Molecular approaches for the identification of novel egg components

J. GAUTRON1*, F. NAU2, K. MANN3, C. GUERIN-DUBIARD2, S. REHAULT1, M.T. HINCKE4 and Y. NYS1

1*Unité de Recherches Avicoles, Fonction et Régulation des Protéines de l’œuf, INRA, 37380 Nouzilly, France, 2UMR INRA -AgroCampus Rennes 1253 Science et Technologie du Lait et de l'Oeuf, 65 rue de Saint-Brieuc, 35042 Rennes Cedex, France, 3Max-Planck-Institut für Biochemie, Abteilung für Proteomics und Signaltransduktion, Martinsried, Germany, 4Department of Cellular and Molecular Medicine, University of Ottawa, Ottawa, ON, K1H 8M5, Canada

*Corresponding author: gautron@tours.inra.fr

The avian egg is a nutritious food and also a major source of biologically active compounds that are beneficial for human health. These biologically active molecules are widely used by pharmaceutics, cosmetic and food industries. Egg proteins were previously studied using classical biochemical techniques such as chromatographic and electrophoretic separation and Edman sequence analysis. The development of molecular biology in the late 80’s and the recent publication of the chicken genome sequence are major scientific advances leading to identification and characterization of a number of minor egg components that were not previously identified. Using recent data on the characterisation of egg white and eggshell matrix proteins, we have illustrated in this review the recent developments in the biochemistry of the egg (proteomics) and in the molecular biology of the egg (cDNA and ESTs libraries, bioinformatic analysis of the chicken genome, transcriptomics). These methods allowed the identification of hundreds of minor egg protein components with potential applications for industry. Functional studies to study the biological activities of these novel egg protein components and exploit their potential will form the next frontier for egg science.

Keywords: Laying hens; eggs; proteins; genes; identification

Introduction

The chicken egg is a container for the extra-uterine development of the embryo using the natural reserve of the egg components. Its original role as an embryonic chamber leads one to suppose that it contains many compounds essential for life. Consequently, the avian egg is a major source of active molecules usable by the pharmaceutical industry as well as by the cosmetic and food industries (Nau et al., 2002; Mine and Kovacs-Nolan, 2006). Egg components possess a broad range of biological activities. The egg contains molecules with antibacterial, antiviral, anti-hypertensive, anticancer, antioxidant, antigenic, cryoprotective, immunomodulating and anti-adhesive activities that have been intensively reviewed (Nau et al., 2002; Mine and Kovacs-Nolan, 2006). In addition to these properties, the pathogen-free egg is also a basic human nutrient (Nys, 2001). These remarkable properties mainly rely on the proteins present in the egg yolk, white and shell.

The fractionation of the proteins of albumen and yolk was been initiated more than 50 years ago. The major proteins of albumen and yolk have been separated and purified using ammonium sulphate precipitation, chromatographic and electrophoretic techniques (Li-Chan et al, 1995). This catalogue of classical biochemical techniques was reinforced by molecular biology tools in the 80’s. Despite these efforts, the composition of egg is still not completely understood. Only the major egg proteins have been identified. They only represent a minor percentage of the large variety of egg proteins and of the potentialities of egg components.
During the last ten years, the results of genomic studies have dramatically transformed biological studies, including egg science. The recent development of high-throughput methods used in combination with the now available chicken genomic sequence (International Chicken Genome Sequencing Consortium, 2004) represents promising openings for the characterization of minor components. Some of these molecules may be a source of active compounds with specific properties to benefit human and animal health. In this review, we will describe our recent contributions in the characterisation of egg white and eggshell matrix proteins to illustrate the current and potential methods used to identify egg proteins.

**Egg biochemistry**

**Purification by chromatographic and electrophoretic methods**

Liquid chromatography and electrophoresis are well known techniques to separate and characterize the protein components of a composite mixture. These methods were largely developed in the last century to resolve the main egg white proteins (Li-Chan et al, 1995). Recently, these separation techniques and their refinements have also been used to characterise eggshell matrix proteins and lower abundant egg white proteins.

The eggshell matrix proteins were separated after removal of minerals from the eggshell. The polypeptides contained in extra- and intramineral extracts, and in the precursor milieu (uterine fluid) where eggshell calcification takes place, were analysed by 1D-electrophoresis (SDS-PAGE). A complex array of distinct bands is demonstrated in the soluble intra- and extramineral compartments (Hincke et al., 1992; Gautron et al., 1996) and in the uterine fluid that showed different patterns between the 3 stages of the eggshell calcification process (initial, growth and terminal) (Gautron et al, 1997). N-terminal sequencing of the electrophoretic bands allowed the identification of egg white proteins (ovalbumin, lysozyme and ovotransferrin) as well as of specific shell matrix proteins (Hincke, 1995; Hincke et al, 1999, 2000; Gautron et al., 2001a, 2001b). N-terminal and internal amino acid sequencing of protein bands revealed that a number of them did not correspond to previously identified sequences. Efficient purification schemes (diethylaminoethyl (DEAE), hydroxyapatite and Carboxymethyl-sepharose FF columns) were developed to purify to homogeneity Ovocleidin-17 (Hincke et al., 1995) and Ovocalyxin-32 (Hincke et al., 2003) from the eggshell extracts.

Different chromatographic techniques have been used to separate hen egg white proteins, allowing determination of the protein composition of this biological fluid (Awade et al., 1994; Castellani et al., 2003; Croguennec et al., 2000, 2001; Nau et al., 1999, Hiidenhovi et al., 1999). Recently developed supercritical technologies were powerful tools to separate egg fractions, notably the egg yolk (Aro et al., 2004). Nevertheless, these approaches failed to resolve all egg proteins, especially the minor ones. In that aim, electrophoretic separations appeared as promising complementary methods. SDS-PAGE was for a long time the only technique described for the study of egg white proteins. But isoelectrofocusing (IEF), and even more 2D-PAGE allowed the separation of much more numerous proteins, some of them present in very low quantities (Désert et al., 2001; Nau et al., 2003a; Nau et al., 2005).

More recently, affinity chromatography on gelatine-agarose beads and gelatinzymography have been used to purify and characterize the activity of a metalloprotease from egg white and yolk (Rénault S, personal communication). This purification method and gelatin casein zymographies are commonly used to study metalloproteases (Das et al., 2004, Hu et al., 2004).

**Proteomics**

Proteins are the functional units of biological processes and consequently their study is of great interest. The term proteome refers to the complete set of proteins produced by a given cell or organism under defined conditions and proteomics is the study of the proteome, aiming at the identification of all proteins, their post-translational modifications and their interactions. The proteome is not static, but can change with time or under the influence of environmental conditions. The study of proteomes usually involves mass spectrometry-based high-throughput methods for protein identification (Mann and Steen, 2004). Typically the proteins of a given cell or tissue are extracted and degraded with specific proteases. This highly complex mixture is then introduced into a mass spectrometer usually
via a capillary HPLC column that is directly coupled to the mass spectrometer inlet. The peptides separated on this column elute from a fine tip and the eluate is then vaporized and the peptides ionized by applying a strong electric potential. This process is called electrospray ionization (ESI). In the mass spectrometer the masses of the peptides and of their fragments, created in a process called collision-induced dissociation (CID) brought about by the collision of the peptide ions with an inert gas, are determined as precisely as possible. For the identification of proteins, the sequences of peptides are determined by comparing the pattern of fragment ions to the theoretical fragment ion pattern predicted by dedicated computer programs from in silico digestion and fragmentation of sequences stored in databases. Such an approach of course depends on the availability of the sequences to be analyzed.

The recent publication of 90-95% of the chicken genome (International Chicken Genome Sequencing Consortium, 2004), makes it possible to explore the egg proteome, or egg compartment sub-proteomes, using mass spectrometry-based high-throughput methods as shown recently for the acid-soluble organic matrix of the chicken calcified eggshell layer (Mann et al., 2006) and for the egg white (Guérin-Dubiard et al., 2006). LC-MS/MS (mass spectrometric sequence analysis after separation of the peptides by liquid chromatography) of peptides obtained by direct cleavage of the organic matrix or after pre-fractionation of the proteins by 1-dimensional SDS-gel electrophoresis identified altogether 520 different proteins as constituents of the eggshell matrix, including all matrix proteins known before. Calculating the exponentially modified Protein Abundance Index (emPAI; Ishihama et al., 2005) for the proteins it was possible to divide them into three abundance groups. The highly abundant group of 32 proteins contained, among others, all known specific eggshell proteins, known as ovocleidins and ovocalyxins. In addition to specific proteins and proteins which also occur in other tissues but are also produced by eggshell gland cells probably on purpose to play a role in shell matrix assembly, the matrix protein mixture also contained components which may be derived from decaying cells and basement membranes lining the oviduct and leftovers from secretion processes. These proteins may have been included into the mineralized shell just because they were present in the uterine fluid during mineralization. The complex mixture contained, among many others, egg white proteins, extracellular growth factors and many other signal transduction chain components, lipid-binding proteins, immune system-related and antimicrobial proteins, proteins also occurring in body fluids, such as serum albumin, hemopexin or vitamin D-binding protein, and also some previously uncharacterized proteins of unknown origin and function.

For hen egg white proteomic analysis, different strategies and methods have been used to bypass some technical problems due to specific egg white properties. The egg white proteins were separated by 2D-PAGE using different pH ranges to increase resolution power of the technique, but previous chromatographic fractionation was also used to improve detection of minor proteins. Many proteins were identified with peptide mass fingerprinting (PMF). The excised protein spots from 2-D gels were cleaved in gel with specific proteases and the masses of the peptides were measured using MALDI-TOF, a mass spectrometric method which is less suitable for MS/MS sequencing. However, sequence analysis is not used in PMF. The pattern of experimentally determined masses of a given set of peptides (forming the fingerprint) is then used to search sets of mass profiles generated by theoretical fragmentation of a protein database for matches. However, the confidence level with PMF is often not as high as with MS/MS methods and nano LC-MS/MS was necessary in many cases to confirm the PMF-MALDI-TOF analysis for an unambiguous identification (Guérin-Dubiard et al., 2005; Guérin-Dubiard et al., 2006). Altogether 16 proteins were identified, two of which, Tenp and VMO-I, have not been identified as egg white proteins before, and which also occur in the shell matrix. Many of these proteins produced more than one spot in electrophoretic separation. This may be due to differences in posttranslational modifications.

Molecular biology of the egg

cDNA and EST libraries

The development of functional genomic tools provides alternative methods to identify egg proteins. A cDNA library is a set of the mRNAs contained within a cell, organism or tissue and consequently contains the coding region of a genome that will be translated into proteins and then possibly subjected to post-translational modification. Consequently, a complete cDNA library gives the total of the proteins expressed in this tissue, cell or organ. The cDNA library is prepared by isolating the total
mRNA from the tissues or organs of interest. Because mRNAs are fragile and consequently it is difficult to work with them, the enzyme reverse transcriptase is used to synthesize a DNA strand complementary to each mRNA molecule (cDNA). The DNA molecules are then inserted into vectors and cloned. These reverse transcribed mRNAs inserted in bacterial vectors are collectively known as the cDNA library. Unidirectional sequencing of cDNAs yields short cDNA sequences (200–500 nucleotides) which are known as Expressed Sequence Tags (ESTs). Databanks of EST sequences are a comprehensive catalog of global or tissue-specific mRNA sequences expressed in an organism. EST libraries have been prepared from a large variety of organs under different physiological situations that could modify gene expression. For chicken, 529,525 EST sequences are reported in the Tigr database (release 10.0 January 28, 2005) and 588,288 in the database of “Expressed Sequence Tags” (dbEST) (release 051906, May 2006).

A cDNA library of pooled RNA from chicken uterus harvested at the middle phase of shell calcification has been successfully used for the identification of novel eggshell matrix proteins that were previously unidentified (Hincke et al, 1999; Gautron et al, 2006). Expression screening of this library with specific polyclonal antisera raised to partially purified eggshell matrix proteins allowed clones with corresponding cDNA sequences to be identified. After purification, the plasmid that contained the cDNA coding for the matrix proteins was sequenced. The conceptual amino acid sequence was compared to partial amino acid sequencing data for proteins present in uterine fluid and eggshell extracts. This method allowed the identification of two novel eggshell matrix proteins, ovocleidin-116 (Hincke et al., 1999) and ovocalyxin-36 (Gautron et al., 2006).

The available EST sequences can be compared to partial protein or nucleotide sequences from egg components. This method was successfully used to identify a 32 kDa band abundant in uterine fluid at the terminal phase of shell calcification (Gautron et al, 2001b). Limited dissolution of the outer shell with HCl produced an extract that was enriched in the 32 kDa protein which was further purified by liquid chromatography (Hincke et al., 2003). This purified protein was microsequenced to yield the sequences of the N-terminus and internal peptides. Database searching using tblastN with these sequences allowed the identification of several corresponding expressed sequence tags (ESTs). They could be assembled to obtain a full length cDNA sequence containing all peptides previously sequenced. This protein is novel and was named ovocalyxin-32 (OCX-32) (Gautron et al., 2001b).

As well, internal peptide sequencing, achieved from an excised 2D-PAGE spot, enabled the identification of a member of lipocalin protein family, CALgamma protein (Pagano et al., 2003), in hen egg white by comparison with translated proteins in data banks (Nau et al., 2003a). But sometimes tryptic digestion of protein separated by 2D-PAGE can reveal unknown peptide sequences, because the databanks are not yet completed. That is what happened with HEP21 protein, a new egg white minor protein revealed, cloned and characterized using an efficient combination of 2D-PAGE and molecular biology techniques (Nau et al., 2003b).

Use of the chicken genomic sequence

The publication of the chicken genome resource (International Chicken Gene Sequence Consortium, 2004) was a major event for the scientific community. Poultry scientists are of course enthusiastic, but this genome is also particularly interesting for evolution studies. It is the first sequenced bird genome and it will enable us to understand animal and human evolution (Jensen, 2005). Poultry is also one of the major animal productions in the world that involve breeding programs. The chicken genome sequence publication with the accompanying paper on the chicken polymorphisms (International Chicken Polymorphism Map Consortium, 2004) will help researchers to identify specific markers usable in the breeding industry (Jensen, 2005; Chaves et al, 2005; Dunn, 2005; Vignal and Besbes, 2005). The chicken genome sequence and the existence of a large number of expressed sequence tags (ESTs) provides a large opportunity to identify proteins of interest that will be deposited in the egg. One example of the use of the functional genomic information is the study of the expression of defensins in chicken (Xiao et al., 2004). Defensins are a large family of antimicrobial peptides that are capable of killing pathogens and thus play a critical role in the first line of host defense. A genome-wide screen of the chicken sequence was used for comparative analysis of the defensins among chicken, mouse and human (Xiao et al, 2004). These proteins all share a consensus defensin motif that was used to screen the chicken genome. Thirteen different chicken
defensins, designated as gallinacins, were described. Specific primers were designed to study their expression in various tissues or organs including oviduct and ovary where the egg is formed. Gallinacin-10, -11 and -12 were found to be specifically expressed in oviduct tissues indicating that they could be deposited in the egg to play a role in the defense against pathogens. A similar approach was used to identify three potential cathelicidins, also involved in host defense against pathogens (Xiao et al., 2006). This strategy seems very promising and should be more generally applied in a near future to identify relevant chicken proteins.

Transcriptomics

Only a very small percentage of the DNA in vertebrate genomes encodes proteins because the exons of most genes are separated by much longer non-coding introns. The genome also contains vast amounts of noncoding "junk" DNA, so even when the complete sequence of a genome is known, it is often difficult to focus on particular genes. One approach to solve the problem is to examine a transcriptome of the organism, which is all the mRNA molecules (or transcripts) in one cell type or a population of cells for a given set of physiological circumstances. Transcriptomics depict the expression level of genes, often using techniques capable of sampling thousands of different mRNA molecules at a time. This technique uses DNA arrays which are a collection of microscopic DNA spots attached to a solid surface, such as glass or plastic membrane forming an array. In general, arrays are described as macroarrays or microarrays, according to the size of the sample spots. Macroarrays contain hundreds to 1-2 000 sample spots. A microarray typically contains thousands of genes on a surface that is as large as a stamp. DNA arrays provide a medium for matching known and unknown DNA samples based on base-pairing rules. RNAs extracted from 2 conditions that have to be compared (different tissues or physiological stages) are reverse transcribed into cDNA which are then labeled with fluorescent probes. cDNAs of the first condition are labeled with a green fluorescent dye or with a red fluorescent for cDNAs of the second situation. The labeled cDNAs (probe) are then mixed and put on slides for hybridization. After washing, the fluorescence present on the surface of the glass is measured. When a gene is only expressed in the first condition, it will appear in green and in red when only expressed in the second condition. When a gene is expressed under both conditions to be compared, green and red colors are mixed to appear more or less yellow. The next step is to use bioinformatics tools first to link the expressed genes to the identified clones present on the arrays, then to analyze the differential expression of genes. After completion, a list of genes expressed under only one of the two conditions chosen or under both conditions will be available. Differentially expressed genes are then classified according to already described functions, and consequently give information on the biological role of the genes expressed and of their translated proteins.

Transcriptomics analyses of the hen oviduct are currently being developed in two European programs (SABRE for Cut Edge for Substainable Animal Breeding, 2006-2010, and RESCAPE for Reducing Egg Susceptibility to Contaminations in Avian Production in Europe, 2006-2009). Microarrays will be use to identify novel proteins expressed in relation with the deposition and synthesis of egg components. In a first step, genes specifically expressed in the magnum (organ of egg white deposition), the white isthmus (organ of eggshell membranes deposition) and in uterus (organ of eggshell calcification) will be analysed. Then the gene’s expression in organs will be studied at various physiological stages to determine the differentially expressed genes related to the deposition of proteins in each of these egg compartments. Differentially expressed genes will be classified according to already described functions and consequently will give us information on the biological role of the genes expressed during the egg deposition.

Conclusion

Until recently, the total number of egg proteins that was already characterized was less than 50 and was limited to the most abundant ones. The recent development of high-throughput methods and the availability of the chicken genome sequence have already allowed the identification of hundreds of novel minor components of the egg. It is obvious that this list of egg proteins and genes coding for them will increase and this screening step might be achieved in a very near future. Novel egg compounds with potential biological active properties have been identified. Several alternatives are
available to assess the physiological functions of proteins, through the analysis of their biological activities (enzymatic or antimicrobial activities for example). These conventional techniques can easily be combined to bioinformatic analysis of protein sequences. The latter gives numerous details starting with the physicochemical characterization of proteins (isoelectric point, molecular mass, post-translational modification…) to the identification of bioactive domains or motifs that may be useful to predict properties and activities of proteins. The study of functions of proteins and of their potential for industrial application is a great challenge for the next years.

References


