

A COMPUTATIONAL LABEL-FREE METHOD FOR THE QUANTITATIVE ANALYSIS OF *LUPINUS ALBUS* STORAGE PROTEINS USING HPLC-CHIP-MS/MS

Francesca Brambilla^{1*}, Donatella Resta¹, Ilana Isak¹, Giovanna Boschin², Marco Zanotti³ and Anna Arnoldi¹

¹Laboratory of Food Chemistry and Mass Spectrometry, Institute of Endocrinology, University of Milan, Via Balzaretti 9, I-20133 Milan, Italy

²Department of Agri-Food Molecular Sciences (DISMA), University of Milan, Via Celoria 2, I-20133 Milan, Italy

³Agilent Technologies Italia SpA, Cernusco sul Naviglio, Italy

* Corresponding author's email: francesca.brambilla@unimi.it

ABSTRACT

In shotgun proteomics, a complex mixture of proteins with different chemical and physical behavior is digested with trypsin to obtain a mixture of peptides suitable to be analyzed with HPLC-Chip-MS/MS. As far as quantitative proteomics is concerned, the label-free quantification techniques are able to obtain relative protein abundances directly from high-throughput proteomics analyses in order to profile the differential protein expression and the relative abundance of peptides contained in different samples. In this investigation we have developed a differential label-free method for the simultaneous qualitative characterisation of the main seed storage proteins and relative quantification of the bioactive proteins in the total protein extracts of different *Lupinus albus* cultivars. In particular our method enables to detect the differential expression of the two main classes of bioactive lupin proteins, vicilins and gamma-conglutin, by optimizing two different bioanalytical parameters. The main features of this method are the use of HPLC-Chip technology for the chromatographic separation and of the Spectrum Mill MS Proteomics Workbench software for processing raw MS/MS data.

KEYWORDS

differential analysis, HPLC-Chip-MS/MS analysis, *Lupinus albus*, shotgun proteomics

INTRODUCTION

In recent years, mass spectrometry-based quantitative proteomics has become an important tool in the field of food quality control with the aim to trace and quantify the differential expression of specific proteins such as bioactive or allergenic proteins. In particular advances in mass spectrometry, especially the ion trap device, and the introduction of specific software have led to the development of 'shotgun proteomics' (Gao *et al.* 2003). It refers to the use of bottom-up proteomics techniques in which the extracted protein

content of a sample is digested prior to chromatographic separation and mass spectrometry analysis. Typically, liquid chromatography (LC) is coupled with tandem mass spectrometry (MS/MS) resulting in high-throughput peptide analysis. The experimental MS/MS peptide spectra are searched against a protein database to identify the proteins of the sample starting from their sequence. This method reduces reliance on the previous fractionation of the mixture protein to be analyzed.

In this work, we have developed a HPLC-Chip-MS/MS label-free method based on 'shotgun proteomics' for the simultaneous qualitative characterisation and relative quantification of all storage proteins in the total protein extracts (TPEs) of *Lupinus albus* seeds of different cultivars (cv. Adam, Arés, Lucky, and Multitalia). Our attention has been focalised on the main bioactive lupin seed storage proteins. They are vicilins (or beta-conglutin) and gamma-conglutin. Vicilins are an heterogeneous class of different protein isoforms which derive from the same precursors named vicilin-like protein and beta-conglutin precursor in the NCBI database. They seem to be the hypocholesterolemic component among lupin seed proteins due to their high sequence homology with the alpha' subunit of soybean beta conglucinin (Wait *et al.* 2005). Gamma-conglutin is a homogeneous protein constituted by a light chain and a heavy chain linked each other by disulfide bonds. It is a major lupin allergen and a few literature data show that it have a potential hypoglycaemic activity (Magni *et al.* 2005).

For the differential approach, that aims to evaluate the abundance ratios of the two target bioactive proteins in different cultivars without the use of labeled reagent, we have develop two different computational parameters according to the different nature of the vicilins and gamma conglutin. They are respectively the N-MEAN, that take in account all vicilin peptides identified in the three replicates of all the four cultivars, and the N-ACRP, based on gamma-conglutin Common and Reproducible Peptides (CRP), i.e. those peptides identified simultaneously in all replicates of each cultivar and in all four cultivars.

Our LC-MS/MS system is the microfluidic nano HPLC-Chip, a hybrid between a HPLC and a nano-electrospray ionisation source, since the chip incorporates an enrichment column, an analytical column and a nano-electrospray capillary needle coupled with an Ion Trap device. The use of this innovative nano-technology allowed a dramatic increase in sensitivity and reproducibility of the lupin proteins detection with respect to the use of a conventional ESI source.

MATERIALS AND METHODS

SAMPLES

Total Protein Extracts (TPEs) from mature dry seeds of *L. albus* cv. Adam, Arès, Lucky and Multitalia were obtained with the following experimental procedure. Briefly, the storage proteins were extracted from the defatted flour obtained by lupin seeds of the four cultivars (cv. Adam, Arès, Lucky and Multitalia) in Tris-HCl 100 mM, NaCl 0.5 M, pH 8.2; the suspension was centrifuged at 10,000 rpm for 45 min at 7°C. The protein content of the supernatant was assessed with the colorimetric method of Bradford. TPEs were all spiked with the same amount of bovine serum albumin (BSA) (1 ng BSA: 10 ng TPE), digested with trypsin, without any previous fractionation, to a mixture of peptides suitable for the HPLC-Chip chromatography coupled with a data dependent acquisition of the MS/MS fragmentation spectra (Fig. 1).

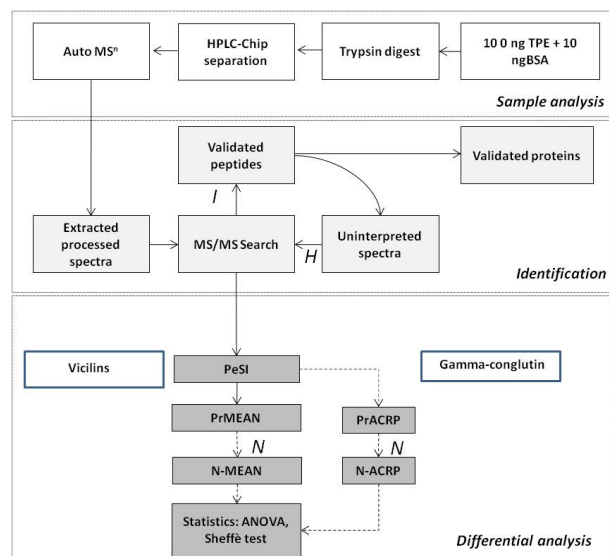


Fig. 1. Flow scheme of the differential label-free method for relative quantification of bioactive lupin proteins in seeds of different cultivars of *L. albus*.

HPLC-CHIP-MS/MS ANALYSIS

A constant amount of each spiked TPEs (110 ng of protein = 100 ng TPE + 10 ng BSA) was injected three times onto a HPLC-Chip-MS/MS system consisting of a 1200 Series liquid chromatograph, a HPLC-Chip Cube MS interface, and a SL Ion Trap mass spectrometer (all Agilent Technologies, Palo Alto, CA, USA). The chromatographic Chip incorporated a 40 nL enrichment

column, a 43 mm × 75 μm analytical column packed with Zorbax 300SB-C18 5-μm particles and a nanospray needle. The MS/MS analysis were performed in Auto (MSn) mode (Fig. 1).

DATABASE SEARCH FOR PROTEIN IDENTIFICATION

Figure 1 shows the experimental strategy used in this work to obtain simultaneously qualitative and quantitative information. The whole National Center for Biotechnology Information non redundant (NCBI) database was searched using the new version of Agilent Spectrum Mill software (Rev A.03.03). The Spectrum Mill Data Extractor program prepares MS/MS data files for processing; it extracts high-quality experimental fragmentation spectra from raw MS/MS data files and rejects spectra that are too noisy or do not represent peptides. The preprocessed spectra are used by Spectrum Mill to interrogate the NCBI protein database in the MS/MS search; NCBI is searched for tryptic peptides. In the first stage of MS/MS search, the identity mode is used to look for sequences that are identical to the peptide used to generate the MS/MS spectrum (unmodified peptides). The validation of results is as stringent as possible to assure a high quality match and subsequent protein identification reducing the risk of false positives. The resulting validated protein sequences are used as hits in the second MS/MS search step that is performed in the homology mode to search unvalidated MS/MS spectra. The homology mode looks both for the main variable modification among a limited list of modified amino acids (modified peptides: acetyl K, oxidised M, pyroglutamic acid on N-termQ, deaminated N, phosphorylated S, T, Y), and for matches that are consistent with a single amino acid substitution (substituted peptides). The homology mode search is therefore useful for increasing the percentage coverage of validated proteins. Consequently, the homology MS/MS search is an essential step in the flow scheme of our method to assure a satisfactory qualitative and quantitative analysis (Fig. 1).

RELATIVE QUANTIFICATION

The quantitative parameter directly and easily accessible through the software is the Protein Mean Peptide Spectral Intensity (PrMEAN), the mean peak intensity of all peptide precursor ions identified for each protein. This parameter considers the peptide intensities (PeSI) both of Common Reproducible Peptides (CRPs), those peptides identified in all three replicates of the same cultivar and in all cultivars, of Specific Reproducible Peptides (SRPs), those peptides identified in all three replicates of the same cultivar, but not in all cultivars, and of Non-Replicable Peptides (NRPs) those peptides identified only in some replicates of single cultivars.

In order to reduce the data variability due to the contribution of SRPs and NRPs, we developed a new parameter named Protein Average of Common Reproducible Peptides (PrACRP) which is the mean peak intensity of the only CRPs identified for each protein.

In order to comparing different samples, these two parameters were normalised by dividing each of them by the corresponding parameter of the internal standard BSA to give N-MEAN and N-ACRP (Fig. 1). The normalised parameters were submitted to a statistical evaluation to verify whether there were any statistically significant differences in the content of the target bioactive proteins, i.e. gamma-conglutin and vicilins, among the four investigated cultivars.

RESULTS AND DISCUSSION

QUALITATIVE ANALYSIS OF THE STORAGE PROTEINS IN THE TOTAL PROTEIN EXTRACTS FROM LUPIN SEEDS

The main classes of lupin seed storage proteins are: 7S acidic globulins (beta-conglutin or vicilins), a 7S basic globulin (gamma-conglutin), 11S globulins (alpha-conglutin or legumins), and a 2S globulin (delta-conglutin). Gamma-conglutin is an homogeneous protein, composed by a heavy and a light chain linked by disulfide bonds; the vicilins and the legumins have a multigenic origin and appear as complex mixtures of polypeptides with different molecular weights and pIs.

A single HPLC-Chip-MS/MS analysis of the TPE tryptic digest (100 ng of digest proteins) permitted to identify simultaneously all major and minor storage proteins with very satisfactory percentage coverage. The lupin protein sequences successfully identified in the protein database NCBIInr were the following: for the class of vicilins, the beta-conglutin precursor (NCBIInr accession number 46451223) and the vicilin-like protein (NCBIInr accession number 89994190); for gamma-conglutin, conglutin-gamma (NCBIInr accession number 11191819); for the class of legumins, the legumin-like protein (NCBIInr accession number 85361412); for delta-conglutin, the conglutin-delta seed storage protein precursor (NCBIInr accession number 80221495).

In the case of the vicilins, the HPLC-Chip-MS/MS analysis enabled the identification of both the beta-conglutin precursor with a percentage coverage ranging from 32% to 49% and the vicilin-like protein with a percentage coverage ranging from 33% to 42% for the analyzed cultivars. In spite of their high sequence homology, estimated by using the BLAST-P program as 78% identity and 80% positivity, it was possible to identify both common and specific peptides of these sequences. The identified specific peptides of the beta-conglutin precursor were: RGQEQSDQDEGVIVISSK, AIYVVVVDEGEGNYELVGIR and the corresponding missed-cleavage peptide AIYVVVVDEGEG YELVGIRDQQR, LSEGDFVIPAGYPISINASSNLR, NFLAGSKDNVIR, INEGALLPHYNSK, LLGFGI NADENQR, AVNELTFPGSAEDIER, whereas those of the vicilin-like protein were NFLAGSEDNVIR, AIFIV VVGEGNGK and the corresponding missed-cleavage peptide AIFIVVVGEGNGKYELVGIR, LLGFGIN AYENQR, LSEGDFVIPAGYPISVNASSNLR, ELTFPGSAEDIER, LIK_nQQQSYFANALPQQQQ QSEK.

Vicilins are oligomers of 150-170 kDa formed by three similar subunits 40-70 kDa with no disulfide linkages. They are highly heterogeneous and their heterogeneity is due to the expression of multigene families whose individual gene are very closely related (Freitas *et al.* 2007). cDNA-deduced vicilin sequences in databases are not yet complete. In this work, the homology search is performed with the aim of complementing the information that can be obtained by an identity search against the deposited vicilin-like protein and beta-conglutin precursor sequences. Several single aminoacid substitution are successfully identified confirming the presence of high homologous vicilin precursors. Some mutations have been identified in all four cultivars: S469F, I493V, E494D, D492T. All the other single aminoacid mutations are not common to all cultivars according to the high degree of polymorphism within each species of lupin (Freitas *et al.* 2007). Conglutin-gamma was identified with 25% to 33% percentage coverage in all cultivars, a very good result considering that this protein is very resistant to enzymatic digestion. In the gamma-conglutin sequence the homology search did not highlight any single aminoacid substitution confirming that this protein is the lupin protein with the most peculiar and less heterogeneous subunit composition. The percentage sequence coverage of the legumin-like protein was always small falling between 3% and 13%, which may perhaps indicate that the correct sequences have not yet been deposited in the database. The coverage of delta-conglutin was between 27% and 37%.

QUANTITATIVE ANALYSIS OF GAMMA-CONGLUTIN IN THE TPES FROM LUPIN SEEDS

The method allowed the identification of conglutin-gamma with 3 CRPs, belonging either to the light chain (AGIALGTHQLEENLVVFDLRLAR and HSIFEVFT QVFANNVPK) or the heavy chain (TPLMQVP VLLDLNGK). These three peptides were detected in the analyses of all cultivars and in all replicates for the same cultivar and were, therefore, the best candidates to develop a quantitative method for gamma-conglutin. The linearity of the mass spectrometry response for gamma-conglutin PrACRP and PrMEAN was evaluated injecting different amounts of the unspiked TPE tryptic digest of the cultivar Arés in the range from 25 ng to 500 ng. The gamma-conglutin PrACRP was calculated in the range from 50 ng to 500 ng, where these CRPs were all detected. Comparing PrACRP vs PrMEAN in the same range, it is shown that PrACRP showed the best linearity ($R^2=0.999$). It is therefore possible to affirm that the limit of quantification of gamma-conglutin according to the parameter PrACRP is equal to 50 ng of TPE.

Table 1 reports the PeSI averages, the standard deviations, and the corresponding RSD % of gamma-conglutin peptides after triplicate analyses of the four spiked TPES. In spite of the peptide RSD% might be greater than 20% in some cultivars, the corresponding RSD% of PrACRPs ranged between 10% and 20%,

which may be considered a very good result for a bioanalytical measurement. On the contrary, the RSD% of PrMEANs ranged between 4% and 37% (Table 2).

We have chosen the PrACRPs as the most reliable quantitative parameters for the differential analyses of gamma-conglutin and we have normalised this parameter with the corresponding PrACRPs calculated for BSA in order to obtain N-ACRPs (Table 3). One-way ANOVA ($p < 0.05$) and Sheffè test showed that N-ACRP parameter can discriminate different cultivars with respect to their amount in gamma-conglutin and that the cultivar Adam contains less gamma-conglutin than the cultivars Multitalia, Arès, and Lucky.

QUANTITATIVE ANALYSIS OF THE VICILINS IN THE TOTAL PROTEIN EXTRACTS FROM LUPIN SEEDS

The label-free comparative analysis of lupin vicilins is complicated by the very complex nature of this class of proteins, both because they have a multigenic origin and undergo extensive post-transcriptional

modifications, and the availability of protein sequences deposited in the database until now (i.e. the beta-conglutin precursor and vicilin-like protein) that may not be exhaustive. The vicilin PrACRPs, calculated from the intensities of CRPs, was not suitable for the differential analysis of the whole vicilin class, since it does not take in consideration the contribution of SRPs, such as IVEFQSKPNTLILPK that is a main peptide in some cultivars. Thus for the differential analysis of vicilins, it was necessary to consider the contribution of all identified peptides, because they are all essential for the correct quantification of all vicilin isoforms. Consequently, in the case of the vicilins we used Pr-MEAN, which was normalised by using BSA Pr-MEAN to give N-MEAN (Table 3). The results were subjected to statistical evaluation as already described for gamma-conglutin: one way ANOVA indicated that N-MEAN was able to discriminate the cultivars for their vicilin content and the Sheffè test indicated that the cultivar Multitalia contains more vicilins than the other cultivars.

Table 1. PeSI values of the three Common Reproducible Peptides of gamma-conglutin of four *L. albus* cultivars; values were expressed as average \pm standard deviation ($n = 3$) and Relative Standard Deviation percentage.

	AGIALGTHQLEENLVVFDLAR		HSIFEVFTQVFANNVPK		TPLMQVPVLLDLNGK	
	Average \pm S.D.	RSD%	Average \pm S.D.	RSD%	Average \pm S.D.	RSD%
Adam	5.00E+07 \pm 7.91E+06	15.83	5.83E+07 \pm 1.93E+07	33.13	1.80E+07 \pm 8.75E+06	48.55
Arès	5.52E+07 \pm 3.70E+06	6.70	4.38E+07 \pm 7.89E+06	18.03	4.07E+07 \pm 1.94E+07	47.54
Lucky	3.53E+07 \pm 2.05E+07	58.04	1.02E+08 \pm 1.72E+07	16.89	2.60E+07 \pm 2.63E+06	10.11
Multitalia	3.45E+07 \pm 1.00E+07	29.01	6.32E+07 \pm 8.46E+06	13.38	2.15E+07 \pm 6.82E+06	31.74

Table 2. Values of PrACRP and PrMEAN of the gamma conglutin of four *L. albus* cultivars; values were expressed as average \pm standard deviation ($n = 3$) and Relative Standard Deviation percentage.

	PrACRP		PrMEAN	
	Average \pm S.D.	RSD%	Average \pm S.D.	RSD%
Adam	4.21E+07 \pm 8.32E+06	19.77	3.64E+07 \pm 5.11E+06	14.04
Arès	4.65E+07 \pm 4.72E+06	10.13	3.41E+07 \pm 2.36E+06	6.93
Lucky	5.44E+07 \pm 1.16E+07	21.39	4.41E+07 \pm 1.84E+06	4.19
Multitalia	4.47E+07 \pm 7.54E+06	16.86	2.70E+07 \pm 9.96E+06	36.84

Table 3. Values of N-ACRP of gamma-conglutin and N-MEAN of vicilins of four *L. albus* cultivars; values were expressed as average \pm standard deviation ($n = 3$) and Relative Standard Deviation percentage.

	N-ACRP of gamma-conglutin		N-MEAN of vicilins	
	Average \pm S.D.	RSD (%)	Average \pm S.D.	RSD (%)
Adam	0.61 \pm 0.09	14.08	2.34 \pm 0.16	6.90
Arès	0.99 \pm 0.06	5.79	2.46 \pm 0.34	13.76
Lucky	0.91 \pm 0.15	16.82	2.18 \pm 0.54	24.29
Multitalia	0.78 \pm 0.04	4.97	3.56 \pm 0.27	7.58

CONCLUSION

The proposed shotgun-proteomics analysis based on HPLC-Chip-MS/MS allowed a complete characterisation of all lupin seed storage proteins, since minor ones, such as gamma-conglutin and delta-conglutin, were easily identified together with major ones with satisfactory percentage coverages without any previous fractionation of the TPE. As concerns the qualitative characterisation of lupin proteins, the homology mode search has permitted to identify single amino acid substitutions in vicilins. The MS/MS sequencing of single amino acid substituted peptides, starting from the vicilin-like protein and the beta-conglutin precursor, enabled to acquire important knowledge of the heterogeneous sequences of vicilins. Moreover, in the present investigation, we have demonstrated that our internal standard label-free method based on HPLC-Chip-MS/MS is able to profile the different expression of vicilin and gamma-conglutin, in a potentially unlimited number of samples. N-MEAN appears to be the most suitable parameter for profiling the differential expression of vicilins, whereas the normalised parameter N-ACRP appears to be the most suitable one for profiling the differential expression of gamma-conglutin.

This work was supported by Fondazione Cariplo, Project 'Novel Methodologies for the quality control of the production of lupin and lupin based food products' (2005-1803), by the EU Project 'Bioprofibre' (COOP-CT-2006-032075) and by PRIN 'Metodologie innovative per la quantificazione dei principali allergeni in alimenti complessi' (2006075417).

LITERATURE CITED

- Freitas, L.R., A.R. Teixeira and R.B. Ferreira. 2007. Vicilin-type globulins follow distinct patterns of degradation in different species of germinating legume seeds. *Food Chem.* 102: 323.
- Gao, J., G.J. Opitek, M. Friedrichs, A.R. Dongre and S.A. Hefta. 2003. Changes in the protein expression of yeast as a function of carbon source. *J Proteom Res* 2: 643.
- Liu, H., R.G. Sadygov and J.R. Yates. 2004. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem* 76: 4193.
- Magni, C., C. Ballabio, P. Restani, E. Sironi, A. Scarafoni, C. Poesi and M. Duranti. 2005. Two-dimensional electrophoresis and western blotting analyses with anti Ara h 3 basic subunit IgG evidence the cross-reacting polypeptides of *Arachis hypogaea*, *Glycine max* and *Lupinus albus* seed proteomes. *J. Agric Food Chem.* 53: 2275.
- Wait, R., R. Gianazza, D. Brambilla, I. Eberini, S. Morandi, A. Arnoldi and C.R. Sirtori. 2005. Analysis of *Lupinus albus* Storage Proteins by Two-Dimensional Electrophoresis and Mass Spectrometry. *J. Agric. Food Chem.* 53: 4599.