

RESEARCH ARTICLE

Open Access



# Proteomic analysis of quail calcified eggshell matrix: a comparison to chicken and turkey eggshell proteomes

Karlheinz Mann\* and Matthias Mann

## Abstract

**Background:** Eggshell mineralization in commercially important species such as chicken, turkey or quail is of interest as a general model of calcium carbonate biomineralization. Knowledge of proteins and molecular mechanisms in eggshell assembly may also pave the way to manipulation of thickness of the calcified layer or other features. Comparison of eggshell matrix proteomes of different species may contribute to a better understanding of the mineralization process. The recent publication of the quail genome sequence now enables the proteomic analysis of the quail shell matrix and this comparison with those of chicken and turkey.

**Results:** The quail eggshell proteome comprised 622 identified proteins, 311 of which were shared with chicken and turkey eggshell proteomes. Forty-eight major proteins (iBAQ-derived abundance higher than 0.1 % of total identified proteome) together covered 94 % of total proteome mass. Fifteen of these are also among the most abundant proteins in chicken and turkey eggshell matrix. Only three proteins with a percentage higher than 1.0 % of the total had not previously been identified as eggshell matrix proteins. These were an uncharacterized member of the latexin family, an uncharacterized protease inhibitor containing a Kunitz domain, and gastric intrinsic factor. The most abundant proteins were ovocleidin-116, ovalbumin and ovocalyxin-36 representing approximately 31, 13 and 8 % of the total identified proteome, respectively. The major phosphoproteins were ovocleidin-116 and osteopontin. While osteopontin phosphorylation sites were predominantly conserved between chicken and quail sequences, conservation was less in ovocleidin-116.

**Conclusions:** Ovocleidin-116 and ovocalyxin-36 are among the most abundant eggshell matrix proteins in all three species of the family Phasianidae analyzed so far, indicating that their presently unknown function is essential for eggshell mineralization. Evidence for other chicken eggshell-specific proteins in quail was inconclusive. Therefore measurement of additional eggshell proteomes, especially from species of different families and preferentially from outside the order Galliformes, will be necessary.

## Background

The biological function of the avian eggshell is to provide protection against mechanical impact, to form a first line of defense against microbial infection, to regulate water and gas exchange and to provide calcium for the developing embryo [1]. The shell is formed in the eggshell gland (uterus) and consists to approximately 95 % of calcium carbonate in the form of calcite. The calcified layer is pervaded by an organic matrix consisting of proteins and proteoglycans, which is thought to

provide a 3-dimensional network guiding and controlling the mineralization process [2]. The most studied avian eggshell is that of the chicken, probably because of commercial importance and easy accessibility. Among the first identified chicken eggshell matrix proteins were major egg white proteins, such as ovalbumin [3], lysozyme [4], and ovotransferrin [5]. These proteins were shown by immunohistochemical methods to be part of the calcified matrix rather than surface contaminants. Egg white proteins are produced and secreted predominantly in the magnum section of the oviduct [2]. Soluble remnants of egg the white assembly may migrate with the egg into the eggshell gland, where they are eventually

\* Correspondence: mann@biochem.mpg.de  
Max-Planck-Institut für Biochemie, Abteilung Proteomics und Signaltransduktion, Am Klopferspitz 18, D-82152 Martinsried, Germany

incorporated into the mineralizing matrix. In addition, messages for lysozyme and ovotransferrin are also present at much lower concentration in white isthmus [4, 5]. These messages can even be detected in red isthmus and uterus after extensive amplification [4, 5]. At present it is unknown what percentage of these egg white proteins found in the eggshell matrix may be contributed by these tissues. Other proteins are produced by eggshell gland epithelial cells but apparently not in other sections of the oviduct and not in other selected tissues analyzed. The first of these so-called eggshell-specific proteins was ovocleidin-17 (OC17) [6], subsequently shown to belong to the C-type lectin-like family of proteins [7]. This was followed by cloning and characterization of ovocleidin-116 (OC116) [8–10], which was first identified by its N-terminal sequence in a dermatan sulfate proteoglycan preparation from chicken eggshell [11]. It was subsequently also detected in chicken bone [12, 13] and is thus not strictly eggshell-specific. Other members of this group with possible eggshell-specific distribution are ovocalyxin-32 (OCX32), a member of the latexin family of carboxypeptidase inhibitors [14], and ovocalyxin-36 (OCX36) [15], a member of the BPI/LBP/PLUNC family of anti-microbial proteins [16]. Other ovocalyxins occasionally mentioned in publications but poorly characterized are ovocalyxins-25 and -21. OCX-25 contains protease inhibitor domains and OCX-21 is apparently identical to gastrophilin-2, a constituent of the gastric secretome [17]. Two other eggshell matrix proteins of widespread distribution, osteopontin and glypican-4, are induced in eggshell gland epithelia by the mechanical strain exerted upon entry of the egg into the gland [18, 19]. The exact role of these proteins remains unknown at present, but *in vitro* studies have shown that uterine fluid as well as isolated eggshell components may influence calcite crystallization *in vitro* and thus may also control eggshell mineralization *in vivo* [20–24]. Other matrix proteins may participate in the egg's anti-microbial defense during and after egg production [25, 26].

The almost complete sequencing of the chicken genome [27] provided the possibility to identify more eggshell matrix proteins using high-throughput mass-spectrometry-based proteomics. The first compartment of the egg to be analyzed by such methods was the acid-soluble fraction of the chicken eggshell calcified layer [28, 29], leading to identification of 528 proteins in a wide abundance range. Measurement of the acid-insoluble matrix [30, 31], the cuticle [32, 33], and the soluble fraction of the eggshell membranes in conjunction with the innermost eggshell calcified layer (mammary cones) [34] yielded several additional protein identifications. Soluble eggshell membrane proteins were also analyzed at different stages of chicken embryo development [35]. In addition, the proteome of the uterine fluid bathing

the egg during shell mineralization was compared to the proteome of the calcified shell [36]. In that study a total of 577 proteins were identified in uterine fluid and 466 in eggshell, with an overlap of 244. The proteomes of uterine fluid at different stages of eggshell mineralization has also been compared [37]. In sum, these studies identified a total of 675 eggshell proteins [37] and yielded important information on the distribution of shell proteins in different compartments and on the temporal sequence of their appearance in the uterine fluid. Proteomic studies were complemented by transcriptomic studies aiming at identifying genes expressed in uterus with possible importance for eggshell production [38–42].

Compared to chicken, the eggshell proteomes of other species were less well explored. This was due to the lack of comprehensive sequence databases, still a prerequisite for high-throughput proteomics. The publication of the almost complete turkey genome sequence [43] provided the possibility to compare the eggshell proteomes of chicken and turkey [44]. The turkey eggshell matrix yielded 697 proteins. The overlap with the chicken eggshell proteome was 52 %. However, if only turkey proteins with an abundance of >0.01 % of the total were compared, the overlap increased to 95 %. This indicated that most of the major proteins were conserved between species, but that there were also a few potentially important differences.

In the present report, we use the recently published genome-derived sequence database of Japanese quail [45] to compare another eggshell proteome of the same avian family, the Phasianidae, to the preceding ones. Previously Western blotting analysis of eggshell matrices of several avian species using antibodies against known hen eggshell matrix proteins identified ovotransferrin, osteopontin and ovalbumin as components of the quail eggshell [46]. In addition, ovomucoid and lysozyme were identified by N-terminal sequence analysis of electrophoretically separated matrix proteins [47]. More recently, an unknown 32 kDa protein of the cuticle was characterized by a short N-terminal sequence [48], and the calcified matrix was shown to contain a protein with sequence similarity to ovocleidin-116 [49]. Ovocleidin-116, ovocleidin-17, ovocalyxin-32, clusterin, cystatin, lysozyme, osteopontin, ovalbumin, ovoinhibitor, ovomucoid and ovotransferrin were also identified in quail eggshell matrix in a previous proteomic study using the chicken sequences for protein identification [50]. Here we show that the number of quail eggshell matrix proteins is similar to that of chicken and turkey and explore similarities and differences among the major matrix proteins.

## Materials and methods

### Matrix and peptide preparation

Fresh quail eggs were bought at a local market. The broken shells were emptied, cleaned under a jet of water,

and washed with 5 % EDTA at 6 °C for 30 min to facilitate mechanical removal of the cuticle and the membranes. The cuticles were then removed by brushing under a jet of de-ionized water, and pieces of calcified shell were stripped off the wet membranes. The dried pieces of calcified eggshell were demineralized in 50 % acetic acid (20 ml/g of shell) at 4–8 °C for 15 h with constant stirring. The turbid mixture was dialyzed (Spectra Por 6, cut off 2000; Spectrum Europe/Carl Roth GmbH, Karlsruhe, Germany) against 2 × 10 vol. of 10 % acetic acid and 2 × 10 vol. of 5 % acetic acid, and freeze dried.

SDS-PAGE was done using pre-cast 4–12 % Novex Bis-Tris gels in the MOPS buffer system using reagents and protocols supplied by the manufacturer (Invitrogen, Carlsbad, CA). The kit sample buffer was modified by adding β-mercaptoethanol to a final concentration of 2 %, and the sample was suspended in 30 μl sample buffer/100 μg of organic matrix and heated to 70 °C for 10 min. Samples were centrifuged to sediment insoluble material and gels were loaded with the dissolved proteins of 100 μg of matrix per lane and stained with colloidal Coomassie (Invitrogen) after electrophoresis. Gels were cut into 20 slices for in-gel reduction, carbamidomethylation and digestion with trypsin [51]. Peptides were cleaned with C18 Stage-Tips [52] before mass spectrometric analysis.

#### LC-MS and data analysis

Peptide mixtures were analyzed by in-line nanoflow liquid chromatography using the EASY-nLC system (Proxeon Biosystems, Odense, Denmark; now part of Thermo Fisher Scientific) with 15 cm capillary columns of an internal diameter of 75 μm filled with 3 μm Reprosil-Pur C18-AQ resin (Dr. Maisch GmbH, Ammerbuch-Entringen, Germany). The gradient consisted of 5–30 % acetonitrile in 0.5 % acetic acid at a flow rate of 250 nl/min for 85 min, 30–60 % acetonitrile in 0.5 % acetic acid at a flow rate of 250 nl/min for 5 min and 60–80 % acetonitrile in 0.5 % acetic acid at a flow rate of 250 nl/min for 7 min. The eluate was electrosprayed into an LTQ Orbitrap Velos (Thermo Fisher Scientific, Bremen, Germany) through a Proxeon nanoelectrospray ion source. The Orbitrap Velos was operated in an HCD top 10 mode essentially as described [53] at a resolution of 60,000 for full scans and of 7500 for fragment measurement (both specified at *m/z* 400). The dynamic exclusion time was 120 s.

Raw files were processed using the Andromeda search engine-based version 1.5.1.6 of MaxQuant (<http://www.maxquant.org/>) [54–56] with enabled second peptide identification, iBAQ quantification and match between runs (match time window 0.5 min; alignment time window 20 min) options. We used the predicted gene database of *Coturnix japonica* [45] ([\[genome.org/japanese\\\_quail.html?lang=en\]\(http://genome.org/japanese\_quail.html?lang=en\); 30810 entries; downloaded November 2014\) combined with a \*Coturnix\* subset of the UniProtKB database \(Release 2014\\_9, 813 entries\). In some trial searches, the sequences of chicken ovocleidins and ovocalyxins were added. The corresponding reversed databases and the sequences of common contaminants possibly introduced during sample preparation and handling were appended to the database. Carbamidomethylation was set as a fixed modification. Variable modifications were oxidation \(M\), N-acetyl \(protein\), pyro-Glu/Gln \(N-term\), and phosphorylation \(Phospho \(STY\)\). In some trial searches, hydroxyproline was added as a variable modification. Initial peptide mass tolerance and allowed MS/MS mass deviation were 20 ppm. Two missed cleavages were allowed and the minimal length required for a peptide identification was seven amino acids. The protein false discovery rate \(FDR\) and PSM \(peptide spectral match\) FDR were set to 0.01. The minimal peptide score for unmodified and modified peptides was set to 60. Identifications with less than three sequence-unique peptides were in each case validated using the MaxQuant Expert System software \[57\] considering the assignment of major peaks, occurrence of uninterrupted \*y\*- or \*b\*-ion series of at least four consecutive amino acids, preferred cleavages N-terminal to proline bonds, the possible presence of \*a2/b2\* ion pairs and immonium ions, and mass accuracy. We only accepted protein identifications with at least two sequence-unique peptides occurring in at least two replicates with a total of three peptides. Exceptions were protein identifications sharing peptides with very similar proteins and fragments of proteins obviously belonging together. The iBAQ \(intensity-based absolute quantification\) \[58\] option of MaxQuant was used to calculate, based on the sum of peak intensities, the approximate share of each protein in the total identified proteome mass, and thus discerning major from minor proteins.](http://www.nodai-</a></p></div><div data-bbox=)

Sequence homology searches were performed with FASTA (<http://www.ebi.ac.uk/Tools/sss/fasta/>) [59] against current releases of Uniprot Knowledgebase (UniProtKB). We also searched the quail sequence databases for some proteins of particular interest using the Local Blast function [60] of BioEdit Sequence Alignment Editor version 7.2.5 from <http://www.mbio.ncsu.edu/bioedit/bioedit.html> and with the TBLASTN function of the BLAST site provided by the NODAI Genome Research Center ([http://www.nodai-genome.org/blast/Japanesequail\\_scaf\\_1000/blast/](http://www.nodai-genome.org/blast/Japanesequail_scaf_1000/blast/)). We also used Kalign (<http://www.ebi.ac.uk/Tools/msa/kalign/>) [59] and Clustal Omega for sequence alignments (<http://www.ebi.ac.uk/Tools/msa/clustalo/>) [61], InterPro (<http://www.ebi.ac.uk/interpro/>) [62] for domain predictions, and the Venn diagram plotter (<http://omics.pnl.gov/>

software/VennDiagramPlotter.php) for preparing Venn diagrams. Kinase motifs apart from FAM20C consensus phosphorylation motifs were predicted using NetPhosK (<http://www.cbs.dtu.dk/services/NetPhosK/>) [63].

## Results and discussion

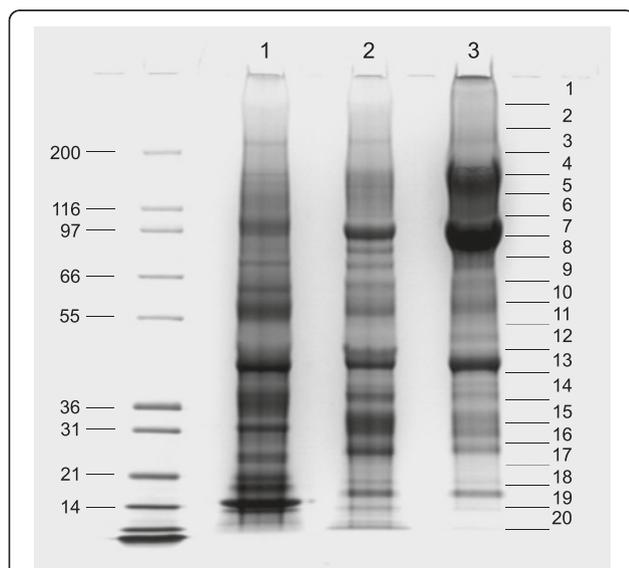
In this study, we used the total PAGE sample buffer-soluble matrix of quail eggshell without separation of acid-soluble and acid-insoluble fractions because only approximately 10 % of the quail eggshell matrix is acid-soluble. Comparison of this total matrix fraction to the quail acid-soluble fraction alone showed that the band pattern remained the same but that some higher molecular weight bands were enriched in the total matrix (Fig. 1, lanes 2 and 3). In addition lane 1 of Fig. 1 shows the chicken eggshell acid-soluble matrix for comparison with the quail acid-soluble quail matrix (Fig. 1, lane 2). Both matrices showed similar complexity but low-molecular weight bands below 21 kDa were less prominent in quail than in chicken. In chicken this mobility range includes major matrix proteins, such as ovocleidin-17 and lysozyme.

To ensure we obtained a representative average shell proteome, we pooled the calcified layers of eighteen washed eggshells (12.85 g) and performed demineralization and proteome analysis in triplicate. The yield of quail eggshell matrix was 37 mg/g of air-dried eggshell calcified layer. Proteins were separated by SDS-PAGE and the gels

were sliced into 20 sections (Fig. 1) for in-gel reduction, carbamidomethylation and digestion with trypsin.

Processing the resulting 60 raw-files with MaxQuant yielded 1173 identified protein groups (Additional file 1: Table S1). Protein groups collect protein entries that cannot be distinguished by the obtained MS evidence (however, both proteins and protein groups are used interchangeably in the manuscript). The corresponding peptide data are shown in Additional file 2: Table S2. The amino acid sequences of identified accessions were compared to sequences contained in the UniProt Knowledgebase using FASTA. The results indicated that the NODAI quail sequence database obviously contained single entries comprising several unrelated proteins as well as different entries likely containing fragments of the same protein. The first possibility is illustrated by entry 713 (Fig. 2). The sequence of its first 230 amino acids is nearly identical to the sequence of bone sialoprotein 2. This sequence region comprised three identified sequence-unique peptides. The sequence of amino acids 231–991 was very similar to chicken ovocleidin-116 and contained 45 identified sequence-unique peptides. The remainder of the sequence of 713 was very similar to osteopontin and the four peptides identified in this part of the entry (Fig. 2) were also contained in the *Coturnix japonica* osteopontin entry Q9I832\_COTJA. An example of different entries containing fragments of the same protein is illustrated using accessions 14241 and 14979, which both matched to different regions of pentraxin (Q5UMH8) without overlap in matching regions (Fig. 3). Finally, entries 15278 and part of 3793 (3793b) are an example of entries already indicating their origin from one and the same protein by partially shared sets of identified peptides (Fig. 4). Both quail database entries were highly similar to cathepsin D (CATD\_CHICK).

After elimination of identifications not matching the criteria detailed above (Materials and methods, LC-MS and data analysis) and tentatively grouping together protein fragments matching identical UniProt Knowledgebase accessions, we accepted 622 identified protein groups (Additional file 3: Table S3). However, this number remained tentative, because fragment assembly and grouping may contain errors and must be confirmed by determination of complete gene or mRNA sequences. Furthermore, proteins appear to be missing from the genome-derived database as indicated by identification of quail Uniprot protein entries not contained in the genome-derived database. The number of identified proteins is also lower than that of chicken (675; [37]) and turkey (697; [44]). Furthermore, identities between quail sequences and similar chicken or turkey sequences were frequently much lower than expected (Additional file 3: Table S3) and



**Fig. 1** PAGE separation of eggshell matrices. Lane 1, acid-soluble chicken eggshell matrix. Lane 2, acid-soluble quail eggshell matrix. Lane 3, total quail eggshell matrix. 100  $\mu$ g of matrix were separated per lane. Molecular weight of markers is shown on the left in kDa. Gel sections used for in-gel digestion and analysis are indicated on the right

SIAL_CHICK	1	MR	SALLLACLLATASAFSVRSWLRARRAGDSEENAVLKSRRHYLYRYAYPPLHRYKGS
713	1	MRT	ALLLACLLATASAFSVRSWLRARRAGDSEENAVLKSRRHYLYRYAYPPLHRYKGS
SIAL_CHICK	61	SSEEEGDGSEEEEEEGGAPSHAGTQAAGEGLTLGDVGPGGDAASAHQD	CKGGQKGT
713	59	-----	-----CKGGQKGTQGDSSG
SIAL_CHICK	121	DESDSEEEEEEEEEEEEEVEEQDVSVNGTSTNTTAETPHGNNTVAEEEEDDDEEEEE	
713	72	DESDSEEEEEEEEEEEEEVEEQDVSVNGTSTNTTAETPRGNSTAAAAEEEEEEEEEGEE	
SIAL_CHICK	181	EEEEEEAEATTAATAAQDEVTTLGDEQRSEVTTAGEQWEYEVTVGARGDEGPTESSYGD	
713	132	EEEEEEAEATTAATAAQDEATTLGDDQRFATTAGEQWEYEVTAGARGDEGPTEGSYGD	
SIAL_CHICK	241	QEPARGDSYRAYEDEYGYKGHGYDMYQDYYNQ-----	
713	192	QEGPPRGDTYRNYEDEYGYKGHGYDIYQDYYYSHSFYSINYPFNYSNTPSNLTLANTS	
OC116_CHICK	1	-----	-----MRATLFLCLC
713	252	VNILGYPIIYPTDYFIDHPLEYSSNTDTFLPAGATSSQGKRKLPWTAPGCNTNHGFVFK	
OC116_CHICK	10	LCLLG--TVLPT-----FVSLPARARGNCP-----GQHQLLK--GCNTKHGFYIFQ	
713	312	FMHSYLMRRNQTVQVKEEGDHQGTIHGRWLKGV DGEALGQGVGSSHPEDKDSPKPHSHI	
OC116_CHICK	53	YIYSHLMQKNQTVQVKEEGDHQGTIHGHWLKGV DGEAPGQAGSSHPEDKDSPKTSHSI	
713	372	TPVSKVGGRALRPFSTGDSISVYPTSTSMEGSGDMGSMLLGEIINGEDGLPQSTHPGEPHG	
OC116_CHICK	113	TPASKGEGRALRPGIGDSNSVYPTSTSVESGDMGSMILLGEIINGEDGLPQSTHPGEPHG	
713	432	DNGGNGVLVDGAVTAGRERATNSEGAGGEGSSHATVPDQOQAGTMVTDGSAITSVTDSA	
OC116_CHICK	173	DGDGGNGVLVDGAVTAGRERASGSEGAGSEGGSHAPVPDQOQAGTMVTDGSAITSVTDSA	
713	492	ITSVTKKEDVHVD AEDIGEFAYIPYVDAATITRQDGETHVS PKDEVQIFIGRANIQVGE	
OC116_CHICK	233	ITSVTKKEDVHVDTEGIDEFAYIPDVAVITITRQDGETHISPEDEVKIFIGRANIQVGE	
713	552	NDGSAGSTGATSEANVIPAAVNARPOGYPEESA SMATVHHGDSATSRPGDHPSAGNSGDG	
OC116_CHICK	293	NDSVGSAGATSEANVIPVVTVRPQGHPEESATMATLHHGDSVTSRPFVGHPSVGNNSGDG	
713	612	ATAIHSGQLEAPSPRESTDGEATVTTAAGSQSGKGRSGQRPVKGHSLPATMTTQARGT	
OC116_CHICK	353	ATEIHSGQLEAPSPWESTGGDATVTVAVGVQSGKGRSGQRALGKHS LPATMTTRGGRGT	
713	672	ASSGLTMGDCSTAASTPSWKGSQVVTAGQEGSREVTAGPKRQRARIQQEVAPAPDVEGG	
OC116_CHICK	413	ASSGLTIGDCSTAASTPSRKGSHVVSAGQGESGEVGTAGPERQRARVQEVAPARGVGG	
713	732	MVVPEGHRARIQQEVAPAPDVEGGMVVPEGHGARTQPGVASAPSTVETEAPERHRSRAQQ	
OC116_CHICK	473	MVVPEGHRARVQEVAPARGVVGGMVVPEGHGARTQPEVASAPSTVGKAAAPERHRNRAQQ	
713	792	GVAPVHSMERETVTPERQRARMQPSARMQQAARPEVAPAPSTAGGIVAPEGHRARVWPG	
OC116_CHICK	533	EVAPVPSMVVETVAPERHRARVREESARLQQAARPEVAPAPSTGGRIVAPGHRARVWPG	
F1NSM8_CHICK	1	-----	-----MKLAF
713	852	AAPAPGVIRVARPAPSRVYNGDKRDAMGKSTDVPRDPWVWGSTSPQTQHTRGSMVAGSFS	
OC116_CHICK	593	AAPAPGVVGVARPAPSKAYNGDKRVAIGKSTDVPRDPWVWGSAHPQAQHTRGSTVAGGFA	
F1NSM8_CHICK	7	CL-----CFISIAAAMPVSKSRQ-HAISASSEKY-----DPRSHHTRHYQDH	
713	912	HLHGGQRMGSLAELERSQVEQVROADRLRLHERAVYGLSGVGGPLQPLAAHTDPWSADS	
OC116_CHICK	653	HLHRGQRLGLTMEHSRQVEQVRHADRLRLHERAVYGLSGVGGPLQPPAVHTDPWSADS	
F1NSM8_CHICK	50	VDSQSQEHLQQTQNDLASLQQTHYSSEENADVPEQDFPDI PSKSQEAVDDDDDDNDSN	
713	972	SQS-SEGRWGSRSNSGEEDGETHYSSEENADVPEQDFPDI PSKSQETVDDDDDDNDSN	
OC116_CHICK	713	SQS-SEGRWGSHSDSHEEDGEVIRGYPYGRQSL-----	
F1NSM8_CHICK	110	DTDESDEVVTFDFPTEAPVTPFNRGDNAGRGDSVAYGFRAKAHVVKASKLRKAARKLIEDD	
713	1031	DTDESDEVVTFDFPTEAPVTPFNRGDNAGRGDSVAYGFRAKAHVVKASKLRKAARKLIEDD	
F1NSM8_CHICK	170	ATAEVGDSQLAGLWLPKESREQDSRELAQHQS VENDSRPFDSPVVDGGDSKASAGVDSR	
713	1091	ATTEVGDSQPAGLWLPKESREQNSRELPQHQS VENDSRPFDSPVVDGGDSKASAGVDSR	
F1NSM8_CHICK	230	ESLASASAVDTSNQTLESAEDAEDRHSIENNEVTR	
713	1151	ESLASAPAVDTSNQTLESAEDAEDRHSIENNEVTR	

**Fig. 2** Analysis of accession 713 amino acid sequence. The sequence of quail accession no. 713 is aligned to chicken bone sialoprotein 2 (SIAL\_CHICK), chicken ovocleidin-116 (OC116\_CHICK) and chicken osteopontin (F1NSM8\_CHICK). Identical amino acids are shaded yellow. Sequence regions covered by identified peptides are shown in bold red letters

in almost all instances this was due to gaps in the respective sequence alignments caused by shorter quail sequences. The overlap of the quail shell proteome with turkey and chicken shell proteomes was approximately 63 and 60 %. Approximately 50 % of the

proteins/protein groups were identified in all three species (Fig. 5).

As before [44, 64] we used intensity-based absolute quantification (iBAQ) as implemented in MaxQuant to discern major from minor proteins (Additional file 3:

Q5UMH8_CHICK 14241	1	MLPGGVLSLLMLFCVFRASASVLDDEGDDYDLMYVNLNDEIEGPVLGTEESASDCDCQREHT
	1	-----ASASVLDDEDDYDLMYVNLNDEIEGPVLGTEESASDCDCQREHT
Q5UMH8_CHICK 14241	61	EWDKLFIMLENSQMKENMMLQALEDEGLKADVQAIRAELNQLTAGLAGTFFSSTVQQLTSHA
	44	EWDKLFIMLENSQMKENMMLQALEDEGLKEADVQAIRAELSLSAGLAGTFFSSAVQQLTSHA
Q5UMH8_CHICK 14241	121	LAQLEQALQRNGDQPEDAERLCESQGGKQVLEQVLQLSHNVSARLQGLESSWLWRWAEREAQ
	104	LAQLEQALQRNRDQPEDAERLCSEFKQGGKQVLEQVLQLSHIRVHQAGPAGELLAEMGRGGGP
14979	1	-----YETAILFPMR
Q5UMH8_CHICK 14241	181	ETAFQLQDQKLGHSREDNLLNLTWKELOQTRSELKASQKWSQHLLPAGYETAILFPMR
	164	RNSFPAPTGA-----WPG-QRGGPSL-----EHL-----MERAADQSR
14979	11	SRKI--FGSIHPTAAMTLLTSFTVCVWVKATEVLDKTVFSYGTKLNPYEIQLYLSHQSAV
Q5UMH8_CHICK 14241	241	SRKI--FGSIHPTAAMTLLSSFTVCVWVKATEALDKTIIFSYGTKLNPYEIQLYLSRQSAV
	198	TEGI SEVGC PAPTASR-----
14979	69	LAVGNAQHKVTARDVVVPGKWIHLGCSWSSDNGTASLWVDGALTATAVGIADTHIVPDGG
Q5UMH8_CHICK	299	LAVGSAQHKLAARDVVVPGKWIHLGCTWSSDNGTASLWVDGELTATAVGIADTHIVPDGG
14979	129	ILQIGQEKNGCCIGGGFDEALAFSGKLTGLNLWDRVLSSTEEIAAQGGEDACSSRGNIVGW
Q5UMH8_CHICK	359	ILQVGQEKNGCCVGGGFDEALAFSGKLTGLNLWDRVLSSTEEIAAQGGEDACSSRGNIVGW
14979	189	GVTEILPYGGAQYMS
Q5UMH8_CHICK	359	GVTEVLLYGGAQYVS

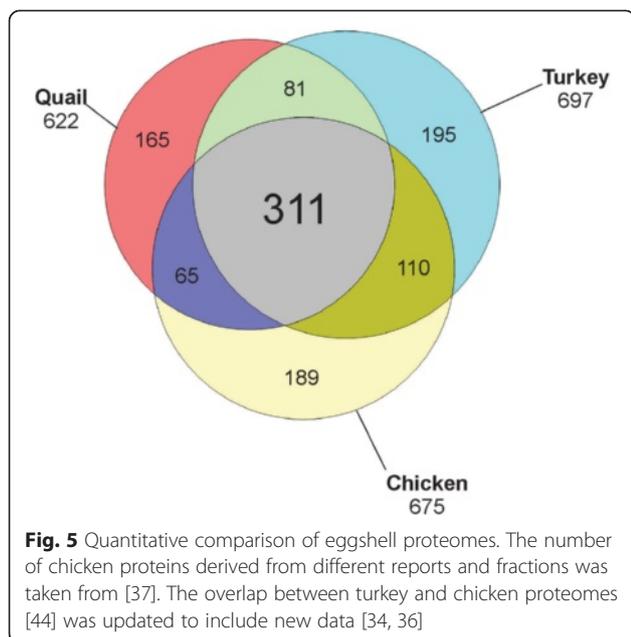
**Fig. 3** Alignment of pentraxin fragments. Quail entries 14241 and 14979 matched to different sequence regions of pentraxin (Q5UMH8\_CHICK) strongly indicating that both were fragments of the same protein. Identical amino acids are shaded yellow. Sequence regions covered by identified peptides are shown in bold red letters

Table S3). To obtain more correct abundances for entries obviously containing sequences from different proteins, these were dissected into different database entries distinguished by small letters following the original accession number (Additional file 4: Figure S1). We also joined entries tentatively identified as fragments of the same protein to form a single accession

according to the differentially shaded groups in Additional file 3: Table S3. This table contains abundances before and after database modification. Using threshold of  $\geq 0.1\%$  of the total identified proteome to differentiate between minor and major proteins, we obtained 48 major proteins that together covered 94 % of the proteome (Table 1).

3793b	220	-VQRGLLHLII-----AIASX	IPLTKFTSTRRMLTEVGSEIPDMNAITQFLKFKLGFA
CATD_CHICK	1	MAPRGLLVLLLALVGPCALIR	IPLTKFTSTRRMLTEVGSEIPDMNAITQFLKFKLGFA
15278	1	-----IPLTKFTSTRRMLTEVGSEIPDMNAITQFLKFKLGFA	
3793b	52	DLAEPTEILKNYMDAQYYGEIGIGTPPQK	FTVVFDTGSSNLWVPSVHCHLLDIACLLHH
CATD_CHICK	61	DLAEPTEILKNYMDAQYYGEIGIGTPPQK	FTVVFDTGSSNLWVPSVHCHLLDIACLLHH
15278	38	DLAEPTEILKNYMDAQYYGEIGIGTPPQK	FTVVFDTGSSNLWVPSVHCHLLDIACLLHH
3793b	112	KYDASKSSTYVENGTEFAIHYGTGSLSGFLSQDVT	-----
CATD_CHICK	121	KYDASKSSTYVENGTEFAIHYGTGSLSGFLSQDVT	TLGNLKIKNQIFGEAVKQPGITFIA
15278	98	KYDASKSSTYVENGTEFAIHYGTGSLSGFLSQDVT	TLGNLKIKNQIFGEAVKQPGITFIA
3793b	148	--FDGILGMAFPRI	ISVDKVTFFDNVMQQLIEKNIFS FYLN RDPTA QPGGELLGGTDP
CATD_CHICK	181	AKFDGILGMAFPRI	ISVDKVTFFDNVMQQLIEKNIFS FYLN RDPTA QPGGELLGGTDP
15278	158	AKFDGILGMAFPRI	ISVDKVTFFDNVMQQLIEKNIFS FYLN-----
3793b	206	KYYSGDFS	WVNVTRKAYQVHMDSVDVANGLTLCKGGCEAIVDTGTSLITGPTKEVKELO
CATD_CHICK	241	KYYSGDFS	WVNVTRKAYQVHMDSVDVANGLTLCKGGCEAIVDTGTSLITGPTKEVKELO
3793b	266	SAIGAKPLIKQYVIPC	EKISSLPVVTMLGKPKYQLTGEQYVFKVSAQGETICLSGFSG
CATD_CHICK	301	TAIGAKPLIKQYVISC	DKISSLPVVTMLGKPKYQLTGEQYVFKVSAQGETICLSGFSG
3793b	326	LDVPPP	GGPLWILGDVFIGPYYTVFDRDND SVGFAKSA
CATD_CHICK	361	LDVPPP	GGPLWILGDVFIGPYYTVFDRDND SVGFAKCV

**Fig. 4** Alignment of cathepsin D fragments. Quail database entry 15278 and part of 3793 (3793b) contained overlapping sequences and were both highly similar to chicken cathepsin D. Sequences confirmed by MS/MS sequencing are shown in red. Peptides identified only in one of the entries and not in the other one are underlined. Peptides matching only entry 3793b align to a gap in the alignment of this entry to chicken cathepsin D, and peptides matching only entry 15278 align to gaps in 3793b. Because of the very high sequence identity of both entries to chicken cathepsin D and the occurrence of several sequenced peptides shared by both entries we suggest that both entries are incomplete and likely belong to the same protein. Identical amino acid positions are shaded yellow



### Eggshell-specific proteins

A group of major eggshell proteins generally referred to as eggshell-specific proteins was once thought to occur specifically in the uterus because they were neither found in other sections of the oviduct nor in a few selected other tissues. These are the ovocleidins and ovocalyxins and they will be discussed in this section.

The first eggshell-specific protein to be detected in the chicken eggshell as a major matrix component was ovocleidin-17 (OC17) [6]. Sequence information of this C-type lectin-like protein was not contained in the published genome sequence of chicken but was known from direct sequence analysis of the isolated protein [7]. A full length clone containing the OC17 sequence was isolated and characterized only very recently and its mRNA expression level was found to correlate negatively with eggshell strength [65]. Furthermore, the message was found to be expressed in uterus and, at a lower level, in the preceding oviduct section, the isthmus, but not in 13 other tissues tested. However, its possible function in eggshell mineralization remains ill-defined at present and may include antimicrobial activity [25] and direct interaction with calcite surfaces or carbonate ions [22, 23, 66]. We could not identify turkey OC17 in our previous study [44], however, identification of a few peptides after addition of the chicken sequence to the searched database and previous detection of a protein in turkey eggshell matrix in Western blotting analysis using anti-chicken OC17 antiserum [46] indicated its presence in that eggshell matrix, but not in the genome sequence database. Homology searching of the quail sequence database with the chicken sequence did not yield a significant match and addition of the chicken sequence to

the database searched by MaxQuant did not identify any matching peptide. This contradicts the previous identification of OC17 in quail eggshell using such a trans-species approach [50], but agrees with Western blotting results [46]. Very thick eggshells, such as the shells of ostrich [67], emu and rhea [68] all have two major proteins of this family in their shell matrix. Chicken and goose with their much thinner shell both only have one of these proteins [28, 69]. The apparent absence of an OC17 homolog in the even thinner shell of quail may thus indicate a connection between number or concentration of these proteins and eggshell thickness.

Ovoclein-116 was by far the most abundant protein of the quail and turkey eggshell matrix (Table 1), and also one of the most abundant chicken eggshell matrix components [28]. Most of the quail protein sequence was contained in the middle region of accession 713 (713b; Fig. 2). However, a second entry, accession 2298, was even more similar to chicken OC-116 than accession 713. However, we did not identify any peptides shared between these sequences. Closer examination showed that two overlapping, very abundant, peptides contained in 2298 (168 times identified; Additional file 3: Table S3) were located in the first 40 amino acids, preceded by a predicted secretion signal peptide. The N-terminus of the predicted mature protein was similar, but not identical, to a short N-terminal sequence published previously [49] (accession 2298 in Additional file 4: Figure S1). The remainder of the amino acid sequence of accession 2298 contained only two identified low-abundance peptides identified just four times altogether. In summary, the evidence indicated that accession 2298 contained the N-terminus of OC-116 missing in accession 713. We consequently combined this part of the sequence (aa1-40; 2298a) into one entry termed 713b to obtain a better estimate of OC-116 abundance. OC116 was mainly localized to the palisade layer [8], in agreement with its predominant secretion during the growth phase of eggshell mineralization [37]. The protein was also identified in chicken bone [12, 13], thus establishing a link between these two different biomineralization systems in this species. OC116 was suggested to have a mammalian counterpart, matrix extracellular phosphoprotein (MEPE), belonging to the secretory calcium-binding phosphoprotein (SCPP) group of proteins [70] or small integrin-binding ligand N-linked glycoproteins (SIBLING), a group of proteins including dentin matrix protein, osteopontin, dentin sialoprotein and other proteins important for skeletal and dental mineralization and remodeling [71, 72]. In fact, chicken OC116 is N-glycosylated [10] and phosphorylated [29]. While the phosphorylation status of the turkey protein was not determined, quail OC116 was also a major phosphoprotein of the eggshell matrix (see below).

**Table 1** Major proteins ( $\geq 0.1$  %) of the quail eggshell calcified layer

Protein	Accession (possible turkey and chicken homologs in brackets) <sup>1</sup>	iBAQ %	iBAQ % Turkey	Chick abundance (emPAI)
Ovocleidin-116 <sup>a g</sup>	2898a;713b (G1N6E1;OC116)	46.40	31.30	High (65.3) <sup>2a</sup>
Ovalbumin <sup>i</sup>	20482;Q6V115;20852 (G1MYL1;G1MYK6;OVAL)	15.97	12.52	High (85.6) <sup>2a</sup>
Ovocalyxin-36 <sup>a g, t</sup>	21806 (G1N6M8;Q53HW8)	7.80	6.00	High (23.2) <sup>2a</sup>
Uncharacterized member of the latexin family	11366a	6.32	-	-
Uncharacterized proteinase inhibitor (I1)	23208	3.47	-	-
Avidin/avidin-related protein <sup>i</sup>	29257 (G1MSZ9;AVID)	1.95	2.67	Intermediate (2.2) <sup>2a</sup>
EGF-like repeat and discoidin I-like domain-containing protein 3 (EDIL3) <sup>a i</sup>	25038;28103;20199 (G3UT43; F1NCN3)	1.65	3.70	High (83.3) <sup>2a</sup>
Tsukushin <sup>a g, t</sup>	7859 (G1NRF2;F1NDH7)	1.51	0.25	Intermediate (3.9) <sup>2a</sup>
Alpha-2-antiplasmin (SERPINF2) <sup>a</sup>	5112;6916 (G1N0Z7;F1NAR5)	0.60	0.80	Low-intermediate <sup>2b,c,d</sup>
Gastric intrinsic factor	8574 (G1MW98;R4GLQ4)	0.42	-	-
Serum albumin <sup>g, t</sup>	16345;19800;4290a (G1NCR2;ALBU;F2Z4L6)	0.39	2.57	High (113.5) <sup>2a</sup>
Lactadherin (MFG8) <sup>i</sup>	9870;11245;14016 (G1N944;E1C0K5)	0.36	1.25	High (15.9) <sup>2a</sup>
Glutathione peroxidase (GPX3) <sup>i, t</sup>	19686 (G1N365;F1NPJ8)	0.36	0.04	Intermediate (2.2) <sup>2a</sup>
Actin(s) <sup>i</sup>	7084 (G1NS52;ACT5;ACTG1)	0.36	-	High (9.0) <sup>2a</sup>
Uncharacterized/similar to mucin(-5 AC)	4106;10751;6295;6355 (G1N8Z1;E1C037)	0.35	0.15	Low (1.4) <sup>2a</sup>
Uncharacterized/Ovostatin-like (OVSTL)	15509;25086;455a;9460 (G1NK51;F1NEW8)	0.34	0.05	Low (1.5) <sup>2a</sup>
Prostate stem cell antigen (PSCA) <sup>a t</sup>	20430 (G1NJY7;F1NXM7)	0.34	0.18	Low (0.8) <sup>2a</sup>
Extracellular fatty acid-binding protein (EXFAB)	Q9I9P7 (G1MQ16;E1C0K1)	0.34	3.82	High (26.8) <sup>2a</sup>
Pigment epithelium-derived factor/SERPINF1 <sup>i</sup>	5860 (G1N131;E1C7H6)	0.31	0.64	Intermediate (6.0) <sup>2a</sup>
Clusterin <sup>a i</sup>	6174;P14018 (G1NMM6;Q9YGP0)	0.31	0.31	High (42.7) <sup>2a</sup>
Serpin E2/Glia-derived nexin	6327 (G3URX3;E1BWU2)	0.31	0.18	High (11.1) <sup>2a</sup>
Osteopontin <sup>a</sup>	Q9I832;713c (G1N6D8;G1N6D8)	0.28	0.15	High (1.89) <sup>2a</sup>
Apolipoprotein D (APOD)	28649 (G1N591;Q5G8Y9)	0.27	0.04	High (18.3) <sup>2a</sup>
Glypican-4 <sup>a</sup>	6640 (G1MVZ0;F1NAU1)	0.23	0.12	Intermediate (6.3) <sup>2a</sup>
Semaphorin-3G <sup>g, t</sup>	621a (G1MXI6;F1NQ93)	0.22	0.15	Low (1.7) <sup>2a</sup>
Similar to Bactericidal/ permeability-increasing protein-like 2/BPI fold-containing family C protein <sup>a i, t</sup>	28084;20375 (IPI_CHICK00571823)	0.21	-	High (16.8) <sup>2a</sup>
Pentraxin (PTX3/PPTX)	14241;14979 (G1NDX8;Q5UMH8)	0.21	0.01	Low (0.6) <sup>2a</sup>
Ovalbumin-related protein X <sup>i</sup>	30155 (G1MZH2;R9TNA6)	0.21	<0.01	Intermediate (3.8) <sup>2a</sup>
Ovotransferrin (TRFE) <sup>i</sup>	2596;18629;24968 (G1MNV5;E1BQC2;Q4ADJ7)	0.19	2.22	High (22.9) <sup>2a</sup>
Cathepsin D	15278;3793b (G1N8P0;CATD)	0.18	0.22	Intermediate (5.3) <sup>2a</sup>
Cathepsin B	8924;17826 (G1NN37;F1N9D8)	0.18	0.09	High (24.8) <sup>2a</sup>
Sulfhydryl oxidase 1 (QSOX1) <sup>i</sup>	2299;5467 (G1MV84;QSOX1;F1NYK2)	0.18	1.08	Intermediate (4.7) <sup>2a</sup>
Apolipoprotein A-I <sup>a g, t</sup>	12041;P32918 (G1MVX1;APOA1)	0.17	0.05	Intermediate (2.2) <sup>2a</sup>
Lysozyme C <sup>g</sup>	P00701 (LYSC;LYSC)	0.14	0.64	High (128.2) <sup>2a</sup>
Fibronectin (FN1)	141;A0A060PIY8 (G1MWJ4;F1NJT3;F1NJT4)	0.13	0.13	High (18.9) <sup>2a</sup>
IGFBP5	Q9DGI2 (-;F1ND88)	0.13	-	Low (0.6) <sup>2a</sup>
FAM20C kinase/DMP4 <sup>a</sup>	21781;14673 (G1MSD1;E1C4X0)	0.12	0.27	Intermediate (2.3) <sup>2a</sup>
Proactivator polypeptide (PSAP)	2306;1756 (G1MZV1;E1BSP1)	0.12	0.06	High (9.0) <sup>2a</sup>
Polyubiquitin/ubiquitin <sup>t</sup>	11531;2058 (G1NR52;Q91021;UBB)	0.12	0.12	High (9.0) <sup>2a</sup>

**Table 1** Major proteins ( $\geq 0.1$  %) of the quail eggshell calcified layer (*Continued*)

Regenerating islet-derived protein 4 (REG4)	8715a (G1MZE6;E1BZV4)	0.11	0.38	Low (0.7) <sup>2a</sup>
Cystatin <sup>f</sup>	14227 (G1N522;CYT)	0.11	1.67	High (45.4) <sup>2a</sup>
Ovomucoid <sup>f</sup>	P01003;15709 (IOVO;IOVO)	0.11	0.05	Intermediate (2.2) <sup>2a</sup>
UDP-glucuronic acid decarboxylase 1	11592 (G1NPP3;E1BV28)	0.10	0.07	Intermediate (2.2) <sup>2a</sup>
Uncharacterized/similar to mucin	5306;6950;1527 (G1N931;E1C037)	0.10	0.15	Low (1.4) <sup>2a</sup>
Insulin-like growth factor-binding protein 7 (IGFBP7)	18344 (G1N8M9;F1NVP4)	0.10	0.07	Intermediate (3.6) <sup>2a</sup>
Ganglioside GM2 activator <sup>a, g, t</sup>	24235 (G1N3B0;Q5ZK02)	0.10	-	High (12.9) <sup>2a</sup>
Synthenin-1 (SDCBP) <sup>f</sup>	24420 (G1NEY6;Q5ZHM8)	0.10	-	Intermediate (2.8) <sup>2a</sup>
Similar to procollagen C-endopeptidase enhancer	30796 (G1N443;F1NH70)	0.10	-	-

<sup>f</sup>in this order; turkey entries start with G. <sup>2a</sup>[28]; <sup>2b</sup>[31]; <sup>2c</sup>[32]; <sup>2d</sup>[36]. <sup>a</sup>gene expression up-regulated in chicken uterus upon mineralization [42] or upon sexual maturation of the hen [38]. Highest abundance in chicken uterus fluid during <sup>i</sup>initial phase, <sup>g</sup>growth phase, <sup>t</sup>terminal phase [37]

Ovocalyxin-36 (OCX36) is related to lipopolysaccharide-binding proteins (LPS), bactericidal permeability-increasing proteins (BPIP) and Plunc family proteins [15]. Immunofluorescence staining indicated a distribution of the protein throughout the eggshell and the membrane with highest intensity at the interface between membranes and mammillary cones [15]. Its expression was greatest in the uterus during eggshell mineralization and its distribution was restricted to isthmus and uterus [15]. Its similarity to the LPS/BPI/PLUNC family of proteins and its ability to bind to bacterial lipopolysaccharide suggested a function in egg antimicrobial defense [15, 16, 73, 74]. A sequence with approximately 70 % identity to chicken OCX36 was among the most abundant proteins of the quail eggshell proteome (Additional file 3: Table S3). Thus OCX36 is a major component of the eggshell proteomes of all three species analyzed (Table 1), indicating an essential role in eggshell production.

Ovocalyxin-32 (OCX32) is a member of the latexin family of carboxypeptidase inhibitors [14], with unknown function in the eggshell matrix. It was most abundant in uterine fluid during the initial phase of mineralization [37]. We found OCX32 neither in turkey eggshell matrix [44] nor in quail matrix, despite the addition of the chicken sequence to databases for possible cross-species identification with peptides of identical sequence. This is in contrast to the previously reported presence of OCX32 in quail eggshell using the chicken sequence [50]. However, we identified a highly abundant quail database entry (accession 11366; Additional file 3: Table S3) with high similarity to latexin. Closer examination of the sequence and the distribution of identified peptides indicated that the highly abundant peptides all matched to the N-terminal 71 amino acids of this entry (11366a), which was not very similar to latexin but showed a weak similarity to chicken OCX32 (C7G541\_CHICK; Additional file 3: Table S3; accession 11366 in Additional file 4: Figure S1) and

contained a predicted proteinase inhibitor I47 (latexin) domain. The remainder of the sequence of this entry with high similarity to latexin (11366b) was represented by only one peptide identified only once. In conclusion, evidence for the presence of OCX32 in the quail eggshell matrix was inconclusive, but we unequivocally identified a highly abundant fragment of a protein belonging to the latexin family (Table 1) that may have a function similar to that of OCX32.

Two less characterized so-called eggshell-specific proteins occasionally appearing in the literature are ovocalyxin-25 (OCX25) and ovocalyxin-21 (OCX21). OCX25 was described as a protease inhibitor with WAP and Kunitz domains [37]. A Kunitz protease inhibitor-type domain was also predicted for the very abundant uncharacterized protein in quail accession 23208 (Table 1). The gene id given for chicken OCX25 [37], LOC771972, was linked to five chicken protein sequences of different length that all showed blocks of almost complete sequence identity separated by gaps in the alignments to the uncharacterized chicken protein F1NPR2. The N-terminal 50 amino acids of accession 23208 (length 120aa) also showed a weak similarity to the same chicken protein. This was also the accession retrieved by searching the quail sequence database with the sequence of LOC771972 and accession 23208 was 45 % identical to aa178-311 of LOC771972 (isoform X3, XP\_004947248.1; Fig. 6). However, we are not sure whether this is sufficient to establish the protein encoded in entry 23208 as a homolog of OCX25.

OCX21 was among the most abundant proteins of the chicken eggshell matrix (IPI00574331.1) [28]. This protein is identical to gastroke-2 (E1C2G7\_CHICK) [32, 37], a protein of the gastric mucosal secretome and is therefore not truly eggshell-specific. We did not identify this protein in quail eggshell, although the quail sequence database contained two accessions (29506, 30653) with more than 90 % sequence identity to chicken and turkey OCX21/gastroke-2 sequences.

LOC771972	1	MLNPAGRSPDVAQGPQKSGPYIRSAECGRCEGSAMEPCPLLLLLLLLLLAPCGRGWAVSGR
LOC771972	61	APAETARTDGRSVLPGEFVKCHKTSMTAGSPCNCVSSKCPGSGTGCQEQEQEKDDFHP
23208	1	-----MGT
LOC771972	121	RTDPTTNCVNNCRDDGNCGRSRKCHIRCPFRCPQVPVPGKSRGSLGTMTAGTFPMGM
23208	4	<b>SPSPSPSAAKPDT</b> SPEK- <b>IPHTIGCCNSTHSSDTNCPKHRWCSPPLRRIFGVSDTIR</b> ---
LOC771972	181	<b>SPSPSPSAARPD</b> TYPKK <b>KVPHIIGCCNSTCSSD</b> DFPNHLRC <b>QPMRRSSRITVALSLLD</b>
23208	60	- <b>NNWG</b> ----- <b>CHLDLERGPKK</b> YNC <b>RNYSSGSRSCQTQI</b> - <b>VGCKNFKHFRTMQE</b> CO
LOC771972	241	LGC <b>WCS</b> DPEK <b>L</b> CRLI <b>PE</b> HR <b>L</b> C <b>R</b> K <b>R</b> TYCYACIPAL <b>RSC</b> RVFVHSS <b>CGGN</b> ANN <b>FRT</b> LA <b>EC</b> CO
23208	110	<b>QACKHG</b> ESWV-----
LOC771972	301	<b>QVCKHG</b> ESW <b>ARVGAQSRVKDAQGSTVCAW</b> GKQP <b>VLLGLPLSGLPGALPHRAVLRWPGSSS</b>
LOC771972	361	HGCSAAVHVGATRGLEAFLGDSNQRTAPSWELGLSSA

**Fig. 6** Alignment of quail accession 23208 to LOC771972 (OCX25). The isoform if LOC771972 shown is X3 (XP\_004947248.1). Identical amino acids are shaded yellow. Sequence regions covered by identified peptides are shown in bold red

Therefore, this major chicken eggshell protein is not part of the quail eggshell proteome. Remarkably, however, all three species contain among their major shell matrix proteins a protein of the gastric secretome (see below).

#### Other major proteins

Egg white proteins, such as ovalbumin, lysozyme C and ovotransferrin were among the first eggshell proteins identified [3–5] and were invariably found among the major proteins of eggshell matrices [28, 44] (Table 1). Among its major components quail eggshell matrix contained ovalbumin, avidin, ovostatin-like protein, ovalbumin-related protein X, ovotransferrin, lysozyme C, cystatin and ovomucoid. All of these were also identified in chicken and turkey eggshell matrices at different abundances (Table 1), but ovalbumin was always among the most abundant eggshell matrix components, in agreement with its very high abundance in egg white. No specific function in the mineralization process at the molecular level has been reported for these proteins so far. Some of these proteins show antimicrobial activity mediated by different mechanisms, such as degradation of bacterial cell walls, inhibition of microbial proteases or sequestration of components essential for microbial growth [75]. Therefore, these proteins are thought to contribute to general egg defense against microbial contamination, which can also happen during shell growth in the uterus. However, because several of these proteins were also shown to influence calcium carbonate mineralization *in vitro*, a direct effect on eggshell mineralization cannot be excluded at present. Many of these egg white proteins are very abundant and therefore even weak binding of calcite could affect mineralization. Egg white proteins are predominantly produced in the magnum section of the oviduct and not in uterus [38, 42] and may migrate together with the unfinished egg from the site of their synthesis to the uterus. However, the messages for lysozyme and

ovotransferrin [4, 5] were also detected at much lower levels in white isthmus, red isthmus and uterus tissues. At present it is not known whether these messages are translated into protein, and, if they are, what percentage of these abundant shell matrix proteins is derived from these alternative sources.

Major proteins that we identified only in quail eggshell matrix were an uncharacterized member of the latexin family (accession 11366a) and an uncharacterized proteinase inhibitor containing a predicted Kunitz domain (accession 23208). While the first may have a similar function as OCX32, the latter may be functionally related to OCX25. Another major new eggshell matrix protein was gastric intrinsic factor (GIF). GIF is part of the gastric mucosal secretome and binds cobalamin (vitamin B12). Interestingly each of the three proteomes considered in this report contained among its most abundant proteins a protein of gastric origin that was not identified in the respective other species' eggshell proteome. In chicken this was gastrokine-2/OCX21 [28] and in turkey this was trefoil factor 2 (TFF2) [44]. The role of these gastric proteins in eggshell production and the origin of the eggshell matrix forms remain unknown at present.

EGF-like repeats and discoidin I-like domains 3 protein (EDIL3/Del1) and tsukushin were among the most abundant proteins with an iBAQ of more than 1 % (Table 1) in all three species analyzed. EDIL3 is a widespread extracellular calcium-binding protein that was most abundant in chicken uterine fluid during the initial phase of eggshell mineralization [37]. No connection of this protein to mineralization or eggshell production is known at present. This is also true for tsukushin, a member of the small leucine-rich proteoglycan family [76]. This protein was highly abundant in quail and turkey eggshell matrix (Table 1) but comparatively less abundant in chicken eggshell matrix. In chicken uterus fluid it reached the highest abundance during growth

**Table 2** Phosphoproteins and phosphosites

Accession	Peptide	Number of P-sites	Probability <sup>2</sup>	Mod/unmod
<b>10704/17712</b> (Golgi phosphoprotein 4) <sup>1a</sup>	232IQVSpSHEESQAPGLLLSDLK <sub>251</sub>	1	0.806	3/3
	8VNAQMDDVQVHpsYPK <sub>22</sub>	1	1	1/7
<b>11366</b> (latexin family)	51IMFEHpsSREDLSYNVAQVK <sub>68</sub>	1	0.998	1/48
<b>13580</b> (HSPA13) <sup>1a</sup>	96NSGDAPsSDEENKFVK <sub>109</sub>	1	1	2/7
<b>14227</b> (cystatin) <sup>1a</sup>	97SSADLQpsSCEFHDPEMAK <sub>114</sub>	1	1	6/11
<b>18232</b> (PPIB)	59GDGTGKpsSIYGDRFPDENFK <sub>78</sub>	1	0.989	1/1 <sup>3</sup>
<b>188</b> (VIT1; not phospho domain)	1721ESVLSDPsSGVSEYEKDNK <sub>1738</sub>	1	0.97	3/5
<b>2265</b> (β-amyloid 751) <sup>1a</sup>	411AVIQHFQEKVEpsSLEQEAANER <sub>431</sub>	1	1	1/1
	420VEpsSLEQEAANER <sub>431</sub>	1	1	2/6
<b>3917</b> (secretogranin-2) <sup>1b</sup>	115MPLGHYEDp[SS]RDSPFK <sub>130</sub>	1	[0.89]	1/1
	483RVPVPASepsSDLQEQLEQAIR <sub>504</sub>	1	0.961	4/10
<b>5112;6916</b> (SERPINF2) <sup>1a</sup>	24LADTWETYGTSPSISTpsPETGDEESPGDK <sub>52</sub>	1	0.95	1/20
	53ATAGAVpsSCHEQEPSGK <sub>68</sub>	1	1	1/17
	284YPLSWFpTLESQDIQVAK <sub>300</sub>	1	0.958	1/115
	55IpsSEEGEGEEK <sub>65</sub>	1	1	1/3
	55IpsSEEGEGEKNCDLTWK <sub>72</sub>	1	1	7/16
	55IpsSEEGEGEKNCDLTWKK <sub>73</sub>	1	1	1/4
<b>5116</b> (Fibroleukin)	110Dp[SS]NEFLPPNAETPAEIQDNR <sub>131</sub>	1	[1]	1/1
	102LQADDNQERDp[SS]NEFLPPNAETPAEIQD NR <sub>131</sub>	1	[1]	2/2
<b>5306</b> (uncharacterized mucin)	170QIECQAEDYPEIpsSIEQVQWQCDVHYG LVCK <sub>200</sub>	1	1	21/51
<b>5467</b> (QSQX1) <sup>1a</sup>	259EHFSPGNLLLDYAIPITpsSGEEAAASAR <sub>286</sub>	1	0.931	9/23
	311EGETGRPGpsSEL <sub>323</sub>	1	0.968	3/4
<b>5586</b> (DNAJC3)	212QIEpsSAEEFIR <sub>221</sub>	1	1	2/30
<b>621</b> (Semaphorin-3G) <sup>1a</sup>	197WEADpsSREHTIVSR <sub>210</sub>	1	0.993	5/8
	553DILHLASpsSADR <sub>563</sub>	1	0.976	23/53
	553DILHLAps[SS]ADRQR <sub>565</sub>	1	[1]	2/2
<b>713</b> (713a; bone sialoprotein 2) <sup>1b</sup>	28AGDpsSEENAVLK <sub>38</sub>	1	1	3/3
<b>713</b> (713b; OC116) <sup>1a</sup>	498KEDVHVAEDIGEFAPYIPYVDAATITR <sub>524</sub>		0.94	14/655
	499EDVHVAEDIGEFAPYIPYVDAATITR <sub>524</sub>	1	0.98	48/512
	545ANIQVGENDGsSAGpsSpTGATSEANVIPAVNAR <sub>575</sub>	2	0.993; 0.829	23/1111
	545ANIQVGENDGsSAGsTGATpSEANVIPAVNAR <sub>575</sub>	1	0.814	240/1111
	649SGQRPVGHKpsSLPATMTTQGAR <sub>669</sub>	1	0.984	2/13
	657HpsSLPATMTTQGAR <sub>669</sub>	1	1	2/308
	692GSQWVTAGQEGpsSREVTAGPK <sub>712</sub>	1	0.916	3/3
	787SRAQQGVAPVHpsSMER <sub>801</sub>	1	1	1/22
	787SRAQQGVAPVHpsSMERETVTPER <sub>808</sub>	1	0.998	18/29
	789AQQGVAPVHpsSMER <sub>801</sub>	1	1	125/251
	789AQQGVAPVHpsSMERETVTPER <sub>808</sub>	1	0.999	202/240
	881STDVPRDPWWGpsSTSPQTQHTR <sub>902</sub>	1	0.915	43/499
	887DPWWGpsSTSPQTQHTR <sub>902</sub>	1	0.948	15/224
	903GSMVAGpsSFLHLHGGQR <sub>918</sub>	1	0.977	112/538
	903GSMVAGpsSFLHLHGGQR <sub>918</sub>	2	1; 1	21/538

**Table 2** Phosphoproteins and phosphosites (Continued)

	919MGpSLAELER <sub>927</sub>	1	1	46/353
	919MGSLAELERpSRQVEQVR <sub>935</sub>	1	1	8/8
<b>7172</b> (Nucleobindin-2) <sup>1a</sup>	101AKMDpSVQDTGIDHQALLK <sub>118</sub>	1	1	1/7
<b>7717</b> (β-amyloid 695) <sup>1a</sup>	215AVIQHFQEKVEpSLEQEANER <sub>235</sub>	1	1	1/1
	224VEpSLEQEANER <sub>235</sub>	1	1	2/5
<b>Q6V115</b> (Ovalbumin) <sup>1a</sup>	57VHFDFKLPFGFDpSIEAQCGTSANVHSSLR <sub>84</sub>	1	1	122/487
	57VHFDFKLPFGFDpSIEAQCGpT <sub>S</sub> ANVHSSLR <sub>84</sub>	2	1;0.875	8/487
	57VHFDFKLPFGFDpSIEAQCGTSANVHSSLRDILNQITK <sub>93</sub>	1	1	3/20
	62LPGFGDpSIEAQCGTSANVHSSLR <sub>84</sub>	1	1	300/733
	62LPGFGDpSIEAQCGT <sub>p</sub> SANVHSSLR <sub>84</sub>	2	1;0.976	41/733
	265LpTEWTSSIMEER <sub>277</sub>	1	1	1/614
	341DVVGpSAEAGVDATEEFR <sub>357</sub>	1	1	1415/1419
	341DVVGpSAEAGVDATEEFRADHPFLFCVK <sub>367</sub>	1	1	66/66
<b>Q9DG12</b> (IGFBP5) <sup>1b</sup>	74EHEEPTpT <sub>S</sub> EMTEETYPPK <sub>91</sub>	1	0.809	1/6
	74EHEEPTp[TS]EMTEETYPPK <sub>91</sub>	1	[0.962]	2/6
	150HMEASLQELKpSSQR <sub>163</sub>	1	0.802	4/5
	206LPGTDYLSGDLQCHAFDpSS <sub>N</sub> VE <sub>227</sub>	1	0.806	3/4
<b>Q9I832</b> (Osteopontin) <sup>1a</sup>	22SKQHAI <sub>S</sub> A <sub>p</sub> [SS]EEKYDPR <sub>38</sub>	1	[0.958]	1/5
	22SKQHAI <sub>S</sub> A <sub>p</sub> SpEEKYDPR <sub>38</sub>	2	0.913; 0.913	4/5
	24QHAI <sub>S</sub> A <sub>p</sub> [SS]EEKYDPR <sub>38</sub>	1	[1]	21/45
	24QHAI <sub>S</sub> A <sub>p</sub> SpEEKYDPR <sub>38</sub>	1	0.973	10/45
	24QHAI <sub>p</sub> SA <sub>p</sub> SpEEKYDPR <sub>38</sub>	2	1; 0.951	3/45
	24QHAI <sub>S</sub> A <sub>p</sub> SpEEKYDPR <sub>38</sub>	2	1;1	7/45
	24QHAI <sub>S</sub> A <sub>p</sub> SpEEKpYDPR <sub>38</sub>	3	0.967;0.967;0.967	1/45
	24QHAI <sub>p</sub> SA <sub>p</sub> SpEEKYDPR <sub>38</sub>	3	1;1;1	3/45
	133GDNAGRGDpSVAYGFR <sub>147</sub>	1	1	5/33
	139GDpSVAYGFR <sub>147</sub>	1	1	1/54
	164KLIEDDApT <sub>S</sub> IEDGDSQPAGLWWPK <sub>186</sub>	1	0.779	1/4
	164KLIEDDAp[TT]EDGDpSQPAGLWWPK <sub>186</sub>	2	[1];1	1/4
	164KLIEDDATTEDGDSQPAGLWWPKepSR <sub>189</sub>	1	1	5/10
	164KLIEDDApT <sub>S</sub> IEDGDSQPAGLWWPKepSR <sub>189</sub>	2	0.912;1	1/10
	164KLIEDDAp[TT]EDGD <sub>S</sub> QPAGLWWPKepSR <sub>189</sub>	2	[0.875];1	1/10
	164KLIEDDAp[TT]EDGDpSQPAGLWWPKepSR <sub>189</sub>	3	[1];1;1	2/10
	164KLIEDDApT <sub>S</sub> IEDGDpSQPAGLWWPKepSR <sub>189</sub>	3	0.927;1;1	1/10
	165LIEDDApT <sub>S</sub> IEDGDSQPAGLWWPK <sub>186</sub>	1	0.838	1/1
	165LIEDDApT <sub>S</sub> IEDGDSQPAGLWWPKepSR <sub>189</sub>	2	0.946;1	3/4
	165LIEDDApT <sub>S</sub> IEDGDpSQPAGLWWPKepSR <sub>189</sub>	3	0.819; 0.795;1	1/4
	190EQNpSRELPHQpSVENDpSRPK <sub>209</sub>	3	1;1;1	2/3
	190EQNpSRELPHQSVENDpSRPK <sub>209</sub>	2	0.996; 0.998	1/3
	195ELPQHQPpSVENDpSRPK <sub>209</sub>	2	1;1	35/49
	195ELPQHQPpSVENDSRPK <sub>209</sub>	1	1	4/49
	195ELPQHQS <sub>V</sub> ENDpSRPK <sub>209</sub>	1	1	8/49
	210FDpSREVDGGDSK <sub>221</sub>	1	1	1/4

**Table 2** Phosphoproteins and phosphosites (*Continued*)

210FD <u>p</u> SREVDGGD <u>p</u> SK <sub>221</sub>	2	1;1	3/4
214EVDGGD <u>p</u> SKASAGVD <u>p</u> SR <sub>229</sub>	2	1;0.998	3/4
214EVDGGD <u>S</u> KA <u>p</u> SAGVD <u>p</u> SR <sub>229</sub>	2	0.988;1	1/4
230ESQQGSVPAVDASNQ <u>p</u> TLESAEDAEDR <sub>254</sub>	1	0.955	1/7

<sup>1</sup>identified as phosphoprotein before in quail or other species; <sup>1a</sup>chicken or quail, <sup>1b</sup>other species. <sup>2</sup>only the highest localization probability in the sequence is shown. <sup>3</sup>S<sub>66</sub> also occurred 9 times unmodified in the cleavage product <sub>66</sub>SIYGD<sup>R</sup>FPDENFK<sub>78</sub>. Underlined amino acids in the peptides shown represent other possible phosphorylation sites with lower probabilities or lower frequency. The number of total phosphorylated peptides always includes peptides containing sites with lower localization probabilities. Square brackets delimit neighboring amino acids if no unequivocal localization of the phosphorylation site(s) was possible

and terminal phase [37]. The expression of both genes was up-regulated when an unfinished egg was present in the uterus [42] indicating a specific, but presently unknown, function.

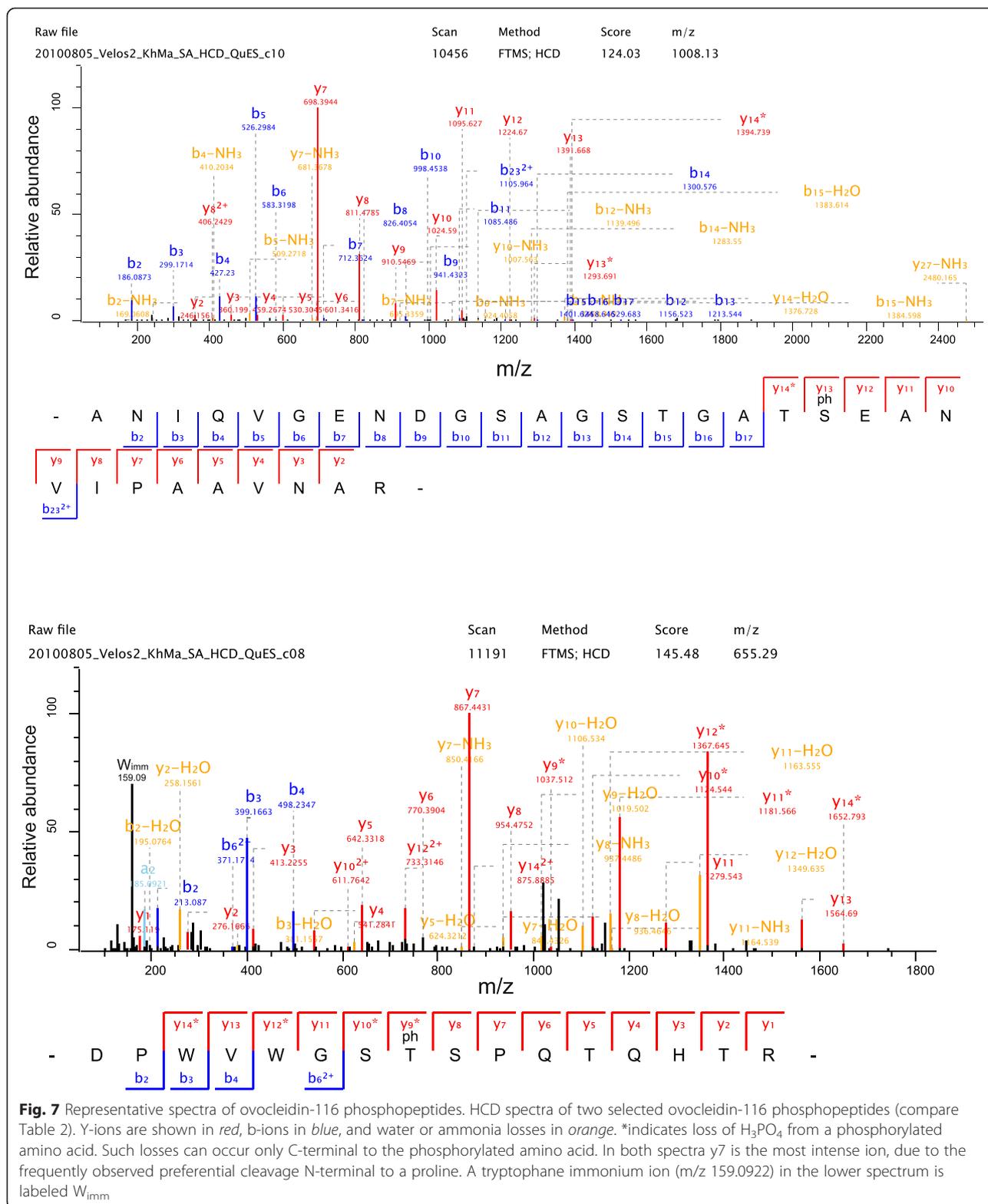
The list of major proteins (Table 1) contains two proteins that were previously shown to be induced and secreted in the uterus upon entry of the egg before the advent of large scale transcriptomic studies, and may be supposed to have some specific function in shell production. These are osteopontin and glypican-4 [18, 19]. Osteopontin is a member of the SIBLING family of mineralization-related secreted phosphoproteins [71, 72]. It was identified as a major protein of the shell matrix of all three phasianid species [28, 29, 44] (Table 1). The involvement of this multifunctional protein in mineralization processes is reviewed in [72, 77, 78]. Its activity is generally inhibitory and depends on phosphorylation. The localization of osteopontin predominantly at the surface of mammillary cones and eggshell pores and at the margins of calcite columns of the palisade layer supports an inhibitory function in eggshell mineralization by binding to selected crystal surfaces [24]. Glypican-4 is one of several related cell surface heparane sulfate proteoglycans with a GPI anchor and may be released from the surface by Notum protein (present in the quail eggshell matrix as a minor protein; Additional file 3: Table S3). Glypicans have been implicated in various regulatory processes at the cell surface [79], but no connection to mineralization events has been reported so far. Glypican-4 was identified in all three species analyzed, but was a major protein only in turkey and quail eggshell matrix (Table 1).

Also among the most abundant eggshell proteins in all three species were serum albumin, lactadherin (milk fat globule membrane protein 8), extracellular fatty acid-binding protein, the extracellular chaperone clusterin, extracellular serpin E2, the extracellular matrix protein fibronectin, and ubiquitin. Nothing is known about their specific role, if any, in eggshell production. Lactadherin and clusterin were most abundant in chicken uterus fluid during the initial phase of mineralization, while ubiquitin was most abundant in the terminal phase and albumin in the growth phase [37].

### Phosphoproteins

We showed previously that higher energy collisional dissociation (HCD) fragmentation, also used in the present report, is well suited to determine peptide phosphorylation sites [80]. Applied to low-complexity proteomes such as those of biomineral matrices, this technique can yield useful information about major phosphoproteins and their phosphorylation sites without prior enrichment of modified peptides [64]. Because phosphorylation was reported previously to potentially affect biomineralization processes (reviewed in [81–83]) we included it among the variable modifications used for MaxQuant search of our raw files.

In this way, we identified 21 phosphoproteins with a total of 56 different phosphorylation sites each with a localization probability in the sequence of >0.75 [84] (Table 2; only highest probability shown). This was less than in a previous study using phosphopeptide enrichment prior to analysis of the chicken eggshell phosphoproteome, which enables the identification of even traces of phosphorylation [29]. However, such sites may be less important for protein function, at least in proteins without catalytic or regulatory function. We compared the number of repeatedly identified phosphorylated versions of peptides to the number of non-phosphorylated forms of these peptides. This indicated that most sites were phosphorylated only partially (Table 2). The highest site occupancy, close to 100 %, was found for Ser<sub>346</sub> of ovalbumin. This site and a second one (S<sub>69</sub>), with an estimated site occupancy of approximately 25 %, were already detected previously in several avian species including chicken [85]. In addition, we detected three previously unknown phosphorylation sites at Thr<sub>76</sub>, Ser<sub>77</sub> and Thr<sub>266</sub> of ovalbumin with very low site occupancy (Table 2). Ser<sub>80</sub> in cystatin [86], corresponding to Ser<sub>103</sub> in accession 14227 is a previously known phosphorylation site in a quail protein. However, we could not confirm the complete site occupancy reported previously (Table 2). As in chicken [29], the major phosphoproteins of the quail eggshell were OC116 with 10 different phosphosites (Fig. 7) and site occupancies up to approximately 80 %, and osteopontin with 14 different phosphosites and occupancy of up to 75 %.



While nothing is known about the specific function of OC116 phosphorylation, osteopontin is one of the few matrix proteins better characterized in this regard. This multifunctional glycosylated phosphoprotein generally

inhibits mineralization in a phosphorylation-dependent way. This was shown by comparison of phosphorylated osteopontin to enzymatically de-phosphorylated or unphosphorylated recombinant osteopontin in *in vitro*

crystallization assays using different minerals, such as calcium phosphate [87], calcium oxalate [88], or calcium carbonate [89]. Calcification of cultured human smooth muscle cells, used as a model of vascular calcification, was also inhibited by phosphorylated osteopontin, but not by an enzymatically de-phosphorylated form [90]. The extent of phosphorylation may also play a role. Moderately phosphorylated osteopontin, such as bone osteopontin, inhibited mineral formation in a hydroxyapatite crystallization assay, while the more heavily phosphorylated milk osteopontin rather promoted mineralization [91]. Eleven of fourteen quail osteopontin phosphorylation sites were previously identified in chicken eggshell osteopontin by mass spectrometric analysis after phosphopeptide enrichment [29] or by Edman degradation of peptides isolated from metabolically  $^{32}\text{P}$ -labeled cultured chicken osteoblast osteopontin [92] (Fig. 8). Differences may be species specific, tissue specific, or due to differences in isolation and detection methods.

Comparison between quail and chicken ovocleidin-116 phosphorylation sites provided a different result (Fig. 9). Only four of the twelve quail OC116 phosphorylation sites were shared between the species, although eight of the amino acids phosphorylated in the quail protein were conserved in the chicken sequence. The higher number of phosphorylation sites in chicken OC116 may be due to phosphopeptide enrichment prior to analysis, enabling the identification of traces of phosphopeptides, possibly with no functional importance [93]. Overall, phosphorylation sites in OC116 were less well conserved than osteopontin phosphorylation sites. It is possible that the overall extent of phosphorylation is more important in OC116 than conservation of exact positions.

A total of 12 of the phosphoproteins were previously identified as such in chicken [29], and three in mammals (Table 2). More than half of the phosphorylation sites

(29 sites) agree with the consensus motif for phosphorylation by the secreted kinase FAM20C (S/T-X-[D,E,pS]), a kinase that appears to be identical with Golgi casein kinase and is known to modify many biomineralization-related extracellular proteins, such as the members of the SIBLING cluster [94, 95]. We also identified this kinase among the major quail eggshell matrix proteins (Table 1). Other kinase motifs detected in phosphopeptides were those recognized by casein kinase 2 (CH-IL, four sites), protein kinase A (PKA, three sites), protein kinase C (PKC, two sites), pyruvate kinase (PK1, one site), ribosomal S6 kinase (RSK9, one site), and casein kinase 1 (CK1, one site). The phosphorylation site data, such as best score, mass error, and site probability as determined by MaxQuant are in Additional file 5: Table S4, which also contains data of peptides with lower site probability than 0.75 that were not included into Table 2.

## Conclusions

The quail eggshell matrix proteome shares 50 % of its 622 identified proteins with the shell proteomes of chicken and turkey. Nevertheless we did not find the quail homologs of several major proteins of chicken and turkey eggshell matrix, including members of the so-called eggshell-specific group of proteins. Only ovocleidin-116 and ovocalyxin-36 were unequivocally identified as major eggshell proteins in all three species analyzed and these may therefore be essential for eggshell mineralization in general (Table 1). Osteopontin, like OC116 a member of the SIBLING cluster of biomineralization-related proteins, and a major phosphoprotein in all three eggshell matrices (Table 1), may also be essential. For almost all major proteins no specific function in eggshell mineralization is known. However, there are many suggestions for possible functions based on known enzymatic or binding activities of

Q9I832_COTJA	1	MKLTLLCLCFISITAAWVSKSKQHAI	SASS	EKEYDPRSHHTRYHQDHVDSQSQEHLQQ
OSTP_CHICK	1	MKLAFLCLCFISITAAWVSKSRQHAI	SASS	EKEYDPRSHHTRYHQDHVDSQSQEHLQQ
		***	:	*****
Q9I832_COTJA	61	TQNDLASLQQTHYSSEENADVPEQPDFDPVPSKSKQETVDDDDDDNDSNDTDESDEVFTD		
OSTP_CHICK	61	TQNDLASLQQTHYSSEENADVPEQPDFDIPSKSQEAVDDDDDDNDSNDTDESDEVVTD		
		*****	:	*****
Q9I832_COTJA	121	FPTEAPVAPFNRGDNAGRGS	S	VAYGFRAKAHVVKASKIRKAARKLIEDDAT
OSTP_CHICK	121	FPTEAPVTPFNRGDNAGRGS	S	VAYGFRAKAHVVKASKLRKAARKLIEDDAT
		*****	:	*****
Q9I832_COTJA	181	GLWWPKESREQNSRELPHQHS	S	VENDSRPKFDSREVDGGDSKASAGVDSRESQGSVPAVDA
OSTP_CHICK	181	GLWLPKESREQDSRELAHQHS	S	VENDSRPRFDSPEVGGGDSKASAGVDSRESLAGRSVAVDT
		***	*****	*****
Q9I832_COTJA	241	SNQTLESAEDAEDRHSIENNEVTR		
OSTP_CHICK	241	SNQTLESAEDAEDRHSIENNEVTR		
		*****	:	*****

**Fig. 8** Comparison of phosphorylation sites in quail and chicken osteopontin. Phosphorylated amino acid residues of eggshell osteopontin are highlighted by yellow shading. Phosphorylation sites identified in metabolically  $^{32}\text{P}$ -labeled chicken osteoblast osteopontin [92] are underlined



It also contains other relevant data such as scores, charge states, and mass accuracy. (XLSX 2472 kb)

**Additional file 3: Table S3.** All accepted identifications with annotations derived from FASTA searches. Proteins and protein groups predicted to belong together are indicated by alternate white and grey background. The table also contains UniProt accession numbers of similar chicken and turkey proteins with FASTA-derived percentages of sequence identity and e-values, number of peptides and iBAQ-derived abundances. (DOCX 419 kb)

**Additional file 4: Figure S1.** Alignments of quail accessions predicted to contain sequences (or partial sequences) of different proteins with the most similar chicken or turkey protein sequences. Also included is the location of identified peptides. (DOCX 143 kb)

**Additional file 5: Table S4.** Curated version of the MaxQuant output file Phospho (STY) sites with phosphopeptide sequences, localization probabilities, scores, mass errors and other relevant data. Also included are phosphopeptides with site probabilities lower than 0.75, which were not included into Table 2. (XLS 99 kb)

### Abbreviations

Aa: Amino acid; FDR: False discovery rate; HCD: Higher-energy collision-induced dissociation; iBAQ: Intensity-based absolute quantification; OC: Ovocleidin; OXC: Ovocalyxin; PAGE: Polyacrylamide gel electrophoresis; SIBLING: Small integrin-binding ligand N-linked glycoproteins.

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

KM conceived the study, performed sample preparation and data acquisition. MM supplied methodological expertise. Both authors took part in the design of the study and were critically involved in manuscript drafting. Both authors read and approved the final manuscript.

### Acknowledgements

We thank Gaby Sowa for preparing the capillary columns for LC-MS/MS, Korbinian Mayr and Igor Paron for keeping the mass spectrometers in excellent condition, Markus Räschle for help with setting up Bioedit programs, and Ryouka Kawahara-Miki (Genome Research Center of the Nodai Research Institute) for his response to our questions.

Received: 26 May 2015 Accepted: 11 August 2015

Published online: 27 August 2015

### References

- Solomon SE. The eggshell: strength, structure and function. *Brit Poult Sci.* 2010;51 Suppl 1:52–9.
- Hincke MT, Nys Y, Gautron J, Mann K, Rodríguez-Navarro AB, McKee MD. The eggshell: structure, composition and mineralization. *Front Biosci.* 2012;17:1266–80.
- Hincke MT. Ovalbumin is a component of the chicken eggshell matrix. *Connect Tissue Res.* 1995;31:227–33.
- Hincke MT, Gautron J, Panhéleux M, Garcia-Ruiz J, McKee MD, Nys Y. Identification of lysozyme as a component of eggshell membranes and eggshell matrix. *Matrix Biol.* 2000;19:443–53.
- Gautron J, Hincke MT, Panhéleux M, Garcia-Ruiz JM, Boldic T, Nys Y. Ovotransferrin is a matrix protein of the hen eggshell membranes and basal calcified layer. *Connect Tissue Res.* 2001;42:255–67.
- Hincke MT, Tsang CPW, Courtney M, Hill V, Narbaitz R. Purification and immunochemistry of a soluble matrix protein of the chicken eggshell (ovocleidin-17). *Calcified Tissue Int.* 1995;56:578–83.
- Mann K, Siedler F. The amino acid sequence of ovocleidin-17, a major protein of the avian eggshell calcified layer. *Biochem Mol Biol Int.* 1999;47:997–1007.
- Hincke MT, Gautron J, Tsang CPW, McKee MD, Nys Y. Molecular cloning and ultrastructural localization of the core protein of an eggshell matrix proteoglycan, ovocleidin-116. *J Biol Chem.* 1999;46:32915–23.
- Mann K, Hincke MT, Nys Y. Isolation of ovocleidin-116 from chicken eggshells, correction of its amino acid sequence and identification of disulfide bonds and glycosylated Asn. *Matrix Biol.* 2002;21:383–7.
- Nimtaz M, Conradt HS, Mann K. LacdiNAc (GalNAc $\beta$ 1-4GlcNAc) is a major motif in N-glycan structures of the chicken eggshell protein ovocleidin-116. *Biochim Biophys Acta.* 2004;1675:71–80.
- Carrino DA, Rodríguez JP, Caplan AI. Dermatan sulfate proteoglycans from the mineralized matrix of the avian eggshell. *Connect Tissue Res.* 1997;36:175–93.
- Horvat-Gordon M, Yu F, Burns D, Leach RM. Ovocleidin (OC-116) is present in avian skeletal tissues. *Poult Sci.* 2008;87:1618–23.
- Bardet C, Vincent C, Lajarille M-C, Jaffredo T, Sire J-Y. OC-116, the chicken ortholog of mammalian MEPE found in eggshell, is also expressed in bone cells. *J Exp Zool.* 2010;314B:653–62.
- Gautron J, Hincke MT, Mann K, Panhéleux M, Bain M, McKee MD, et al. Ovocalyxin-32, a novel chicken eggshell matrix protein. Isolation, amino acid sequencing, cloning and immunohistochemical localization. *J Biol Chem.* 2001;42:39243–52.
- Gautron J, Murayama E, Vignal A, Morisson M, McKee MD, Réhault S, et al. Cloning of ovocalyxin-36, a novel chicken eggshell protein related to lipopolysaccharide-binding proteins, bactericidal permeability-increasing proteins and Plunc family proteins. *J Biol Chem.* 2007;282:5273–86.
- Gautron J, Réhault-Godbert S, Pascal G, Nys Y, Hincke MT. Ovocalyxin-36 and other LBP/BPI/Plunc-like proteins as molecular actors of the mechanisms of the avian egg natural defences. *Biochem Soc Trans.* 2011;39:971–6.
- Nys Y, Gautron J, Garcia-Ruiz JM, Hincke MT. Avian eggshell mineralization: biochemical and functional characterization of matrix proteins. *Comptes Rendus Palevol.* 2004;3:549–62.
- Lavelin I, Meiri N, Pines M. New insight in eggshell formation. *Poult Sci.* 2000;79:1014–7.
- Lavelin I, Meiri N, Einat M, Genina O, Pines M. Mechanical strain regulation of the chicken glypican-4 gene expression in the avian eggshell gland. *Am J Physiol Integr Comp Physiol.* 2002;283:R853–61.
- Gautron J, Bain M, Solomon SE, Nys Y. Soluble matrix of hen's eggshell extracts changes in vitro the rate of calcium carbonate precipitation and crystal morphology. *Brit Poult Sci.* 1996;37:853–66.
- Dominguez-Vera JM, Gautron J, Garcia-Ruiz JM, Nys Y. The effect of avian uterine fluid on the growth behavior of calcite crystals. *Poult Sci.* 2000;79:901–7.
- Marin-Garcia L, Frontana-Urbe BA, Reyes-Grajeda JP, Stojanoff V, Serrano-Posada HJ, Moreno A. Chemical recognition of carbonate anions by proteins involved in biomineralization processes and their influence on calcite crystal growth. *Crystal Growth Design.* 2008;8:1340–5.
- Hernández-Hernández A, Gómez-Morales J, Rodríguez-Navarro AB, Gautron J, Nys Y, Garcia-Ruiz JM. Identification of some active proteins in the process of hen eggshell formation. *Crystal Growth Design.* 2008;8:4330–9.
- Chien Y-C, Hincke MT, Vali H, McKee MD. Ultrastructural matrix-mineral relationships in avian eggshell, and effects of osteopontin on calcite growth *in vitro*. *J Struct Biol.* 2008;163:84–99.
- Wellman-Labadie O, Lakshminarayanan R, Hincke MT. Antimicrobial properties of avian eggshell-specific C-type lectin-like proteins. *FEBS Lett.* 2008;582:699–704.
- Wellman-Labadie O, Picman J, Hincke MT. Antimicrobial activity of cuticle and outer eggshell protein extracts from three species of domestic birds. *Brit Poultry Sci.* 2008;49:133–43.
- International Chicken Genome Sequencing Consortium. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Science.* 2004;432:695–716.
- Mann K, Maček B, Olsen JV. Proteomic analysis of the acid-soluble organic matrix of the chicken calcified eggshell layer. *Proteomics.* 2006;6:3801–10.
- Mann K, Olsen JV, Maček B, Gnad F, Mann M. Phosphoproteins of the chicken eggshell calcified layer. *Proteomics.* 2007;7:106–15.
- Mikšik I, Eckhardt A, Sedláková P, Mikulíková K. Proteins of insoluble matrix of avian (*Gallus gallus*) eggshell. *Connect Tissue Res.* 2007;48:1–8.
- Mikšik I, Sedláková P, Lacinová K, Pataridis S, Eckhardt A. Determination of insoluble avian eggshell matrix proteins. *Anal Bioanal Chem.* 2010;397:205–14.
- Rose-Martel M, Du J, Hincke MT. Proteomic analysis provides new insight into the eggshell cuticle. *J Proteom.* 2012;75:2697–706.

33. Mikšik I, Ergang P, Pácha J. Proteomic analysis of chicken eggshell cuticle membrane layer. *Anal Bioanal Chem*. 2014;406:7633–40.
34. Rose-Martel M, Smiley S, Hincke MT. Novel identification of matrix proteins involved in calcitic biomineralization. *J Proteomics*. 2015;116:81–96.
35. Kaweewong K, Garnjanagoonchorn W, Jirapakkul W, Roytrakul S. Solubilization and identification of hen eggshell membrane proteins during different times of chicken embryo development using the proteomic approach. *Protein J*. 2013;32:297–308.
36. Sun C, Xu G, Yang N. Differential label-free quantitative proteomic analysis of avian eggshell matrix and uterine fluid proteins associated with eggshell mechanical property. *Proteomics*. 2013;13:3523–36.
37. Marie P, Labas V, Brionne A, Harichaux G, Hennequet-Antier C, Nys Y, et al. Quantitative proteomics and bioinformatics analysis provide new insight into protein function during avian eggshell biomineralization. *J Proteomics*. 2015;113:178–93.
38. Dunn IC, Wilson PW, Lu Z, Bain MM, Crossan CL, Talbot RT, et al. New hypotheses on the function of the avian shell gland derived from microarray analysis comparing tissue from juvenile and sexually mature hens. *Gen Comp Endocr*. 2009;163:225–32.
39. Jonchère V, Réhault-Godbert S, Hennequet-Antier C, Canau C, Sibut V, Cogburn LA, et al. Gene expression profiling to identify eggshell proteins involved in physical defense of the chicken egg. *BMC Genomics*. 2010;11:55.
40. Jonchère V, Brionne A, Gautron Y, Nys Y. Identification of uterine ion transporters for mineralization precursors of the avian eggshell. *BMC Physiol*. 2012;12:10.
41. Fan Y-F, Hou Z-C, Yi G-Q, Xu G-Y, Yang N. The sodium channel gene family is specifically expressed in hen uterus and associated with eggshell quality traits. *BMC Genet*. 2013;14:90.
42. Brionne A, Nys Y, Hennequet-Antier C, Gautron J. Hen uterine gene expression profiling during eggshell formation reveals putative proteins involved in the supply of minerals or in the shell mineralization process. *BMC Genomics*. 2014;15:220.
43. Dalloul RA, Long JA, Zimin AV, Aslam L, Beal K, Blomberg LE, et al. Multi-Platform next-generation sequencing of the domestic turkey (*Meleagris gallopavo*): Genome Assembly and analysis. *PLoS Biol*. 2010;8, e1000475.
44. Mann K, Mann M. The proteome of the calcified layer organic matrix of turkey (*Meleagris gallopavo*) eggshell. *Proteome Sci*. 2013;11:40.
45. Kawahara-Miki R, Sano S, Nunome M, Shimamura T, Kuwayama T, Takahashi S, et al. Next-generation sequencing reveals genomic features of Japanese quail. *Genomics*. 2013;101:345–53.
46. Panhéleux M, Bain M, Fernandez MS, Morales I, Gautron J, Arias JL, et al. Organic matrix composition and ultrastructure of eggshell: a comparative study. *Brit Poultry Sci*. 1999;40:240–52.
47. Lakshminarayanan R, Loh XJ, Gayathri S, Sindhu S, Banerjee Y, Kini RM, et al. Formation of transient amorphous calcium carbonate precursor in quail eggshell mineralization: an in vitro study. *Biomacromolecules*. 2006;7:3202–9.
48. Rahman MA, Moriyama A, Iwasawa A, Yoshizaki N. Cuticle formation in quail eggs. *Zool Sci*. 2009;26:496–9.
49. Iwasawa A, Rahman MA, Roy TK, Moriyama A, Yoshizaki N. Morphological and histochemical changes in the uterus epithelium during eggshell formation in quail. *J Poultr Sci*. 2010;47:183–9.
50. Portugal SJ, Cooper HJ, Zampronio CG, Wallace LL, Cassey P. Can museum egg specimens be used for proteomic analysis? *Proteome Sci*. 2010;8:40.
51. Shevchenko A, Tomas H, Havlis J, Olsen JV, Mann M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat Protoc*. 2006;1:2856–60.
52. Rappsilber J, Mann M, Ishihama Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat Protoc*. 2007;2:1896–906.
53. Olsen JV, Schwartz JC, Griep-Raming J, Nielsen ML, Damoc E, Denisov E, et al. A dual pressure linear ion trap-Orbitrap instrument with very high sequencing speed. *Mol Cell Proteomics*. 2009;8:2759–69.
54. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized ppb-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnol*. 2009;26:1367–72.
55. Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M. Andromeda – a peptide search engine integrated into the MaxQuant environment. *J Proteome Res*. 2011;10:1794–805.
56. Tyanova S, Temu T, Carlson A, Sinitcyn P, Mann M, Cox J. Visualization of LC-MS/MS proteomics data in MaxQuant. *Proteomics*. 2015;15:1453–6.
57. Neuhauser N, Michalski A, Cox J, Mann M. Expert system for computer-assisted annotation of MS/MS spectra. *Mol Cell Proteom*. 2012;11:1500–9.
58. Schwanhäusser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, et al. Global quantification of mammalian gene expression control. *Nature*. 2011;473:337–42.
59. Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paern J, et al. A new bioinformatics analysis tools framework at EMBL-EBI (2010). *Nucleic Acids Res*. 2010;38(Suppl):W695–9.
60. Altschul SF, Thomas TL, Schaffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997;25:3389–402.
61. Sievers F, Wilm A, Dineen DG, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*. 2011;7:539.
62. Hunter S, Jones P, Mitchell A, Apweiler R, Attwood TK, Bateman A, et al. InterPro in 2011: new developments in the family and domain prediction database. *Nucleic Acids Res*. 2011;40:D306–12.
63. Blom N, Sicheritz-Ponten T, Gupta R, Gammeltoft S, Brunak S. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics*. 2004;4:1633–49.
64. Mann M, Edsinger E. The *Lottia gigantea* shell matrix proteome: re-analysis including MaxQuant iBAQ quantitation and phosphoproteome analysis. *Proteome Sci*. 2014;12:28.
65. Zhang Q, Liu L, Zhu F, Ning Z, Hincke MT, Yang N, et al. Integrating de novo transcriptome assembly and cloning to obtain chicken ovocleidin-17 full-length cDNA. *Plos One*. 2014;9, e93452.
66. Freeman CL, Harding JH, Quigley D, Rodger PM. Simulations of ovocleidin-17 binding to calcite surfaces and its implications for eggshell formation. *J Phys Chem*. 2011;115:8175–83.
67. Mann K, Siedler F. Ostrich (*Struthio camelus*) eggshell contains two different C-type lectin-like proteins. Isolation, amino acid sequence, and post-translational modifications. *Biochim Biophys Acta*. 2004;1696:41–50.
68. Mann K, Siedler F. Amino acid sequences and phosphorylation sites of emu and rhea eggshell C-type lectin-like proteins. *Comp Biochem Physiol B*. 2006;143:160–70.
69. Lakshminarayanan R, Valiyaveetil S, Rao VS, Kini RM. Purification, characterization, and in vitro mineralization studies of a novel goose eggshell matrix protein, ansocalcin. *J Biol Chem*. 2003;278:2928–36.
70. Kawasaki K, Weiss KM. Evolutionary genetics of vertebrate tissue mineralization: the origin and evolution of the secretory calcium-binding phosphoprotein family. *J Exp Zool (Mol Dev Evol)*. 2006;306B:295–316.
71. Rowe PSN. The chicken or the egg: PHEX, FGF23 and SIBLINGs unscrambled. *Cell Biochem Funct*. 2012;30:355–75.
72. Staines KA, MacRae VE, Farquharson C. The importance of the SIBLING family of proteins on skeletal mineralization and bone remodeling. *J Endocrinol*. 2012;214:241–55.
73. Cordeiro CMM, Esmaili H, Anshah G, Hincke MT. Ovocalyxin-36 is a pattern recognition protein in chicken eggshell membranes. *PLoS One*. 2013;8, e84112.
74. Kovacs-Nolan J, Cordeiro C, Young D, Mine Y, Hincke MT. Ovocalyxin-36 is an effector protein modulating the production of proinflammatory mediators. *Vet Immunol Immunop*. 2014;160:1–11.
75. Abdou AM, Sato K, Kim M. Functional proteins and peptides of hen's egg origin. INTECH Open Access Publisher;2013. <http://www.intechopen.com/books/bioactive-food-peptides-in-health-and-disease/functional-proteins-and-peptides-of-hen-s-egg-origin>
76. Dellet M, Hu W, Papadaki V, Ohnuma S. Small leucine rich proteoglycan family regulates multiple signaling pathways in neural development and maintenance. *Dev Growth Differ*. 2012;54:327–40.
77. Giachelli CM, Steitz S. Osteopontin: a versatile regulator of inflammation and biomineralization. *Matrix Biol*. 2000;19:615–22.
78. Sodek J, Ganss B, McKee MD. Osteopontin. *Crit Rev Oral Biol Med*. 2000;11:279–303.
79. Filmus J, Capurro M, Rast J. Glypicans. *Genome Biol*. 2008;9, e224.
80. Nagaraj N, D'Sousa RCJ, Cox J, Olsen JV, Mann M. Feasibility of large-scale phosphoproteomics with higher energy collisional dissociation fragmentation. *J Proteome Res*. 2010;9:6786–94.
81. Veis A, Sfeir C, Wu CB. Phosphorylation of the proteins of the extracellular matrix of mineralized tissues by casein kinase-like activity. *Crit Rev Oral Biol Med*. 1997;8:360–79.
82. George A, Veis A. Phosphorylated proteins and control over apatite nucleation, crystal growth, and inhibition. *Chem Rev*. 2008;108:4670–93.

83. Alvares K. The role of acidic phosphoproteins in biomineralization. *Connect Tissue Res.* 2014;55:34–40.
84. Olsen JV, Blogoev B, Gnäd F, Mačák B, Kumar C, Mortensen P, et al. Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell.* 2006;127:635–48.
85. Henderson JY, Moir AJ, Fothergill LA, Fothergill JE. Sequences of sixteen phosphoserine peptides from ovalbumins of eight species. *Eur J Biochem.* 1981;114:439–50.
86. Gerhartz B, Engh RA, Mentele R, Eckerskorn C, Torquato R, Wittmann J, et al. Quail cystatin: isolation and characterization of a new member of the cystatin family and its hypothetical interaction with cathepsin B. *FEBS Lett.* 1997;412:551–8.
87. Hunter GK, Kyle CL, Goldberg HA. Modulation of crystal formation by bone phosphoproteins: structural specificity of the osteopontin-mediated inhibition of hydroxyapatite formation. *Biochem J.* 1994;300:723–8.
88. Hunter GK, Grohe B, Jeffrey S, O'Young J, Sørensen ES, Goldberg HA. Role of phosphate groups in inhibition of calcium oxalate crystal growth by osteopontin. *Cells Tissues Organs.* 2009;189:44–50.
89. Hincke MT, St.Maurice M. Phosphorylation-dependent modulation of calcium carbonate precipitation by chicken eggshell matrix proteins. In: Goldberg M, Boskey A, Robinson C, editors. *Chemistry and biology of mineralized tissues.* Rosemont: American Academy of Orthopaedic Surgeons; 2000. p. 13–7.
90. Jono S, Peinado C, Giachelli CM. Phosphorylation of osteopontin is required for inhibition of vascular smooth muscle cell calcification. *J Biol Chem.* 2000;275:20197–203.
91. Gericke A, Quin C, Spevak L, Fujimoto Y, Butler WT, Sørensen ES, et al. Importance of phosphorylation for osteopontin regulation of biomineralization. *Calcif Tissue Int.* 2005;77:45–54.
92. Salih E, Ashkar S, Gerstenfeld LC, Glimcher MJ. Identification of the phosphorylated sites of metabolically <sup>32</sup>P-labeled osteopontin from cultured chicken osteoblasts. *J Biol Chem.* 1997;272:13966–73.
93. Lienhard G. Non-functional phosphorylations? *Trends Biochem Sci.* 2008;33:351–2.
94. Ishikawa HO, Xu A, Ogura E, Manning G, Irvine KD. The Raine syndrome protein FAM20C is a Golgi kinase that phosphorylates bio-mineralization proteins. *PLoS One.* 2012;7, e42988.
95. Tagliabracci VS, Engel JL, Wen J, Wiley SE, Worby CA, Kinch LN, et al. Secreted kinase phosphorylates proteins that regulate biomineralization. *Science.* 2012;336:1150–3.
96. Marie P, Labas V, Brionne A, Harichaux G, Hennequet-Antier C, Rodriguez-Navarro AB, et al. Quantitative proteomics provides new insights into chicken eggshell matrix protein functions during the primary events of mineralization and the active calcification phase. *J Proteomics.* 2015;126:140–54.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

