Proteomic analysis of different extracts from barley grains

J. Chmelík[1], P. Řehulka[1], M. Střelcová[2], V. Kubáň[2], C. Mayrhofer[3], G. Allmaier[3]

[1]Institute of Analytical Chemistry, Academy of Sciences of the Czech Republic, Brno, Czech Republic
[2]Mendel University of Agriculture and Forestry in Brno, Czech Republic
[3]Institute for Analytical Chemistry, University of Vienna, Austria

ABSTRACT

Identification of several proteins extracted with various solvents from barley grains was performed by the proteomic strategy based on a combination of gel electrophoresis, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) and bioinformatics. The mass spectrometry-based strategy designed for protein identification exhibits good sensitivity and rapid protein identification in comparison to other methods. In order to simplify the protein mixture extracted from grains, we used sequential extractions of grains with various solvents (based on modified Osborne system). Several proteins (e.g. β-amylase, B3-hordein) were identified in this way. It seems that this approach can become an important tool for proteomics of cereals.

Keywords: proteomics; protein identification; barley; mass spectrometry; grain extraction

Cereals are cultivated grasses that are grown throughout the world providing food for man and farm stock. Barley (*Hordeum vulgare* L.) is one of major world crops (ranks among the top ten crops and is fourth among the cereals). Barley also serves as an important experimental or model plant species for numerous studies in malting and brewing chemistry, plant breeding methodology, genetics, biotechnology etc. (Nilan and Ullrich 1993). Understanding of the function of biological systems requires knowledge of their chemical composition. The state-of-life of a cell at any given time is defined by its protein composition, i.e., its proteome. That is why it is necessary to have a rapid and an efficient tool for characterization of the proteome.

There are several ways of protein identification, however, majority of them are very time consuming procedures. A modern and efficient tool is based on comparison of theoretically derived molecular masses of peptides from known DNA sequences and primary protein structures in databases with experimentally obtained molecular masses of enzyme-generated peptides of separated proteins. The mass spectrometry-based strategy designed for protein identification exhibits good sensitivity and this approach allows even detailed structure/function analyses.

Separation techniques (e.g. one-dimensional gel electrophoresis used in this work), mass spectrometry (e.g. matrix-assisted laser desorption/ionization time-of-flight mass spectrometry employed in this work), and bioinformatics (e.g. database searching) became important tools for protein and proteome analysis (Aebersold and Goodlett 2001). These methods have been successfully applied for the identification of proteins in numerous biological systems (Cordwell et al. 2001, Mann et al. 2001) including barley leaves and grains (Chmelík et al. 2001). The advantages of this proteomic method in comparison to other methods are mainly in smaller amounts of the samples necessary and speed of protein identification. The limited amount of the protein is determined mainly by sensitivity of the staining procedure used to visualize protein samples in the separation gels. Even spots visualized by very sensitive staining procedures (e.g. silver staining) contain enough protein material (about 10 ng) for identification. The *in-gel* digestion and peptide extraction and purification take about 24 hours. However, because it is possible to process several protein spots simultaneously and subsequent MALDI-TOF measurements are not too time demanding (about 10 min for one protein digest without sample purification and preparation). Identification of several proteins can be done within 48 hours. It is much faster than it is possible by other methods.

Cereal grains have a long storage life under favorable conditions because they are harvested at relatively low moisture content and comprise stable components. The principal energy sources within the grains are protected from infestation by outer coverings. The largest morphological component of all grains is the endosperm, and approximately 80% of this is starch. Other important compound classes are the storage proteins, which make the next largest contribution to endosperm dry weight. Proteins are important both as nutrients and as active chemicals because they include enzymes that, although making a small contribution to grain weight, can have a marked effect on grain quality.

The classification system of seed (grain) proteins is based largely on the pioneering work of Osborne, who recognised that seed proteins differ in their solubility properties (Osborne 1895). Although the classical Osborne scheme can be criticised as being insufficiently

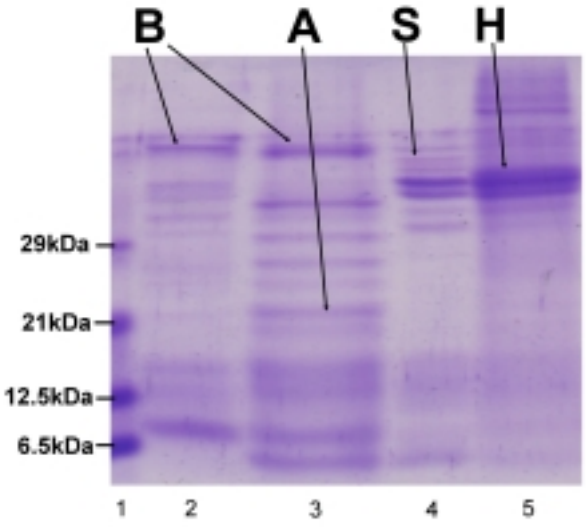

Figure 1. SDS gel electrophoresis of barley extracts from grains; the identified proteins are labeled: B = β-amylase, A = α-amylase/subtilisin inhibitor precursor, S = signal recognition particle 54kDa protein 1, H = B3-hordein; lane 1 contains protein standard markers, lane 2 to lane 5 contain proteins from water, salt, ethanol and NaOH extracts

rigorous in modern terms, it provides a useful way of classifying seed proteins and it is still used by most workers in this field (Cooke 1984). There are thus considered four categories of proteins occurring in seeds:

1) albumins, which are soluble in water and comprise mostly enzymatic proteins;
2) globulins, which are soluble in dilute salt solutions and generally occur in protein bodies (i. e., they can be considered as storage proteins in the strict sense);
3) prolamins, which are soluble in aqueous ethanol solutions and are also found in protein bodies as true storage proteins;
4) glutelins, which are soluble in alkaline or acid solutions, or in detergents and are probably mainly structural proteins, although some of them may have metabolic functions.

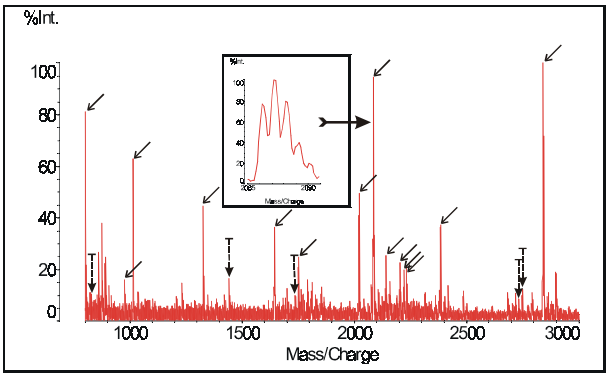

Figure 2. MALDI-TOF MS spectrum of a peptide mixture from *in-gel* tryptic digest of a protein band B, lane 2 (identified as β-amylase) isolated by gel electrophoresis from barley grain extract; the arrow-marked peptides were used for database searching, the peaks marked with dashed-arrows are peptides found after reevaluation of the spectra; an image of one isotopically resolved peak is shown as an inset

The extraction procedure not only differentiates various categories of proteins occurring in seeds but also simplifies the protein mixture for characterization. For high throughput proteomics, a combination of gel electrophoresis with peptide mass fingerprinting based on molecular mass measurement of enzyme-generated peptides by mass spectrometry is very useful. These methods together with database searching were used for identification of proteins extracted from barley grains. Protein mixtures were isolated by sequential extraction with various solvents from barley grains, then separated using one-dimensional SDS polyacrylamide gel electrophoresis (SDS-PAGE) and subsequently *in-gel* digested with trypsin. Exact peptide molecular masses (precision of measurements was ≤ 200 ppm with external calibration) were obtained by means of matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) (Karas and Hillenkamp 1988) and these were used as an input into the database-searching programs. This strategy is not always successful because the success depends mainly on the presence of the protein sequence in the database. In the case of barley, only part of the genome is known and there exist many variants of barley. In spite of this complication, several proteins were identified.

MATERIAL AND METHODS

For the identification of grain proteins, applying a combination of SDS-PAGE and mass spectrometry followed by database searching, the barley variety Monaco was selected.

The extraction procedure is based on differences in solubility of grain proteins in various solvents. Barley grains (1 g) were milled and then extracted consecutively in four solvents. First, the milled sample was extracted with distilled water (10 ml). The suspension was stirred at laboratory temperature for 20 min and then centrifuged (7000 × g, 15 min). The filtrated supernatant was used as the extract 1 (albumin fraction). The remaining insoluble sample was mixed with aqueous 5% (w/v) NaCl solution (10 ml), the extraction procedure was repeated, and the extract 2 was collected (globulin fraction). After following extractions with aqueous 70% (v/v) ethanol and aqueous 0.2% NaOH solution the extract 3 (prolamin fraction), and the extract 4 (glutelin fraction) were obtained.

Gel electrophoresis, *in-gel* digestion of separated proteins and determination of peptide masses were described elsewhere (Chmelík et al. 2001). A simple purification step using ZipTip C18 technology (Millipore, Bedford, MA, USA) was employed and all positive ion MALDI mass spectra were acquired in the linear modus with either Kompact MALDI IV tDE or AXIMA-CFR (Kratos Analytical) TOF instrument. The determined exact molecular masses of the tryptic peptides in the region m/z 800 to 3000 were submitted into the search program

Table 1. Summary of matched masses and protein coverage of identified β-amylase

m/z submitted	MH^+ matched	Delta ppm	Start	End	Peptide sequence	Modifications
1016.7300	1016.5569	170.2539	411	418	(K)LFGFTYLR(L)	
1326.9100	1326.6694	181.3633	384	394	(R)YDPTAYNTILR(N)	
1753.1300	1752.8921	135.7337	419	433	(R)LSNQLVEGQNYVNFK(T)	
2022.1400	2021.9317	103.0171	111	128	(R)DVGTRDPDIFYTDGHGTR(N)	
2086.3000	2086.0722	109.2181	129	146	(R)NIEYLTLGVDNQPLFHGR(S)	
2206.2900	2206.0901	90.5955	439	457	(R)MHANLPRDPYVDPMAPLPR(S)	1Met-ox
2222.2100	2222.0851	56.2297	439	457	(R)MHANLPRDPYVDPMAPLPR(S)	2Met-ox
2841.2100	2841.2869	–27.0511	217	242	(K)AAAAAVGHPEWEFPNDVGQ YNDTPER(T)	

The matched peptides cover 21% (115/535AA's) of the protein

ProteinProspector – MS-Fit (http://prospector.ucsf.edu/ucsfhtml3.4/msfit.htm) and SwissProt (ver. 9.30.2001) was used as an annotated protein searching database. The lower mass region (< m/z 800) was not suitable for peptide detection because a lot of signal noise was caused by the formation of matrix-related ions.

RESULTS AND DISCUSSION

The SDS gel electropherograms of protein mixtures extracted by various solvents from the grains are shown in Figure 1. There were identified four proteins: β-amylase (labeled as B), α-amylase/subtilisin inhibitor precursor (A), signal recognition particle 54kDa protein 1 (S), B3-hordein (H). β-amylase was mostly present in the water-containing fractions, but it was also found to a minor extent in salt fraction. Other identified proteins were not found in more than one fraction.

After *in-gel* digestion of the separated proteins by trypsin, the isolated and purified peptides were transferred onto a MALDI target covered with the matrix α-cyano-4-hydroxycinnamic acid (Vorm et al. 1994). An example of the MALDI-TOF spectrum of peptides from the aqueous extract (the separated protein spot is labeled as B in Figure 1) is shown in Figure 2. The marked peptides were used for protein identification and Table 1 summarizes the database search results. We employed a simple reevaluation procedure for the analysis of acquired spectrum for this protein. After identification of this protein as β-amylase, we found corresponding amino acid sequence in protein database (Table 2) and then we made theoretical trypsin cleavage of this protein using web accessible program PeptideMass (http://www.expasy.ch/tools/peptide-mass.html). We compared theoretical masses of ionized peptides (as $[MH]^+$) with the spectrum and we found some other low intensity peaks belonging to peptides from β-amylase (they are labeled with dashed arrow in Figure 2). The coverage of primary protein sequence then increased from 21% up to 28%.

The other labeled proteins were identified in the same way. The results are supported by the analysis shown in Table 3. Besides these identified proteins, several proteins from so-called hordein group were identified and identification of other proteins is under development.

Our results proved that the combination SDS PAGE followed by *in-gel* digestion and MALDI-TOF MS is a suitable strategy for the identification of barley proteins. The mass spectrometry-based strategy designed for protein identification exhibits good sensitivity and this approach will allow both identification of barley varieties and detailed structure/function analyses (it includes the investigation of protein composition at different stages of the plant life). Further improvement of the extraction procedure and the use of larger gels (the present size was 100 × 100 × 0.7 mm) will be necessary for these purposes. A MS/MS analysis (post-source decay

Table 2. The results of reevaluation of spectra data of identified β-amylase (only peptides found after reevaluation are shown in this table); the $[MH]^+$ ion at m/z 2750.27 can be correlated to two peptides, corresponding to two theoretical peptide values (m/z 2750.36 and 2750.43); the data were acquired in the linear TOF mode, which allowed no differentiation

Reevaluated masses	MH^+ matched	Delta ppm	Start	End	Peptide sequence
832.4600	832.5079	–57.4993	325	331	(R)TIARMLK(R)
1442.9800	1442.7062	189.7912	371	383	(R)EGLNVACENALPR(Y)
1739.0900	1738.8737	124.3689	395	410	(R)NARPHGINQSGPPEHK(L)
2736.3400	2736.4123	–26.4126	395	418	(R)NARPHGINQSGPPEHKLFGFTYLR(L)
2750.2700	2750.3572	–31.7007	371	394	(R)EGLNVACENALPRYDPTAYNTILR(N)
2750.2700	2750.4306	–58.3949	411	433	(K)LFGFTYLRLSNQLVEGQNYVNFK(T)

The peptides corresponding submitted and revaluated masses cover 28% (151/535AA's) of the protein

Table 3. Summary of identified proteins

LA	EX	SC	PE	M/pI	PO	AC	NA
B	H_2O	2.91e + 004	8/14 (57%)	59647.9/5.58	HORVU	P16098	β-amylase (1,4-α-D-glucan maltohydrolase)
A	NaCl salt	8.47e + 007	13/30 (43%)	22164.1/7.77	HORVU	P07596	α-amylase/ subtilisin inhibitor precursor (BASI)
S	ethanol	3.09e + 003	8/16 (50%)	54512.2/9.33	HORVU	P49968	signal recognition particle 54kDa protein 1 (SRP54)
H	NaOH	1.57e + 003	5/21 (23%)	30195.7/7.74	HORVU	P06471	B3-hordein (fragment)

LA – labels of proteins on the gel (see Figure 1)
EX – extraction medium
SC – MOWSE score based on Pappin et al. (1993): Curr. Biol., *3*: 327–332.
PE – ratio of matched and submitted peptides
M – relative molecular mass of the identified protein
pI – isoelectric point of the protein based on computing from primary structure
PO – protein origin
AC – accession number of protein in SwissProt database
NA – protein name

in MALDI-TOF MS) of peptides from *in-gel* digestion should also be employed for better protein identification.

This work was supported by the grant 521/99/1576 from the Grant Agency of the Czech Republic and an Austrian-Czech Republic Collaboration Grant (II/4).

REFERENCES

Aebersold R., Goodlett D.R. (2001): Mass spectrometry in proteomics. Chem. Rev., *101*: 269–295.

Chmelík J., Řehulka P., Mayrhofer C., Allmaier G. (2001): Proteomics – a new and fast tool for variety identification of barley varieties. Kvasný Prům., *46*: 158–163.

Cooke R.J. (1984): The characterization and identification of crop cultivars by electrophoresis. Electrophoresis, *5*: 59–72.

Cordwell S.J., Nouwens A.S., Walsh B.J. (2001): Comparative proteomics of bacterial pathogens. Proteomics, *1*: 461–472.

Karas M., Hillenkamp F. (1988): Laser desorption ionization of proteins with molecular masses exceeding 10000 daltons. Anal. Chem., *60*: 2301–2303.

Mann M., Hendrickson R.C., Pandey A. (2001): Analysis of proteins and proteomes by mass spectrometry. Ann. Rev. Biochem., *70*: 437–473.

Nilan R.A., Ullrich S.E. (1993): Barley: Taxonomy, origin, distribution, production, genetics, and breeding. In: Barley: Chemistry and technology, MacGregor A.W., Bhatty R.S. (eds.): Amer. Assoc. Cereal Chem., St. Paul.

Osborne T.B. (1895): The proteids of barley. J. Amer. Chem. Soc., *17*: 539–567.

Vorm O., Roepstorff P., Mann M. (1994): Improved resolution and very high-sensitivity in MALDI TOF of matrix surfaces made by fast evaporation. Anal. Chem., *66*: 3281–3287.

Received on December 11, 2001

ABSTRAKT

Proteomická analýza různých extraktů z obilek ječmene

Pomocí proteomické metody založené na kombinaci gelové elektroforézy, hmotnostní spektrometrie a bioinformatiky bylo identifikováno několik bílkovin (např. β-amyláza, B3-hordein) izolovaných z obilek ječmene pomocí různých rozpouštědel. Tato metoda prokázala dobrou citlivost a rychlost při identifikaci bílkovin ve srovnání s jinými metodami. Pro získání jednodušších směsí bílkovin byla použita postupná extrakce různými rozpouštědly (na základě modifikovaného Osbornova systému). Výsledky ukazují, že by se tento postup mohl stát důležitým nástrojem pro studium proteomiky obilovin.

Klíčová slova: proteomika; identifikace bílkovin; ječmen; hmotnostní spektrometrie; extrakce obilek

Corresponding author:

RNDr. Josef Chmelík, CSc., Ústav analytické chemie AV ČR, Veveří 97, 611 42 Brno, Česká republika,
tel.: + 420 5 32 29 01 83, fax: + 420 5 41 21 21 13, e-mail: chmelik@iach.cz