Talanta 82 (2010) 1412–1420

Contents lists available at ScienceDirect

Talanta

journal homepage: www.elsevier.com/locate/talanta

Decision peptide-driven: A free software tool for accurate protein quantification using gel electrophoresis and matrix assisted laser desorption ionization time of flight mass spectrometry

Hugo M. Santos[a], Miguel Reboiro-Jato[b], Daniel Glez-Peña[b], J.D. Nunes-Miranda[c], Florentino Fdez-Riverola[b], R. Carvallo[b], J.L. Capelo[a,c,*]

[a] REQUIMTE, Departamento de Química, Faculdade de Ciências e Tecnologia, FCT, Universidade Nova de Lisboa, 2829-516 Caparica, Portugal
[b] SING Group, Informatics Department, Higher Technical School of Computer Engineering, University of Vigo, Ourense, Spain
[c] Bioscope Group, Physical Chemistry Department, Science Faculty, University of Vigo, Ourense, Spain

ARTICLE INFO

Article history:
Received 23 February 2010
Received in revised form 30 June 2010
Accepted 3 July 2010
Available online 3 August 2010

Keywords:
Protein quantification
Gel electrophoresis
MALDI-TOF-MS
^{18}O-labeling
DPD software

ABSTRACT

The decision peptide-driven tool implements a software application for assisting the user in a protocol for accurate protein quantification based on the following steps: (1) protein separation through gel electrophoresis; (2) in-gel protein digestion; (3) direct and inverse ^{18}O-labeling and (4) matrix assisted laser desorption ionization time of flight mass spectrometry, MALDI analysis. The DPD software compares the MALDI results of the direct and inverse ^{18}O-labeling experiments and quickly identifies those peptides with paralleled loses in different sets of a typical proteomic workflow. Those peptides are used for subsequent accurate protein quantification. The interpretation of the MALDI data from direct and inverse labeling experiments is time-consuming requiring a significant amount of time to do all comparisons manually. The DPD software shortens and simplifies the searching of the peptides that must be used for quantification from a week to just some minutes. To do so, it takes as input several MALDI spectra and aids the researcher in an automatic mode (i) to compare data from direct and inverse ^{18}O-labeling experiments, calculating the corresponding ratios to determine those peptides with paralleled losses throughout different sets of experiments; and (ii) allow to use those peptides as internal standards for subsequent accurate protein quantification using ^{18}O-labeling. In this work the DPD software is presented and explained with the quantification of protein carbonic anhydrase.

1. Introduction

Protein quantification methods based on protein separation by 1D-gel electrophoresis and matrix assisted laser desorption ionization time of flight mass spectrometry, MALDI, have a number of drawbacks that make reliable quantification difficult. The differences in the yields of protein digestion obtained in different sets of in-gel digestions, the variation observed in the recovery yields of individual peptides within a set of in-gel digest, the biased losses of peptides that might occur during the post-digestion sample processing of in-gel digests when are used ZipTip pipette tips to clean the sample or the speed vacuum pump to dry down and preconcentrate the sample. The aforementioned drawbacks can cause loss of peptides ranging in between 30 and 90% depending on the amount of sample loaded in the gel and on the type of peptide studied [1–3].

If for a given protein was possible to identify a certain number of peptides that had low and paralleled loses through a typical proteomic workflow entailing 1D-gel protein separation and in-gel protein digestion, then such peptides would allow robust and accurate protein quantification. The experimental method that could allow to extract and to identify the number of peptides that remains constant in expression level through a typical in-gel digestion workflow should be based in a peptide differential analysis. A variation of the method proposed by Wang et al. and called "inverse labeling" can be used to do such analysis [4]. With this procedure it is easily detected if a peptide is randomly loosed, or the observed yields of individual peptides vary strongly within a set of in-gel digests. This methodology can be used to unambiguously verify the yield of peptides obtained during in-gel protein digestion at different concentrations, and thus clearly illustrates which peptides can be used for quantification through a given dynamic range of differential quantification.

The application of the "inverse labeling" methodology requires the use of mass spectrometry. MALDI can be used for this purpose. The comparison of MALDI spectra, however, it makes this approach for protein quantification tedious and time-consuming.

* Corresponding author at: Bioscope Group, Physical Chemistry Department, Science Faculty/Facultad de Ciencias, As Lagoas, E-32004 Ourense, Spain.
Tel.: +34 610 835 903; fax: +34 988 387 001.
E-mail address: jlcapelom@uvigo.es (J.L. Capelo).

0039-9140/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.talanta.2010.07.007

To speed the treatment of data the software "Decision Peptide Driven", DPD, has been developed based on previous software developed for medical applications [5] as a computer tool to extract and to identify the peptides that remains constant in expression level through different sets of a typical in-gel digestion workflow. The present manuscript described in detail the software tool DPD, explaining through a real example how to use it. This software is freely source code available, and it can be run as a multiple platform.

2. Materials and methods

2.1. Apparatus

Gel electrophoresis was performed with an electrophoresis system, model Mini-PROTEAN Tetra Cell, from Biorad (Hercules, CA, USA), following the manufacturer instructions. Protein digestion and labeling were done in safe-lock tubes of 0.5 mL from Eppendorf (Hamburg, Germany). A minicentrifuge, model Spectrafuge-mini, from Labnet (Madrid, Spain), and a minicentrifuge-vortex, model Sky Line, from ELMI (Riga, Latvia) were used throughout the sample treatment, when necessary. A vacuum concentrator centrifuge from UniEquip (Martinsried, Germany) model UNIVAPO100H with a refrigerated aspirator vacuum pump model Unijet II was used for (i) sample drying and (ii) sample pre-concentration. Milli-Q natural abundance ($H_2{}^{16}O$) water was obtained from a Simplicity™ from Millipore (Milan, Italy). An ultrasonic bath, model Transsonic TI-H-5, from Elma (Singen, Germany) with control of temperature and amplitude was used to speed up the gel washing, the protein reduction and the protein alkylation steps, and a sonoreactor model UTR200, from Dr. Hielsher (Teltow, Switzerland), was used to accelerated the enzymatic digestion step. All materials were used without further purification. α-Cyano-4-hydroxycinnamic acid, α-CHCA, *puriss* for MALDIfrom Fluka (Buchs, Switzerland) was used as MALDI matrix. ProteoMass Peptide MALDI-MS Calibration Kit (MSCAL2) from Sigma was used as mass calibration standard for MALDI-TOF-MS.

2.2. Standards and reagents

Reduction and alkylation were carried out, respectively, with D,L-dithiothreitol (DTT, 99%) and iodoacetamide (IAA) from Sigma. The following reagents were used during sample digestion: ammonium bicarbonate buffer (AmBic, pH 8.5, ≥99.5%) and formic acid (FA, ∼98%) from Fluka (Buchs, Switzerland); and labeling: ammonium acetate (>99.0%) from Fluka, calcium chloride (∼97%, anhydrous) from Sigma, Mag-Trypsin from Clontech (USA) and $H_2{}^{18}O$ (97 atom%) from ISOTEC™ (Miamisburg, USA). Trifluoroacetic acid (TFA, 99%) was obtained from Riedel-de Haäen (Seelze, Germany).

2.3. In-gel protein digestion

Ultrasonic in-gel enzymatic digestion was done according to the ultrafast proteolytic digestion protocol previously developed in our laboratory [6,7]. Protein bands were manually excised from the gel and placed in safe-lock tubes of 0.5 mL. Gel pieces were washed, first with AmBic 25 mM/acetonitrile (100 μL) and then with acetonitrile (100 μL), in an ultrasonic bath operating at 35 kHz (60% amplitude) for 5 min for each step. Then, the gel pieces were dried in a vacuum concentrator centrifuge for 5 min. Protein reduction and alkylation steps were done as follows: disulfide bonds from cysteine residues were reduced with DTT in an ultrasonic bath operating at 35 kHz (60% amplitude) for 5 min at room temperature, and then, the reduced cysteines were blocked with IAA in an ultrasonic bath operating at 35 kHz (60% amplitude) for 5 min at room temperature. After reduction and alkylation steps, the gel was submitted again to the washing procedure in the same way as described above, followed by another dry step of 10 min. Afterward, the dried gel pieces were incubated with trypsin (375 ng in 25 μL) in an ice bath for 60 min to rehydrate the gel and to allow enzyme penetration into it. Subsequently, in-gel protein digestion was performed in a sonoreactor operating at 50% amplitude for 4 min. Next, trypsin activity was stopped by the addition of 20 μL of formic acid 5% (v/v).

2.4. ^{18}O-labeling: the decoupled procedure

For the ^{18}O-labeling, the digested peptides were reconstituted with 10 μL of 25 mM calcium chloride and 10 μL of (acetonitrile 20% v/v + 50 mM ammonium acetate, pH 6.75). Then the samples were vacuum dried again, and after evaporation the dried samples were reconstituted in 5 μL of natural abundance water or 97% ^{18}O-enriched water and 5 μL of a 5% suspension of Mag-Trypsin in $H_2{}^{16}O$ or $H_2{}^{18}O$ were added. The digested peptides were labeled during 15 min of vortexing and centrifugation and finally trypsin were removed by a magnetic separation. A detailed explanation of this procedure can be found elsewhere [8].

2.5. Inverse ^{18}O-labeling of peptides

Proteins were separated by 1D-PAGE and then submitted to the protocols described in Sections 2.3 and 2.4 and then the inverse ^{18}O-labeling protocol as described by Wang et al. [4] was then used.

2.6. Quantification of peptides

Quantification of peptides through ^{18}O was done with the mathematical algorithm for deconvolution described by Yao et al. [8] Eq. (1). Reduction of the spectra to a centroided plot was done using the centroiding option function of the Data Explorer™ software (version 4.0) from Applied Biosystems. This function is an advanced peak filtering method that improve mass spectral data quality and reduce data file size. Profile data, in which many points are used to delineate a mass spectral peak, is converted into mass-centroided data by a data compression algorithm. The centroided mass peak is located at the weighted center of mass of the profile peak. The normalized area of the peak provides the mass intensity data.

$$\left(\frac{^{16}O}{^{18}O}\right) = \frac{I_0}{I_4 - (M_4/M_0)I_0 + (1-(M_2/M_0))I_2 - (1-(M_2/M_0))(M_2/M_0)I_0} \quad (1)$$

where M_0, M_2 and M_4 correspond to the theoretical relative intensities of the monoisotopic peak and the monoisotopic peaks with masses 2 and 4 Da higher, respectively; and I_0, I_2 and I_4 are the measured relative intensities of the first, the third and the fifth peaks in the isotopic cluster.

2.7. Case study

To explain how to work with the DPD program we have follow a real example based in the standard protein carbonic anhydrase.

2.8. MALDI analysis

Prior to MALDI analysis, the sample was mixed with the matrix solution. α-CHCA matrix was used throughout this work and was prepared as follows: 10 mg of α-CHCA was dissolved in 1 mL of Milli-Q water/acetonitrile/TFA (1 mL/1 mL/2 μL). Then, 4 μL of the aforementioned matrix solution was mixed with 4 μL of sample and the mixture was shaken in a vortex for 30 s. One microliter of each sample was hand-spotted on a well of a MALDI sample

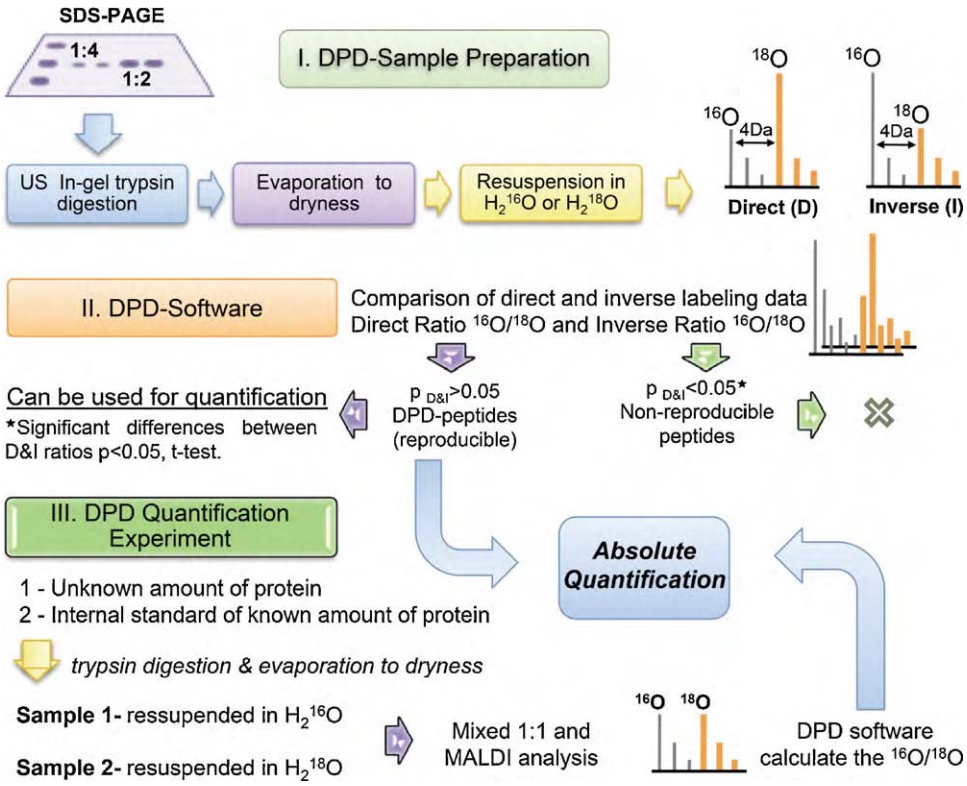
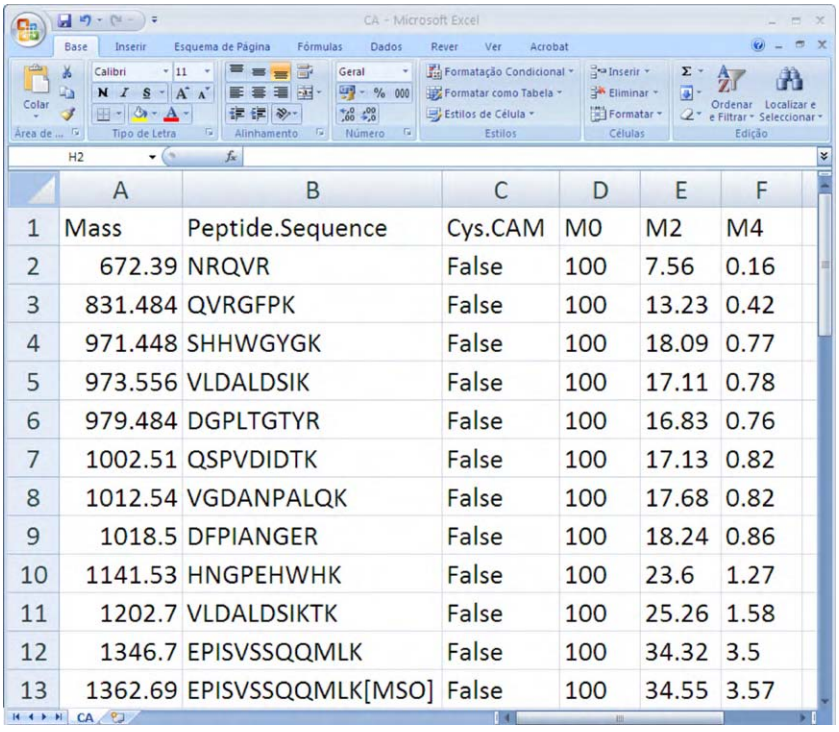


Fig. 1. Experimental workflow for the identification of DPD (decision peptide-driven) peptides and subsequent protein quantification.

plate and was allowed to dry. A MALDI system model Voyager DE-PRO Biospectrometry Workstation equipped with a nitrogen laser radiating at 337 nm from Applied Biosystems (Foster City, CA) was used to acquire the PMFs. Measurements were done in the reflector positive ion mode, with a 20 kV accelerating voltage, 75.1% grid voltage, 0.002% guide wire and a delay time of 100 ns. Two close external calibrations were performed with the monoisotopic peaks of the bradykinin, angiotensin II, P14R, and ACTH peptide fragments (*m*/*z*: 757.3997, 1046.5423, 1533.8582, and 2465.1989, respectively). Monoisotopic peaks were manually selected from



	A	B	C	D	E	F
1	Mass	Peptide.Sequence	Cys.CAM	M0	M2	M4
2	672.39	NRQVR	False	100	7.56	0.16
3	831.484	QVRGFPK	False	100	13.23	0.42
4	971.448	SHHWGYGK	False	100	18.09	0.77
5	973.556	VLDALDSIK	False	100	17.11	0.78
6	979.484	DGPLTGTYR	False	100	16.83	0.76
7	1002.51	QSPVDIDTK	False	100	17.13	0.82
8	1012.54	VGDANPALQK	False	100	17.68	0.82
9	1018.5	DFPIANGER	False	100	18.24	0.86
10	1141.53	HNGPEHWHK	False	100	23.6	1.27
11	1202.7	VLDALDSIKTK	False	100	25.26	1.58
12	1346.7	EPISVSSQQMLK	False	100	34.32	3.5
13	1362.69	EPISVSSQQMLK[MSO]	False	100	34.55	3.57

Fig. 2. In silico file – excel CSV – containing peptide masses, the theoretical peptide sequence assigned to each mass, carbamide methylations present (yes-true or no-false) and the isotopic mass distribution for the protein carbonic anhydrase.

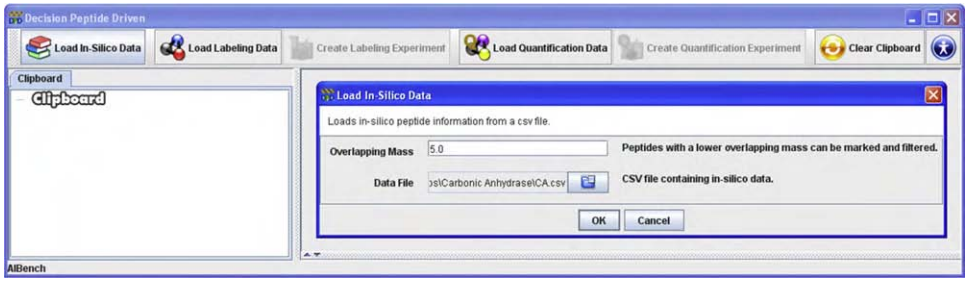


Fig. 3. By clicking in "load in-silico data" the data created as described in Sections 3.1.1 and 3.1.2 is introduced in the form of a CSV excel sheet.

each of the spectra obtained. Mass spectral analysis for each sample was based on the average of 500 laser shots.

2.9. Software distribution

Current versions of the software and their supporting user manuals are freely available for downloaded and use, without restriction, via the internet at http://sing.ei.uvigo.es/DPD. This program operates on excel comma-separated values, CSV, files with centroid mass and relative intensity data extracted from the Data Explorer™. The program was developed based on previous work related to cancer diagnosis [9] and on the suggestions given by the Bioscope group.

2.10. Decision peptide-driven experimental workflow

A schematic diagram illustrating the sequential steps of the sample treatment workflow is presented in Fig. 1. In brief, 1 and 2 μg, four replicates each, are loaded and separated by 1D-SDS-PAGE. The bands are then excised from the gel and the proteins are in-gel tryptic digested with the aid of ultrasonic energy as reported by Galesio et al. [6]. The pool of peptides thus obtained is then dried and finally, reconstituted in normal water or in 97% ^{18}O-water [7,10–14]. Following the pipeline of Fig. 1, after protein separation, in-gel protein digestion and peptide labeling, the next step is to perform the so-called inverse labeling protocol [4]. With this procedure two converse labeling experiments are performed in parallel as follows. In the "direct" labeling, the sample is reconstituted in normal water whilst its counterpart of higher amount (i.e. 1:2) is reconstituted in ^{18}O-water as described in Section 2.4. In the "inverse" method the labeling is done conversely. Finally, an equal sample volume of non-labeled and its labeled counterpart are mixed and analysed through MALDI-TOF-MS. The ^{16}O/^{18}O peak ratios (MALDI relative peak intensity) are used then in the final step of the workflow, as it is shown in Fig. 1. In this step, the software DPD (Decision Peptide Driven) is used to find out which peptides are adequate for protein quantification within a given accuracy. In brief, this software compares the labeled to unlabeled ratios of the same peptides obtained

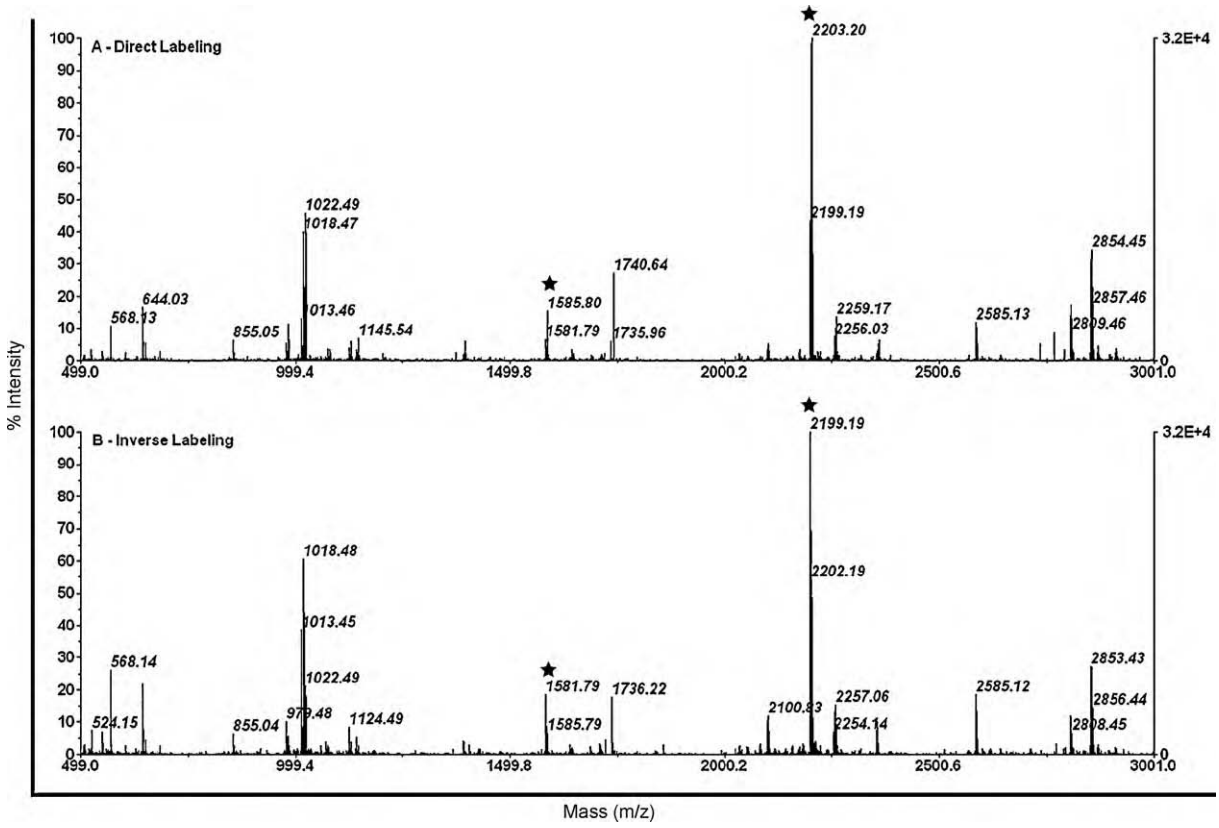


Fig. 4. MALDI spectra of carbonic anhydrase obtained for the direct (1 μg of unlabeled protein and 2 μg of ^{18}O-labeled protein, samples were mixed in a 1:1 ratio before MALDI analysis) and inverse labeling (2 μg of ^{18}O-labeled protein and 1 μg of unlabeled protein, samples were mixed in a 1:1 ratio before MALDI analysis). The symbol ("⋆") denotes the peptides assigned by the DPD software.

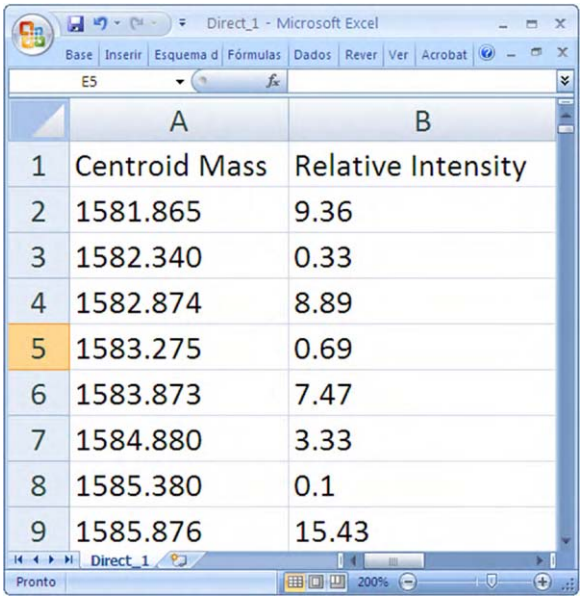


Fig. 5. CSV excel sheet in which one column contents the peak mass values whilst other contents their corresponding MALDI intensities.

in the "direct" and "inverse" methods. Only those peptides having the "direct" and "inverse" ratios within a given p significance level (t-test) are selected for quantification.

3. Results and discussion

The DPD software is explained in detail in the following sections. The data which is needed to enter in each step of the software, the reasons why, and the information which is obtained once each step is completed is described below. To run the program, real data is provided in supplementary material section I. This example is based in the quantification of carbonic anhydrase.

3.1. Preparing theoretical data

3.1.1. How to obtain the in silico protein digestion

After 1D-gel electrophoresis separation, the protein is in-gel digested with the aid of an enzyme, generally trypsin. This process will generate a pool of peptides which can be previously known, because the enzymes used to digest proteins do the cleavage always in the same residues of the peptide chain. For instance, the enzyme trypsin cleavages the proteins in the amino acids arginine and lysine, if they are not followed by a proline residue. In other words, if the protein to be quantified and the enzyme to be used are known, the pool of peptides expected can be obtained in advance. Nowadays there are powerful software tools that can provide the above-mentioned theoretical pool of peptides (http://www.expasy.org/sprot/). When the DPD program is used, the first step consists in the introduction of this theoretical pool of peptides. This is necessary because the program will compare the masses of those theoretical peptides with the masses of the peptides obtained using MALDI. This comparison has the goal to assign the experimental masses obtained with the MALDI with their corresponding (theoretical) peptides. A step-by-step description of how to obtain the in silico protein digestion is provided in supplementary material-part II.

3.1.2. How to obtain the isotopic mass distribution of the peptides

The in silico digestion of the protein provides a list of theoretical peptides. The IMD, isotopic mass distribution, of those peptides, M_0, M_2, M_4, is required for the subsequent calculation of the ratios between the ^{18}O-labeled and the non-labeled peptides in following parts of the workflow, as it is shown in Eq. (1) (see above). For each peptide the IMD is calculated in an automated mode in the following webpage http://prospector.ucsf.edu/cgi-bin/msform.cgi?form=msisotope. A step-by-step description of how to obtain the isotopic mass distribution of the peptides is provided in supplementary material-part III.

3.2. Finding out the DPD peptides

3.2.1. First step: loading the in silico data

With the data obtained in Sections 3.1.1 and 3.1.2 a CSV excel file is created containing six columns. The first column has the theoretical masses of the peptides, the second column contains the theoretical peptide aminoacid sequences, the third column contains the possibility of carboxyamidomethylations on cysteine due to alkylation. The fourth, fifth and sixth columns, contains the IMD of the M_0, M_2 (M_0 + 2 Da) and M_4 (M_0 + 4 Da) masses respectively obtained for each peptide. Fig. 2 shows an example of an in silico file from carbonic anhydrase. When the DPD program is started the interface shown in Fig. 3 appears in the screen. By clicking in the "*load in-silico data*" button we are asked to introduce the CSV excel file containing the in silico information. In our example the

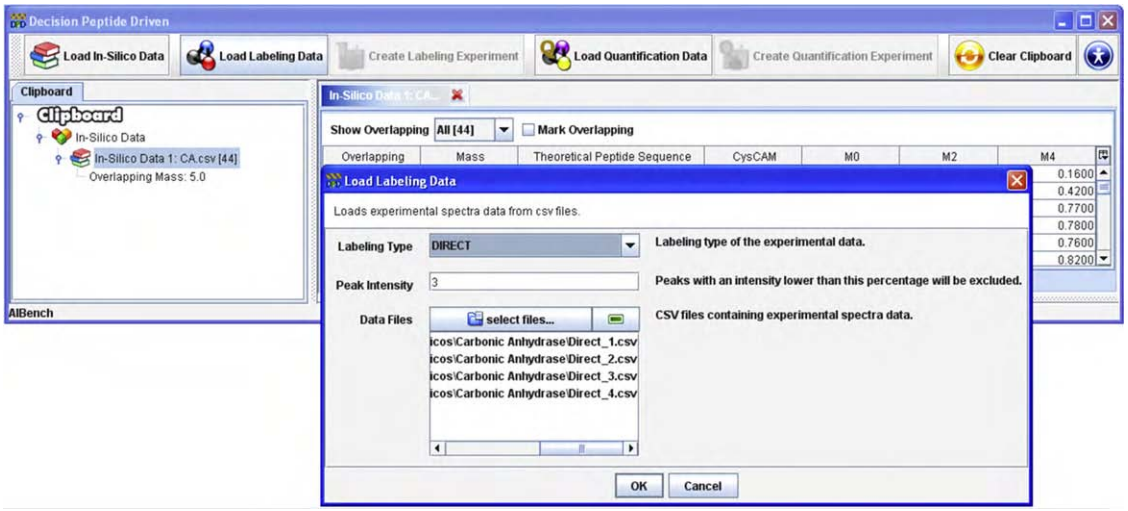


Fig. 6. By clicking in "load labeling data" the data created as described in Section 3.3 for the direct (n = 4) and inverse (n = 4) experiments are introduced in the form of a CSV excel sheet. The user is also asked to introduce a value for the minimum peak intensity for which a peak will be considered as different from noise.

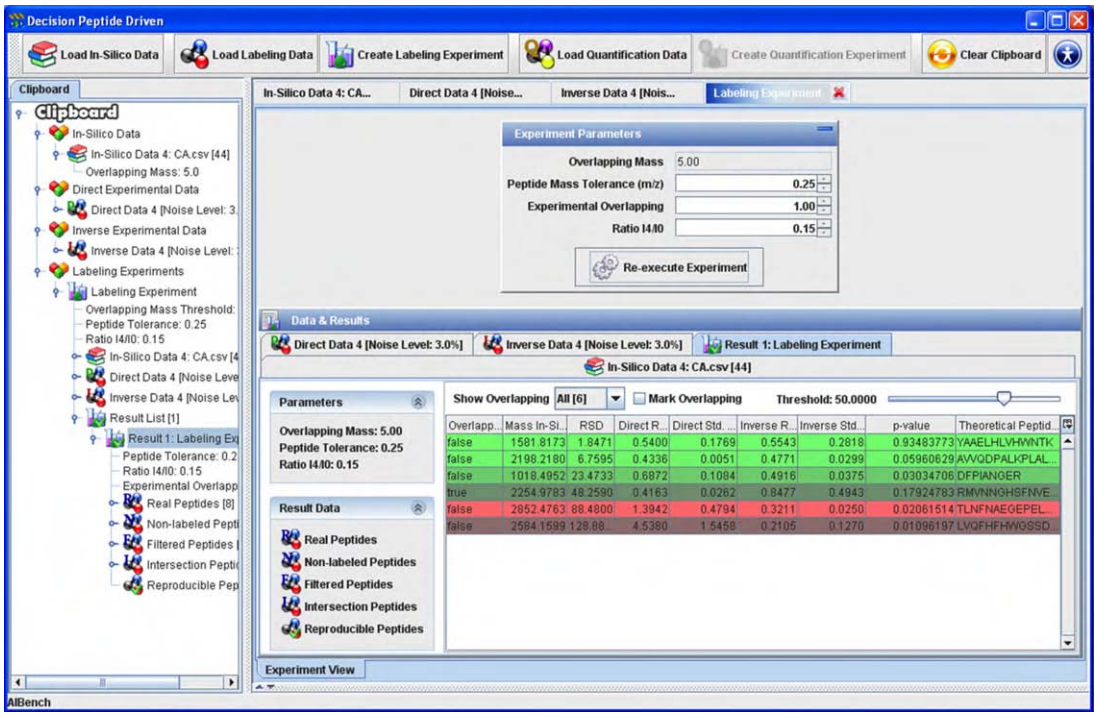


Fig. 7. This figure shows the result of the execution which is a report containing the values of the input parameters and the generated output data including: (i) a list of peptides identified in the raw data input files, (ii) a list of non-labeled peptides, (iii) a list of filtered peptides, (iv) a list of intersection peptides and finally (v) a list of reproducible peptides, those that can be used for quantification.

file to input is the one named as "in silico carbonic anhydrase" given in supplementary material, and prepared as described in Sections 3.1.1 and 3.1.2. In addition, the user must introduce an "overlapping mass" value that indicates the mass difference considered critical between two peptides whose isotopic mass distribution can potentially be overlapped, in our case 5 Da is the value chosen. This concept can be explained with the simplest case of overlapping as follows: the in silico digestion of carbonic anhydrase predicts the occurrence of YGDFGTAAQQPDGLAVVGVFLK (2253.16 m/z) and RMVNNGHSFNVEYDDSQDK (2254.98 m/z), their isotopic mass distribution are overlapped, after isotopic labeling, if both are present in the experimental data. Nevertheless, could also happen that the peak 2253.16 m/z is present in the experimental data but not the peak 2254.98 m/z or vice versa. In this case, the peptide virtually could be used for quantification. Therefore those peaks will be assigned as potential overlapped ("true") by the program and they must be checked in the spectrum to assess whether overlapping occurs or not.

3.2.2. *Second step: loading experimental data*

It was explained in the experimental section that two converse experiments are done to identify the peptides that have paralleled loses in different sets of experiments. Of each set of direct ($n = 4$) and inverse ($n = 4$) labeling experiments, MALDI spectra are obtained, showing the typical pattern of labeled and non-labeled peptides (see Fig. 4). Those spectra are converted in a CSV excel file in which one column contents the peak mass values whilst other contents their corresponding intensities, as shown in Fig. 5. By clicking in "load labeling data", see Fig. 6, the program asks for the introduction of the files corresponding to the direct labeling. In our example, those files are named in supplementary material as direct 1, direct 2, direct 3 and direct 4. Once the direct files have been loaded, then the files corresponding to the inverse labeling are also loaded. In supplementary material the corresponding files are named as inverse 1, inverse 2, inverse 3 and inverse 4. The DPD program has an algorithm that allows the user to choose the minimum peak intensity from which the masses will be considered different of

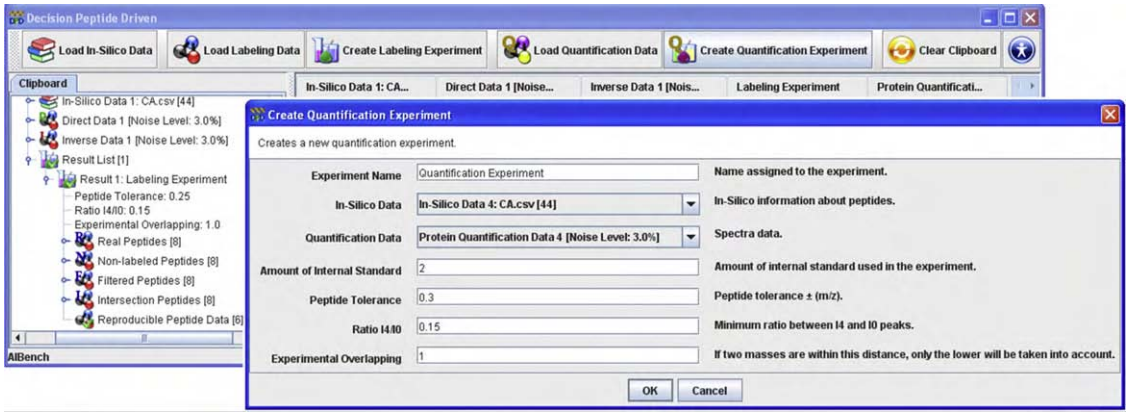


Fig. 8. DPD application showing a ready-to-run quantification experiment where the user-defined parameters can be specified.

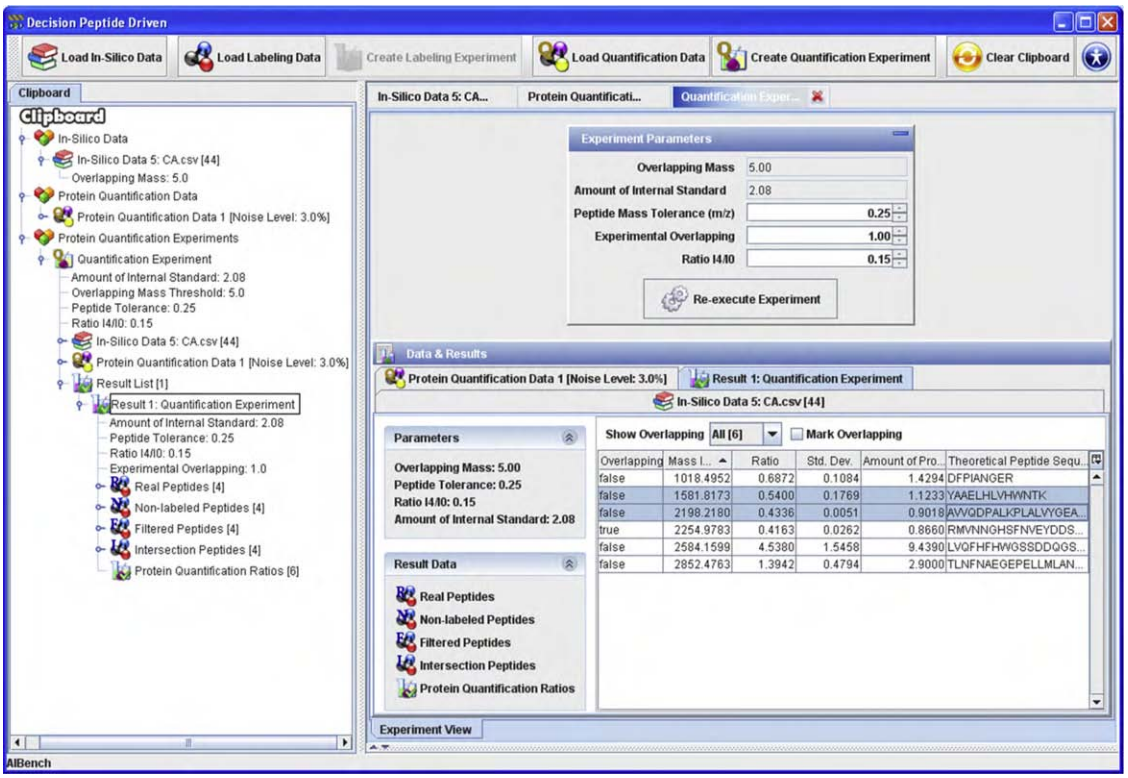


Fig. 9. Protein quantification results computed by the DPD application once an experiment is executed.

instrumental noise. In other words, it is possible to select the minimum intensity from which a peak will be considered as belonging to a peptide. This value will depend on the quality of the MALDI spectra. In our example the recommended value is 3%, which represents the percentage of the maximal relative peak intensity.

3.2.3. Third step: creating a label experiment

At this stage, the list of in silico peptide masses and the list of experimental m/z values (corresponding most likely to peptides) of the direct and inverse experiments have been loaded in the program. In the next step, by clicking in "create labeling experiment" a dialog box will appear and the user is asked to define an error margin via the "peptide tolerance" box. This parameter is instrument dependent and is directly linked to the accuracy that it is expected to obtain in the MALDI system. In our conditions, a typical value to be assigned to this parameter is 0.25 (corresponding to an accuracy of 0.25 Da). Once a experimental peptide mass is matched with its theoretical value, M_0, the system checks if the spectrum also contains that mass plus 2 Da, M_2, and 4 Da, M_4, corresponding to one and two ^{18}O incorporations, respectively. The ratio between the intensities of M_0 and M_4 (ratio I_4/I_0 in the dialog box) is used to discharge natural occurring peptides. This is because the peptides that we are comparing correspond to mixtures of labeled and non-labeled peptides, the intensity ratios between M_0 and M_4 are different of the natural occurring M_0 and M_4 ratios. A typical value assigned to this parameter is 0.15. Every time that M_0, M_2 and M_4 are found within the given values of 0.25 for peptide tolerance and 0.15 for ratio M_0/M_4, the peptide is selected as a candidate to be considered a reproducible peptide. Otherwise, the mass is discharged.

3.2.4. Fourth step: intersect peptides

At this stage, the lists of masses from the direct and inverse experiments include only the peptide masses that (i) have matched the corresponding in silico mass, (ii) that have the masses corresponding to the 2 and 4 Da shift caused by the single and double ^{18}O incorporation, within a given peptide mass tolerance of 0.25 Da and a experimental overlapping of 1.00 Da and (iii) that have a I_4/I_0 ratio over a given threshold, 0.15 in this case. Now the direct and inverse lists are compared to select the common masses, this is, the masses corresponding to peptides that are found in both direct and inverse experiments.

In addition, a labeling ratio ($^{16}O/^{18}O$) is calculated as explained in Eq. (1) for these peptides taking into account the isotope mass distributions M_0, M_2 and M_4, and the corresponding intensities measured in the experimental data I_0, I_2 and I_4, respectively. By following this criterion, DPD software generates a list of common peptides for both direct and inverse samples, along with their corresponding non-label to label ratios. The ratio must reflect the relation between the amount of label and non-label protein as established at the beginning of the experiment. In our example, the amount of protein labeled was twice the amount of protein non-labeled, therefore the direct ratio is 0.5 (1/2), whilst the inverse ratio is 2 (2/1). To facilitate the comparison the DPD program shows the inverse ratio as (inverse ratio)$^{-1}$, thus the expected values in our case are 0.5 for both ratios. In addition, the average amount of protein ratios is given with their corresponding relative standard deviations, RSDs. Finally, to find the peptides that have similar ratios in the direct and in the inverse method, it is needed to click in the "execute experiment" tool bar. Now the program asks which threshold level of RSD is required. The program compares the medium values and provides a relative standard deviation, $RSD_{D\&I}$, that arise the difference in % between the medium values. Thus, the peptides with direct and inverse values within a chosen difference (for instance less than 10%) can be easily selected for quantification, as showed in Fig. 7. In addition, the program also provides a p value. This p is obtained by comparing the direct and inverse ratios for each peptide. If $p > 0.05$, then both values can be considered

Table 1
Peptides assigned by the DPD software as candidates for quantification of carbonic anhydrase. Direct labeling: 1.04 μg of unlabeled protein and 2.08 μg ^{18}O-labeled protein, samples were mixed in a 1:1 ratio before MALDI analysis. The expected direct ratio is 0.5 corresponding to the unlabeled/labeled ratio (1.04/2.08). Inverse labeling: 1.04 μg of ^{18}O-labeled protein and 2.08 μg of unlabeled protein, samples were mixed in a 1:1 ratio before MALDI analysis. Software parameters: peak intensity −3; Peptide mass tolerance ±0.25 Da; experimental overlapping −1; ratio I_4/I_0 −0.15.

Mass in silico	$RSD_{D\&I}$ (%)	Direct ratio (0.5)[a]	Inverse ratio (0.5)[a]	$p_{D\&I}$ value	Peptide sequence – in silico	$RSD_{D\&I}$ (%)	Direct – μg of protein (1.04)[a]	Inverse – μg of protein (2.08)[a]
1581.82	0.8	0.54 ± 0.18	0.55 ± 0.28	0.97	YAAELHLVHWNTK	2–10%	1.0 ± 0.2	2.0 ± 0.2
2198.22	5.8	0.43 ± 0.01	0.48 ± 0.03	0.18	AVVQDPALKPLALVYGEATSR			
1018.50	23.6	0.69 ± 0.11	0.49 ± 0.04	0.03	DFPIANGER	20–40%*	1.1 ± 0.4	1.7 ± 0.6
2098.88	48.3	0.42 ± 0.03	0.9 ± 0.5	0.18	MVNNGHSFNVEYDDSQDK			
2852.48	91.1	1.4 ± 0.5	0.3 ± 0.1	0.02	TLNFNAEGEPELLMLANWRPAQPLK	>90%*	6 ± 5	4.2 ± 1.1
2584.16	128.9	4.5 ± 1.6	0.2 ± 0.1	0.01	LVQFHFHWGSSDDQGSEHTVDR			

[a] Expected values.
* Significant differences between D&I ratios ($p < 0.05$, t-test).

statistically non-different, and then the peptide can be considered valuable for protein quantification,

In addition, the software allows the user to change the experiment parameters in order to perform multiple analyses on the same data. Every time the user selects the "Execute Experiment" button, a new result is added to the experiment containing all the information generated during the process. In this manner the variables "peptide mass tolerance", "experimental overlapping" and ratio I_4/I_0 can be changed as desired. Before starting the execution, the user must specify the RSD threshold parameter used to highlight those peptides that are useful (under the statistical threshold) and those that are invalid (over the threshold). In our example the peptides recovered with a RSD between direct and inverse methods below 10% correspond to peptides YAAELHLVHWNTK and AVVQDPALKPLALVYGEATSR. If the RSD chosen is changed to 50% the peptides are now YAAELHLVHWNTK, AVVQDPALKPLALVYGEATSR, DFPIANGER and RMVNNGHSFNVEYDDSQDK. It is noteworthy that only two peptides can be used to accurately (below 10%) quantify the protein.

3.3. Quantification experiment

The user can proceed to load the data to be used for protein quantification through the "Load Quantification Data" toolbar button (please be sure that the file containing the in silico data of the protein has been introduced). The *noise level* (as peak intensity) parameter and the files containing the spectra are introduced in this step. Following our example, the recommended peak intensity value is 3 and the files to be introduced are named as Direct_1, 2, 3 and 4 corresponding to MALDI data of four independent samples of carbonic anhydrase provided in supplementary material.

Now, by clicking in "create quantification experiment" the *amount of internal standard* used for quantification, *peptide tolerance*, *ratio I4/I0* and the *experimental overlapping* are introduced as displayed in Fig. 8. In our case, the following parameters were introduced: peptide tolerance, 0.25; ratio I_4/I_0, 0.15; amount of internal standard 2.08 μg. The detailed description of how this sample was treated is explained in the experimental section. Once the quantification experiment was done, a list of peptides and the corresponding calculated amount of protein are generated by clicking in "execute experiment".

As in the labeling experiment, the application allows the user to change the experiment parameters in order to perform multiple analyses using the same data. Every time the user selects the "Execute Experiment" button, a new result is added to the experiment containing all the information generated during the process. The result of the execution is a report containing the values of the input parameters and the generated output data including: (i) a list of peptides identified in the raw data input files, (ii) a list of non-labeled peptides, (iii) a list of filtered peptides, (iv) a list of intersected peptides and (v) the protein quantification ratios. Fig. 9 shows the results of the experiment in the working zone panel.

Now it is necessary only to check the peptides that have been previously identified as the DPD peptides, this is, the peptides that can be used for quantification. In our example using peptides YAAELHLVHWNTK (1581.82 m/z) and, AVVQDPALKPLALVYGEATSR (2198.21 m/z) the amounts of protein calculated are 1.12 and 0.90 μg respectively, corresponding to the amount of protein loaded into the gel. Table 1 shows the amount of protein found for the different *peptides* selected in this work.

4. Merits and limitations

The present program has been developed specifically for proteins separated by 1D-gel electrophoresis. However the program can be potentially used for proteins separated through 2D-gel electrophoresis or for proteins separated by HPLC. Another merit is the possibility to adapt the program to other type of labeling. The main limitation of this program is that it has been developed for MALDI ionization systems.

5. Conclusions

We have developed a friendly software to help in an automated mode to identify those peptides that have paralleled loses through a typical proteomic workflow. The use of such peptides allow robust and accurate quantification of proteins using 1D-gel electrophoresis an matrix assisted laser desorption ionization time of flight mass spectrometry. Those peptides have been named as "decision peptide driven", DPD peptides. The software presented in this work allows for the identification of DPD peptides and is based in a series of steps entailing different algorithms that perform in an automated mode a peptide differential analysis to extract and to identify the number of peptides that remains constant in expression level through different sets of a typical in-gel digestion workflow as the one described in this work. The DPD software saves times, allowing the user to accurate quantify proteins in an automated mode, overcoming the long time needed when the treatment of data is done manually. In addition the DPD software has a wizard easy to follow for its installation. Furthermore, the interface has been done in an easy-to-follow mode, and therefore the skills required for any potential operator are reduced to know how to apply the sample treatment procedure described in supplementary material section.

The installation wizard is available from the DPD web site as an executable file that depends on the final user operating system: Windows, Linux or MAC. By executing the setup file, the installation wizard will be automatically launched. If the user does not have the required Java Runtime Environment (JRE) installed in the computer, the installation wizard will first install this component, and then it will continue with the DPD installation. The user has to simply

follow the instructions on the screen to successfully complete the installation.

Acknowledgements

H.M. Santos acknowledges the doctoral grant SRFH/BD/38509/2007 provided by FCT (Fundação para a Ciência e a Tecnologia - Portugal). This work was partially funded by the project *Research on Translational Bioinformatics* (08VIB6) from University of Vigo. D. Glez-Peña was supported by a María Barbeito contract from Xunta de Galicia. M. Reboiro-Jato was supported by a PhD grant from Xunta de Galicia. Dr. José-Luis Capelo-Martínez is grateful to the Xunta de Galicia (Spain) for the program Isidro Parga Pondal and the University of Vigo for financial support under projects InOu-Univ.Vigo 2009-K915 and K914. Xunta de Galicia (Spain) is also acknowledged for financial support under project 09CSA043383PR-2009. The research findings here reported are protected under patent pending PCT/IB2006/052314 and PT 103 303.

Contributions: H.M. Santos conceived the idea and developed the experiments under the guidance of J.L. Capelo. J.L. Capelo obtained financial support to develop the experiments. H.M. Santos and J.L. Capelo, interpreted experimental data. J.D. Nunes-Miranda developed part of the experiments under the guidance of H.M. Santos. F. Fdez-Riverola and R. Carbalho obtained financial support for the software development. H.M. Santos, D. Glez-Peña, and M. Reboiro-Jato, developed the DPD software, under the supervision of F. Fdez-Riverola, H.M. Santos, J.L. Capelo and R. Carbalho. All authors contributed equally to write the manuscript.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.talanta.2010.07.007.

References

[1] J. Havlis, A. Shevchenko, Anal. Chem. 76 (2004) 3029.
[2] I.I. Stewart, T. Thomson, D. Figeys, Rapid Commun. Mass Spectrom. 15 (2001) 2456.
[3] K. Speicher, O. Kolbas, S. Harper, D. Speicher, J. Biomol. Tech. 11 (2000) 74.
[4] Y.K. Wang, Z. Ma, D.F. Quinn, W.F. Emil, Anal. Chem. 72 (2001) 3742.
[5] D. Glez-Peña, M. Reboiro-Jato, P. Maia, M. Rocha, F. Díaz, F. Fdez-Riverola, Comput. Methods Programs Biomed. 98 (2010) 191–203.
[6] M. Galesio, D.V. Vieira, R. Rial-Otero, C. Lodeiro, I. Moura, J.L. Capelo, J. Proteome Res. 7 (2008) 2097.
[7] R. Carreira, C. Lodeiro, M.S. Diniz, I. Moura, J.L. Capelo, Proteomics 9 (2009) 1.
[8] X. Yao, A. Freas, J. Ramirez, P.A. Demirev, C. Fenselau, Anal. Chem. 73 (2001) 2836.
[9] D. Glez-Peña, F. Díaz, J.M. Hernández, J.M. Corchado, F. Fdez-Riverola, BMC Bioinformatics 10 (2009) 187.
[10] C.A. White, O. Nicomedus, A. Emili, J. Proteome Res. 7 (2009) 3653.
[11] X.D. Yao, C. Afonso, C. Fenselau, J. Proteome Res. 2 (2003) 147.
[12] J.L. Capelo, R.J. Carreira, L. Fernandes, C. Lodeiro, H.M. Santos, J. Simal, Talanta 80 (2010) 1476.
[13] R.J. Carreira, R. Rial-Otero, D. López-Ferrer, C. Lodeiro, J.L. Capelo, Talanta 76 (2008) 400.
[14] H.M. Santos, C. Mota, C. Lodeiro, I. Moura, I. Isaac, J.L. Capelo, Talanta 77 (2008) 870.