

RESEARCH

Open Access

In-depth analysis of the chicken egg white proteome using an LTQ Orbitrap Velos

Karlheinz Mann*, Matthias Mann

Abstract

Background: Hen's egg white has been the subject of intensive chemical, biochemical and food technological research for many decades, because of its importance in human nutrition, its importance as a source of easily accessible model proteins, and its potential use in biotechnological processes. Recently the arsenal of tools used to study the protein components of egg white has been complemented by mass spectrometry-based proteomic technologies. Application of these fast and sensitive methods has already enabled the identification of a large number of new egg white proteins. Recent technological advances may be expected to further expand the egg white protein inventory.

Results: Using a dual pressure linear ion trap Orbitrap instrument, the LTQ Orbitrap Velos, in conjunction with data analysis in the MaxQuant software package, we identified 158 proteins in chicken egg white with two or more sequence unique peptides. This group of proteins identified with very high confidence included 79 proteins identified in egg white for the first time. In addition, 44 proteins were identified tentatively.

Conclusions: Our results, apart from identifying many new egg white components, indicate that current mass spectrometry technology is sufficiently advanced to permit direct identification of minor components of proteomes dominated by a few major proteins without resorting to indirect techniques, such as chromatographic depletion or peptide library binding, which change the composition of the proteome.

Background

The avian egg white functions as a shock-absorber, keeps the yolk in place, constitutes an antimicrobial barrier, and provides water, protein and other nutrients to the developing embryo. Besides these biological roles it is an inexpensive source of high quality protein for food industries, contains proteins of pharmaceutical interest, and proteins that have found widespread use in biomedical research and protein chemistry [1-6]. Therefore, it is no surprise that egg white has been the target of proteomic studies previously. Raikos et al. [7] used 2D electrophoresis to separate the proteins and MALDI-TOF-based peptide mass fingerprinting to analyze the spots. Seven proteins were identified. 2D electrophoresis, peptide mass fingerprinting and LC-MS/MS using a quadrupole-TOF mass spectrometer were used to identify sixteen proteins in a more advanced study [8]. We have reported the high confidence identification of

78 proteins in egg white using a workflow consisting of SDS-PAGE to separate proteins, coupled to LC-MS/MS and MS³ with an LTQ-FT mass spectrometer [9]. The use of combinatorial hexapeptide libraries [10] in conjunction with LC-ESI-IT-MS/MS allowed the identification of 148 egg white proteins, demonstrating the power of this novel technology to detect minor components even in samples dominated by a few major proteins [11]. Bead-coupled peptide libraries are thought to "equalize" the proteome by providing similar numbers of binding sites to each of the different proteins contained in a proteome. However, it was shown recently that, in contrast to the previously proposed mode of action, the beneficial effect of the peptide beads does not appear to be mediated by specific interaction but is instead dominated by simple hydrophobic effects [12].

Samples, such as egg white, where ovalbumin, ovotransferrin and ovomucoid make up approximately 75% of the total protein, are traditionally difficult to analyze in depth by mass spectrometry, because the peptides of these few proteins tend to dominate the full mass

* Correspondence: mann@biochem.mpg.de
Max-Planck-Institut für Biochemie, Abteilung Proteomics und
Signaltransduktion, Martinsried, Germany

spectra and are selected for fragmentation by MS/MS over and over again. This difficulty has been addressed by the above-mentioned peptide ligand library bead or hydrophobic bead technology [10-12]. However, disadvantages of the peptide library technology include that it is only amenable to soluble proteins and that the composition of the proteome is modified in an unknown and unpredictable way, which makes it impossible to determine the absolute quantity of the proteins. Since the publication of those studies, new developments in instrumentation and peptide identification software occurred, which raised the possibility that in-depth investigation of the egg white proteome would not have to rely on enrichment technologies any more. In the present report we used a novel dual pressure linear ion trap instrument, the LTQ Orbitrap Velos [13]. This new generation of mass spectrometers has increased sensitivity and scan speed as compared to the LTQ-FT used in our previous study [9]. The LTQ Orbitrap Velos is fast enough to isolate and fragment ten or more peaks simultaneously with the acquisition of one high resolution mass full scan spectrum. For evaluation of spectra and database searches we used the MaxQuant software, which is particularly suited for the use of high-resolution MS data and yields very high mass accuracy and peptide identification rates [14-16].

Materials and methods

Preparation of peptides

Proteins were separated by PAGE with pre-cast 4-12% Novex Bis-Tris gels in MES buffer, using reagents and protocols supplied by the manufacturer (Invitrogen, Carlsbad, CA). The kit sample buffer was modified by adding SDS and β -mercaptoethanol to a final concentration of 5% and 2%, respectively, and the sample was suspended in 40 μ l sample buffer/100 μ g of egg white protein and boiled for 5 min. Gels were stained with colloidal Coomassie (Invitrogen) after electrophoresis. Three lanes loaded with 100 μ g of protein were used in each of three separate experiments. The gels were cut into 24 slices for in-gel digestion with trypsin [17] and the peptides were cleaned with Stage Tips [18] before mass spectrometric analysis.

LC-MS and data analysis

Peptide mixtures were analyzed by on-line nanoflow liquid chromatography using the EASY-nLC system (Proxeon Biosystems, Odense, Denmark, now part of Thermo Fisher Scientific) with 15cm capillary columns of an internal diameter of 75 μ m filled with 3 μ m 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 250nl/min for 85min, 30-60% acetonitrile in 0.5%

acetic acid at a flow rate of 250nl/min and 60-80% acetonitrile in 0.5% acetic acid at a flow rate of 250nl/min for 7min. The eluate was electrosprayed into an LTQ Orbitrap Velos (Thermo Fisher Scientific, Bremen, Germany) through a Proxeon nanoelectrospray ion source. The LTQ Orbitrap Velos was operated in a CID top 10 mode essentially as described [13]. The resolution was 30,000 (1 experimental data set) and 60,000 (2 experimental data sets) for the Orbitrap whereas fragment spectra were read out at low resolution in the LTQ. Ion trap and orbitrap maximal injection times were set to 25ms and 500ms, respectively. The ion target values were 5000 for the ion trap and 1000000 for the orbitrap. Raw files were processed using version 1.1.0.45 of MaxQuant (<http://www.maxquant.org/>). For protein identification the ipi.CHICK protein database v3.65 (<http://www.ebi.ac.uk/IPI/IPI-chicken.html>) was combined with the reversed sequences and sequences of widespread contaminants, such as human keratins. Carbamidomethylation was set as fixed modification. Variable modifications were oxidation (M), N-acetyl (protein) and pyro-Glu/Gln (N-term). Initial peptide mass tolerance was set to 7ppm and fragment mass tolerance was set to 0.5 Da. Two missed cleavages were allowed and the minimal length required for a peptide was seven amino acids. Two unique peptides were required for high-confidence protein identifications. These could also be derived from different experimental data sets. The peptide and protein false discovery rates (FDR) were set to 0.01. The maximal posterior error probability (PEP), which is the probability of each peptide to be a false hit considering identification score and peptide length, was set to 0.01. Proteins identified in two of three experimental data sets were accepted. Tentative identifications with only one unique peptide, or two (or more) unique peptides in only one experimental data set, were manually validated considering the assignment of major peaks, occurrence of uninterrupted y- or b-ion series of at least 3 consecutive amino acids, preferred cleavages N-terminal to proline bonds, the possible presence of a2/b2 ion pairs and mass accuracy. The Protein-Prospector MS-Product program (<http://prospector.ucsf.edu/>) was used to calculate the theoretical masses of fragments of identified peptides for manual validation. The exponentially modified protein abundance index (emPAI) provides an estimate of the absolute abundance of a protein from the ratio of observed to observable peptides [19] and was used to differentiate between major and minor proteins. The emPAI calculation considered the preset modifications, miss-cleavages and different charge states. Usually only unique peptides were counted, but in the case of substantial overlap, i.e. almost identical proteins, these were grouped together and the emPAI was calculated for the protein with highest sequence coverage.

Results and discussion

Egg white proteins were separated by PAGE and gels were cut into 24 sections for in-gel digestion (Figure 1) followed by mass spectrometric analysis of the resulting peptides on a high resolution instrument with fast sequencing speed. Three repetitions of the experiment resulted in seventy-two raw-files that yielded a total of approximately 61,500 peptides identified and accepted with a peptide posterior error probability (PEP) of <0.01 and a preset false discovery rate (FDR) of 0.01. Of these, 1,373 peptides were sequence-unique. The average absolute mass deviation was 1.2ppm. By searching of a chicken protein sequence database and by accepting only protein identifications with two sequence-unique peptides occurring in at least two of three experimental data sets, 158 proteins were identified (Additional file 1: Egg white proteins identified with two or more unique peptides). If approximately equal conditions are used between the present study and the peptide library-based study [11] by also considering proteins identified with single peptides occurring in at least two experimental data sets, or proteins identified by two or more unique peptides in only one experimental data set, 44 more

proteins can be added to the list (Additional file 2: Tentatively identified egg white proteins), resulting in a total of 202 possibly identifications. Additional protein data, such as UniProt and RefSeq accession codes, number of identified peptides, sequence coverage, and protein PEP scores for accepted proteins (without contaminants) are provided in Additional file 3: Protein data. These results compared favorably with those obtained with peptide ligand library beads [11], where 68 proteins were identified with two or more sequence-unique peptides and a total of 148 proteins were obtained by accepting unique single peptide hits from different experiments (Figure 2). Furthermore, our study conservatively groups proteins with very similar sequences together and counts them as one "protein group", even when unique peptides pointed at the presence of isoforms or very similar proteins possibly encoded in different genes. Thus, the

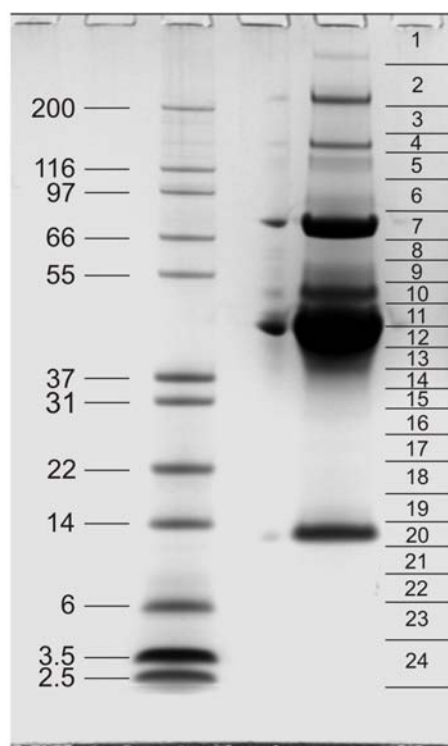


Figure 1 PAGE separation of chicken egg white proteins. Left, the marker proteins are labeled with their molecular weight in kDa. Right, slices used for in-gel digestion are indicated. Overloaded gels show additional bands in the low molecular weight region [9].

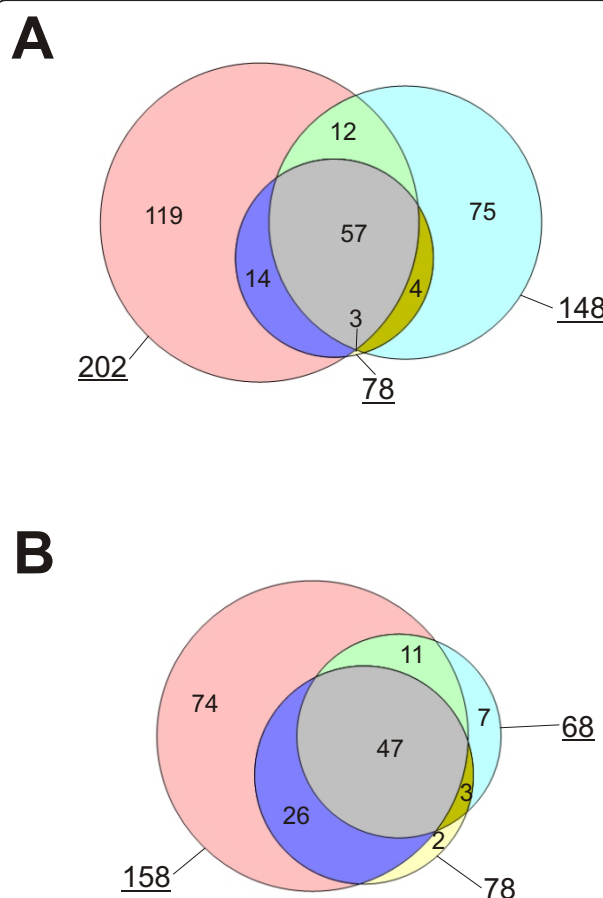


Figure 2 Overlap between recent egg white proteomic studies. **A**, total number of identified proteins; **B**, protein identifications with 2 or more sequence unique peptides (or with confirmation by MS³ in [9]). Underlined numbers indicate sum of identified proteins in the present report (202/158), in [11] (148/68) and in [9] (78/78). Venn diagrams were drawn and calculated using the Venn Diagram Plotter of <http://omics.pnl.gov/software/VennDiagramPlotter.php>.

number of identified proteins is probably higher. A representative example is ovotransferrin, which seemed to represent a mixture of several forms containing many shared and a few unique peptides. Unique peptide data for accepted proteins (without contaminants), such as sequences, PEP scores, and distribution among gel sections are shown in Additional file 4: Peptide data.

Several previously identified proteins [9] were not identified immediately in the new egg white proteome. However, searching the new database version, IPIchick v3.65, with peptide sequences responsible for the previous identification of these proteins indicated that this was in many cases due to changes in the database. Thus, for instance, the ovosecretoglobulin sequence was no longer joined to a channel protein sequence in IPI00575434 but appeared with a new accession number, IPI00847051. Other proteins changed name. Thus, chondrogenesis-associated lipocalin (IPI00600353) is now lipocalin-type prostaglandin synthase D. The only proteins that could not be identified again in the present study were HMG-1 (IPI00595982), a hypothetical protein (IPI00597019), histone H1 (IPI00597019), 60S ribosomal protein L27 (IPI00577674) and poly(ADP-ribosyl) polymerase 1 (IPI00588387). The first two proteins were previously identified predominantly (HGM-1) or exclusively (Hypothetical protein) by in-solution tryptic cleavage, which was not performed in the present study. Three of these proteins, HMG-1, histone H1, and poly(ADP-ribosyl) polymerase were, however, confirmed in a recent study [11]. Therefore, the reason for their absence in the present study is not clear, but as these proteins are unlikely to play functional roles in egg white, their inclusion in egg white preparations may vary. Keratins were excluded from our results because they usually shared all or most peptides with common contaminants. Only few of the new egg white proteins identified using peptide ligand library beads [11] were also detected in the present study. These were nine proteins in the group of identifications with >2 unique peptides (Additional file 1: Egg white proteins identified with two or more unique peptides) and four among the tentatively identified proteins (Additional file 2: Tentatively identified egg white proteins).

Reassuringly, only two new protein identifications were contained among the 30 most abundant egg white proteins (Additional file 1: Egg white proteins identified with two or more unique peptides). This group of proteins contained 79 proteins that were not identified as egg white components previously. The new egg white proteins included several typical major yolk residents, such as apovitellenin-I, vitellogenin-1 to -3 and apolipoprotein B. These proteins are synthesized in the liver, carried to the ovary via the blood circulation, taken up

by oocytes via receptor-mediated processes, and incorporated into the globular fraction of egg yolk [20]. Because the egg yolk was not damaged during mechanical separation of egg white and yolk, these proteins do not seem to be simple contaminants. Rather, residual protein not taken up by the egg cell may be liberated from the ovary together with the egg and migrate with the egg into the oviduct, mixing with egg white proteins secreted in the magnum section. In line with this suggestion, apovitellenin-I and vitellogenins have also been identified in the eggshell organic matrix [21]. This indicates that the oviduct fluid in the eggshell gland still contained these proteins. A few representative peptide fragmentation spectra for some of these proteins are shown in Figure 3. However, many of the new proteins present at low abundance are proteins normally found in intracellular compartments (Additional file 1: Egg white proteins identified with two or more unique peptides; Additional file 2: Tentatively identified egg white proteins). Golgi and ER proteins may have reached the oviduct fluid as by-products of the secretion of major egg white proteins. Other intracellular proteins may have come from damaged, leaky cells of the epithelium lining the oviduct, or from organelles, such as lysosomes, which occur in egg white [22]. Analysis of previously known subcellular locations of proteins identified in egg white shows a decrease in secreted proteins from approximately 64% in the whole proteome to 37% among the new proteins and 18% in tentative identifications (Figure 4). This is accompanied by a similar increase in intracellular proteins, indicating that we have now reached a depth of proteome characterization beyond which it may become difficult to identify functional egg white components. Therefore, minor specific egg white proteins of interest, such as MMP-2, may preferentially be enriched by specific methods before analysis [23]. However, the search for minor components in egg white remains of importance, because very low-abundance proteins, such as bone morphogenetic protein 1 (Additional file 1: Egg white proteins identified with two or more unique peptides) may have a biological role, for instance in early embryonic development.

Conclusions

Our results indicate that current state of the art mass spectrometry technology is sufficiently advanced to permit direct mining of minor components of proteomes dominated by a few major proteins without the necessity to resort to broad specificity protein enrichment techniques, such as peptide ligand library tools, that change the proteome and render absolute quantification impossible. In addition we have significantly expanded the previously known egg white protein inventory.

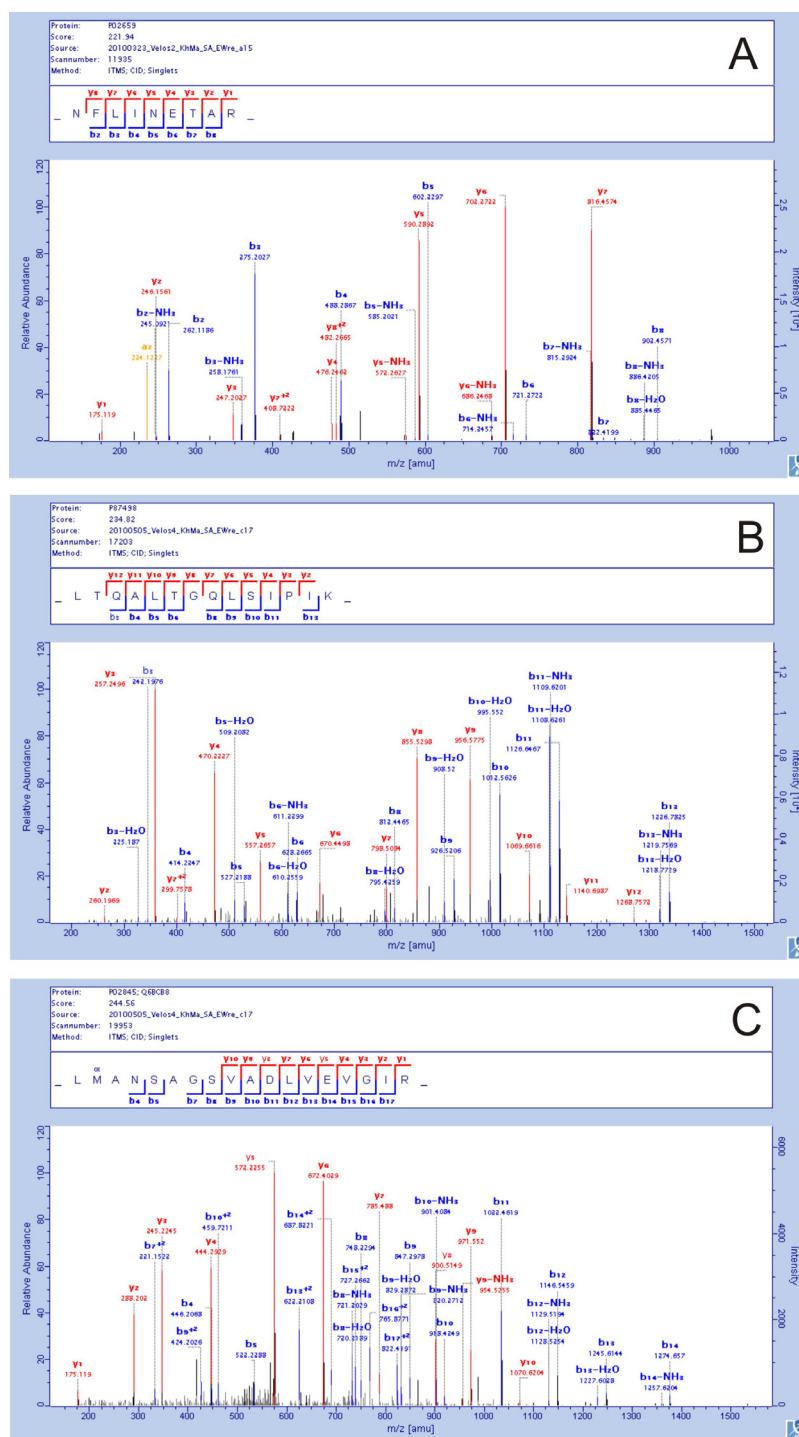


Figure 3 Typical MaxQuant-annotated spectra instrumental in the identification of new egg white proteins. **A**, apovitellenin-1 peptide corresponding to sequence positions 75-83. The precursor peptide mass of this doubly charged peptide was determined with a mass error of 0.35 ppm. The peptide posterior error probability (PEP) was 1.1×10^{-15} and the MaxQuant score was 221. This is an example of a short peptide with almost uninterrupted b- and y-ion series. **B**, vitellogenin-1 peptide corresponding to sequence positions 95-108. The precursor mass of this doubly charged peptide was determined with a mass error of 0.18 ppm. The PEP was 1.22×10^{-28} and the MaxQuant score was 234. The most intense y-ion, y3, indicates the well known preferential cleavage N-terminal to proline. **C**, spectrum of a longer peptide corresponding to sequence positions 666-683 of vitellogenin-2. The precursor mass of this triply charged peptide was determined with a mass error of 0.53 ppm. PEP was 5.2×10^{-42} and the score was 244. The fragmentation pattern shows long uninterrupted b- and y-ion series, but as frequently seen with CID fragmentation of longer peptides, the sequence coverage is not complete.

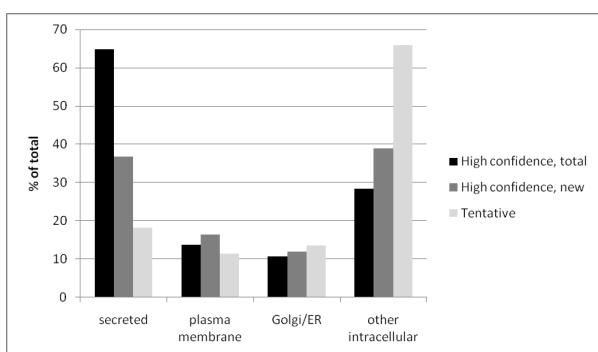


Figure 4 Subcellular location of proteins identified in egg white. Subcellular location of proteins as taken from the UniProt database (<http://www.ebi.ac.uk/uniprot/>) entry of the identified proteins, or of similar proteins identified by searching the database using FASTA (<http://www.ebi.ac.uk/Tools/fasta33/>), and from signal sequence predictions using SignalP (<http://www.cbs.dtu.dk/services/SignalP/>) and InterProScan (<http://www.ebi.ac.uk/Tools/InterProScan/>). Proteins occurring in more than one cellular compartment are counted in each category.

Additional material

Additional file 1: Egg white proteins identified with two or more unique peptides. Docx-file containing a list of proteins identified with high confidence.

Additional file 2: Tentatively identified egg white proteins. Docx-file showing a list of proteins identified with one unique peptide in two of three experimental sets and proteins identified with 2 or more peptides in only one experimental set.

Additional file 3: Protein data. Xls-file containing MaxQuant output data such as protein scores, number of peptides, Uniprot and Refseq accession codes, and sequence coverage.

Additional file 4: peptide data. Xls-file containing MaxQuant output data such as peptide sequences, peptide scores, peptide mass and slice distribution.

Abbreviations

CID: Collision-induced Decomposition; ESI: Electrospray Ionization; FDR: False Discovery Rate; FT: Fourier transform; IT: Ion Trap; LC: Liquid Chromatography; MALDI: Matrix-assisted Laser Desorption Ionization; MES: 2-(N-Morpholino) ethanesulfonic acid; PAGE: Polyacrylamide Gel Electrophoresis; PEP: Posterior Error probability; SDS: Sodium Dodecyl Sulphate; TOF: Time-of-flight.

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. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Received: 10 December 2010 Accepted: 7 February 2011
Published: 7 February 2011

References

1. Stevens L: **Egg white proteins.** *Comp Biochem Physiol B* 1991, **100**:1-9.

2. Mine Y: **Recent advances in the understanding of egg white protein functionality.** *Trends Food Sci Technol* 1995, **6**:225-232.
3. Mine Y, Kovacs-Nolan J: **Biologically active hen egg components in human health and disease.** *J Poult Sci* 2004, **41**:1-29.
4. Mine Y, Kovacs-Nolan J: **New insights in biologically active proteins and peptides derived from hen egg.** *World Poult Sci* 2006, **62**:87-95.
5. Anton M, Nau F, Nys Y: **Bioactive egg components and their potential uses.** *World Poult Sci* 2006, **62**:429-438.
6. Mine Y: **Egg proteins and peptides in human health - Chemistry, bioactivity and production.** *Curr Pharmaceut Design* 2007, **13**:875-884.
7. Raikos V, Hansen R, Campbell L, Euston SR: **Separation and identification of hen egg protein isoforms using SDS-PAGE and 2D gel electrophoresis with MALDI-TOF mass spectrometry.** *Food Chem* 2006, **99**:702-710.
8. Guérin-Dubiard C, Pasco M, Mollé D, Désert C, Croguennec T, Nau F: **Proteomic analysis of hen egg white.** *J Agric Food Chem* 2006, **54**:3901-3910.
9. Mann K: **The chicken egg white proteome.** *Proteomics* 2007, **7**:3558-3568.
10. Boschetti E, Righetti PG: **The art of observing rare protein species in proteomes with peptide ligand libraries.** *Proteomics* 2009, **9**:1492-1510.
11. D'Ambrosio C, Arena S, Scaloni A, Guerrier L, Boschetti E, Mendieta ME, Citterio A, Righetti PG: **Exploring the chicken egg white proteome with combinatorial peptide ligand libraries.** *J Proteome Res* 2008, **7**:3461-3474.
12. Keidel EM, Ribitsch D, Lottspeich F: **Equalizer technology - Equal rights for disparate beads.** *Proteomics* 2010, **10**:2089-2098.
13. Olsen JV, Schwartz JC, Griep-Raming J, Nielsen ML, Damoc E, Denisov E, Lange O, Remes P, Taylor D, Splendore M, Wouters ER, Senko M, Makarov A, Mann M, Horning S: **A dual pressure linear ion trap-Orbitrap instrument with very high sequencing speed.** *Mol Cell Proteomics* 2009, **8**:2759-2769.
14. Cox J, Hubner NC, Mann M: **How much peptide sequence information is contained in ion trap tandem mass spectra?** *J Am Soc Mass Spectrom* 2008, **19**:1813-1820.
15. Cox J, Mann M: **MaxQuant enables high peptide identification rates, individualized ppb-range mass accuracies and proteome-wide protein quantification.** *Nature Biotechnol* 2009, **26**:1367-1372.
16. Cox J, Mann M: **Computational Principles of determining and improving mass precision and accuracy for proteome measurements in an orbitrap.** *J Am Soc Mass Spectrom* 2009, **20**:1477-1485.
17. Shevchenko A, Tomas H, Havlis J, Olsen JV, Mann M: **In-gel digestion for mass spectrometric characterization of proteins and proteomes.** *Nature Protocols* 2006, **1**:2856-2860.
18. Rappsilber J, Mann M, Ishihama Y: **Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips.** *Nature Protocols* 2007, **2**:1896-1906.
19. Ishihama Y, Oda Y, Tabata T, Sato T, Nagasu T, Rappsilber J, Mann M: **Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein.** *Mol Cell Proteom* 2005, **4**:1265-1272.
20. Burley RW, Vadehra DV: *The Avian Egg* New York: John Wiley and Sons; 1989.
21. Mann K, Macek B, Olsen JV: **Proteomic analysis of the acid-soluble organic matrix of the chicken calcified eggshell layer.** *Proteomics* 2006, **6**:3801-3810.
22. Yoon J, Park J-M, Kim K-J, Kim Y-H, Min J: **Antimicrobial activity of the cell organelles, lysosomes, isolated from egg white.** *J Microbial Biotechnol* 2009, **19**:1364-1368.
23. Réhault-Godbert S, Gautron J, Labas V, Belghazi M, Nys Y: **Identification and characterization of the precursor of chicken matrix metalloproteinases 2 (pro-MMP-2) in hen egg.** *J Agric Food Chem* 2008, **56**:6294-6303.
24. Emanuelsson O, Brunak S, von Heijne G, Nielsen H: **Locating proteins in the cell using TargetP, SignalP, and related tools.** *Nature Protocols* 2007, **2**:953-971.

doi:10.1186/1477-5956-9-7

Cite this article as: Mann and Mann: In-depth analysis of the chicken egg white proteome using an LTQ Orbitrap Velos. *Proteome Science* 2011 **9**:7.

Egg white proteins identified with two and more unique peptides

No.	IPI Accession ¹	Protein	Experiment	Unique ³ peptides	Usual location	emPAI
1	00583974 ^{2a,b}	Ovalbumin; N-term acetylated	1-3	37	secreted	251187.6
2	00683271 ^{2a,b} 00578012 ^{2a,b} 00821645	Ovotransferrin 78kDa protein 105kDa protein share 86 peptides with ovotransferrin	1-3	98 5	secreted	37274.9
3	00600859 ^{2a,b}	Lysozyme C	1-3	20	secreted	2309.1
4	00597129 ^{2a,b}	Ovomucoid	1-3	19	secreted	773.3
5	00587313 ^{2a,b} 00681776 ^{2a}	Ovoinhibitor 53Kda protein; share 24 peptides	1-3	25 1	secreted	561.3
6	00573738 ^{2a,b}	Ovalbumin-related protein Y; shares 1 peptide with IPI00585021	1-3	27	secreted	507.0
7	00585021 ^{2a,b}	Similar to ovalbumin-related protein Y (OVA-X); N-term. acetylated	1-3	30	secreted	427.1
8	00600265 ^{2a,b}	Apolipoprotein D	1-3	7	secreted	371.8
9	00847051 ^{2a}	Ovosecretoglobulin/10kDa protein	1-3	3	secreted	315.2
10	00601265 ^{2a,b} 00822987	Avidin 17kDa protein	1-3	9	secreted	267.3
11	00589747 ^{2a,b} 00818180	Ovostatin 164kDa protein; share 81 peptides	1-3	82 4	secreted	250.2

12	00595847 ^{2a,b}	Similar to MGC82112 ; similar to α 2-macroglobulin-like 1	1-3	82	secreted	219.4
13	00819356 ^{2a,b}	Clusterin/49kDa protein	1-3	20	secreted	205.9
14	00584841 ^{2a,b}	Hep21	1-3	7	secreted	157.5
15	00576782 ^{2a,b}	Cystatin	1-3	5	secreted	145.8
16	00600069 ^{2a,b}	Extracellular fatty acid-binding protein	1-3	8	secreted	137.9
	00594673	21kDa protein (80% identical to ExFABP)		4		
	00818310	Protein				
	00819914	20kDa protein				
17	00589360 ^{2a,b}	Glutathione peroxidase 3	1-3	9	secreted	132.4
	00822052	22kDa protein				
18	00600353 ^{2a,b}	Lipocalin-type prostaglandin D synthase/chondrogenesis –associated lipocalin	1-3	7	ER, Golgi, plasma membrane, secreted	132.3
19	00584997 ^{2a,b}	Ovomucin α -subunit	1-3	126	secreted	101.6
20	00598113 ^{2a,b}	Ovoglycoprotein	1-3	8	secreted	99.0
21	00580509	Apovitellenin-1	1-3	5	secreted	99.0
22	00578016 ^{2a,b}	Dickkopf-related protein 3	1-3	18	secreted	65.6
23	00574804 ^{2a,b}	Gallinacin-11 (VMO-II; β -defensin-11)	1-3	6	secreted	62.1
24	00598229 ^{2a,b}	Tenp	1-3	15	secreted	60.1
	00819823	44kDa protein; share 11 of 15 peptides		4		
25	00574195 ^{2a,b}	Serum albumin (α -livetin)	1-3	30	secreted	46.5

26	00597482 ^{2a,b}	(Similar to) ovomucin β -subunit	1-3	53	secreted	30.6
27	00596061 ^{2a,b}	VMO-I	1-3	10	secreted	27.5
28	00681274 ^{2a,b}	Gallin	1-3	3	secreted	20.5
	00822809	Similar to meleagrins				
29	00573473 ^{2a,b}	Hypothetical protein; partial hemopexin	1-3	8	secreted	16.8
30	00847036	8kDa protein/secretory trypsin inhibitor	1-3	2	secreted	9.0
31	00577639 ^{2a}	Hypothetical protein; partial serpin domain; some similarity to part of plasma protease C1 inhibitor	1-3	2	secreted	9.0
32	00589837 ^{2a}	Similar to autotoxin isoform 1/ ectonucleotide pyrophosphatase/ phosphodiesterase family member 1	1-3	25	secreted/plasma membrane	9.0
33	00590040 ^{2a,b}	Lymphocyte antigen 86	1-3	5	secreted	9.0
34	00593624 ^{2b}	Similar to Kazal-type serine protease inhibitor	1-3	2	secreted	9.0
35	00655503 ^{2a,b}	Actin, cytoplasmic 1; shares most of its peptides with other chicken actin entries and bovine actin, a common contaminant, which is 100% identical.	1-3	15	cytoplasm	9.0
	00837485	Actin, cytoplasmic 2;		1		
	00582169	Actin, alpha skeletal muscle		1		
		Share 14 of 15 and 6 of 7 peptides with IPI00655503				
36	00591488 ^{2a,b}	Ig μ chain C region	1-3	8	secreted	7.7
37	00810881 ^{2a,b}	Similar to complement component 8, γ polypeptide	1-3	8	secreted	7.4

38	00594746 ^{2a,b}	Riboflavin-binding protein	1-3	8	secreted	7.1
39	00571581 ^{2a,b}	Sulfhydryl oxidase 1	1-3	19	Golgi membrane	7.1
40	00576977 ^{2a}	12kDa protein	1-3	5	nucleus	6.2
	00819077	12kDa protein				
	00572919	Histone H4				
	00598581	Histone H4 typeVIII				
41	00587452 ^{2a}	Similar to oncoprotein-induced transcript 1	1-3	8	secreted	5.8
42	00602226	Vitellogenin-2	1-3	50	secreted	5.4
	00591299	205kDa protein; share 49 of 50 peptides		1		
43	00586548 ^{2a,b}	Angiopoietin-like protein 3	1-3	13	secreted	5.1
	00601120	54kDa protein, 97.6% identical, share 11 peptides		1		
44	00581002 ^{2a,b}	Ubiquitin	1-3	2	cytoplasm	4.6
	00821431	52kDa protein				
45	00578622 ^{2b}	Ovocalyxin-32	1-3	2	secreted	4.2
	00821633					
46	00866802 ^{2a,b}	Polymeric immunoglobulin receptor	1-3	13	plasma membrane, secreted	4.1
47	00818415	24kDa protein	1-3	5	ER lumen	4.0
	00583337 ^{2a,b}	Peptidyl-prolyl-cis/trans isomerase B				
48	00600127 ^{2b}	Similar to bactericidal permeability- increasing protein	1-3	10	secreted	4.0
49	00597109 ^{2a,b}	Aminopeptidase Ey	1-3	19	plasma membrane, secreted	3.9
50	00597264 ^{2b}	Hypothetical protein; Pleiotrophin	1-3	3	secreted	3.6
	00573025	15kDa protein; Midkine				
51	00593541 ^{2b}	Renin/prorenin receptor	1-3	7	plasma membrane	3.6

52	00597105 ^{2a}	Similar to METRNL (meteorin-like) protein	1-3	7	secreted	3.3
53	00580429	13kDa protein; similar to junction plakoglobin; shares 1 of 5 peptides with IPI00579090 (β -catenin)	1-3	4	cytoplasm, plasma membrane	3.2
54	00595270 ^{2a,b}	TIMP-2	1-3	5	secreted	3.2
55	00583507	Putative uncharacterized protein; similar to syntenin-1	1-3	4	ER membrane, plasma membrane	3.2
56	00822431	Histone H2B variants	1-3	3	nucleus	3.2
57	00574703 ^{2a}	10kDa protein; Ig λ chain	1-3	3	secreted	3.0
	00683203 ^{2a}	10kDa protein; shares 1 of 2 peptides		1		
58	00598054 ^{2a}	Immunoglobulin J polypeptide	1-3	3	secreted	3.0
59	00595834 ^{2a}	FAMC3 protein	1-3	4	secreted	2.8
60	00600609	13kDa protein; Ig superfamily	1-3	2	secreted	2.7
	00684262	13kDa protein		1		
	00812261 ^{2a}	Similar to Ig α chain		1		
	00594704	7kDa protein		1		
	00819112	15kda protein		1		
	00819112	15kDa protein		1		
61	00580810 ^{2a,b}	Similar to tazarotene-induced gene 2	1,3	5	secreted	2.6
62	00575584	α -Enolase; shares 2 peptides with IPI00592520 (γ -Enolase)	1-3	5	cytoplasm	2.4
63	00576813	Putative uncharacterized protein/similar to palmitoyl protein thioesterase 1	1,3	5	lysosome	2.2
64	00599307	60S acidic ribosomal protein P1; N-term acetylated	1,3	3	cytoplasm	2.2

65	00587380 ^{2a,b} 00582445	Cytotactin/tenascin; share 27 of 28 peptides	1-3	28	secreted	2.1
66	00589382 ^{2a,b}	Hypothetical protein; similar to acyloxyacyl hydrolase	1-3	10	secreted	2.0
67	00680973 ^{2b} 00583269 ^{2b} 00599848 00602067	CEPU-1, isoform1 CEPU-1 CEPU-1, isoform 2 Neural secreted glycoprotein	1-3	4	secreted	1.9
68	00822414	15kDa selenoprotein	1-3	2	ER lumen	1.7
69	00589985	Elongation factor 1- α 1	1-3	5	cytoplasm	1.7
70	00581589 ^{2a,b} 00593580 ^{2a,b}	135kDa protein; similar to PIT 54 similar to deleted in malignant brain tumors 1; share 5 peptides	1-3	12 1	secreted/membrane	1.5
71	00820456 ^{2a} 00595705	16kDa protein/Epididymal secretory protein E1 (Niemann-Pick disease type C2 protein)	1-3	3	secreted	1.5
72	00591843 00820086	Vitellogenin-1 211kDa protein; share 30 of 31 peptides	1-3	31 1	secreted	1.5
73	00578632 00820116	14-3-3 Protein ζ (zeta) 24kDa protein Share 1 of 5 peptides with other 14-3-3 proteins	1-3	4	cytoplasm	1.4
74	00596201 ^{2a,b}	Olfactomedin-like protein 3	1-3	7	secreted	1.4
75	00588294 ^{2a,b}	Similar to transmembrane protease, serine 9	1-3	16	plasma membrane	1.4
76	00574658 ^{2a,b}	TIMP-3	1-3	4	secreted	1.3
77	00599180	Hypothetical protein; leucine-rich repeat-containing protein	1-3	3	plasma membrane	1.3

78	00580166 ^{2a,b}	Glutathione S-transferase 2	1-3	5	cytoplasm	1.3
79	00593199	β-2-Microglobulin	1,3	2	secreted	1.2
80	00577274	RGD-CAP; similar to transforming growth factor-induced protein iq-h3	1-3	8	secreted	1.2
81	00587277	Similar to SDF3 (PEDF-like, serpin-family)	1-3	8	secreted	1.2
82	00775749 00819590	Apolipoprotein B 51kDa protein; shares 5 peptides with Apo B	1-3	75 1	secreted	1.1
83	00595826 ^{2a,b} 00822310 00819712	Golgi apparatus protein 1 130kDa protein 120kDa protein	1-3	15	ER membrane	1.1
84	00820677 00578917	60kDa protein Cochlin	1-3	7	secreted	1.0
85	00600823	Proteasome subunit α type-7	1-3	3	cytoplasm	1.0
86	00580765 ^{2a,b}	Apolipoprotein A-I	1-3	5	secreted	0.9
87	00585627 ^{2a,b}	Similar to bactericidal/permeability-increasing protein like-2	1,3	4	secreted	0.9
88	00588908 ^{2a}	Similar to ovulatory protein 2	1-3	3	secreted	0.9
89	00588868	Eukaryotic initiation factor 4A-II	1,3	4	cytoplasm	0.9
90	00681344 ^{2b}	Similar to Kunitz-like protease inhibitor	1-3	3	secreted	0.9
91	00590375 ^{2b}	Glucose-regulated 78kDa protein; HSP 70 family, shares 2 of 8 peptides with IPI00582091	1-3	6	ER membrane (lumen)	0.9

92	00829402	ASPIC/cartilage acidic protein 1	1-3	9	secreted	0.9
93	00582452	Triosephosphate isomerase	1-3	3	cytoplasm	0.9
94	00597853	Proteasome subunit, α -type	2,3	3	cytoplasm	0.8
95	00602683	20kDa protein; similar to putative Agr-rich protein mutated in early tumors	1,3	2	secreted	0.8
96	00584676	Similar to corticotrophin-releasing factor-binding protein	1,3	3	secreted	0.8
97	00685019 ^{2a,b}	24kDa protein	1-3	3	secreted	0.8
	00577021	Ig λ chain C-region				
98	00820598	26kDa protein; FAMB3 family	1,2	2	secreted	0.8
	00810914	Hypothetical protein; FAMB3 family				
	00571613	26kDa protein; FAM3B family				
99	00813945	Similar to SLIT2 isoform 1	1-3	12	secreted	0.7
	00883281	169kDa protein				
	00814509 ^{2b}	Similar to SLIT2 isoform 2				
100	00574055 ^{2a,b}	Procollagen-lysine, 2-oxoglutarate 5-dioxygenase 1	1-3	7	ER lumen	0.7
101	00572503	Similar to α -mannosidase class 2B, member 2	1-3	7	secreted	0.7
102	00822785 ^{2a,b}	22kDa protein/ADP-ribosylation factor 5	1,3	4	Golgi, cytoplasm	0.7
103	00581527	Calmodulin; N-term acetylated	1-3	2	cytoplasm	0.7
	00823322					
104	00596673	Protein disulfide-isomerase	1-3	6	ER membrane, plasma membrane	0.7
105	00598551	Similar to prosomal P27K protein/proteasome subunit α -type	1-3	2	cytoplasm	0.6
106	00572470	Annexin I	1-3	2	nucleus, cytoplasm, plasma membrane	0.6

107	00593654	Hypothetical protein; similar to di-N-acetylchitobiase	1-3	3	lysosome	0.6
108	00585747	Elongation factor 2	1,3	8	cytoplasm	0.6
109	00580059	Peroxioredoxin-1	1-3	2	cytoplasm	0.6
110	00812933 00582091	Similar to natural killer cell enhancing HSP70; shares 2 of 6 peptides with IPI00590375 (78kDa glucose-regulated protein)	1,3	4	cytoplasm, cell surface	0.6
111	00598223 ^{2b}	Similar to decoy receptor 3	2,3	2	secreted	0.5
112	00575271	Protein disulfide isomerase A3	1-3	5	ER lumen	0.5
113	00822110 00579245	105kDa protein; similar to insulin-degrading enzyme	2,3	5	cytoplasm, plasma membrane	0.5
114	00580985	Calreticulin	1-3	3	ER lumen	0.5
115	00590239	40S ribosomal protein SA/laminin receptor 1	2,3	2	plasma membrane, cytoplasm, nucleus	0.5
116	00572461	α -Actinin-4	2,3	8	cytoskeleton	0.5
117	00590085	Hypothetical protein; similar to galactocerebrosidase	1,3	2	lysosome	0.5
118	00593765	Putative uncharacterized protein; similar to gastric intrinsic factor/transcobalamin	1-3	2	secreted	0.5
119	00603219 00819358	Tumor necrosis factor-related apoptosis-inducing ligand	1,3	2	plasma membrane	0.4
120	00573327	Vitamin D-binding protein	1,3	2	secreted	0.4
121	00577371	Hypothetical protein; thiol protease family	1-3	3	secreted, lysosome	0.4

122	00599064	Putative uncharacterized protein; similar to N-acylsphingosine amidohydrolase	1,3	3	lysosome	0.4
123	00574951	Semaphorin-3C	2,3	5	secreted	0.4
124	00570770	Endoplasmin; shares 1 of 6 peptides with IPI00820593	1-3	5	ER lumen	0.4
125	00602986 ^{2a}	Similar to desmoplakin isoform II isoform 2	1-3	15	cell junction	0.4
126	00818934	184kda protein; vitellogenin-3	1,3	12	secreted	0.4
	00823141					
127	00575767	Catalase	1-3	4	peroxysome	0.4
128	00602351	Hepatic α -amylase	1-3	4	secreted	0.4
129	00586802	Hypothetical protein; transitional ER ATPase	1-3	5	cytoplasm	0.4
130	00593455	Cadherin-1	1-3	3	plasma membrane	0.3
	00575514	Hypothetical protein; similar to cadherin				
131	00601768 ^{2a,b}	Angiotensin-converting enzyme	1,3	6	secreted	0.3
132	00582281	Similar to β -hexosaminidase β -chain	1,3	5	lysosome	0.3
133	00584670	Similar to prostatic acid phosphatase	1,3	2	secreted, plasma membrane, lysosome	0.3
134	00592552	Bone morphogenetic protein 1	1-3	5	secreted	0.3
135	00576308 ^{2a}	Similar to carboxypeptidase D	1-3	7	plasma membrane	0.3
136	00597023	Hypothetical protein; similar to legumain	2,3	2	lysosome	0.3
137	00596315	Guanine nucleotide-binding protein subunit β 2-like 1	2,3	2	cytoplasm	0.3

138	00570697 00818683	Histidine ammonia lyase 71kDa protein	1-3	3	cytoplasm	0.3
139	00582635 ^{2a}	Similar to aminopeptidase A	1-3	5	plasma membrane	0.3
140	00819050	Similar to ceruloplasmin	1-3	3	secreted	0.2
141	00577039	Annexin A2	1-3	2	cytoplasm, plasma membrane, secreted	0.2
142	00682229	66kDa protein; similar to UDP-N-acetyl- α - D-galactosamine:polypeptide N- acetylgalactosaminyltransferase 4	1,3	2	Golgi membrane	0.2
143	00579090 00822777	β -catenin 75kDa protein share 1 of 4 peptides with IPI00580429	1-3	3	plasma membrane	0.2
144	00583980	195kDa protein; C3 and PZP-like, alpha-2- macroglobulin domain containing protein 8	1-3	5	secreted	0.2
145	00818621 00591231	57kDa protein; similar to ER-Golgi intermediate compartment 53kDa protein	1,3	2	ER/Golgi membrane	0.2
146	00589584	RAB-GDP dissociation inhibitor	1,3	3	cytoplasm	0.2
147	00587219	Lamin-A	1-3	2	cytoplasm/cytoskeleton	0.2
148	00581368	Ovocleidin-116	1,2	2	secreted	0.2
149	00819496 00573261	98kDa protein; similar to cell death 6- interacting protein	1,2	4	intracellular	0.2
150	00821924	Aminopeptidase, puromycin sensitive	1,3	2	cytoplasm, nucleus	0.1
151	00681096	Similar to complement C4-1	2,3	2	secreted	0.1
152	00596586	HSP90- α ; shares 1 of 3 peptides with IPI00820593 (HSP90- β)	1-3	2	cytoplasm	0.1

153	00570607 00819130	Hypoxia up-regulated protein 1 111kDa protein	1-3	2	ER lumen	0.1
154	00591552 00820593	Heat shock cognate protein HSP90 β 84kDa protein Share 1 peptide of 5 with IPI00570770 (endoplasmin), and 1 peptide of 5 with IPI00596673 (HSP90 α)	1-3	3	cytoplasm	0.1
155	00585604	189kDa protein; similar to mucin 5	2,3	3	secreted	0.1
156	00583368 00818136	Similar to procollagen-lysine, 2- oxoglutarate 5-dioxygenase; 83kDa protein	1,3	2	ER membrane (lumen)	0.1
157	00575808 00822939	Putative uncharacterized protein; ATP- citrate synthase	1,3	2	cytoplasm	0.1
158	00572165	Myosin-9	1,3	4	cytoskeleton	0.1

Entries are ordered according to decreasing emPAI. ¹, only the IPI accession number of the leading protein (the protein with most identified peptides) of a group of similar proteins is given. Complete data are shown in Supplementary file 3. ^{2a,b}, previously identified in [9]^a or [11]^b. ³, protein group unique peptides.

Tentatively identified egg white proteins

No.	IPI Accession ¹	Protein	Experiment	Unique Peptides ³	Location	emPAI
1	00683104 ^{2b}	Histone H2A	1-3	1	nucleus	2.2
	00590686 ^{2a}	Histone H2A-V share 1 peptide of 2 with many HA2 variants		1		
2	00883091	Plakoglobin fragment	1-3	1	cell junction	2.2
3	00576073	Hypothetical protein; similar to THO complex subunit 4	1	2	nucleus, cytoplasm	1.6
4	00574447	Histone H3.3	1,3	1	nucleus	1.2
	00651224	Histone H3				
	00575899	16kDa protein		1		
	00684128	16kDa protein				
5	00683452	Histone H3.2; entries share 1 of 2 peptides	3	4	cytoskeleton	1.0
	00587398	Actin-related protein 3				
6	00591852	14-3-3 protein beta/alpha	1-3	2	cytoplasm	0.9
	00577739	14-3-3 protein theta/gamma		1		
	00579092	14-3-3 Protein epsilon		1		
	00820692	29kDa protein		1		
	00818270	26kDa protein		1		
7	00812585 ^{2a}	Similar to immunoglobulin λ chain	1,2	1	secreted	0.8

8	00603587 00822443	Hypothetical protein/ α -tubulin 8 44kDa protein/ α -tubulin 2 share 3-4 peptides with each other and bovine α -tubulin 1, a common contaminant	1,3	1 1	cytoskeleton	0.7
9	00580626 00592950 00603718	β -Tubulin 3/7 β -Tubulin 1/2 β -tubulin 4; share most peptides	1,3	1 1 1	cytoskeleton	0.7
10	00590301	Voltage-dependent anion channel 1	3	4	plasma membrane	0.7
11	00600819	Collagen α -1 (X) chain	1,3	1	secreted	0.6
12	00592949	Proteasome subunit, α -type	3	2	cytoplasm	0.6
13	00593872	Proteasome α -2 subunit	3	2	cytoplasm	0.6
14	00581952	Putative uncharacterized protein/tropomyosin	3	2	cytoskeleton	0.4
15	00572756 ^{2b}	Ovocleidin-17	1-3	1	secreted	0.4
16	00577421	HSP-60, mitochondrial	1	2	mitochondrion	0.4
17	00580968	T-complex protein 1, subunit eta	1	3	cytoplasm	0.4
18	00651336	Putative uncharacterized protein/heterogeneous nuclear	1	2	nucleus	0.3
19	00591422 00571469	F-actin-capping protein subunit beta isoforms 1 and 2	3	2	cytoskeleton	0.3
20	00576279	Putative uncharacterized protein; similar to eIF6	2,3	1	cytoplasm, nucleus	0.3
21	00586516 ^{2b} 00684903	Similar to human P5 isoform 2 (protein disulfide isomerase)	2,3	1	secreted/membrane	0.3

22	00589224	Similar to maspin (serpin protease inhibitor family)	1-3	1	cytoplasm, secreted	0.3
23	00601033	L-Lactate dehydrogenase B chain	1,3	2	cytoplasm	0.3
24	00576601	Putative uncharacterized protein; similar to heterogeneous nuclear riboprotein	1,3	2	nucleus	0.3
25	00594653	Glyceraldehyde-3-phosphate dehydrogenase; shares 1 of 2 peptides with the bovine enzyme (common contamination)	1-3	1	cytoplasm	0.3
26	00573727 ^{2b} 00820954	Putative uncharacterized protein; similar to nucleobindin 49kDa protein; similar to nucleobindin	3	2	cytoplasm, plasma membrane, secreted	0.2
27	00573643 00571478 00820847	Hypothetical protein Ras-related protein Rab-6a 21kDa protein	1-3	1	Golgi apparatus	0.2
28	00682113 00574064	58kDa protein Pyruvate kinase muscle isozyme	3	3	cytoplasm	0.2
29	00581163	Similar to carboxypeptidase E	1	2	secreted	0.2
30	00597100	Hypothetical protein; similar to transcriptional adapter 2B	1-3	1	nucleus	0.2
31	00592330	Putative uncharacterized protein/HSP70 protein 13	1,3	2	ER	0.2
32	00584371	Proactivator polypeptide	1,2	2	lysosome	0.2
33	00575877 00822490	122kDa protein Contactin-1	3	2	plasma membrane	0.1
34	00581158	Complement C3	1	2	secreted	0.1
35	00583347	Similar to N-acetylglucosaminyl transferase V	1,3	1	Golgi membrane	0.1

36	00592520	γ -Enolase; shares 2 of 3 peptides with α -enolase	1-3	1	cytoplasm	0.1
37	00593450	Hypothetical protein/UDP-glucose ceramide glucosyltransferase-like 1	1	2	ER lumen	0.1
38	00594476	DNAJ homolog subfamily C, member 3	1,2	1	ER	0.1
39	00594512	Similar to UTP14, U3 small nucleolar	1,2	1	nucleus	0.1
	00822376	ribonucleoprotein, homolog A/15kDa protein				
40	00597128	Similar to UDP-N-acetyl- α -D-galactosamine: Polypeptide N-	1-3	1	Golgi membrane	0.1
41	00604251	Similar to tRNA splicing endonuclease 54 homolog	1-3	1	nucleus	0.1
42	00573263	Similar to bullous pemphogoid antigen 1	2,3	1	cytoskeleton	<0.1
	00820606	620kDa protein				
43	00786274	PAR3	1,2	1	cell junctions	<0.1
44	00591842	Similar to fat tumor suppressor 1	1	2	plasma membrane	<0.1

This table contains egg white proteins identified with one unique peptide in several experiments, with several unique peptides in one experimental set, or with only one of two unique peptides validated manually. Entries are ordered according to decreasing emPAI. ¹, only the IPI accession number of the leading protein (the protein with most identified peptides) of a group of similar proteins is given. Complete data are shown in Supplementary file 3. ^{2a,b}, previously identified in [9]^a or [11]^b. ³, group unique peptides.