Rapid pre-gel visualization of proteins with mass spectrometry compatibility

Marco A Mata-Gómez [1], Matthew T Yasui [2] and Robert Winkler [1, 3,*]

[1] Departamento de Biotecnología e Ingeniería de Alimentos, Centro de Biotecnología-FEMSA, Tecnológico de Monterrey, Campus Monterrey, Ave. Eugenio Garza Sada 2501 Sur, 64849 Monterrey N.L., México.

[2] School of Chem/Bio/Envr Engineering, Oregon State University, 102 Gleeson Hall, Corvallis, OR 97331-2702, U.S.A.

[3] Departamento de Biotecnología y Bioquímica, CINVESTAV Unidad Irapuato, Km. 9.6 Libramiento Norte Carr. Irapuato-León, 36821 Irapuato Gto., México

* Corresponding author: Robert Winkler, Tel: +52-462-6239-635,

e-mail: robert.winkler@ira.cinvestav.mx

Despite all of the prophecies of doom, gel electrophoresis is still prevalent in modern proteomic workflows. However, the currently used protein staining methods represent a serious bottleneck for a quick subsequent protein analysis using mass spectrometry. Substituting traditional protein stains by pre-gel derivatization with visible and mass spectrometry compatible reagents eliminates several processing steps and drastically reduces the sample preparation time. A defined chemistry permits seamless integration of such covalent protein staining methods into standardized bioinformatic pipelines. Using Uniblue A we could covalently stain simple to complex protein samples within 1 minute. Protein profiles on the gels were not compromised and MS/MS based sequence coverages higher than 80% could be obtained. In addition, the visual tracking of covalently stained proteins and peptides facilitates method development and validation. Altogether, this new chemo-proteomic approach enables true “at-line” analysis of proteins.

Undoubtedly, sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) of proteins is among the most important tools for biochemistry. Whereas the original protocol of Ulrich Laemmli[1] has not been substantially altered since 1970, the combination with other methods enhanced its possible applications to near infinity. In 1996, Shevchenko et al. introduced a procedure for in-gel digestion of proteins for mass spectrometry (MS), which provided unprecedented opportunities for the characterization of proteins[2]. Later, multidimensional chromatography/ MS based protein identification (MudPID), introduced in 2001 by Yates and coworkers[3], was expected to replace gel/ MS based work-flows. However, at the present, pre-separation of protein samples by one-dimensional GE, followed by LC-MS/MS identification of tryptic digests from gel slices, is still considered as one of the most capable strategies for proteome research[4, 5].

But whereas the gel electrophoresis and LC-MS parts of such proteomic workflows are in the time scale of 1-2 hours each, significantly more time is spent for protein staining and post-

electrophoresis sample work-up, which typically includes de-staining, reduction, alkylation, tryptic digestion and extraction of the peptides.

Since the compatibility of silver stain with mass spectrometry is still questioned, staining with colloidal Coomassie is currently the method of choice. Considering the quickest protocols, 3 hours are necessary for colloidal Coomassie staining[6], and another 4 hours for preparing selected gel pieces for MS[4]. Many tedious manual steps are necessary, which increase the risk of contamination. Automation is possible, but its costs are high and the reliability and flexibility of robots is sometimes not satisfactory.

A couple of pre-gel covalent fluorescent dyes for proteomic applications are available on the market, but they demand special hardware for the examination of the gels and are expensive[7].

Some authors also describe the covalent pre-gel staining of proteins with visible stains, such as dabsyl chloride[8], Remazol dyes[9] and Uniblue A[10]. However, those studies were focused on the preparation of molecular weight standards, and no one examined the suitability of visible pre-gel staining methods for proteomic workflows.

After some theoretical considerations and initial testing of several protein reactive dyes, Uniblue A seemed to be our most promising candidate, due to its solubility in water, commercial availability with adequate purity and low price. Additionally, its blue color aids in achieving a sufficient optical contrast. Uniblue A contains a single vinyl sulfone group that may react with primary amines via nucleophilic addition (see **Fig. 1**). Covalently modified residues will have a defined monoisotopic mass shift of 484.0399 Da.

We discovered that sufficient covalent pre-gel staining of the protein with Uniblue A can be obtained within only 1 minute at 100 °C (see Online Methods, **Fig. 1b**).

Further steps of the protocol include quenching of excess Uniblue A, reduction and alkylation (see Online Methods). Altogether, the sample preparation for the SDS-PAGE can be completed in less

than 10 minutes. The apparent molecular weights of pre-stained and un-labeled Coomassie stained proteins are in agreement (see **Fig. 1c**). Hence, the electrophoretic mobility of the proteins is not changed significantly by their covalent staining, which is in congruence with previous studies employing dabsyl chloride[8] or Remazol dyes[9]. Presumably, the small appendices do not contribute to the binding of SDS. Sensitivity and resolution are slightly reduced for pre-stained proteins, but protein patterns of pre-stained and un-labeled Coomassie stained proteins are comparable, as demonstrated for the *Escherichia coli* disintegrate **(Fig. 1d).** For subsequent mass spectrometric analyses, staining intensity and resolution are perfectly adequate.

For the work-up of gel pieces de-staining, reduction and alkylation can be skipped, since those steps are already integrated into the SDS-PAGE sample preparation. Gel pieces only need to be shrunk in acetonitril (ACN), dried and rehydrated in trypsin solution. The tryptic digestion was performed for 30 minutes at 60 °C. Tryptic peptides were extracted by ACN/0.1% tri fluoro acetic acid (1:1 v/v), dried in a vacuum centrifuge, re-dissolved in 0.1 % formic acid and analyzed by NanoLC-MS/MS. In comparison to the current best-in-class methods, the staining time could be reduced from 3 hours to less than 10 minutes, and the sample work-up time from 4 hours to about 2 hours. In total, the required sample processing time was condensed to less than a third, and the manual handling steps could be significantly reduced, which reduces the risk of contamination. No stain particles are present, which reduces the chance of blockages as may occur in the NanoLC analysis of Coomassie stained samples.

Raw MS/MS data were converted into mzXML and evaluated automatically. In short, the search was performed against a concatenated target-decoy database[11] using the Open Mass Spectrometry Search Algorithm[12] (OMSSA). The peptide hits were validated by PeptideProphet[13] and ProteinProphet[14]. After this automatic processing, the raw data and identification results could be easily converted into valid PRoteomics IDEntifications database[15] (PRIDE) XML, using the

PRIDE converter tool[16], and uploaded to the repository. Covalent derivatization with Uniblue A has been added by the PRIDE team as a protein modification (PSI-MOD) for the ontology lookup service (OLS) with the comma separated value (CSV) term MOD: 01659.

The covalent modification influences the protein and peptide properties in various ways.

Tagged proteins and peptides display color in the visible spectrum and their fate can be tracked visually. This allows for the direct monitoring of sample processing steps, such as extraction and re-dissolution of peptides. This feature facilitates optimization and validation of sample preparation methods in proteomics.

The additional sulfate group increases the solubility of derivatized proteins and peptides, which supports their extraction, especially in cases of very hydrophobic species.

By allowing for different potential derivatization sites in the database search, we could prove that the reaction was highly selective for lysine. Moreover, only a fraction of the lysines were derivatized. No other residue, such as cysteine, seems to be affected by the reaction. This is crucial for efficient database searches, since only one potential modification site, namely +484.0399 Da at lysine, has to be considered.

No C-terminal lysine with Uniblue A modification was found, which suggests, that the modification inhibits tryptic cleavage. Therefore, the method could also be employed to generate longer tryptic peptide fragments.

During mass spectrometric analysis, tagged and un-tagged peptides exhibited slightly different behavior. In general, the Uniblue A modification has a tendency to reduce the charge state of the molecules in positive ionization mode due to its negative sulfate group. **Fig. 2** compares the fragmentation spectrum of a doubly charged Uniblue A derivatized peptide versus the spectrum of a triply charged untagged peptide with the same sequence. Both spectra were found in the same sample.

In this example, the N-terminus of the tryptic peptide is derivatized. The mass shift allows the clear assignment of the N-terminal fragment ions a_1-NH_3 and b_1, which otherwise would be outside the mass range of the analyzer. Whereas the position of the C-terminal y-ions was not affected, all N-terminal a/b-series ions were shifted, which facilitates the assignment of peaks b_{10} to b_{14}. Additionally, the signal-to-noise ratio of N-terminal ions was significantly improved. Altogether more fragment ions can be assigned automatically for the Uniblue A derivatized peptide.

The "peptidic diversity" is increased by different chromatographic properties between un-derivatized and labeled peptides, defined mass shifts and different ionization behavior. This reduces the sensitivity of the mass spectrometric analysis, but by principle might help to increase the sequence coverage, especially when it comes to short tryptic fragments.

To examine the overall performance of this rapid covalent derivatization protocol in comparison to the standard Coomassie based strategy, we compared the identification results after PeptideProphet/ProteinProphet validation (**Tab. 1**). Both methods yield identification results which comply with strict acceptance criteria. All proteins were identified with a ProteinProphet probability of 1.0000. At least 6 unique peptides were proven and the MS/MS based sequence coverage was at least 25% in all procedures.

The reduced number of identified peptides when using only Uniblue A is probably caused by matrix suppression effects during the mass spectrometry, since the samples are washed for less time compared to the Coomassie protocol. This was confirmed by the analysis of samples which were first derivatized with Uniblue A and after electrophoresis stained with Coomassie. For two of the three samples, the double staining led to a dramatically increased number of validated peptides, whereas in only one case the number remained about the same. This demonstrates that Uniblue A derivatization *per se* does not interfere negatively with mass spectrometry based protein

identification. In fact, optimized protocols that address sample-to-sample variation and matrix suppression might even boost possible sequence coverage results.

To prove the suitability of the method for complex samples, we applied the covalent derivatization to disintegrates of *Escherichia coli* cells producing a recombinant protein (see Online Methods). Uniblue A derivatized and Coomassie stained samples exhibit the sample protein profile (see **Fig. 1d**), underlining the suitability of the method e.g. for expression clone screening. The supposed recombinant protein at approximately 50 kDa (theoretical molecular weight from sequence: 50,871 Da) was cut and subjected to NanoLC-MS/MS, yielding an excellent MS/MS based sequence coverage above 80% in both cases. Since for some parts of the sequence the data are complementary, the combined MS/MS sequence coverage reaches 92.0% (**Supplementary Fig. 1**).

The covalent pre-gel derivatization represents a novel approach for the rapid visualization of proteins in biochemical analysis. We demonstrated full mass spectrometry compatibility, enabling its use in modern proteomic workflows. The additional protein modification was successfully integrated into a state of the art data processing pipeline, which underlines the vast potential of such chemo-proteomic strategies in research and industry.

METHODS

Methods and any associated references are available in the online version of the article at http://www.nature.com/naturebiotechnology/.

Accession codes. PRIDE: see Table 1 and Online Methods

Note: Supplementary information is available on the Nature Methods website.

ACKNOWLEDGEMENTS

We greatly appreciate the support by the OMSSA team of the National Center for Biotechnology Information (NCBI) and the team of the European Bioinformatic Institute (EBI). This work was generously supported by the research chair of bioengineering and nano-bioparticles (grant CAT

161). The Rituximab sample was kindly donated from the research chair for bio-pharmaceuticals (grant CAT 122).

PATENT APPLICATION

Patent pending: MX/a/2009/013417

AUTHOR CONTRIBUTIONS

R.W. developed the methodology, designed experiments, carried out mass spectrometry measurements, evaluated the data and wrote the manuscript; M.A.M.-G. and M.T.Y. performed and optimized protein staining and gel electrophoresis experiments.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

References

1. Laemmli, U.K. *Nature* **227**, 680-685 (1970).

2. Shevchenko, A., Wilm, M., Vorm, O. & Mann, M. *Anal. Chem.* **68**, 850-858 (1996).

3. Washburn, M., Wolters, D. & Yates, J. *Nat. Biotechnol.* **19**, 242-247 (2001).

4. Shevchenko, A., Tomas, H., Havlis, J., Olsen, J. & Mann, M. *Nat Protoc* **1**, 2856-2860 (2006).

5. de Godoy, L., Olsen, J., de Souza, G., Li, G., Mortensen, P. et al. *Genome Biol.* **7**, R50 (2006).

6. Westermeier, R. *Proteomics* **6 Suppl 2**, 61-64 (2006).

7. Miller, I., Crawford, J. & Gianazza, E. *Proteomics* **6**, 5385-5408 (2006).

8. Parkinson, D. & Redshaw, J. *Anal. Biochem.* **141**, 121-126 (1984).

9. Compton, M., Lapp, S. & Pedemonte, R. *Electrophoresis* **23**, 3262-3265 (2002).

10. Bogoev, R. & Kang, D. *United States Patent Application 20090178926* , (2009).

11. Elias, J. & Gygi, S. *Nat. Methods* **4**, 207-214 (2007).

12. Geer, L., Markey, S., Kowalak, J., Wagner, L., Xu, M. et al. *J. Proteome Res.* **3**, 958-964 (2004).

13. Keller, A., Nesvizhskii, A., Kolker, E. & Aebersold, R. *Anal. Chem.* **74**, 5383-5392 (2002).

14. Nesvizhskii, A., Keller, A., Kolker, E. & Aebersold, R. *Anal. Chem.* **75**, 4646-4658 (2003).

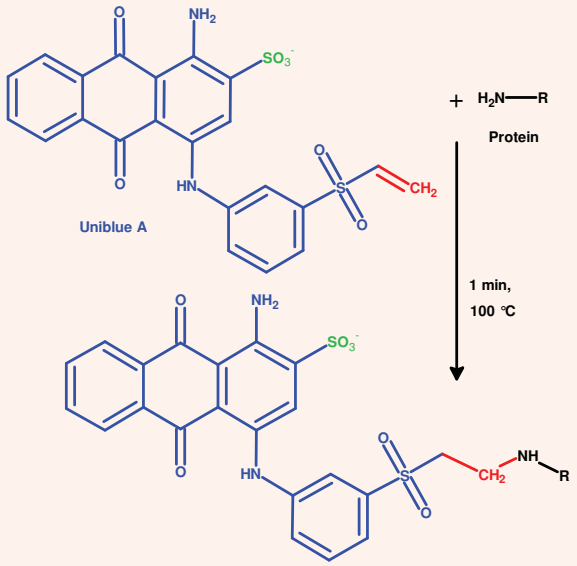
15. Vizcaíno, J.A., Côté, R., Reisinger, F., Foster, J., Mueller, M. et al. *Proteomics* **9**, 4276-4283 (2009).

16. Barsnes, H., Vizcaino, J., Eidhammer, I. & Martens, L. *Nat. Biotechnol.* **27**, 598-599 (2009).

Figure 1 a) Covalent staining of proteins via nucleophilic addition of Uniblue A. The vinyl sulfone group (red) reacts with primary amines, preferably on lysine residues. The sulfate group (green) supports the solubility of the dye and affects the ionization properties of the labeled peptide during mass spectrometry measurements. b) SDS-PAGE gel showing pre-stained marker and Uniblue A (Uni A) derivatized Rituximab antibody chains. The staining was achieved within 1 minute. The third lane contains the equal concentration of un-derivatized Rituximab sample (nat). c) The gel after subsequent staining with Coomassie, now also revealing the un-derivatized Rituximab sample. Staining intensity and protein profiles are comparable. d) *E. coli* TOP10, transformed with pMAL-c4x and auto-induced. Uniblue A (Uni A) derivatized and un-derivatized (nat) disintegration sample display comparable protein profiles after Coomassie staining. The assumed recombinant protein band was cut and subjected to nanoLC-MS/MS analysis, confirming the identity with >80% MS/MS based sequence coverage in both samples.

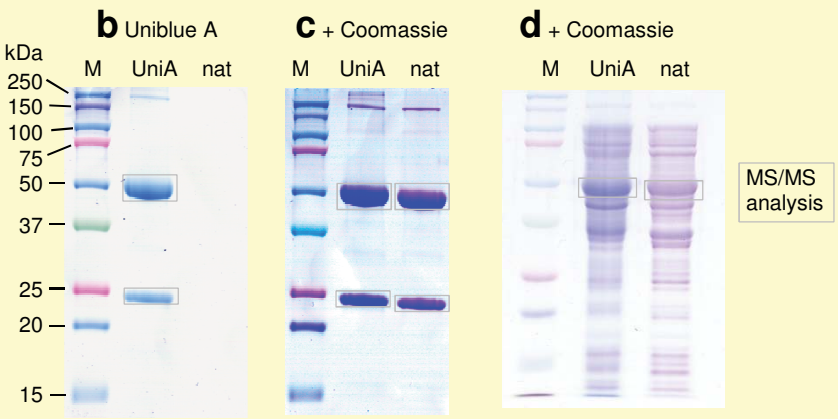
Figure 2 The direct comparison of MS/MS fragmentation spectra of Uniblue A derivatized (top) versus un-labeled (bottom) peptide KVPQVSTPTLVEVSR displays significantly increased signal intensities for the derivatized N-terminal ions (a- and b- series, shown in red). The defined mass shift for modified residues of 484.0399 *m/z* allows the detection of fragments otherwise outside the measuring range (fragments a_1-NH_3 and b_1). Thus the MS/MS based sequence coverage for individual peptides is improved.

a 1-Minute covalent staining of proteins by nucleophilic addition



SDS-PAGE of Uniblue A stained/ unstained samples

Commercial Rituximab antibodies **_E. coli_ TOP10/ pMAL-c4x**



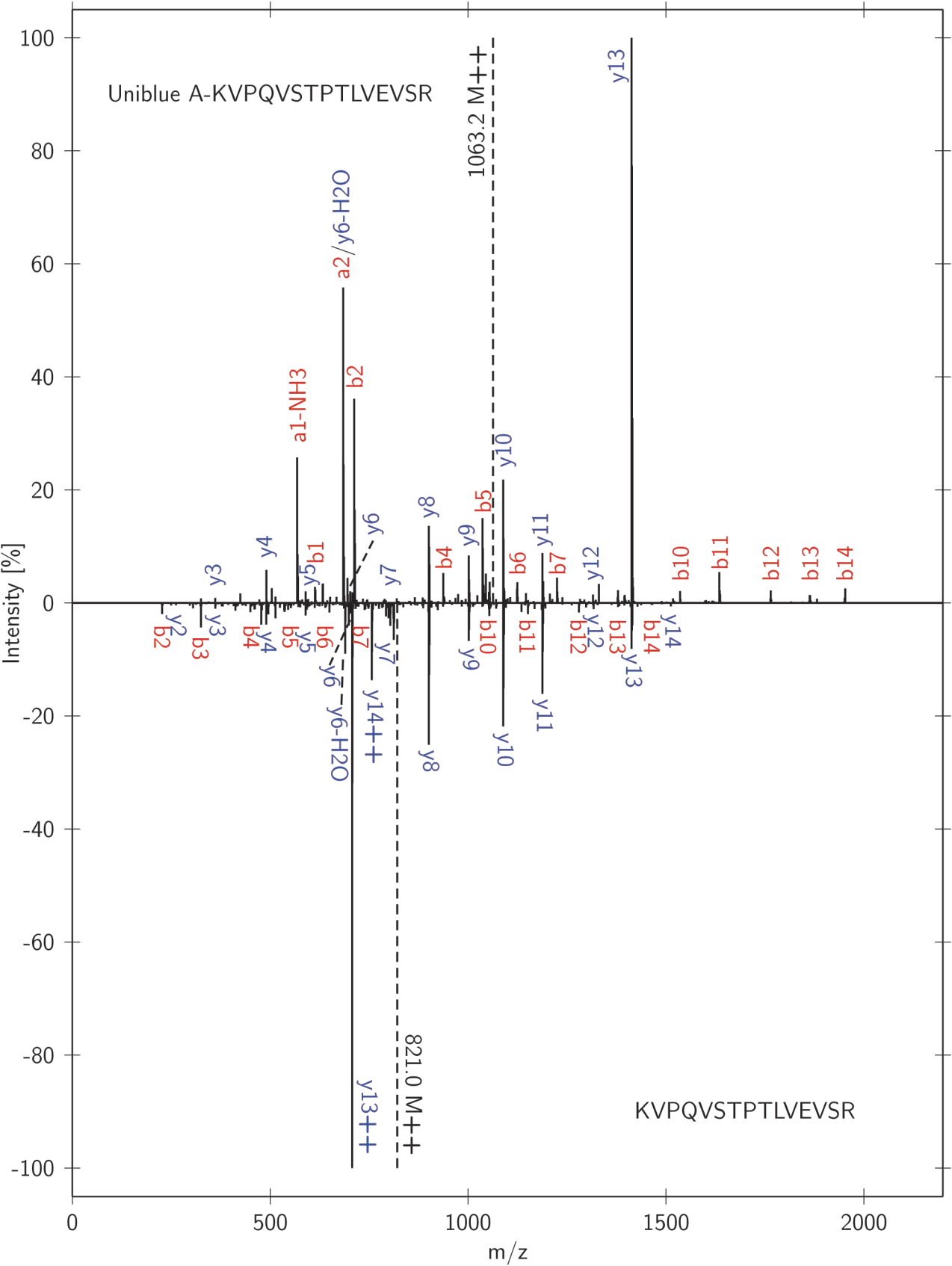

Uniblue A-KVPQVSTPTLVEVSR
KVPQVSTPTLVEVSR
Intensity [%]
m/z
100
80
60
40
20
0
-20
-40
-60
-80
-100
0
500
1000
1500
2000
1063.2 M++
821.0 M++
a2/y6-H2O
a1-NH3
y13
y13++
y6-H2O
y14++
y8
y10
y11
b2
b5
y9
y6
y4
b1
y5
y7
b4
b6
b7
y12
b10
b11
b12
b13
b14
y2
y3
b3
b8
b9
b14

Table 1 NanoLC-MS/MS identification results for gel bands of proteins

	Bovine serum albumin			Rituximab, heavy chain			Rituximab, light chain		
	Uni A	Uni A + Coom	Coom	Uni A	Uni A + Coom	Coom	Uni A	Uni A + Coom	Coom
MS/MS spectra	2898	2851	2778	2905	2818	2852	2887	2818	2869
ProteinProphet Probability	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
MS/MS Sequence coverage	32.5%	52.1%	40.2%	27.7%	26.0%	41.6%	33.8%	58.1%	62.9%
Total validated peptides	40	115	93	18	13	99	20	55	125
Unique peptides	17	51	41	10	10	25	6	24	35
PRIDE accession #	12567	12565	12564	12571	12569	12568	12575	12573	12572

Proteins were either covalently labeled with Uniblue A (Uni A) before electrophoresis or stained with Coomassie (Coom) after electrophoresis. Also sequential staining with both methods was applied (Uni A + Coom).
The bands of interest were cut, tryptically digested and subjected to nanoLC-MS/MS identification (see Online Methods).