., 2014; Ramm et al., 2014; Schumacher et al., 2014], cryptic female choice [Gasparini and Pilastro, 2011; Løvlie et al., 2013], and sexual conflict [Clark et al., 2009; Sirot et al., 2014]. Prominent examples of male reproductive genes under postcopulatory sexual selection are primate and rodent protamines and semenogelins, whereby the first replace histones in the sperm head [Wyckoff et al., 2000; Ramm et al., 2008; Lüke et al., 2014] and the latter polymerize to a copulation plug in the female genital tract [Jensen-Seaman and Li, 2003; Dorus et al., 2004; Ramm et al., 2008]. Also, vertebrate sperm proteins like zonadhesin that bind to the zona pellucida or the oocyte are known for elevated rates of sequence divergence [Swanson et al., 2003; Herlyn and Zischler, 2006; Dorus et al., 2010; Claw et al., 2014]. In fact, the closer male reproductive proteins are functionally related to fertilization, the higher seems to be their rate of sequence evolution. In contrast, proteins expressed in the male reproductive tract show considerable sequence conservation when they are only indirectly involved in reproduction [Dean et al., 2009; Ramm et al., 2009; Dorus et al., 2010; Schumacher et al., 2014].

The interactions that a protein engages in belong to the main factors determining the rate of sequence evolution. Thus, interactions can imply an increase in substitution rates as known for coevolving proteins in phages and bacteria, and for receptor-ligand systems in the reproduction of diverse metazoans [Swanson and Vacquier, 2002; Herlyn and Zischler, 2008; Paterson et al., 2010]. In the first case, the evolutionary race is probably driven by selective advantages for phages and bacteria with higher infectivity and resistance, respectively [Paterson et al., 2010]. Sperm competition, female choice, and polyspermy avoidance could in turn explain elevated fixation rates in reproductive proteins [Anderson and Dixson, 2002; Swanson and Vacquier, 2002; Herlyn and Zischler, 2007; Clark et al., 2009]. However, apart from such systems, interactions usually have a constraining impact on sequence evolution [Fraser et al., 2002; Krylov et al., 2003; Avila-Herrera and Pollard, 2015]. Thereby, the detailed impact of a protein-protein interaction (PPI) on sequence evolution might be modulated by various biophysical and expressive factors [Bloom and Adami, 2003; Jordan et al., 2003; Drummond et al., 2005; Franzosa and Xia, 2009; Wilke and Drummond, 2010]. But in general, additional PPIs imply greater functional constraint and thus stronger evolutionary conservation [reviewed in Jancura and Marchiori, 2012]. Accordingly, the number of PPIs is a main constraining factor in the evolution of mammalian testicular and sperm proteins too [e.g., Schumacher et al., 2017; Schumacher and Herlyn, 2018].

The ultimate cause behind the conserving effect of PPIs could be that highly connected proteins, even when expressed in the male reproductive tract, gain viability importance through their commitments in nonreproductive tissues and organs [Jeong et al., 2001; Krylov et al., 2003; Hahn and Kern, 2005; Schumacher et al., 2017]. Yet, the proximate cause seems to be that any amino acid exchange in a highly connected protein is more likely to fall into an interacting region [Fraser et al., 2002; Lovell and Robertson, 2010; Avila-Herrera and Pollard, 2015]. Thereby, the area involved in binding may well extend beyond the directly interacting amino acids [Kastritis et al., 2014]. In any case, new residues have usually a negative impact on one or more existing PPIs or lead to costly, sometimes even toxic, misinteractions [Zhang et al., 2008; Wilke and Drummond, 2010; Yang et al., 2012; Chen et al., 2013; Montiel-García et al., 2016]. The new residue may occasionally be neutral or slightly disadvantageous on its own but increases the probability of a compensatory substitution at another amino acid site [Avila-Herrera and Pollard, 2015]. But only rarely it will have a favorable effect on protein functionality, and the corresponding mutation will spread through positive selection [Gong et al., 2009; Lovell and Robertson, 2010; Qian et al., 2011].

The principles outlined above for interactions between proteins also apply to amino acids that interact within an individual protein. This may be sites that are neighbored in the primary sequence or adjacent in the folded protein [Süel et al., 2003; Buck and Atchley, 2005; Chakrabarti and Panchenko, 2009, 2010; Hopf et al., 2014], but also distal sites connected by a chain of coevolving residues [Burger and van Nimwegen, 2010]. Some of the coevolving amino acid sites will determine protein functionality, e.g., by binding to a substrate, a regulatory ligand, or an interacting protein, while others will have higher importance for the foldability, structure, and stability of a protein [Chakrabarti and Panchenko, 2009, 2010; Marks et al., 2011; Morcos et al., 2011; Sandler et al., 2014]. However, whether they have more functional or structural relevance, most exchanges will have a negative impact on the interaction with the interacting residues of the intramolecular network [Benkovic and Hammes-Schiffer, 2003; Süel et al., 2003]. Thus, highly connected sites within proteins have less freedom to acquire a new amino acid as detailed above for entire proteins. Probably, for this reason, amino acid sites

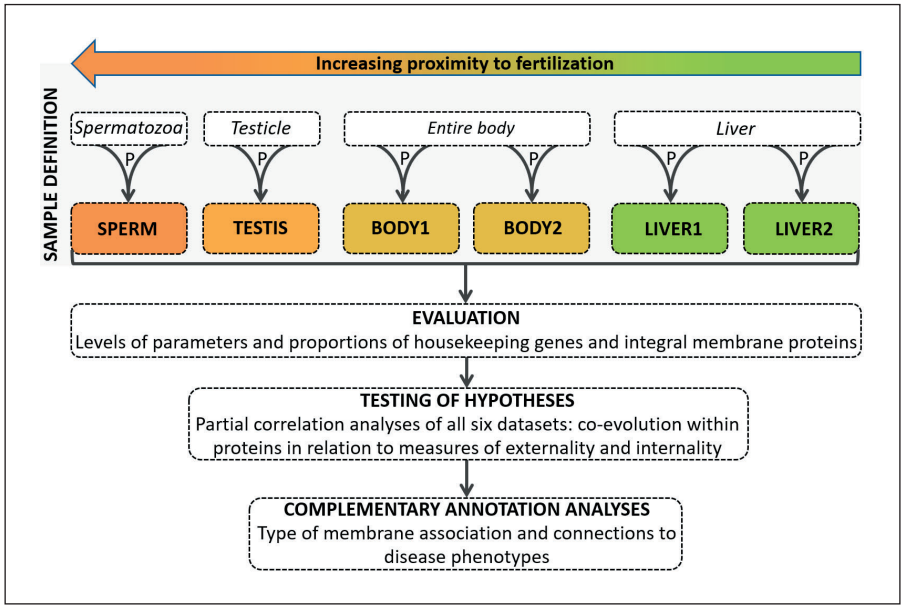

Fig. 1. Study workflow. The proteins in our 6 sets were expressed in human spermatozoa (SPERM), testicle (TESTIS), entire body (BODY1, BODY2), and liver (LIVER1, LIVER2), thus reflecting differential functional proximity to fertilization. Subsequent data evaluation ensured similar proportions of proteins encoded by housekeeping genes and proteins with transmembrane domains and similar parameter levels across the protein sets studied. Analyses of annotations completed the study. More details are given in Materials and Methods. P, proteins and parameters.

with many intramolecular connections are evolutionarily more conserved than less linked ones [del Sol et al., 2006; Lee et al., 2008; Zhou et al., 2008; Chakrabarti and Panchenko, 2009, 2010; Hopf et al., 2015]. However, compensatory exchanges at coevolving sites will occasionally enable the maintenance of structural-functional integrity [Buck and Atchley, 2005; Camps et al., 2007; Chakrabarti and Panchenko; 2009]. Even if both levels, the PPI network and the intramolecular network, were initially looked at separately here, they naturally interact with each other [Saraf and Maranas, 2003; Sandler et al., 2014].

Intramolecular coevolution of amino acid sites has repeatedly been studied for improving predictions on protein structure and function [Kyte and Doolittle, 1982; Dee et al., 2002; Worth et al., 2009; Sandler et al., 2014]. A focus was thereby on bacterial proteins, of which crystallographic data were available [e.g., Marks et al., 2011]. To our knowledge, however, patterns of intramolecular coevolution have not yet been compared between male reproductive proteins and proteins with a stronger somatic relevance. This includes a lack of information on possible differences in the intramolecular coevolution of sperm and testicular proteins. Such differences seem possible, if not likely, given previous reports of increasing substitution rates with growing proximity to fertilization [Dean et al., 2009; Ramm et al., 2009; Dorus et al., 2010; Schumacher et al., 2014]. Those investigations, though, focused on general substitution rates, while the question of coevolution within proteins was not addressed. Correspondingly, it is unknown to date whether the share of functional and structural coevolution within proteins might vary depending on how closely they are involved in fertilization. For elucidating this question, we compared patterns of intramolecular coevolution in mammalian protein sets representing different levels of functional proximity to fertilization. In particular, we analyzed the relationships between the proportion of covarying sites per protein and parameters that approximate the extent to which a protein is involved either with itself or with other proteins. In addition, we compared patterns of coevolution between subsets of sperm proteins representing a gradient of fertilization proximity at the cellular level.

Materials and Methods

Protein Sets

We generated 6 random sets of 77 human proteins each, representing different proximity to fertilization (Fig. 1). With TESTIS and SPERM, 2 sets reflected male reproduction. Moreover, we generated 2 protein sets giving entire body (BODY1, BODY2). Two additional ones represented liver (LIVER1, LIVER2) as an organ which is comparably far from reproduction. We also considered it advantageous that liver is known for the expression of genes showing signatures of rapid divergence [Blekhman et al., 2014]. The first set, SPERM, was drawn from a compilation of the human sperm proteome [Amaral et al., 2014]. The only condition for further consideration of a random-selected gene was its determination by LC-MS/MS (online suppl. Table 1; for all online suppl. material, see www.karger.com/doi/10.1159/000509584). The second set, TESTIS, contained proteins with testicular expression, as verified

by consultation of the Tissue Atlas at www.proteinatlas.org (visited between October 2018 and January 2019; see also LC-MS/MS study by Zhang et al. [2015a]). For generating BODY1 and BODY2, we selected 2 mutually exclusive sets of protein-coding genes from the according list at www.genenames.org (Human Genome Nomenclature Committee). We additionally compiled 2 non-overlapping sets of liver proteins, LIVER1 and LIVER2. These proteins were retrieved from one of the most complete compilations, a LC-MS/MS-based murine liver proteome [Ding et al., 2016]. Expression of the human orthologues in liver was secured by consultation of the Tissue Atlas once more. Matching of gene and protein symbols and IDs within and between species was conducted with the aid of Ensembl's data mining tool BioMart (Ensembl Genes 94), accepting only unique hits. If an approved symbol of a human orthologue did not connect to an entry in the Tissue Atlas, we searched HGNC, GeneCards, and STRING for aliases and successively used these as search items.

Parameter Acquisition

For each protein considered, we collected the following 5 parameters (Fig. 1): (i) Node degree (interactivity, connectivity); we reconstructed a PPI network of the human body to capture the pleiotropy of each protein as fully as possible [compare Gong et al., 2009]. The PPIs were collected from BioGrid (version 3.4.141; thebiogrid.org) and from IntAct, APID, MINT, DIP-IMEx, MatrixDB, and InnateDB-IMEx databases in October 2016, using the PSICQUIC plugin in Cytoscape v. 3.4.0 [Shannon et al., 2003]. As we were primarily interested in the potential restriction of a protein through physical interaction with other proteins, we derived the individual node degree for each sampled protein, i.e., the number of direct PPIs per protein (online suppl. Table 2). This was done with the NetworkAnalyzer plugin in Cytoscape. (ii) Proportion of covarying sites; coevolution analysis used amino acid alignments that we generated with the aid of PhyleasProg v. 3.1 (August to October 2017), specifying fine computation level [Busset et al., 2011]. Ensembl orthologues called by PhyleasProg were aligned with PRANK [Löytynoja and Goldman, 2010] and pruned from uncertainly aligned positions with Gblocks [Talavera and Castresana, 2007]. With the species considered, we aimed at a tradeoff between sufficient variability for the analysis of intramolecular coevolution on the one hand, and the avoidance of larger alignment sections being deleted due to too high sequence divergence on the other. For achieving this, we opted out Glires (rodents and lagomorphs) and solely included orthologues from well-annotated genomes of Laurasiatheria and Primates: cat (*Felis catus*), cow (*Bos taurus*), dog (*Canis familiaris*), pig (*Sus scrofa*), and sheep (*Ovis aries*), as well as common chimpanzee (*Pan troglodytes*), gorilla (*Gorilla gorilla*), human (*Homo sapiens*), Rhesus macaque (*Macaca mulatta*), and common marmoset (*Callithrix jacchus*). Using the Laurasiatheria-Primates alignments and the according Ensembl species tree, coevolution analyses were carried out with the aid of the CAPS2 server (www.caps.tcd.ie). The corresponding algorithm seeks for covariation in BLOSUM-corrected amino acid distances at individual alignment positions, whereby potential stochastic and phylogenetic dependencies are removed [Fares and McNally, 2006; Fares and Travers, 2006]. We finally set the number of coevolving amino acid sites as identified by CAPS2 in relation to the total number of amino acids per human protein, thus obtaining an approximation of the proportion of covarying amino acids. (iii) Hydropathy index; we loaded the amino acid sequences of human orthologues into GRAVY (www.gravy-calculator.de) to derive the average hydrophobicity of the contained amino acids, which is equivalent to the hydropathy index of a protein. (iv) dN (nonsynonymous substitution rate); following others [Lin et al., 2007], we focused on dN for studying sequence evolution. In detail, we derived mean dN values from human, murine, and porcine orthologues as retrieved from Ensembl 94 database with the aid of the BioMart data mining tool. We used Ensembl protein stable IDs as query terms and only accepted unique hits. (v) Gene expression level; this parameter was included as a variable to be corrected against in partial correlation analyses [compare Drummond et al., 2005; Worth et al., 2009; Yang et al., 2012]. We especially corrected for the transcript level, since this is more closely related to substitution rate than protein abundance [Eames and Kortemme, 2007]. In particular, we derived medians of expression levels in 27 human tissues from previously published RNA-seq data [Kryuchkova-Mostacci and Robinson-Rechavi, 2015].

Housekeeping Genes and Transmembrane Proteins

For each protein, we determined if it is encoded by a housekeeping gene or not. Using BioMart (Ensembl Genes 96), we first retrieved Ensembl proteins stable IDs for the housekeeping genes reported by Eisenberg and Levanon [2013]. These were genes, for which more than half of the exons of at least 1 RefSeq transcript showed largely constant expression levels in 16 human tissues. Subsequently, we matched the corresponding IDs with the Ensembl proteins stable IDs contained in our 6 protein sets. For the identification of multi-pass and single-pass membrane proteins, we manually checked annotations at the UniProtKB website (www.uniprot.org; November 2019) for reviewed entries referring to human orthologues.

Statistical Tests and Correction for Parallel Testing

Initial tests ensured that downstream correlation analyses were not compromised by confounding factors (Fig. 1). For this purpose, we compared parameter levels (see above: i–v) with the Mann Whitney U (MWU) test implemented in SPSS 23 v. 5 (IBM). Distributions of proteins encoded/non-encoded by housekeeping genes, of multi-pass/non-multi-pass membrane proteins, and of transmembrane/non-transmembrane proteins were compared with an interactive χ^2 test (quantpsy.org/chisq/chisq.htm). The corresponding pairwise comparisons were made between SPERM and the other 5 datasets and between SPERM and TESTIS together and the other 4 datasets.

Subsequently, 4 partial correlation analyses were carried out for each set of genes and proteins (Fig. 1). For enabling SPSS to infer 95% confidence intervals (CIs) of correlation coefficients within the framework of rank correlation, parameter values were first converted into ranks for each of the 6 protein sets. In detail, we tested whether the proportion of covarying sites per protein related to the general amino acid substitution rate as given by dN. We also evaluated if our externality proxy, node degree, was negatively associated with dN. Additional correlation analyses addressed the key question of the present study, namely if our gene or protein sets differ in the extent to which intramolecular coevolution is directed outward or inward. To this end, we associated the proportion of covarying sites per protein with node degree and with our proxy of internality, the hydropathy index. We expected the restricting and relaxing effects of more and less PPIs, respectively, to be particularly evident in a set when proteins have a larger share of outward and thus functional

coevolution. A larger proportion of functional coevolution should also be reflected in a decreased propensity of proteins to form internal structures. Correspondingly, an anti-correlation of the proportion of covarying sites with the hydropathy index would reinforce a stronger outward orientation of coevolution.

Altogether, we carried out 40 tests: 10 MWU tests (levels of 5 parameters in SPERM and TESTIS versus the other 4 sets, and in SPERM versus the other 5 sets), 6 χ^2 tests (distributions of housekeeping and non-housekeeping genes, multi-pass and non-multi-pass membrane proteins, and transmembrane and non-transmembrane proteins; each compared in SPERM and TESTIS versus the other 4 sets, and in SPERM versus the other 5 sets), and 24 partial correlations (4 hypotheses tested per each of 6 protein sets). All tests conducted were two-sided. In order to account for parallel testing, we converted *p* values into false discovery rate (FDR) values applying the method of Benjamini and Hochberg [1995].

Descriptive Approaches

For validating the informative nature of hydropathy index, we checked annotations at UniProtKB (as given under UniProt annotation and GO – Cellular component; December 2018) for associations of proteins with membranes. The corresponding data were collected for proteins in SPERM since only in this dataset hydropathy was adversely correlated with the coevolution proxy. We especially focused on the 20 SPERM proteins with highest hydropathy values (because only these contained transmembrane proteins) and their 20 counterparts with lowest hydropathy scores. We continued with an evaluation of UNiProKB annotations (as given under Function and GO – Molecular Function; May 2020) for validating the representative nature of all 6 datasets. For doing so, we gathered terms suggesting greater functional relevance for the uptake and discharge of substances (search items: exocyt, endocyt, vesic, lysosome, proteasome, exocr, secre, secern). Recording also included proteins with increased relevance for signaling (search items: signal, cytokine, plus references to hormonal signaling) and especially for hormonal signaling (search items: hormon, andro, estro, prosta, insul, cortico, steroid, testost, endocr, follicl; exclusively follicle development). Furthermore, we collected references to spermatogenesis, testis, and spermatozoa (search items: spermato, acrosome, test; excluding testost). In addition, we searched the Human-Mouse: Disease Connection database of Mouse Genome Informatics (MGI 6.13; www.informatics.jax.org) for connections between the genes in our 6 sets and phenotypes involving aberrations in the reproductive system and in mortality and aging (exclusively phenotypes reported for transgenic mice). Finally, it should be noted that in the text below, gene symbols will follow protein symbols only if the corresponding acronyms differ.

Results

Similar Parameter Levels and Similar Representation of Housekeeping Genes and Membrane-Associated Proteins

After having collected parameters for our equally sized protein sets, representing human spermatozoa (SPERM), testis (TESTIS), body (BODY1, BODY2), and liver (LIVER1, LIVER2), we tested for potential skews in their composition (Fig. 1). However, the percentage of proteins encoded by housekeeping genes in SPERM (19%) and TESTIS (18%) was within the range found in the other sets, BODY1 (16%), BODY2 (18%), LIVER1 (31%), and LIVER2 (23%) (online suppl. Table 3). Correspondingly, the null hypothesis assuming even distribution of housekeeping and non-housekeeping genes could not be refuted, neither in the comparison of the 2 male reproductive protein sets, SPERM and TESTIS, with the 4 remaining sets, nor in the comparison of SPERM with the other 5 sets (FDR > 0.050 each; χ^2 test). Also, multi-pass membrane proteins alone and transmembrane proteins as a whole (multi- and single-pass) had similar frequencies in SPERM (8 and 16%, respectively), TESTIS (10 and 14%), BODY1 (13 and 18%), BODY2 (14 and 18%), LIVER1 (9 and 14%), and LIVER2 (8 and 17%) (online suppl. Table 3). Therefore, the null hypotheses again could not be rejected in any of the 2 decisive comparison pairs (FDR > 0.050 each; χ^2 test). In addition, expression levels were similar. Thus, the MWU test did not support different levels of expression between the 2 sets of male reproductive proteins and the other 4 protein sets (FDR > 0.050), and also not between SPERM and the other 5 ones (FDR > 0.050, MWU test). The null hypothesis of equality could also not be rejected for dN, node degree, the proportion of covarying sites, and hydropathy index in any of the 2 pairwise comparisons considered relevant (FDR > 0.050 each; MWU test). Thus, skewed protein composition or parameter levels should not have compromised downstream comparisons between protein sets.

Persistent Correlation of dN with the Proportion of Covarying Sites

Any proxy of intramolecular coevolution should positively correlate with the amino acid substitution rate since the first is part of the latter. As an approximation to this question, we correlated the proportion of covarying sites with the nonsynonymous substitution rate, dN. Thereby, we kept constant the potential influence of the other parameters, node degree, hydropathy, and median expression level. In doing so, we found the proportion of covarying sites to be tightly associated with dN in all 6 datasets (FDR < 0.001 each; Table 1). Accordingly, the proportion of covarying sites is indeed informative with respect to intramolecular coevolution. However, the finding also illustrates that it would not be expedient to correct against the influence of one of the 2 evolutionary proxies in correlations of the other one with a third variable. In fact, such a procedure would foreseeably have erased relevant signal. For this reason, we never included both evolution-

Table 1. Results of partial correlation analyses

Protein set	Pair of comparison	Parameters corrected for	Coefficient (95% CI: lower, upper value)	FDR
SPERM	**dN-coevolution**	**hydropathy, expression, node degree**	**0.911 (0.848, 0.954)**	**<0.001**
TESTIS	**dN-coevolution**	**hydropathy, expression, node degree**	**0.918 (0.865, 0.957)**	**<0.001**
BODY1	**dN-coevolution**	**hydropathy, expression, node degree**	**0.899 (0.828, 0.946)**	**<0.001**
BODY2	**dN-coevolution**	**hydropathy, expression, node degree**	**0.913 (0.859, 0.954)**	**<0.001**
LIVER1	**dN-coevolution**	**hydropathy, expression, node degree**	**0.836 (0.748, 0.900)**	**<0.001**
LIVER2	**dN-coevolution**	**hydropathy, expression, node degree**	**0.846 (0.724, 0.932)**	**<0.001**
SPERM	**coevolution-node degree**	**hydropathy, expression**	**–0.420 (–0.592, –0.227)**	**0.001**
TESTIS	**coevolution-node degree**	**hydropathy, expression**	**–0.390 (–0.562, –0.200)**	**<0.010**
BODY1	coevolution-node degree	hydropathy, expression	0.006 (–0.208, 0.213)	>0.050
BODY2	coevolution-node degree	hydropathy, expression	–0.171 (–0.385, 0.054)	>0.050
LIVER1	coevolution-node degree	hydropathy, expression	–0.259 (–0.470, –0.028)	>0.050
LIVER2	coevolution-node degree	hydropathy, expression	–0.113 (–0.348, 0.134)	>0.050
SPERM	**dN-node degree**	**hydropathy, expression**	**–0.413 (–0.591, –0.216)**	**<0.010**
TESTIS	dN-node degree	hydropathy, expression	–0.254 (–0.454, –0.025)	>0.050
BODY1	dN-node degree	hydropathy, expression	0.034 (–0.180, 0.245)	>0.050
BODY2	dN-node degree	hydropathy, expression	–0.205 (–0.414, 0.016)	>0.050
LIVER1	dN-node degree	hydropathy, expression	–0.231 (–0.446, –0.004)	>0.050
LIVER2	dN-node degree	hydropathy, expression	–0.143 (–0.356, 0.091)	>0.050
SPERM	**coevolution-hydropathy**	**node degree, expression**	**–0.354 (–0.529, –0.152)**	**<0.010**
TESTIS	coevolution-hydropathy	node degree, expression	–0.112 (–0.322, 0.107)	>0.050
BODY1	coevolution-hydropathy	node degree, expression	0.010 (–0.223, 0.243)	>0.050
BODY2	coevolution-hydropathy	node degree, expression	0.009 (–0.224, 0.245)	>0.050
LIVER1	coevolution-hydropathy	node degree, expression	0.133 (–0.094, 0.349)	>0.050
LIVER2	coevolution-hydropathy	node degree, expression	0.114 (–0.119, 0.334)	>0.050

For a definition of protein sets (77 proteins each), see Materials and Methods. Bold type highlights significant results. CI, confidence interval; FDR, false discovery rate.

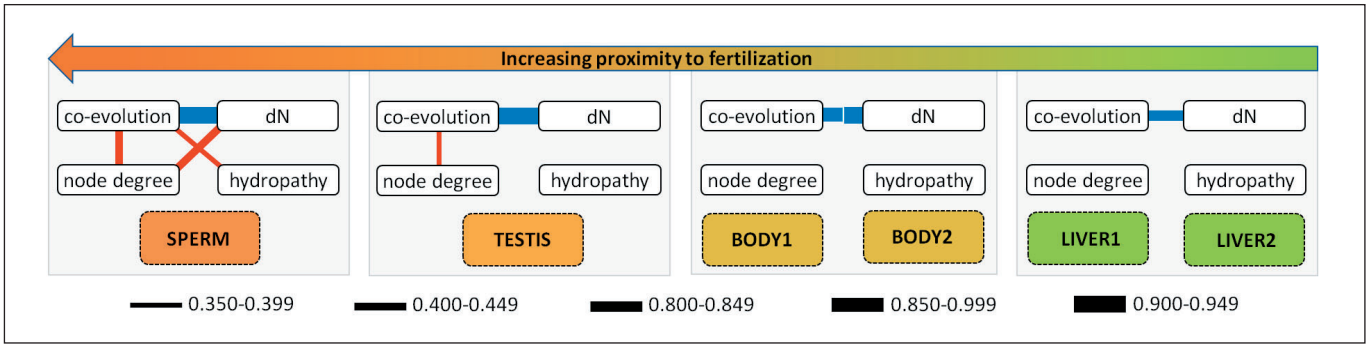

Fig. 2. Partial rank correlations between parameter pairs. Blue bars symbolize positive and red bars negative interrelations between parameters (FDR < 0.05). The scheme illustrates an overall trend for growing absolute values of correlation coefficients (see thickness of bars), accompanied by an increasing number of significant associations, with greater proximity to fertilization (arrow). Detailed values and 95% confidence intervals of correlation coefficients as well as control variables are specified in Table 1. Labels correspond to the parameters correlated, i.e., the number of protein interactants a protein has in a body-wide network (node degree), the proportion of covarying sites per protein which approximates the rate of intramolecular coevolutiuon (co-evolution), the nonsynonymous substitution rate (dN), and average hydrophobicity of the amino acids contained in a human protein (hydropathy). For more details on the parameters, see Materials and Methods.

Table 2. Structural annotations as gathered from UniProtKB: 20 SPERM proteins with high hydropathy indices

HyPa	Ensembl ID of human protein	Gene symbol	UniProt annotation referring to membranes	GO cellular component referring to membranes
0.512	ENSP00000317000	*TM6SF1*	multi-pass membrane protein, lysosomal membrane	lysosomal membrane, integral component of membrane
0.313	ENSP00000266682	*SLC6A15*	multi-pass membrane protein	integral component of membrane, plasma membrane
0.165	ENSP00000351717	*ABCB8*	multi-pass membrane protein, mitochondrial inner membrane	mitochondrial envelope, mitochondrial inner membrane, integral component of membrane, membrane
0.151	ENSP00000394405	*PRXL2B*		ER, extracellular exosome, cytoplasm
0.057	ENSP00000309430	*PIGS*	multi-pass membrane protein, ER membrane	ER membrane, membrane
–0.054	ENSP00000367721	*NUP160*	nuclear pore complex	nuclear envelope, nuclear pore, nuclear pore outer ring
–0.055	ENSP00000364133	*TGFBR1*	cell membrane, single-pass type I, membrane protein, membrane raft	integral component of plasma membrane, plasma membrane, intracellular membrane -bounded organelle, membrane, membrane raft
–0.084	ENSP00000209873	*AAAS*	nuclear pore complex	nuclear envelope, nuclear membrane, nuclear pore, membrane
–0.098	ENSP00000258091	*CCT7*	cytoplasm	
–0.114	ENSP00000321584	*IMPDH2*		peroxisomal membrane, membrane
–0.123	ENSP00000402608	*CPS1*		mitochondrial inner membrane
–0.134	ENSP00000261615	*DPEP1*	apical cell membrane, microvillus membrane	apical plasma membrane, microvillus membrane, plasma membrane, anchored component of cell membrane
–0.138	ENSP00000238561	*ADCK1*	secreted	
–0.138	ENSP00000317159	*CYC1*	mitochondrial inner membrane, single-pass membrane protein, inner membrane side	mitochondrial inner membrane, mitochondrial respiratory chain complex III, integral component of membrane, membrane
–0.165	ENSP00000250244	*AP1M2*	clathrin-coated membrane, peripheral membrane protein, cytoplasmic side	Golgi membrane, trans-Golgi network membrane, lysosomal membrane, clathrin-coated vesicle membrane, cytoplasmic vesicle membrane
–0.168	ENSP00000241041	*PEX16*	peroxisome membrane, multi-pass membrane protein	ER membrane, integral component of peroxisomal membrane, peroxisomal membrane, membrane
–0.169	ENSP00000249269	*PMPCB*		mitochondrial inner membrane
–0.201	ENSP00000332118	*EPHB3*	cell membrane, single-pass type I membrane protein	integral component of plasma membrane, plasma membrane
–0.204	ENSP00000370839	*SGCB*	single-pass type II membrane protein	integral component of plasma membrane
–0.224	ENSP00000220584	*FDFT1*	ER membrane, multi-pass membrane protein	ER membrane, integral component of membrane

For annotations of all 77 proteins contained in the SPERM sample, see online supplementary Table 4. ER, endoplasmic reticulum. HyPa, hydropathy index.

ary parameters in downstream partial correlations, but only one of them.

Specific Patterns in Male Reproductive Proteins

The results of partial correlation analyses were essentially the same within and between both BODY and both LIVER sets (Table 1). However, there were noteworthy differences between these 4 somatically dominated protein sets and our male reproductive protein sets, SPERM and TESTIS, and also between SPERM and the other 5 sets (Fig. 2). In particular, the physical interaction with proteins in the body-wide network, as given by node degree, exclusively anti-correlated with the proportion of covarying sites in both male reproductive protein sets. The FDR values of the corresponding partial correlations were 0.001 for SPERM and <0.010 for TESTIS, while they were >0.050 in BODY1, BODY2, LIVER1, and LIVER2. In addition, the anti-correlation of node degree and dN was exclusively significant in SPERM (FDR < 0.010). SPERM was also the only protein set showing a significant anti-correlation of our coevolution proxy with the hydropathy index (FDR < 0.010). Thus, SPERM proteins with a higher proportion of covarying sites were less hydrophobic and vice versa (Fig. 2; Table 1). As far as results of correlation analyses were significant, the coefficients ranged between 0.354 and 0.918. If 95% CIs were taken into account, the minimum absolute value was still 0.152, while the maximum was 0.957. The corresponding upper and lower values of 95% CIs always had the same sign (Table 1).

Membrane Association in the Most and Least Hydrophobic SPERM Proteins

Evaluation of UniProtKB annotation data confirmed that the hydropathy index approximately gave the propensity of a protein to form internal structures. Thus, single- and multi-pass membrane proteins were present only within the 20 SPERM proteins with highest hydropathy indices, their exact amount being 10 (Table 2; online suppl. Table 4). In contrast, none of the 20 counterparts with the lowest hydropathy indices were integral to membranes (Table 3). Moreover, 14 of the proteins with highest hydropathy indices were known for their associations with the membranes of organelles such as peroxisomes, lysosomes, unspecified vesicles, nucleus, endoplasmic reticulum, and Golgi apparatus. However, only 4 proteins

Table 3. Structural annotations of 20 SPERM proteins with low hydropathy indices as gathered from UniProtKB

HyPa	Ensembl ID of human protein	Gene symbol	UniProt annotation referring to membranes	GO - cellular component referring to membrane
–0.548	ENSP00000355237	*CDC42BPB*	cell membrane, peripheral membrane protein, cytoplasmic side	plasma membrane
–0.570	ENSP00000362590	*TBC1D22B*		
–0.572	ENSP00000415430	*GTSE1*	membrane	
–0.588	ENSP00000340093	*NAPEPLD*	early endosome membrane, Golgi apparatus membrane, peripheral membrane protein, nucleus envelope	early endosome membrane, photoreceptor outer segment membrane, membrane-bounded organelle
–0.606	ENSP00000286800	*BACH1*		
–0.614	ENSP00000260187	*USP2*		membrane
–0.630	ENSP00000303058	*CEP120*		
–0.675	ENSP00000304895	*IRS1*	plasma membrane	intracellular membrane-bounded organelle
–0.696	ENSP00000420854	*EFCAB12*		
–0.708	ENSP00000359799	*DNAJB4*	cell membrane	plasma membrane
–0.752	ENSP00000358045	*ECM1*	extracellular region or secreted	extracellular matrix etc
–0.784	ENSP00000252137	*ESS2*		
–0.817	ENSP00000333666	*ADI1*	cell membrane, peripheral membrane protein, cytoplasmic side	plasma membrane
–0.824	ENSP00000417653	*DBNL*	peripheral membrane protein, cell membrane, cytoplasmic side, clathrin-coated vesicle membrane	
–0.833	ENSP00000333024	*PHF7*		plasma membrane
–0.913	ENSP00000264708	*POMC*	secreted	
–0.923	ENSP00000438262	*TJP2*	cell membrane, peripheral membrane protein	plasma membrane
–0.969	ENSP00000347161	*TSGA10*		
–1.115	ENSP00000342121	*RNF6*		nuclear membrane, intracellular membrane-bounded organelle
–1.233	ENSP00000345917	*LYAR*		

For annotations of all 77 proteins contained in the SPERM sample, see online supplementary Table 4. HyPa, hydropathy index.

Table 4. Representation of functional annotation categories per protein set as retrieved from UniProtKB

Protein set	Uptake and discharge	Signaling	Hormonal signaling	Spermatogenesis, testis, spermatozoa
SPERM	4	12	3	2
TESTIS	6	14	8	4
BODY1	5	14	1	2
BODY2	6	15	4	2
LIVER1	10	10	1	2
LIVER2	8	8	2	0

For definitions of protein sets (77 proteins each) and annotation categories (uptake and discharge, signaling etc.), see Materials and Methods. Online supplementary Table 5 gives the detailed annotations for each individual protein in our 6 protein sets.

of the group with lowest hydropathy indices showed any of these associations. Finally, while 6 sperm proteins in the group with higher hydropathy indices were reported as localizing to the cellular membrane or being secreted, a total of 9 sperm proteins of the group with low hydrophobicity were of such type. Thus, SPERM proteins with higher hydropathy scores tended to reside inside the cell. In contrast, their hydrophilic counterparts localized more frequently to the cell membrane and extracellular compartments.

Protein Functions and Disease Connections

According to UniProtKB annotations, LIVER1 (n = 10; e.g., MON1A) and LIVER2 (8; e.g., ATP7A) each included more proteins engaging in the uptake and discharge of substances than the other protein sets (4–6). Furthermore, with 14, TESTIS contained as many proteins with higher importance for signaling as BODY1 (e.g., AKT2) and only one less than BODY2 (e.g., M3K2/*MAP3K2*). However, when focusing on hormonal signaling, the maximum was shown by TESTIS (8), as exemplified by FSHR and ACTHR (*MC2R*). In contrast, only 1–4 such connections existed for other protein sets. With a total of 4, TESTIS also had the most annotations, which explicitly referred to proteins functioning in testis, spermatogenesis, or spermatozoa (e.g., APOA1, PSB4/*PSMB4*). In contrast, there were only half as many or no such references in the other protein sets, thereunder VATE2 (*ATP6V1E2*) and TSG10 (*TSGA10*) in SPERM (Table 4; online suppl. Table 5). Complementary screening of the Human-Mouse: Disease Connection database at MGI revealed 60 associations of TESTIS genes with

phenotypes affecting mortality and aging (41) and the reproductive system (19) (online suppl. Table 6). There were also more connections to somatic than reproductive system phenotypes in the other protein sets. Nevertheless, compared to TESTIS, their total number (34–46) was persistently lower in SPERM, BODY1, BODY2, LIVER1, and LIVER2.

Discussion

Larger Share of Functional Coevolution with Greater Proximity to Fertilization

Tight associations with dN in all datasets illustrate that the proportion of covarying sites per protein can also be conceived as a coevolution proxy. This was to be expected, since covariation presupposes coevolutionary exchanges which in turn contribute to dN. Nevertheless, the finding corroborates that our coevolution proxy is valid, and this should apply accordingly to the other correlation results obtained under its involvement. This is especially true for the anti-correlation with node degree, which was significant in both male reproductive protein sets only, whereby significance level and nominal coefficient were higher in SPERM than in TESTIS (Fig. 2; Table 1). Thus, the interrelatedness of both variables increased with growing proximity to fertilization, which we interpret as an indication of a shift in the relative shares of natural and sexual selection. In fact, it seems reasonable to assume that natural selection prevails in the body-wide proteome, and the same should be true for the liver proteome [e.g., Bersaglieri et al., 2004; Qiu et al., 2008; Blekhman et al., 2014]. Compared to this, the impact of postcopulatory forms of sexual selection should increase with greater closeness to fertilization [Gasparini and Pilastro, 2011; Løvlie et al., 2013; Lüke et al., 2014; Ramm et al., 2014; Sirot et al., 2014; Zhou et al, 2015]. At the molecular level, present results would be in accordance with a growing proportion of outward coevolution with greater proximity to fertilization. In this explanatory model, additional PPIs in male reproductive proteins imply a higher restriction of coevolving sites engaging in these interactions. In turn, relaxation of functional constraint due to fewer interactions will allow for more interdependent exchanges, which might occasionally pave the way for postcopulatory forms of sexual selection. Thus, in one way or the other, coevolution within male reproductive proteins is determined by external forces.

Confirmation of the Larger Pattern in SPERM Proteins

Proximity to fertilization seems also to be behind the anti-correlation of our coevolution proxy with hydropathy index in SPERM (Fig. 2; Table 1). Thus, a higher hydropathy index implies an elevated propensity of a protein to engage in internal structures [Kyte and Doolittle, 1982; Dee et al., 2002; Clark et al., 2009; Worth et al., 2009]. Accordingly, SPERM proteins with higher hydropathy scores were disproportionately often annotated as membrane proteins, in particular of organelles and other internal structures. In contrast, the least hydrophobic sperm proteins, which also had more coevolving sites, were more frequently reported as being secreted or associated with the cell membrane (Tables 2, 3; online suppl. Table 4). They should therefore be closer to postcopulatory forms of sexual selection known for their potential to increase substitution rates (for references, see above). This might happen against the background of relaxed functional constraint as already mentioned above. In any case, the negative association with hydropathy seems to contribute to the overall picture, according to which the proportion of functional coevolution increases with greater proximity to fertilization. Precisely because this is the case, the patterns seem to be more strongly reflected in the coevolving sites than in the amino acid sites as a whole. After all, the correlation of node degree and dN was only significant in SPERM (Fig. 2; Table 1).

General Characteristics of Protein Sets and TESTIS Specialties

Functional annotations at UniProtKB reinforced the representative nature of our protein sets (Table 4; online suppl. Table 5). This already emerged from very similar frequencies of functional categories in both LIVER and both BODY protein sets, but also in the annotations per se. Thus, the central role of liver in metabolism [Chiang, 2014] was evident in high frequencies of proteins with closer connections to cellular uptake and secretion in both LIVER sets, thereunder MON1A and ATP7A [White et al., 2009; Jin et al., 2020]. In both BODY sets, numerous connections to diverse signaling pathways (e.g., M3K2/*MAP3K2*, AKT2, TKNK/*TAC3*) were in accordance with the complexity of the entire organism [Cheng et al., 2000; Sakamoto et al., 2006; Tusset et al., 2012]. Also, the fact that the annotations for SPERM did not contain specificities is unsurprising: After all, many sperm-expressed proteins exert their multiple functions in diverse tissues [Schumacher and Herlyn, 2018]. Nevertheless, SPERM also exhibited clear connections to re-

production, as exemplified by TSG10 (*TSGA10*), which is mainly expressed in sperm tail, and VATE2 (*ATP6V1E2*; alias *VMA4*), which participates in acrosomal acidification [Sha et al., 2018; Futai et al., 2019]. However, most such associations existed for the protein set representing the organ of spermiogenesis, TESTIS. Among them were proteins like PSB4 (*PSMB4*), which acts in protein degradation, and APOA1, which has been implicated in sperm motility activation [Aakerlöf et al., 1991; Agarwal et al., 2020].

It was less predictable that the number of signaling connections in TESTIS would be in the high range of the BODY protein sets (Table 4; online suppl. Table 5). This could reflect the entanglement of testes in a body-wide crosstalk between organs, as revealed by the even most abundant links to hormonal signaling in TESTIS. In fact, testes are important endocrine glands producing testosterone and inhibin, which are essential for the development of the male phenotype. This occurs primarily under the control of 2 gonadotropins produced in adenohypophysis, follicle stimulating hormone (FSH) and luteinizing hormone, to which testicles are responsive [Dada et al., 1983; Ramaswamy and Weinbauer, 2015]. This responsiveness involves FSH receptor (FSHR) as one of the proteins contained in TESTIS. The receptor for adrenal corticotropin hormone, ACTHR (*MC2R*), is another TESTIS protein illustrating the testicular interaction with other body tissues [O'Shaughnessy et al., 2003]. However, the embedding of germ line into the soma is not confined to the external relationships of testes, but also extends to the internal testicular anatomy. Thus, Sertoli, Leydig, and other somatic cell types provide the structural and physiological framework for spermiogenesis [Weinbauer et al., 2001]. In addition to the annotations, the dual nature of the testicular proteome was evident in the maximum number of connections to disease phenotypes of mortality and aging and the reproductive system in TESTIS (online suppl. Table 6). Consistently, it was previously established that opposing selection pressures left signatures in the genes coding for testicular proteins, depending on their primary importance for reproduction or viability [Schumacher et al., 2017]. The same testicular specificity should thus be the reason for the intermediary results of correlation analyses in TESTIS, ranging between SPERM and the 4 physically dominated protein sets (Fig. 2; Table 1).

Validity of Correlation Analyses Results

Since all datasets included the same number of proteins, size effects cannot have biased the results of present partial correlations. It is also unlikely that differences in the composition of the compared protein sets introduced a skew. In fact, high status genes that are essential for survival are known for overall stronger sequence conservation and higher centrality in PPI networks [Jeong et al., 2001; Hahn and Kern, 2005; Wolf et al., 2006]. However, we found no evidence of an unbalanced distribution of functionally more important genes across the 6 protein sets studied. Rather, similar frequencies of housekeeping genes were found in SPERM and the other 5 protein sets and also in SPERM and TESTIS on one hand and the other 4 protein sets on the other. The corresponding genes were previously selected for their broad and largely constant expression levels [Eisenberg and Levanon, 2013]. Despite the fact that similar criteria have been applied in other studies [e.g., de Jonge et al., 2007], there are alternative concepts for housekeeping genes that further facilitate the inclusion of low-expressed genes [Zhang et al., 2015b]. However, we do not expect that the application of an alternative housekeeping gene concept would change the results. After all, any such change should affect all protein sets to the same extent.

Our male reproductive protein sets together and SPERM alone were also inconspicuous with regard to the frequency of multi-pass membrane and transmembrane proteins (multi-pass and single-pass). Nor was there any indication that the results of correlation analyses could have been biased by different expression levels which are known for their interrelation with substitution rates [Drummond et al., 2005; Worth et al., 2009; Yang et al., 2012]. But above all, we have nevertheless corrected for the influence of expression levels in all partial correlations. Thereby, we especially corrected for the potential impact of transcript level, which is more tightly connected to the substitution rate of a protein than is protein abundance [Eames and Kortemme, 2007]. Furthermore, node degree, average hydrophobicity of amino acids (hydropathy index), and our 2 evolutionary measures had similar levels in our male reproductive and the other sets. This may surprise in particular with regard to the evolutionary parameters, dN and the proportion of covarying sites, since substitution rates have repeatedly been reported to be elevated in male reproductive proteins [Wyckoff et al., 2000; Swanson et al., 2003; Dorus et al., 2010]. However, it has meanwhile become increasingly clear that the majority of male reproductive proteins are actually evolutionarily conserved [Dean et al., 2009; Dorus et al., 2010; Schumacher et al., 2014].

Parameters such as substitution rates, node degree, and expressional level are known for their interrelation

with other factors, which were not directly tested in the present analyses. An increased level of postcopulatory sexual selection, for example, can entail higher substitution rates of male reproductive proteins [Wyckoff et al., 2000; Dorus et al., 2004; Ramm et al., 2008, 2014; Schumacher et al., 2014]. Also, transient PPIs have a lesser conserving effect on sequence evolution than obligate ones [Mintseris and Weng, 2005]. Additionally, higher phyletic age and broader or early onset of expression associate with higher sequence conservation [Zhang and Li, 2004; Good and Nachman, 2005; Toll-Riera et al., 2012; Schumacher and Herlyn, 2018]. Moreover, longer 5′ and 3′ untranslated regions [Worth et al., 2009; Schumacher and Herlyn, 2018] as well as more posttranslational modifications and higher pleiotropy levels [Macek et al., 2008; Jancura and Marchiori, 2012; Schumacher et al., 2013] associate with lowered substitution rates. But these examples also demonstrate that the various untested variables should ultimately be represented in tested variable, in this case dN. Thus, the patterns that emerged from partial correlations improbably reflected an unrecognized bias in an untested variable.

Conclusion

Our coevolution measure enclosed relevance for structure and function and thus for structural-functional integrity [Buck and Atchley, 2005; Taraban et al., 2008; Worth et al., 2009; Kastritis et al., 2011; Talavera et al., 2011; Chakravarty et al., 2013; Erijman et al., 2014]. Nevertheless, current correlation analyses of physically dominated and male reproductive protein sets suggest that the proportion of functional coevolution increases with greater proximity to fertilization (Fig. 2; Table 1). The same pattern is obtained by comparing the annotations of sperm proteins with high and low hydropathy indices (Tables 2, 3). These findings are consistent with previous reports on rising substitution rates with higher outward orientation of proteins [Julenius and Pedersen, 2006; Toll-Riera et al., 2012; Feyertag et al., 2017]. In addition, proteins with greater spatial or functional proximity to fertilization were already known to diverge at elevated rates [Dean et al., 2009; Ramm et al., 2009; Schumacher et al., 2014]. However, these findings were based on substitutions in general, while the present study focused on coevolving sites. Our results are also consistent with, but not equivalent to, previous evidence for the occurrence of coevolution between proteins, even if these proteins were fertilization proteins [Clark et al., 2009]. This is because we studied coevolution within and not between proteins, specifically in mammals and not in abalones as in the latter study. Most importantly, to our knowledge, the relationship between external factors, i.e., protein interactions, and intramolecular coevolution was not otherwise established for mammalian fertilization proteins. Last but not least, it has not been shown that a particular combination of physical and sexual involvements of the testicular proteome is associated with a particular pattern of intramolecular coevolution (Table 4). As it seems, looking at coevolving sites was particularly revealing in this respect.

Acknowledgements

We would like to thank Dr. Jens Blöcher (iomE, Anthropology, University of Mainz) and Dr. David Rosenkranz (Senckenberg Center for Human Genetics, Frankfurt am Main, Germany) for supporting data collection. We additionally thank Karsten Hohkamp, Ph.D. (Smurfit Institute of Genetics, Trinity College Dublin, Ireland) and Dr. Brian Caffrey (Max Planck Institute for Molecular Genetics, Berlin, Germany) for supporting CAPS2 analyses and sharing valuable knowledge on the theoretical background of covariation analysis. Last but not least, we acknowledge that the inclusion of reviewer comments has improved the manuscript.

Statement of Ethics

Ethical approval is not required for this type of research.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Funding Sources

Conductance of the study was funded by the Johannes Gutenberg University of Mainz and the Deutsche Forschungsgemeinschaft, DFG, to the credit of H.H. (project HE 3487/3: Bioinformatic prediction of sperm protein relevance and validation of the parameters network centrality and substitution rate in men and bulls).

Author Contributions

M.K., A.R.A., and H.H. planned the study, collected and interpreted data, and wrote the manuscript. J.S. reconstructed the PPI network and contributed to data collection. J.S., B.B., and H.Z. participated in the interpretation of data and co-wrote the manuscript. H.H. conducted statistical analyses and supervised the study. All authors read and approved the final manuscript.

References

Aakerlöf E, Jörnvall H, Slotte H, Pousette A: Identification of apolipoprotein A1 and immunoglobulin as components of a serum complex that mediates activation of human sperm motility. Biochemistry 30:8986–8990 (1991).

Agarwal A, Panner Selvam MK, Baskaran S: Proteomic analyses of human sperm cells: understanding the role of proteins and molecular pathways affecting male reproductive health. Int J Mol Sci 21: 1621 (2020).

Amaral A, Castillo J, Ramalho-Santos J, Oliva R: The combined human sperm proteome: cellular pathways and implications for basic and clinical science. Hum Reprod Update 20:40–62 (2014).

Anderson MJ, Dixson AF: Sperm competition: motility and the midpiece in primates. Nature 416:496 (2002).

Avila-Herrera A, Pollard KS: Coevolutionary analyses require phylogenetically deep alignments and better null models to accurately detect inter-protein contacts within and between species. BMC Bioinformatics 16:268 (2015).

Benjamini Y, Hochberg Y: Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc 57: 289–300 (1995).

Benkovic SJ, Hammes-Schiffer S: A perspective on enzyme catalysis. Science 301:1196–1202 (2003).

Bersaglieri T, Sabeti PC, Patterson N, Vanderploeg T, Schaffner SF, et al: Genetic signatures of strong recent positive selection at the lactase gene. Am J Hum Genet 74:1111–1120 (2004).

Blekhman R, Perry GH, Shahbaz S, Fiehn O, Clark AG, Gilad Y: Comparative metabolomics in primates reveals the effects of diet and gene regulatory variation on metabolic divergence. Sci Rep 4:5809 (2014).

Bloom JD, Adami C: Apparent dependence of protein evolutionary rate on number of interactions is linked to biases in protein–protein interactions data sets. BMC Evol Biol 3:21 (2003).

Buck MJ, Atchley WR: Networks of coevolving sites in structural and functional domains of serpin proteins. Mol Biol Evol 22:1627–1634 (2005).

Burger L, van Nimwegen E: Disentangling direct from indirect co-evolution of residues in protein alignments. PLoS Comput Biol 6:e1000633 (2010).

Busset J, Cabau C, Meslin C, Pascal G: PhyleasProg: a user-oriented web server for wide evolutionary analyses. Nucleic Acids Res 39:W479–485 (2011).

Camps M, Herman A, Loh E, Loeb LA: Genetic constraints on protein evolution. Crit Rev Biochem Mol Biol 42: 42:313–326 (2007).

Chakrabarti S, Panchenko AR: Coevolution in defining the functional specificity. Proteins 75: 231–240 (2009).

Chakrabarti S, Panchenko AR: Structural and functional roles of coevolved sites in proteins. PLoS One 5:e8591 (2010).

Chakravarty D, Guharoy M, Robert CH, Chakrabarti P, Janin J: Reassessing buried surface areas in protein-protein complexes. Protein Sci 22:1453–1457 (2013).

Chen J, Sawyer N, Regan L: Protein-protein interactions: general trends in the relationship between binding affinity and interfacial buried surface area. Protein Sci 22:510–515 (2013).

Cheng J, Yang J, Xia Y, Karin M, Su B: Synergistic interaction of MEK kinase 2, c-Jun N-terminal kinase (JNK) kinase 2, and JNK1 results in efficient and specific JNK1 activation. Mol Cell Biol 20:2334–2342 (2000).

Chiang J: Liver physiology: metabolism and detoxification, in McManus LM, Mitchell RN (eds): Pathobiology of Human Disease, pp 1770–1782 (Elsevier, San Diego 2014).

Clark NL, Swanson WJ: Pervasive adaptive evolution in primate seminal proteins. PLoS Genet 1:e35 (2005).

Clark NL, Gasper J, Sekino M, Springer SA, Aquadro CF, Swanson WJ: Coevolution of interacting fertilization proteins. PLoS Genet 5:e1000570 (2009).

Claw KG, George RD, Swanson WJ: Detecting coevolution in mammalian sperm-egg fusion proteins. Mol Reprod Dev 81:531–538 (2014).

Dada MO, Campbell GT, Blake CA: A quantitative immunocytochemical study of the luteinizing hormone and follicle-stimulating hormone cells in the adenohypophysis of adult male rats and adult female rats throughout the estrous cycle. Endocrinology 113:970–984 (1983).

Dean MD, Clark NL, Findlay GD, Karn RC, Yi X, et al: Proteomics and comparative genomic investigations reveal heterogeneity in evolutionary rate of male reproductive proteins in mice (*Mus domesticus*). Mol Biol Evol 26: 1733–1743 (2009).

Dee KC, Puleo DA, Bizios R: Protein-surface interactions, in Dee KC, Puleo DA, Bizios R (eds): An Introduction to Tissue-Biomaterial Interactions, pp 37–52 (Wiley, Hoboken 2002).

de Jonge HJM, Fehrmann RSN, de Bont ESJM, Hofstra RMW, Gerbens F, et al: Evidence based selection of housekeeping genes. PLoS One 2:e898 (2007).

del Sol A, Fujihashi H, Amoros D, Nussinov R: Residue centrality, functionally important residues, and active site shape: analysis of enzyme and non-enzyme families. Protein Sci 15:2120–2128 (2006).

Ding C, Li Y, Guo F, Jiang Y, Ying W, et al: A cell-type-resolved liver proteome. Mol Cell Proteomics 15:3190–3202 (2016).

Dorus S, Evans PD, Wyckoff GJ, Choi SS, Lahn BT: Rate of molecular evolution of the seminal protein gene *SEMG2* correlates with levels of female promiscuity. Nat Genet 36:1326–1329 (2004).

Dorus S, Wasbrough ER, Busby J, Wilkin EC, Karr TL: Sperm proteomics reveals intensified selection on mouse sperm membrane and acrosome genes. Mol Biol Evol 27:1235–1246 (2010).

Drummond DA, Bloom JD, Adami C, Wilke CO, Arnold FH: Why highly expressed proteins evolve slowly. Proc Natl Acad Sci USA 102: 14338–14343 (2005).

Eames M, Kortemme T: Structural mapping of protein interactions reveals differences in evolutionary pressures correlated to mRNA level and protein abundance. Structure 15: 1442–1451 (2007).

Eisenberg E, Levanon EY: Human housekeeping genes, revisited. Trends Genet 29:569–574 (2013).

Erijman A, Rosenthal E, Shifman JM: How structure defines affinity in protein-protein interactions. PLoS One 9:e110085 (2014).

Fares MA, McNally D: CAPS: coevolution analysis using protein sequences. Bioinformatics 22:2821–2822 (2006).

Fares MA, Travers SA: A novel method for detecting intramolecular coevolution: adding a further dimension to selective constraints analyses. Genetics 173:9–23 (2006).

Feyertag F, Berninsone PM, Alvarez-Ponce D: Secreted proteins defy the expression level-evolutionary rate anticorrelation. Mol Biol Evol 34:692–706 (2017).

Franzosa EA, Xia Y: Structural determinants of protein evolution are context-sensitive at the residue level. Mol Biol Evol 26:2387–2395 (2009).

Fraser HB, Hirsh AE, Steinmetz LM, Scharfe C, Feldman MW: Evolutionary rate in the protein interaction network. Science 296:750–752 (2002).

Futai M, Sun-Wada GH, Wada Y, Matsumoto N, Nakanishi-Matsui M: Vacuolar-type ATPase: a proton pump to lysosomal trafficking. Proc Jpn Acad Ser B Phys Biol Sci 95:261–277 (2019).

Gasparini C, Pilastro A: Cryptic female preference for genetically unrelated males is mediated by ovarian fluid in the guppy. Proc Biol Sci 278:2495–2501 (2011).

Gong S, Worth CL, Bickerton GR, Lee S, Tanramluk D, Blundell TL: Structural and functional restraints in the evolution of protein families and superfamilies. Biochem Soc Trans 37: 727–733 (2009).

Good JM, Nachman MW: Rates of protein evolution are positively correlated with developmental timing of expression during mouse spermatogenesis. Mol Biol Evol 22:1044–1052 (2005).

Haerty W, Jagadeeshan S, Kulathinal RJ, Wong A, Ravi Ram K, et al: Evolution in the fast lane: rapidly evolving sex-related genes in *Drosophila*. Genetics 177:1321–1335 (2007).

Hahn MW, Kern AD: Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. Mol Biol Evol 22:803–806 (2005).
Herlyn H, Zischler H: Tandem repetitive D domains of the sperm ligand zonadhesin evolve faster in the paralogue than in the orthologue comparison. J Mol Evol 63:602–611 (2006).
Herlyn H, Zischler H: Sequence evolution of the sperm ligand zonadhesin correlates negatively with body weight dimorphism in primates. Evolution 61:289–298 (2007).
Herlyn H, Zischler H: The molecular evolution of sperm zonadhesin. Int J Dev Biol 52:781–790 (2008).
Hopf TA, Schärfe CP, Rodrigues JP, Green AG, Kohlbacher O, et al: Sequence co-evolution gives 3D contacts and structures of protein complexes. Elife 3:e03430 (2014).
Hopf TA, Morinaga S, Ihara S, Touhara K, Marks DS, Benton R: Amino acid coevolution reveals three-dimensional structure and functional domains of insect odorant receptors. Nat Commun 6:6077 (2015).
Jancura P, Marchiori I: A survey on evolutionary analysis in PPI networks, in Cai W (ed): Protein-Protein Interactions – Computational and Experimental Tools, pp 427–456 (InteChOpen, Rijeka 2012).
Jensen-Seaman MI, Li WH: Evolution of the hominoid semenogelin genes, the major proteins of ejaculated semen. J Mol Evol 57:261–270 (2003).
Jeong H, Mason SP, Barabasi AL, Oltvai ZN: Lethality and centrality in protein networks. Nature 411:41–42 (2001).
Jin W, Ma R, Zhai L, Xu X, Lou T, et al: Ginsenoside Rd attenuates ACTH-induced corticosterone secretion by blocking the MC2R-cAMP/PKA/CREB pathway in Y1 mouse adrenocortical cells. Life Sci 245:117337 (2020).
Jordan IK, Wolf YI, Koonin EV: No simple dependence between protein evolution rate and the number of protein–protein interactions: Only the most prolific interactors tend to evolve slowly. BMC Evol Biol 3:1 (2003).
Julenius K, Pedersen AG: Protein evolution is faster outside the cell. Mol Biol Evol 23:2039–2048 (2006).
Kastritis PL, Moal IH, Hwang H, Weng Z, Bates PA, et al: A structure-based benchmark for protein-protein binding affinity. Protein Sci 20:482–491 (2011).
Kastritis PL, Rodrigues JP, Folkers GE, Boelens R, Bonvin AM: Proteins feel more than they see: fine-tuning of binding affinity by properties of the non-interacting surface. J Mol Biol 426:2632–2652 (2014).
Krylov DM, Wolf YI, Rogozin IB, Koonin EV: Gene loss, protein sequence divergence, gene dispensability, expression level, and interactivity are correlated in eukaryotic evolution. Genome Res 13:2229–2235 (2003).
Kryuchkova-Mostacci N, Robinson-Rechavi M: Tissue-specific evolution of protein coding genes in human and mouse. PLoS One 10:e0131673 (2015).
Kyte J, Doolittle RF: A simple method for displaying the hydropathic character of a protein. J Mol Biol 157:105–132 (1982).
Lee BC, Park K, Kim D: Analysis of the residue-residue coevolution network and the functionally important residues in proteins. Proteins 72:863–872 (2008).
Lin YS, Hsu WL, Hwang JK, Li WH: Proportion of solvent exposed amino acids in a protein and rate of protein evolution. Mol Biol Evol 24:1005–1011 (2007).
Lovell SC, Robertson DL: An integrated view of molecular coevolution in protein-protein interactions. Mol Biol Evol 27:2567–2575 (2010).
Løvlie H, Gillingham MAF, Worley K, Pizzari T, Richardson DS: Cryptic female choice favours sperm from major histocompatibility complex-dissimilar males. Proc Biol Sci 280:20131296 (2013).
Löytynoja A, Goldman N: webPRANK: a phylogeny-aware multiple sequence aligner with interactive alignment browser. BMC Bioinformatics 11:579 (2010).
Lüke L, Campbell P, Varea Sánchez M, Nachman MW, Roldan ERS: Sexual selection on protamine and transition nuclear protein expression in mouse species. Proc Biol Sci 281:20133359 (2014).
Macek B, Gnad F, Soufi B, Kumar C, Olsen JV, et al: Phosphoproteome analysis of *E. coli* reveals evolutionary conservation of bacterial Ser/Thr/Tyr phosphorylation. Mol Cell Proteomics 7:299–307 (2008).
Marks DS, Colwell LJ, Sheridan R, Hopf TA, Pagnani A, et al: Protein 3D structure computed from evolutionary sequence variation. PLoS One 6:e28766 (2011).
Mintseris J, Weng Z: Structure, function, and evolution of transient and obligate protein-protein interactions. Proc Natl Acad Sci USA 102:10930–10935 (2005).
Montiel-García DJ, Mannige RV, Reddy VS, Carrillo-Tripp M: Structure based sequence analysis of viral and cellular protein assemblies. J Struct Biol 196:299–308 (2016).
Morcos F, Pagnani A, Lunt B, Bertolino A, Marks DS, et al: Direct-coupling analysis of residue coevolution captures native contacts across many protein families. Proc Natl Acad Sci USA 108:E1293–1301 (2011).
O'Shaughnessy PJ, Fleming LM, Jackson G, Hochgeschwender U, Reed P, Baker PJ: Adrenocorticotropic hormone directly stimulates testosterone production by the fetal and neonatal mouse testis. Endocrinology 144:3279–3284 (2003).
Paterson S, Vogwill T, Buckling A, Benmayor R, Spiers AJ, et al: Antagonistic coevolution accelerates molecular evolution. Nature 464:275–278 (2010).
Qian W, He X, Chan E, Xu H, Zhang J: Measuring the evolutionary rate of protein-protein interaction. Proc Natl Acad Sci USA 108:8725–8730 (2011).
Qiu H, Taudien S, Herlyn H, Schmitz J, Zhou Y, et al: *CYP3* phylogenomics: evidence for positive selection of *CYP3A4* and *CYP3A7*. Pharmacogenet Genomics 18:53–66 (2008).
Ramaswamy S, Weinbauer GF: Endocrine control of spermatogenesis: Role of FSH and LH/testosterone. Spermatogenesis 4:e996025 (2015).
Ramm SA, Oliver PL, Ponting CP, Stockley P, Emes RD: Sexual selection and the adaptive evolution of mammalian ejaculate proteins. Mol Biol Evol 25:207–219 (2008).
Ramm SA, McDonald L, Hurst JL, Beynon RJ, Stockley P: Comparative proteomics reveals evidence for evolutionary diversification of rodent seminal fluid and its functional significance in sperm competition. Mol Biol Evol 26:189–198 (2009).
Ramm SA, Schärer L, Ehmcke J, Wistuba J: Sperm competition and the evolution of spermatogenesis. Mol Hum Reprod 20:1169–1179 (2014).
Sakamoto K1, Arnolds DE, Fujii N, Kramer HF, Hirshman MF, Goodyear LJ: Role of Akt2 in contraction-stimulated cell signaling and glucose uptake in skeletal muscle. Am J Physiol Endocrinol Metab 291:E1031–1037 (2006).
Sandler I, Zigdon N, Levy E, Aharoni A: The functional importance of co-evolving residues in proteins. Cell Mol Life Sci 71:673–682 (2014).
Saraf MC, Maranas CD: Using a residue clash map to functionally characterize protein recombination hybrids. Protein Eng 16:1025–1034 (2003).
Schumacher J, Herlyn H: Correlates of evolutionary rates in the murine sperm proteome. BMC Evol Biol 18:35 (2018).
Schumacher J, Ramljak S, Asif AR, Schaffrath M, Zischler H, Herlyn H: Evolutionary conservation of mammalian sperm proteins associates with overall, not tyrosine, phosphorylation in human spermatozoa. J Proteome Res 12:5370–5382 (2013).
Schumacher J, Rosenkranz D, Herlyn H: Mating systems and protein-protein interactions determine evolutionary rates of primate sperm proteins. Proc Biol Sci 281:20132607 (2014).
Schumacher J, Zischler H, Herlyn H: Effects of different kinds of essentiality on sequence evolution of human testis proteins. Sci Rep 7:43534 (2017).
Sha YW, Sha YK, Ji ZY, Mei LB, Ding L, et al: *TSGA10* is a novel candidate gene associated with acephalic spermatozoa. Clin Genet 93:776–783 (2018).
Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, et al: Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13:2498–2504 (2003).
Sirot LK, Wong A, Chapman T, Wolfner MF: Sexual conflict and seminal fluid proteins: a dynamic landscape of sexual interactions. Cold Spring Harb Perspect Biol 7:a017533 (2014).
Süel GM, Lockless SW, Wall MA, Ranganathan R: Evolutionarily conserved networks of residues mediate allosteric communication in proteins. Nat Struct Biol 10:59–69 (2003).

Swanson WJ, Vacquier VD: The rapid evolution of reproductive proteins. Nat Rev Genet 3: 137–144 (2002).
Swanson WJ, Nielsen R, Yang Q: Pervasive adaptive evolution in mammalian fertilization proteins. Mol Biol Evol 20:18–20 (2003).
Talavera D, Robertson DL, Lovell SC: Characterization of protein-protein interaction interfaces from a single species. PLoS One 6:e21053 (2011).
Talavera G, Castresana J: Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst Biol 56:564–577 (2007).
Taraban M, Zhan H, Whitten AE, Langley DB, Matthews KS, et al: Ligand-induced conformational changes and conformational dynamics in the solution structure of the lactose repressor protein. J Mol Biol 376:466–481 (2008).
Toll-Riera M, Bostick D, Albà MM, Plotkin JB: Structure and age jointly influence rates of protein evolution. PLoS Comput Biol 8:e1002542 (2012).
Torgerson DG, Kulathinal RJ, Singh RS: Mammalian sperm proteins are rapidly evolving: evidence of positive selection in functionally diverse genes. Mol Biol Evol 19:1973–1980 (2002).
Tusset C, Noel SD, Trarbach EB, Silveira LF, Jorge AA, et al: Mutational analysis of *TAC3* and *TACR3* genes in patients with idiopathic central pubertal disorders. Arq Bras Endocrinol Metabol 56:646–652 (2012).
Weinbauer GF, Gromoll J, Simoni M, Nieschlag E: Physiology of testicular function, in Nieschlag E, Behre HM (eds): Andrology. Male Reproductive Health and Dysfunction, 2nd ed, pp 23–62 (Springer, Berlin 2001).
White C, Kambe T, Fulcher YG, Sachdev SW, Bush AI, et al: Copper transport into the secretory pathway is regulated by oxygen in macrophages. J Cell Sci 122:1315–1321 (2009).
Wilke CO, Drummond DA: Signatures of protein biophysics in coding sequence evolution. Curr Opin Struct Biol 20:385–389 (2010).
Wolf YI, Carmel L, Koonin EV: Unifying measures of gene function and evolution. Proc Biol Sci 273:1507–1515 (2006).
Worth CL, Gong S, Blundell TL: Structural and functional constraints in the evolution of protein families. Nat Rev Mol Cell Biol 10:709–720 (2009).
Wyckoff GJ, Wang W, Wu CI: Rapid evolution of male reproductive genes in the descent of man. Nature 403:304–309 (2000).
Yang JR, Liao BY, Zhuang SM, Zhang J: Protein misinteraction avoidance causes highly expressed proteins to evolve slowly. Proc Natl Acad Sci USA 109:E831–840 (2012).
Zhang J, Maslov S, Shakhnovich EI: Constraints imposed by non-functional protein–protein interactions on gene expression and proteome size. Mol Syst Biol 4:210 (2008).
Zhang L, Li WH: Mammalian housekeeping genes evolve more slowly than tissue-specific genes. Mol Biol Evol 21:236–239 (2004).
Zhang Y, Li Q, Wu F, Zhou R, Qi Y, et al: Tissue-based proteogenomics reveals that human testis endows plentiful missing proteins. J Proteome Res 14:3583–3594 (2015a).
Zhang Y, Li D, Sun B: Do housekeeping genes exist? PLoS One 10:e0123691 (2015b).
Zhou T, Drummond DA, Wilke CO: Contact density affects protein evolutionary rate from bacteria to animals. J Mol Evol 66:395–404 (2008).
Zhou T, Wang G, Chen M, Zhang M, Guo Y, et al: Comparative analysis of macaque and human sperm proteomes: Insights into sperm competition. Proteomics 15:1564–1573 (2015).