The data presented here are related to the research paper entitled “Study of a Novel Agent for TCA Precipitated Proteins Washing - Comprehensive Insights into the Role of Ethanol/HCl on Molten Globule State by Multi-Spectroscopic Analyses” (Eddhif et al., submitted for publication) [1]. The suitability of ethanol/HCl for the washing of TCA-precipitated proteins was first investigated on standard solution of HSA, cellulase, ribonuclease and lysozyme. Recoveries were assessed by one-dimensional gel electrophoresis, Bradford assays and UPLC-HRMS. The mechanistic that triggers protein conformational changes at each purification stage was then investigated by Raman spectroscopy and spectrofluorometry.
Finally, the efficiency of the method was evaluated on three different complex samples (mouse liver, river biofilm, loamy soil surface). Proteins profiling was assessed by gel electrophoresis and by UPLC-HRMS.

© 2018 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Specifications Table

| Subject area                      | Chemistry                                      |
|----------------------------------|-----------------------------------------------|
| More specific subject area       | Proteomics, protein purification, protein precipitation, trichloroacetic acid |
| Type of data                     | Tables, Figures                               |
| How data was acquired            | Raman (LabRAM HR800UV confocal microspectrometer, Horiba Jobin Yvon, Kyoto, Japan) |
|                                  | Bradford assay (DC Protein Assay, Biorad)     |
|                                  | Electrophoresis (ImageJ software)              |
|                                  | UPLC-HRMS (Accela LC pumps, Q-Exactive Hybrid Quadrupole-Orbitrap mass spectrometer equipped of an ESI source, Thermo Fisher Scientific, Waltham, MA, USA) |
|                                  | MASCOT search engine (Matrix Science, London, UK; version 2.6.0) and Skyline software (MacCoss Lab, Washington, US; version 3.7.0.10940) |
|                                  | ProteomeXchange Consortium with identifier PXD008110 |
| Data format                      | Raw, analyzed and processed data               |

| Experimental factors             | Proteins extraction was performed on 500 mg of soil, 10 mg of biofilm and 15 mg of mouse liver as starting material according to protocols of Choure et al. [2], Huang et al. [3] and Song et al. [4] respectively. |
| Experimental features            | Proteins were precipitated with 25% (w/v) trichloroacetic acid (TCA). The washing of protein pellet was performed with three different agents (acetone, ethanol, or ethanol/HCl). The mixture was vortexed and kept at −20 °C for 1 h, centrifuged at 16,600 g for 15 min at 4 °C. The resulting pellets were dried in a SpeedVac concentrator, solubilized in a 50 mM of ammonium bicarbonate buffer containing 10 mM of Tris. Proteins were subjected to trypsin digestion for 24 h at 37 °C. Digestion was stopped with formic acid before gel, bradford and mass analysis. |
| Data source location             | Poitiers, France                              |
| Data accessibility               | data are with this article                    |

Value of the data

- Data show a comprehensive evaluation of protein conformational changes throughout TCA precipitation and one single step purification with various solvents.
- Data highlight the efficiency of ethanol/HCl purification for TCA-precipitated proteins.
- Ethanol/HCl represents a quick and inexpensive purification agent for proteomics studies.
- Presence and variability of proteins are potential values to determine which purification method must be used for proteomics investigation.
1. Data

TCA precipitation is one of the most common and robust technique required for protein analyses [5–7]. However it leads to molten globule states which hamper the solubilization of proteins in aqueous buffers for mass spectrometry analysis.

1.1. Comparison of washing agents on standard solutions

A standard solution of HSA, cellulase (exoglucanases and endoglucanases mixture), lysozyme and ribonuclease A, 35 µg mL\(^{-1}\) each, was prepared in high purified water. Proteins were precipitated with 25% (w/v) trichloroacetic acid (TCA) (final concentration). The clean-up of protein pellet was performed following three different approaches: ethanol/HCl (1.25 M; 3.8%), acetone/HCl (0.06 M; 0.2%); acetone/HCl (1.25 M; 3.8%) (Fig. 1).

1.2. Extraction and purification of endogenous proteins from complex sample matrices

See Fig. 2.

1.3. Effects of successive ethanol/HCl washings on proteins recoveries

10 mg of biofilm samples were spiked with the standard solution of HSA, exoglucanase 1 from the mix of cellulase, lysozyme, and ribonuclease A (Rnase). Proteins final concentration was 1 µg mg\(^{-1}\) of matrix to enable HRMS detection of the proteins after the whole process. The mixture was vortexed and left during 24 h at room temperature to favor proteins adsorption on the matrix. After extraction following the published protocol of Huang et al. [3], protein pellets were subjected to one, two or three ethanol/HCl washing(s).

They were then dissolved in 50 mM of ammonium bicarbonate containing 10 mM of Tris (pH 8.5), diluted in a ratio of 1:3 using the same buffer and subjected to trypsin digestion.

![Fig. 1. Standard proteins quantification by Bradford assay and silver-staining on electrophoresis gel. The thin line bars represent standard deviations at the top of the Bradford histogram. For both methods, histograms were constructed from the mean value of three independent assays.](image)
Experiments were performed in triplicate. Fig. 3 gives the mean protein recoveries following the designed approach (Ethanol/HCl) on biofilm matrix after multiple washing steps.

1.4. Understanding the effect of ethanol/HCl on proteins conformation

1.4.1. Spectrofluorometry

To get insights into the role of ethanol/HCl on proteins solubility, their conformational changes were comprehensively investigated, as an extension of the results reported in Ref. [1]. These measures were performed at each purification stage with two spectroscopic techniques: spectrofluorometry and Raman.

Fig. 2. One-dimensional gel electrophoresis of complex matrices (biofilm, soil and mouse liver) after purification following the designed approach versus published protocols on complex matrices. The gel was stained with silver nitrate.

Fig. 3. Proteins recoveries following the designed approach on biofilm sample. The thin line bars represent standard deviations at the top of each column. Each bar shows mean ± s.e.m. from three independent purification assays. Protein recoveries in Tris buffer were determined by UPLC/HRMS in a full scan mode with a resolution of 70.000 and mass range of 200–3000 m/z.

Experiments were performed in triplicate. Fig. 3 gives the mean protein recoveries following the designed approach (Ethanol/HCl) on biofilm matrix after multiple washing steps.
Figs. 4–6 represent the fluorescence emission spectra of lysozyme, HSA and Rnase after TCA precipitation and washing steps (ethanol/HCl, ethanol or acetone).

1.4.2. Raman microspectroscopy

Raman spectrum for Rnase, is presented in Fig. 7. Spectra and curve fitting of the amide I band of proteins corresponding to lysozyme and HSA are presented in Figs. 5 and 6 in Ref. [1], respectively (Figs. 8–11).

Fig. 4. Emission spectra of lysozyme ($\lambda_{\text{exc}} = 400 \text{ nm}$) at different purification steps. Native lysozyme (grey spectrum); Lysozyme-TCA (orange spectrum); Lysozyme-ethanol/HCl (green spectrum); Lysozyme-ethanol (purple spectrum); Lysozyme-acetone (blue spectrum).

Fig. 5. Emission spectra of HSA ($\lambda_{\text{exc}} = 400 \text{ nm}$) at different purification steps. Native HSA (grey spectrum); HSA-TCA (orange spectrum); HSA-ethanol/HCl (green spectrum); HSA-ethanol (purple spectrum); HSA-acetone (blue spectrum).
The unfolding or aggregation of proteins usually involves some dynamic changes in their secondary structures. These changes are mainly monitored by the analysis of the amide I region (1600–1690 cm\(^{-1}\)) which is assumed to be sensitive to α-helical secondary structures [8].

1.5. Extraction and purification of proteins from complex samples: LC-HRMS analysis

We present processed data of UPLC-HRMS analysis of proteins from different samples (mouse liver, river biofilm, soil) after TCA precipitation and solvent purification. The datasets in XML format
can be used to evaluate ethanol/HCl purification for proteins profiling. Table 1 gives the HRMS features of peptides targeted for the standard proteins after in silico trypsic digestion. Table 2 presents endogenous proteins identified in soil, biofilm and mouse liver samples after purification following either the designed approach or published protocols (Mascot identification). Table 3 presents endogenous proteins detected in the mouse liver sample and quantified through Skyline with corresponding peptides and transitions for PRM. Table 4 presents endogenous proteins detected in the biofilm sample and quantified through Skyline with corresponding peptides and transitions for PRM (Table 5).
Fig. 10. Relative integrated intensities of lysozyme amide I contribution from peak #6 assigned to unordered structures (uo), peak #7 (ordered α helices, ho), peak #8 (unordered α helices and β sheets, hu+sh), and peak #9 (turns, tu) as obtained after profile fitting of amide I region of the Raman spectra (Fig. 5, Ref. [1]). Values on top of each bar correspond to the Raman shift on which the contribution peak was centred at the end of the fitting.

Fig. 11. Relative integrated intensities of HSA amide I contribution from peak #1 assigned to unordered structures (uo), peak #2 (ordered α helices, ho), peak #3 (unordered α helices and β sheets, hu+sh), and peak #4 (turns, tu) as obtained after profile fitting of amide I region of the Raman spectra shown in Fig. 6 [1]. Values on top of each bar correspond to the Raman shift on which the contribution peak was centred at the end of the fitting.

Table 1
HRMS features of peptides targeted for the four standard proteins after in silico tryptic digestion.

| Protein name | Peptide sequence | [M+H]^[1] | [M+2H]^[2] | [M+3H]^[3] | [M+4H]^[4] |
|--------------|-----------------|-----------|-----------|-----------|-----------|
| LYSO-1       | FESNFNTQATNR    | 714.8288  | 476.8883  |           |           |
| LYSO-2       | HGLDNYR         | 874.4166  | 437.7119  | 292.1437  |           |
| RNASE-1      | CKPVNTFHVESLADVQAVCS QK | 839.7457 | 630.0611  |           |           |
| RNASE-2      | HIIACEGNPVYPVHFDASV | 1112.5464 | 742.0334  |           |           |
| RNASE-3      | YPNCAVK        | 458.2051  | 298.1293  |           |           |
| HSA-1        | AVMDDFAAFVEK   | 671.8210  | 448.2164  |           |           |
| HSA-2        | LVAASQAALGL    | 1013.5990 | 693.3200  | 448.2164  |           |
| HSA-3        | YLYEIAR        | 927.4934  | 693.3200  | 448.2164  |           |
| EXO-1        | GSCSTSSGVPAPQINESQSPNA K | 1039.4764 | 693.3200  | 448.2164  |           |
| EXO-2        | YGTGVYCDQCPR   | 732.2876  | 488.5275  |           |           |
| EXO-3        | VTSNIK         | 808.4563  | 404.7282  |           |           |
Table 2
Endogenous proteins identified in soil, biofilm and mouse liver after purification following either the designed approach or the published protocols.

| Sample Location                  | Protein name            | Phylum origin                        | Protein coverage (%) | Scorea | GRAVY | MW (Da)b |
|----------------------------------|-------------------------|--------------------------------------|----------------------|--------|-------|----------|
| Soil Extracellular region        | Endoglucanase EG-II     | Hypocrea jecorina                    | 18                   | 161    | –0.19 | 44883    |
|                                  | Xyloglucanase           | Hypocrea jecorina                    | 1                    | 76     | –0.21 | 87307    |
| Soil Extracellular region        | C-phycocyanin alpha chain | *Synechococcus sp,*                 | 17                   | 181    | –0.11 | 17335    |
| Biofilm Cellular thylakoid membrane; Peripheral membrane protein | C-phycocyanin alpha chain | *Synechocystis sp,*                 | 20                   | 209    | –0.12 | 17756    |
| Chloroplast thylakoid membrane; Peripheral membrane protein | Allophycocyanin alpha chain | *Micrastera diplosiphon*          | 11                   | 76     | –0.14 | 17411    |
| Chloroplast thylakoid membrane; Peripheral membrane protein | R-phycoerythrin beta chain | *Porphyridium purpureum*           | 21                   | 117    | 0.25  | 18884    |
| Chromatophore thylakoid membrane; Stromal side | C-phycocyanin alpha chain | *Microastera diplosiphon*          | 21                   | 138    | 0.21  | 19568    |
| Cellular thylakoid membrane; Peripheral membrane protein | R-phycoerythrin beta chain | *Porphyra haitanensis*              | 23                   | 129    | 0.26  | 18810    |
| Chromatophore thylakoid membrane; Stromal side | C-phycocyanin-1 beta chain | *Micrastera diplosiphon*          | 16                   | 64     | 0.17  | 18080    |
| Cellular thylakoid membrane; Peripheral membrane protein | Allophycocyanin subunit alpha 1 | *Nostoc sp,*                      | 17                   | 99     | –0.09 | 17392    |
| Chromatophore thylakoid membrane; Peripheral membrane protein | C-phycocyanin beta chain | *Aglaothamnion neglectum*         | 11                   | 112    | 0.09  | 18290    |
| Stromal side | Ribulose bisphosphate carboxylase large chain | *Trichodesmium erythraeum* | 5                    | 90     | –0.32 | 53615    |
| Ni | Allophycocyanin alpha chain | *Anabaena cylindrica*              | 6                    | 84     | 0.01  | 17128    |
| Cellular thylakoid membrane; Peripheral membrane protein | C-phycocyanin alpha chain | *Pseudanabaena tenuis*            | 18                   | 144    | –0.24 | 17780    |
| Chromatophore thylakoid membrane; Multi-pass membrane protein | Photosystem II CP47 reaction center protein | *Odontella sinensis*             | 8                    | 117    | 0.08  | 56436    |
| NI Ribulose bisphosphate carboxylase large chain | Cyanothecae sp., 9 6 94 89 −0.27 53531 |
| chloroplast Ribulose bisphosphate carboxylase large chain (Fragment) | Calyptrosphaera sphaeroidea tenuistipitata var. liui, 8 10 111 132 −0.10 54442 |
| chloroplast Ribulose bisphosphate carboxylase large chain | Cylindrothece sp., 6 6 109 108 −0.12 54400 |
| chloroplast thylakoid membrane; Peripheral membrane protein; Stromal side chloroplast Ribulose bisphosphate carboxylase large chain | Cyanidium caldarium, 13 16 94 83 −0.04 17574 |
|chloroplast chloroplast | Cyanidium caldarium, 13 16 94 83 −0.04 17574 |
| chloroplast thylakoid membrane; Peripheral membrane protein; Stromal side chloroplast Ribulose bisphosphate carboxylase small chain | Antithamnion sp., 5 5 72 72 −0.58 16247 |
| chloroplast Chloroplast Ribulose bisphosphate carboxylase large chain (Fragment) | Synechocystis sp., 18 29 71 72 −0.19 11128 |
| chloroplast thylakoid membrane; Peripheral membrane protein; Stromal side chloroplast R-phycoerythrin beta chain | Aglaothamnion neglectum Haptofila hirta, 9 10 141 139 −0.11 51098 |
| chloroplast Chloroplast Ribulose bisphosphate carboxylase large chain | Antithamnion sp., 7 7 117 113 −0.12 54372 |
| chloroplast thylakoid membrane; Peripheral membrane protein; Stromal side chloroplast Allophycocyanin beta chain | Thermosynechococcus elongatus Gloeobacter violaceus, 2 2 78 75 0.15 96126 |
| chloroplast Allophycocyanin beta chain | Thermosynechococcus elongatus Gloeobacter violaceus, 2 2 78 75 0.15 96126 |
| chloroplast thylakoid membrane; Peripheral membrane protein; Stromal side chloroplast Photosystem I P700 chlorophyll a apoprotein A2 | Aglaothamnion neglectum Synechocystis sp., 6 6 120 117 −0.29 53084 |
| chloroplast thylakoid membrane; Peripheral membrane protein; Stromal side chloroplast Allophycocyanin beta chain | Galdieria sulphuraria, 16 16 96 73 0.02 17536 |
| Cell inner membrane; Multi-pass membrane protein chloroplast thylakoid membrane; Peripheral membrane protein; Stromal side chloroplast Ribulose bisphosphate carboxylase large chain | Galdieria sulphuraria, 16 16 96 73 0.02 17536 |
| Mouse liver Nucleus, Mitochondrion Carbamoyl-phosphate synthase | Mus musculus 39 33 1637 1268 −0.12 16571 |
| Cytoplasm Arginase-1 | Mus musculus 29 35 300 310 −0.19 34957 |
| Cytoplasmin, Membrane Selenium-binding protein | Mus musculus 31 28 526 405 −0.31 53147 |
| Cytoplasm Argininosuccinate synthase | Mus musculus 32 15 429 191 −0.11 46340 |
| Mitochondrion Glyceraldehyde-3-phosphate dehydrogenase | Mus musculus 31 32 321 298 −0.04 36072 |
| Cytosol Cytosolic 10-formyltetrahydrofolate dehydrogenase | Mus musculus 9 17 139 361 −0.36 99502 |
| Extracellular region 3-ketoacyl-CoA thiolase, mitochondrial | Mus musculus 10 20 137 216 −0.38 42260 |
| Sample Location                           | Protein name                  | Phylogenetic origin | Protein coverage (%) | Score<sup>a</sup> | GRAVY  | MW (Da)<sup>b</sup> |
|------------------------------------------|-------------------------------|---------------------|-----------------------|---------------------|--------|---------------------|
| Serum albumin                           | Mus musculus                  | 15                  | 18                    | 327                 | 349    | −0.09 70700        |
| Alcohol dehydrogenase 1                 | Mus musculus                  | 19                  | 29                    | 161                 | 212    | 0.20 40601         |
| Aspartate aminotransferase, mitochondrial| Mus musculus                  | 15                  | 16                    | 231                 | 215    | −0.23 47780        |
| Carboxylesterase 3B                     | Mus musculus                  | 12                  | 14                    | 201                 | 183    | −0.12 63712        |
| Glycine N-methyltransferase             | Mus musculus                  | 29                  | 19                    | 131                 | 127    | −0.25 33110        |
| Cytochrome P450 2D10                    | Mus musculus                  | 9                   | 2                     | 100                 | 123    | −0.06 57539        |
| Aspartate aminotransferase, cytoplasm   | Mus musculus                  | 7                   | 13                    | 112                 | 115    | −0.25 46504        |
| Adenosylhomocysteinase                  | Mus musculus                  | 27                  | 14                    | 335                 | 120    | −0.07 47780        |
| Fructose-1,6-bisphosphatase 1           | Mus musculus                  | 12                  | 16                    | 117                 | 120    | −0.12 37288        |
| Carboxylesterase 3A                     | Mus musculus                  | 13                  | 9                     | 220                 | 139    | −0.12 63677        |
| Sarcosine dehydrogenase, mitochondrial  | Mus musculus                  | 8                   | 6                     | 182                 | 209    | −0.25 102644       |
| UDP-glucuronosyltransferase 1-1         | Mus musculus                  | 4                   | 8                     | 94                  | 141    | 0.09 60749         |
| Hemoglobin subunit beta-1               | Mus musculus                  | 16                  | 24                    | 111                 | 105    | 0.08 15944         |
| Peroxosomal bifunctional enzyme         | Mus musculus                  | 3                   | 2                     | 98                  | 78     | −0.12 78822        |
| Microsomal glutathione S-transferase    | Mus musculus                  | 17                  | 21                    | 80                  | 87     | 0.15 17597         |
| Cytochrome P450 2F2                     | Mus musculus                  | 6                   | 7                     | 128                 | 130    | −0.13 56141        |
| Pyrethroid hydrolase Ces2a              | Mus musculus                  | 9                   | 5                     | 100                 | 76     | −0.30 57539        |
| Homogentisate 1,2-dioxygenase           | Mus musculus                  | 6                   | 6                     | 81                  | 114    | −0.34 50726        |
| Regucalcin                              | Mus musculus                  | 4                   | 13                    | 72                  | 112    | −0.28 33899        |
| 3-ketoacyl-CoA thiolase B, peroxisomal  | Mus musculus                  | 13                  | 8                     | 116                 | 84     | 0.05 44481         |
| Sorbitol dehydrogenase                  | Mus musculus                  | 6                   | 6                     | 90                  | 89     | 0.06 38795         |
| ATP synthase subunit f, mitochondrial   | Mus musculus                  | 26                  | 26                    | 70                  | 71     | −0.30 10394        |
| ATP synthase subunit alpha, mitochondrial| Mus musculus                  | 14                  | 10                    | 193                 | 160    | −0.10 59830        |
| Urocanate hydratase                     | Mus musculus                  | 2                   | 1                     | 100                 | 76     | −0.14 75227        |
| Fumarlylacetocetase                     | Mus musculus                  | 3                   | 6                     | 75                  | 74     | −0.21 46488        |
| Uricase                                 | Mus musculus                  | 17                  | 11                    | 157                 | 97     | −0.46 35245        |
| Mus musculus                            | Mus musculus                  | 15                  | 13                    | 180                 | 119    | −0.26 39938        |
| Location/Function                        | Enzyme Name                        | Species     | MASCOT Score | Peptide Count | Unique Peptide Count | MW  |
|------------------------------------------|------------------------------------|-------------|--------------|---------------|----------------------|-----|
| Membrane                                 | Fructose-bisphosphate aldolase B   | Mus musculus| 11           | 6             | 104                  | 96  |
|                                          | UDP-glucuronosyltransferase 2B17   |             | 11           | 6             | 104                  | 96  |
| NI                                       | Pyrethroid hydrolase               | Mus musculus| 9            | 7             | 108                  | 89  |
| Cytoplasm                                | 3-hydroxyanthranilate 3,4-dioxygenase |            | 9            | 6             | 90                   | 87  |
| Mitochondrion                            | Hydroxymethylglutaryl-CoA synthase, mitochondrial | Mus musculus| 7            | 6             | 86                   | 70  |
| Mitochondrion                            | Trifunctional enzyme subunit alpha, mitochondrial | Mus musculus| 9            | 7             | 90                   | 81  |
| Endoplasmic reticulum                    | Microsomal triglyceride transfer protein large subunit | Mus musculus| 1            | 1             | 74                   | 80  |
| Membrane                                 | Cytochrome b-c1 complex subunit 2, mitochondrial | Mus musculus| 4            | 4             | 73                   | 76  |

*a MASCOT score greater than 67.
b MW: Molecular weight.
### Table 3
Endogenous peptides and transitions for PRM methods.

| Protein name                        | Abreviation | Peptide          | Precursor (m/z) | Product (m/z) |
|-------------------------------------|-------------|------------------|-----------------|---------------|
| Carbamoyl-phosphate synthase        | CPSM        | TAVDSGIALTNFQVTK | 898.4844        | 950.5306      |
|                                     |             |                  | 837.4465        | 736.3898      |
|                                     |             |                  | 1051.4725       | 722.3138      |
|                                     |             |                  | 591.2733        | 403.2300      |
|                                     |             | VLGTSVESIMATEDR  | 804.4009        | 926.5418      |
|                                     |             |                  | 617.3365        | 390.2096      |
|                                     |             | AFAMTNQILVER     | 696.8688        | 835.4495      |
|                                     |             |                  | 615.3647        | 472.2402      |
|                                     |             | GQNQPVLNITNR     | 677.3653        | 801.4829      |
|                                     |             |                  | 688.3988        | 587.3511      |
|                                     |             |                  | 696.4291        | 355.2340      |
|                                     |             |                  | 468.3180        | 231.1162      |
|                                     |             |                  | 832.4345        | 646.3705      |
|                                     |             |                  | 547.3021        | 491.3073      |
| Argininosuccinate synthase          | ASSY        | EQGYDVIAVLNIGQK  | 891.4571        | 977.5415      |
|                                     |             |                  | 743.4410        | 630.3570      |
|                                     |             |                  | 1085.4972       | 556.3453      |
|                                     |             |                  | 485.3082        | 921.5768      |
|                                     |             |                  | 751.4713        | 541.3344      |
| Selenium-binding protein 2         | SBP2        | GSFVLLDGETFEVK  | 770.8983        | 1037.515      |
|                                     |             |                  | 924.4309        | 809.404       |
|                                     |             |                  | 894.4971        | 821.4338      |
|                                     |             |                  | 708.3498        | 848.4989      |
|                                     |             |                  | 701.4304        | 588.3464      |
|                                     |             |                  | 957.5           | 858.4316      |
|                                     |             |                  | 545.2678        | 261.1598      |
|                                     |             |                  | 875.4469        | 746.4043      |
|                                     |             |                  | 889.4315        | 790.3631      |
|                                     |             |                  | 261.1598        | 261.1598      |
|                                     |             |                  | 261.1598        | 261.1598      |
| Protein name                                    | Abreviation | Peptide                          | Precursor (m/z) | Product (m/z) |
|------------------------------------------------|-------------|----------------------------------|-----------------|--------------|
| Glyceraldehyde-3-phosphate dehydrogenase G3P  | ARG1        | VMEETFSYLLGR                     | 722.8607        | 1214.6052    |
|                                               |             |                                  |                 | 855.4723     |
|                                               |             |                                  |                 | 708.4039     |
|                                               |             | EGLYITEEIYK                      | 679.3479        | 1058.5405    |
|                                               |             |                                  |                 | 895.4771     |
|                                               |             |                                  |                 | 782.3931     |
|                                               |             | VSVVLGDHSLAVGSIHGAR              | 673.3641        | 866.9581     |
|                                               |             |                                  |                 | 817.4239     |
|                                               |             |                                  |                 | 760.8819     |
|                                               |             | SLEIIIGAPFSK                     | 581.3293        | 606.3246     |
|                                               |             |                                  |                 | 556.3341     |
|                                               |             |                                  |                 | 478.266      |

Table 3 (continued)
**Table 4**
Endogenous peptides and transitions for PRM methods.

| Protein name                          | Abreviattion | Peptide | PRM  | Precursor (m/z) Product (m/z) |
|---------------------------------------|--------------|---------|------|------------------------------|
| R-phycoerythrin alpha chain, Porphyra purpurea | PHEA_PORPU  | SVITTISAADAAGR | 717.3834 | 1134.5749 1033.5273 374.2146 715.3846 621.2515 587.3260 920.3956 663.2944 491.2460 |
|                                       |              | FPSSDLESVQGNIQR | 588.6235 | 1033.5273 374.2146 715.3846 621.2515 587.3260 |
|                                       |              | NPGEAGDSQEK | 566.2493 | 1134.5749 1033.5273 374.2146 |
| C-phyocyanin-1 alpha chain, Synechococcus sp. | PHCA1_SYNP6 | TPLTEAVAAADSQGR | 743.8784 | 1175.5651 945.4748 775.3693 1194.6113 1107.5793 790.457 |
|                                       |              | FLSTTELQVAFGR | 727.8855 | 1194.6113 1107.5793 790.457 |
| C-phycoerythrin alpha chain, Synechocystis sp. | PHEA_SYNY1 | TLGLPTAPYVEALSFAR | 602.6647 | 1152.6048 793.4203 664.3777 688.3737 635.2671 560.3151 |
|                                       |              | FPSTSDLESVQGSIQR | 584.2917 | 1152.6048 793.4203 664.3777 688.3737 635.2671 560.3151 |
| C-phycoerythrin alpha chain, Microchaete diplosiphon | PHEA_MICDP | SSVTTTISAADAAGR | 701.3834 | 1116.6008 815.437 374.2146 715.3846 621.2515 587.3260 920.3956 663.2944 491.2460 |
|                                       |              | ALGLPTAPYVEALSFAR | 592.6612 | 1152.6048 793.4203 664.3777 688.3737 635.2671 560.3151 |
|                                       |              | FPSTSDLESVQGSIQR | 584.2917 | 1152.6048 793.4203 664.3777 688.3737 635.2671 560.3151 |

**Table 5**
Total spectrum, peptide and protein counts after purification by our approach versus published protocols on complex matrices.

|                  | Total spectrum count | Peptide count | Protein count |
|------------------|----------------------|---------------|---------------|
| Biofilm-published approach\(^a\) | 932                  | 585           | 195           |
| Biofilm-our approach\(^b\) | 937                  | 424           | 163           |
| Mouse liver-published approach\(^c\) | 1122                 | 1408          | 416           |
| Mouse liver-our approach\(^d\) | 959                  | 1205          | 355           |
| Soil-published approach\(^e\) | 946                  | 293           | 72            |
| Soil-our approach\(^f\) | 932                  | 488           | 128           |

Data from the ProteomeXchange Consortium via the PRIDE [10] repository with the dataset identifier PXD008110 and 10.6019/PXD008110.

\(^a\) Average of three replicates.
\(^b\) Counts of a single replicate.
2. Experimental design, materials and methods

Experimental design and materials and methods have been reported previously [1].

Acknowledgments

This research was carried out with the financial support of the French Ministère de l’Enseignement Supérieur et de la Recherche (5HU66) and the Ligue contre le Cancer (Maj 06-12-2016)

References

[1] B. Eddhif, J. Lange, N. Guignard, Y. Batonneau, S. Papot, C. Geoffroy-Rodier, P. Poinot, Study of a novel agent for TCA precipitated proteins washing - comprehensive insights into the role of ethanol/HCl on molten globule state by multi-spectroscopic analyses, J. Proteomics. 173 (2018) 77–88 (submitted for publication).
[2] K. Chourey, J. Jansson, N. VerBerkmoes, M. Shah, K.L. Chavarria, L.M. Tom, E.L. Brodie, R.L. Hettich, Direct cellular lysis/protein extraction protocol for soil metaproteomics, J. Proteome. Res. 9 (2010) 6615–6622. http://dx.doi.org/10.1021/pr100787q.
[3] H.-I. Huang, W.-Y. Chen, J.-H. Wu, Total protein extraction for metaproteomics analysis of methane producing biofilm: the effects of detergents, Int. J. Mol. Sci. 15 (2014) 10169–10184. http://dx.doi.org/10.3390/ijms15061069.
[4] S. Song, G.J. Hooiveld, W. Zhang, M. Li, F. Zhao, J. Zhu, X. Xu, M. Muller, C. Li, G. Zhou, Comparative proteomics provides insights into metabolic responses in rat liver to isolated soy and meat proteins, J. Proteome Res. 15 (2016) 1135–1142. http://dx.doi.org/10.1021/acs.jproteome.5b00922.
[5] L. Jiang, L. He, M. Fountoulakis, Comparison of protein precipitation methods for sample preparation prior to proteomic analysis, J. Chromatogr. A. 1023 (2004) 317–320. http://dx.doi.org/10.1016/j.chroma.2003.10.029.
[6] E. Fic, S. Kedracka-Krok, U. Jankowska, A. Pirog, M. Dziedzicka-Wasylewska, Comparison of protein precipitation methods for various rat brain structures prior to proteomic analysis, Electrophoresis. 31 (2010) 3573–3579.
[7] D.I. Jacobs, M.S. van Rijssen, R. van der Heijden, R. Verpoorte, Sequential solubilization of proteins precipitated with trichloroacetic acid in acetone from cultured Catharanthus roseus cells yields 52% more spots after two-dimensional electrophoresis, Proteomics 1 (2001) 1345–1350. http://dx.doi.org/10.1002/1615-9861(200111)1:11<1345::AID-PROT1345>3.0.CO;2-F.
[8] A. Rygula, K. Majzner, K.M. Marzec, A. Kaczor, M. Pilarczyk, M. Baranska, Raman spectroscopy of proteins: a review: raman spectroscopy of proteins, J. Raman Spectrosc. 44 (2013) 1061–1076. http://dx.doi.org/10.1002/jrs.4335.
[9] J.A. Vizcaíno, A. Csordas, N. del-Toro, J.A. Dianes, J. Griss, I. Lavidas, G. Mayer, Y. Perez-Riverol, F. Reisinger, T. Ternent, Q. W. Xu, R. Wang, H. Hermjakob, 2016 update of the PRIDE database and related tools, Nucleic Acids Res 44 (2016) D447–D456. http://dx.doi.org/10.1093/nar/gkv1145.