Data Article

Peptidomic data of egg white gastrointestinal digests prepared using the Infogest Harmonized Protocol

Marta Santos-Hernández, Beatriz Miralles, Lourdes Amigo, Isidra Recio*

Instituto de Investigación en Ciencias de la Alimentación, CIAL (CSIC-UAM, CEI UAM+CSIC), Nicolás Cabrera, 9, 28049 Madrid, Spain

ARTICLE INFO

Article history:
Received 9 June 2020
Revised 19 June 2020
Accepted 23 June 2020
Available online 26 June 2020

Keywords:
Egg white protein
Characterization
Mass spectrometry
Maldi-tof
Peptidomics

ABSTRACT

These data are related to the research article entitled “Induction of CCK and GLP-1 release in enteroendocrine cells by egg white peptides generated during gastrointestinal digestion”. In this article, the peptide and free amino acid profile of egg white gastrointestinal in vitro digests is shown. Egg white proteins were digested following the INFOGEST gastrointestinal digestion protocol. Different time points of gastric and intestinal digestion were characterized regarding protein, peptide and amino acid content. Protein degradation was followed by SDS-PAGE where some electrophoretic bands were identified by MALDI-TOF/TOF after tryptic digestion. Moreover, the molecular weight distribution of egg white peptides found at different times of gastrointestinal digestion was performed using MALDI-TOF. Peptides identified from the most abundant egg white proteins by tandem mass spectrometry were represented using a peptide profile tool and raw data are given in table format. These results reveal the protein regions resistant to digestion and illustrate the free amino acid profile of egg white protein at the end of the digestion pro-

DOI of original article: 10.1016/j.foodchem.2020.127188
* Corresponding author.
E-mail address: i.recio@csic.es (I. Recio).

https://doi.org/10.1016/j.dib.2020.105932
2352-3409/© 2020 Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license.
(http://creativecommons.org/licenses/by-nc-nd/4.0/)
cess. These data can be used for nutritional purposes and to identify allergen epitopes or bioactive sequences.

© 2020 Published by Elsevier Inc.
This is an open access article under the CC BY-NC-ND license. (http://creativecommons.org/licenses/by-nc-nd/4.0/)

---

### Specifications Table

| Subject area                  | Biochemistry                                                                 |
|-------------------------------|-----------------------------------------------------------------------------|
| More specific subject area    | Proteomics and biochemistry                                                 |
| Type of data                  | Figures, Tables                                                              |
| How data was acquired         | High-pressure liquid chromatography coupled to electron spray ionization     |
|                               | interface and ion trap. Matrix assisted Laser Desorption/ionization coupled to a |
|                               | time of flight detector                                                     |
| Data format                   | Raw                                                                          |
| Parameters for data collection| Digested samples following INFOGEST protocol were freeze-dried and kept at |
|                               | −20 °C until analysis. Digests were analyzed by mass spectrometry after a    |
|                               | reducing step with dithiothreitol.                                           |
| Description of data collection| MS/MS raw files were processed by using Data Analysis (version 4.0 Bruker    |
|                               | Daltonics) and Biotools version 3.2, and the identification search was achieved|
|                               | using Mascot v2.4                                                           |
| Data source location          | Data is collected and analysed at the Institute of Food Science Research, CIAL |
|                               | (CSIC-UAM), Nicolás Cabrera 9, 28049, Madrid, Spain                         |
| Data accessibility            | Data in this article                                                        |
| Related research article      | M. Santos-Hernández, L. Amigo, Recio, I. “Induction of CCK and GLP-1 release  |
|                               | in enteroeocrine cells by egg white peptides generated during gastrointestinal |
|                               | digestion”.                                                                  |

Value of the data

- The data provide the distribution of the nitrogen fraction into peptides and amino acids at the end of the gastrointestinal digestion of egg white. The profile of free amino acids that can be used for nutritional purposes is also given.
- The here provided proteomic, peptidomic and amino acid profiles of egg white protein digests can be compared with in vivo data or with data obtained in dynamic systems.
- Egg white protein domains resistant to gastrointestinal digestion are provided which could serve to detect allergen epitopes or peptides with biological activities.
- This peptidomic characterization was useful to identify peptides as inducers of incretin hormones, being relevant to control food intake and diabetes.

## 1. Data description

Egg white protein was digested following a harmonized in vitro digestion protocol [1,2] where samples were taken at 30 and 120 min of gastric digestion and 30 and 120 min of gastrointestinal digestion. These digests were centrifuged at 5000 × g over 20 min to separate soluble and insoluble fraction, followed by snap freezing in liquid nitrogen and freeze-dried. All digests were characterized regarding their protein, peptide and amino acid composition.

The distribution of the nitrogen fraction after egg white in vitro gastric and intestinal digestion was assessed by elemental analysis (Fig. 1a) and amino acid analysis (Fig. 1b). Around 90% of the total nitrogen content was collected in the soluble fraction after gastric digestion, determined by elemental analysis, and this percentage increased up to 99% at the end of the intestinal phase (Fig. 1a). Similar percentages were obtained by total amino acid analysis where the soluble fraction represented 91% and 99% at the end of the gastric and the intestinal phase, respectively (Fig. 1b). At the end of the gastric phase, just 1% of the total nitrogen fraction was
Fig. 1. Distribution of the nitrogen content between soluble and insoluble fraction after gastric and intestinal digestion. Digests were centrifuged at 5000×g for 20 min. Supernatant and pellet were freeze-dried and weighted. Nitrogen content in each fraction was determined by a) elemental analysis and b) amino acid analysis. Total and free amino acids were separately determined in the soluble fraction.

Fig. 2. SDS-PAGE protein profiles of egg white protein at different times of in vitro gastrointestinal digestion. MW, molecular weight marker; EW, egg white undigested protein; 30 G and 120 G, 30 and 120 min gastric digestion; 30 I and 120 I, 30 and 120 min intestinal digestion. The numbers correspond to an identified band. 1, Ovomucin; 2, Ovotransferrin; 3, Ovoinhibitor; 4, Ovalbumin-Y; 5, Ovalbumin; 6, Ovomucoid; 7, Lysozyme. The electrophoretic bands with * were identified by MALDI-TOF/TOF after hydrolysis with trypsin.
in the form of free amino acids, increasing up to 27% at the end of the intestinal phase. The difference between total and free amino acids in the soluble part corresponded to the nitrogen fraction in the form of proteins and peptides which ranged from 90% at the end of the gastric phase to 72% at the end of the intestinal phase.

Protein degradation during egg white gastrointestinal digestion was followed by SDS-PAGE (Fig. 2). At the end of the gastric phase, main electrophoretic bands were identified as ovalbumin and ovalbumin-related protein Y and these were still detected at the end of the intestinal phase. Electrophoretic bands with * in Fig. 2 were identified by MALDI-TOF/TOF after in-gel digestion with trypsin.

Egg white protein digests were also characterized by MALDI-TOF and the peptide profile is represented in Fig. 3, which describes the percentage of peptides within a given molecular weight range. As expected, peptides with a molecular weight lower than 2 kDa increased during gastrointestinal digestion and the amount of longer peptides decreased. However, it should be noted that peptides with a molecular weight higher than 6 kDa were still detected at the end of gastrointestinal digest. Peptidomic characterization of the digests was performed by HPLC tandem mass spectrometry (HPLC-MS/MS). Raw data of peptide sequences identified at the end of the gastrointestinal digestion by HPLC-MS/MS are given in Tables 1 to 5. The identified peptides from the major protein fractions, ovalbumin, ovomucoid and ovotransferrin, at different time points during gastrointestinal digestion, are also represented using the peptide profile tool from the Peptigram web-based application (Fig. 4). In these graphs, each vertical bar corresponds to an amino acid identified as part of a peptide sequence, the height of the bar is proportional to the number of peptides overlapping this position and the color intensity is proportional to the sum of the intensities of the peptides overlapping a given position. Under our mass spectrometry conditions, peptides with a molecular weight between 5 and 30 kDa are detected. The blank regions observed in the peptide profile during the gastric phase probably correspond to peptide fragments too long to be solubilized or ionized under our analysis conditions, while blank regions in the intestinal phase are more likely due to short peptides, free amino acids, or peptides with low ionization capacity. It has to be noted that peptide intensity depends on peptide ionization capacity and abundance, and in consequence, peptide intensity cannot directly be transformed to peptide concentration. During the gastric phase, only peptides from ovotransferrin were identified, suggesting a higher susceptibility of this protein to the action of pepsin. Several ovalbumin and ovomucoid peptides were detected after 30 and 120 min intestinal digestion. In addition, the amino acid composition of the identified peptides was analysed by using

![Fig. 3. Molecular weight distribution of egg white peptides found at different times of gastrointestinal digestion. 30 G and 120 G, 30 and 120 min of gastric digestion, respectively; 30 I and 120 I, 30 and 120 min intestinal digestion, respectively.](image-url)
Peptides from ovalbumin, ovomucoid and ovotransferrin identified in egg white gastrointestinal digests represented using the Peptigram Bioware tool. Each vertical bar corresponds to an amino acid identified as part of a peptide sequence. The height of the bar is proportional to the number of the peptides overlapping this position and the color intensity is proportional to the sum of the intensities of the peptides overlapping a given position. Each line corresponds to a different time point: 120 G, 120 min gastric digestion; 30I and 120I, 30 and 120 min intestinal digestion, respectively.

Fig. 4.
Table 1
Ovalbumin-derived peptides identified at the end of egg white gastrointestinal digestion.

| RANGE  | SEQUENCE | RANGE  | SEQUENCE |
|--------|----------|--------|----------|
| 20     | KVHHANENI | 188    | AFKDEDTQAMP |
| 22     | HHANENI   | 190    | KDEDTQAMP  |
| 23     | HANENI    | 200    | RVTEQESKPV |
| 44     | LGAKDST   | 200    | RVTEQESKPVQM |
| 44     | LGAKDSTRT | 200    | RVTEQ     |
| 45     | GAKDST    | 200    | RVTEQESKPVQ |
| 48     | DSTRT     | 205    | ESKPV     |
| 48     | DSTRTQIN  | 213    | YQGL      |
| 54     | IKVVI     | 219    | RVASMASEKM |
| 61     | DKLPG     | 220    | VASMASEKM  |
| 74     | CGTSYN    | 220    | VASMASEKM  |
| 82     | SSLRDLNQQTSDKPVNDV | 230 | ILELP    |
| 83     | SSLRDLNQ  | 230    | ILELPASCTMS |
| 89     | NQITKPNVDV | 230  | ILELPASCTMS |
| 99     | NQITKPNVYS | 232   | ELPEASCTMS |
| 95     | QITKPN    | 232    | ELPEASCTMS |
| 91     | ITKPN     | 243    | LVLLPDEV  |
| 107    | YAAERYPILPE | 244  | VLLPDEVSG |
| 107    | AEERYPILPE | 244   | VLLPDEV    |
| 113    | PIPE      | 244    | VLLPDEV    |
| 116    | PEYQCVEK  | 246    | LPDEV      |
| 116    | PEYQCVKEL | 254    | EQLESII    |
| 121    | CVKEL     | 255    | QLESII     |
| 127    | RGGLEPIN  | 257    | ESIN       |
| 127    | RGGLEPI   | 261    | NFEKLTE    |
| 128    | GGGLEPI   | 269    | TSNVMEER   |
| 136    | QTAADQA   | 271    | SNVMEER    |
| 136    | QTAADQAR  | 286    | MKMEEK     |
| 137    | TAADQA    | 299    | MAMGI      |
| 137    | TAADQREL  | 299    | MAMGITDV   |
| 150    | VESQNGIIRNV | 300   | AMGTDV     |
| 156    | GIIRN     | 302    | GITDV      |
| 161    | VLQPS     | 308    | SSSANGLSIAES |
| 161    | VLQPSVDSQTA | 314  | SGISSA     |
| 161    | VLQPSVDSQ | 314    | SGISSAIEL  |
| 161    | VLQPSVDS  | 314    | SGISSAIEL  |
| 166    | VQPSVDSQTA | 333   | AEINEAGREVGS |
| 167    | VQPSVDSQTM | 336   | NEAGREVGS  |
| 166    | SVDSQ     | 337    | EAGREVGS   |
| 166    | SVDSQTMV  | 354    | SVSEE      |
| 166    | SVDSQTM   | 354    | SVSEEFRADHP |
| 166    | SVDSQTA   | 354    | SVSEEFRADHP |
| 176    | VNAIV     | 360    | RADHP      |
| 188    | AFKDEDTQAMP | 360  | RADHP      |

a Range in the protein. Uniprot accession number: P01012.
b One letter amino acid code is used.

the ExPASy-Protparam tool (Fig. 5). Resistant peptides identified at the end of intestinal digestion by mass spectrometry were rich in serine, valine and in negatively charged residues aspartic and glutamic acid.

Free amino acids released after 120 min of gastric digestion and 30 and 120 min of intestinal digestion were followed by HPLC and post-column ninhydrin derivatization. As shown in Fig. 6, free amino acids were mainly released during intestinal digestion, with phenylalanine, leucine and lysine being the most abundant, followed by arginine, valine and serine.
Table 2  
Ovomucoid-derived peptides identified at the end of egg white gastrointestinal digestion.

| RANGE | SEQUENCE |
|-------|----------|
| 1     | AEVDCS   |
| 21    | VCNKDLRPICGTDGVT |
| 21    | VCNKDLRPI |
| 23    | NKLDRPICTGD |
| 28    | PICGTDGVT |
| 59    | DGECKETVPM |
| 68    | MNCSYAN |
| 86    | LCNAFNP |
| 92    | NIPVCGTDGVT |
| 137   | DCSEYPKD |
| 148   | AEDRPL |
| 149   | EDRP |
| 170   | AVVESNGTLTL |

a Range in the mature form of the protein. Uniprot accession number: P01005.  
b One letter amino acid code is used.

Table 3  
Ovotransferrin-derived peptides identified at the end of egg white gastrointestinal digestion.

| RANGE | SEQUENCE |
|-------|----------|
| 13    | SSPEE   |
| 21    | NNLRLDIQTEQ |
| 24    | RDLTQERISLTCV |
| 29    | QERIS   |
| 66    | EAGLAP  |
| 97    | VVKKGTEFTVNDLQ |
| 105   | TVNLDL |
| 135   | RGAIE   |
| 142   | GIESGS  |
| 148   | IESGVS  |
| 157   | SASCVPAGAT |
| 160   | CVPGA   |
| 178   | PTKC    |
| 214   | NENAPDQ |
| 214   | NENAPDQKDE |
| 230   | DGSRQPVDN |
| 232   | SRQPV   |
| 234   | QPVVDN  |
| 291   | KDPVLKD |

a Range in the mature form of the protein. Uniprot accession number: P02789.  
b One letter amino acid code is used.

2. Experimental design, materials, and methods

2.1. Distribution of nitrogen content

Supernatant and pellet from gastric and intestinal digests were freeze-dried and weighted. Protein content in each fraction was measured by elemental analysis and by total amino acid analysis after acid hydrolysis with HCl 6N at 110°C for 24h. In addition, free amino acids were also determined in the soluble part according to the method previously published [3]. The difference between total and free amino acids was ascribed to proteins and peptides.
### Table 4

Lysozyme-derived peptides identified at the end of egg white gastrointestinal digestion.

| RANGE$^a$ | SEQUENCE$^b$ |
|-----------|--------------|
| 18        | 25           | DNYRGYSL |
| 34        | 44           | FESNFNTQATN |
| 39        | 43           | NTQAT |
| 44        | 52           | NRRTDGSTD |
| 45        | 52           | RRTDGSTD |
| 47        | 52           | TDGSTD |
| 64        | 72           |ure MCGRTPGS |
| 72        | 83           | SRNLCNPSCAL |
| 85        | 90           | SSDITA |
| 85        | 89           | SSDIT |
| 97        | 105          | KVSDNGNM |
| 99        | 105          | VSINGNM |
| 112       | 122          | GTDVQA |

$^a$ Range in the mature form of the protein. Uniprot accession number: P00698.

$^b$ One letter amino acid code is used.

### Table 5

Mucin-5B-derived peptides identified at the end of egg white gastrointestinal digestion.

| RANGE$^a$ | SEQUENCE$^b$ | RANGE$^a$ | SEQUENCE$^b$ |
|-----------|--------------|-----------|--------------|
| 34        | 38           | 1064      | 1073         | EACHKVNVIP |
| 98        | 102          | 1165      | 1169         | GCCYPE |
| 125       | 129          | 1216      | 1223         | YPLNETIY |
| 125       | 131          | 1235      | 1241         | FCPGNNM |
| 134       | 141          | 1326      | 1330         | EAL |
| 147       | 154          | 1384      | 1388         | CLGEE |
| 200       | 204          | 1528      | 1532         | GITE |
| 272       | 277          | 1580      | 1584         | KSDDA |
| 293       | 298          | 1581      | 1590         | SDDARKRNGE |
| 376       | 380          | 1596      | 1600         | KEMAL |
| 433       | 437          | 1648      | 1660         | PPQPYYEVASR |
| 453       | 457          | 1690      | 1694         | RQTN |
| 454       | 458          | 1721      | 1729         | REVIVDTE |
| 545       | 552          | 1744      | 1751         | PDGNNILN |
| 546       | 551          | 1827      | 1834         | TETCVIECD |
| 550       | 562          | 1827      | 1834         | TETCVIECD |
| 657       | 662          | 1880      | 1884         | KPGAV |
| 753       | 761          | 1880      | 1887         | KPGAVVAK |
| 901       | 908          | 1886      | 1896         | PKSSCEDVCT |
| 923       | 928          | 1910      | 1914         | CVVK |
| 1012      | 1019         | 1913      | 1917         | VKCQT |
| 1044      | 1051         | 1930      | 1942         | |

$^a$ Range in the mature form of the protein Uniprot accession number: Q98UI9.

$^b$ One letter amino acid code is used.

### 2.2. Molecular weight distribution by MALDI-TOF

Samples were diluted with 33% acetonitrile and 0.1% of trifluoroacetic acid prior to spotted into a MALDI target plate with 2,5-dihydroxybenzoic acid matrix. Analyses were performed on an Autoflex SpeedTM (Bruker Daltonic, Bremen, Germany). Mass spectra were acquired in positive reflection mode and were collected from the sum of 1000 on average lasers shots. Monoisotopic peaks were generated by FlexAnalysis software. The monoisotopic peaks were organized and represented in a molecular weight distribution range. All other methods are described in [3].
Fig. 5. Amino acid composition of the identified peptide sequences at the end of the in vitro gastrointestinal digestion of egg white. Repeated sequences are avoided.

Fig. 6. Free amino acid profile in egg white protein digests at different time points. 120G corresponds to 120 min of gastric digestion; 30I and 120I correspond to 30 and 120 min of intestinal digestion, respectively.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.
Acknowledgments

This work has received financial support from project AGL2015–66886-R from the Spanish Ministry of Economy and Competitiveness (MINECO). M. S-H is the recipient of an FPU grant (FPU15–03616) from the Ministry of Education, Culture and Sports.

References

[1] A. Brodkorb, L. Egger, M. Alminger, P. Alvito, R. Assuncao, S. Ballance, T. Bohn, C. Bourlieu-Lacanal, R. Boutrou, F. Carriere, A. Clemente, M. Corredig, D. Dupont, C. Dufour, C. Edwards, M. Golding, S. Karakaya, B. Kirkhus, S. Le Feunteun, U. Lesmes, A. Macierzanka, A.R. Mackie, C. Martins, S. Marze, D.J. McClements, O. Menard, M. Minekus, R. Portmann, C.N. Santos, I. Souchon, R.P. Singh, G.E. Vegarud, M.S.J. Wickham, W. Weitschies, I. Recio, INFOGEST static in vitro simulation of gastrointestinal food digestion, Nat. Protoc. 14 (2019) 991–1014 https://doi.org/10.1038/s41596-018-0119-1.

[2] M. Minekus, M. Alminger, P. Alvito, S. Ballance, T. Bohn, C. Bourlieu, F. Carriere, R. Boutrou, M. Corredig, D. Dupont, C. Dufour, L. Egger, M. Golding, S. Karakaya, B. Kirkhus, S. Le Feunteun, U. Lesmes, A. Macierzanka, A. Mackie, S. Marce, D.J. McClements, O. Menard, I. Recio, C.N. Santos, R.P. Singh, G.E. Vegarud, M.S.J. Wickham, W. Weitschies, A. Brodkorb, A standardised static in vitro digestion method suitable for food - an international consensus, Food Funct 5 (2014) 1113–1124 https://doi.org/10.1039/c3fo60702j.

[3] M. Santos-Hernández, D. Tomé, C. Gaudichon, I. Recio, Stimulation of CCK and GLP-1 secretion and expression in STC-1 cells by human jejunal contents and in vitro gastrointestinal digests from casein and whey proteins, Food Funct 9 (2018) 4702–4713 https://doi.org/10.1039/c8fo01059e.