Available online at www.sciencedirect.com
ScienceDirect
Journal of Chromatography B, 863 (2008) 293–297
JOURNAL OF CHROMATOGRAPHY B
www.elsevier.com/locate/chromb

Short communication

Mapping preferred sites for fluorescent labeling by combining fluorescence and MS analysis of tryptic CNA35 protein digests

Alma M.A. Mingels, Joost L.J. van Dongen, Maarten Merkx *

Laboratory of Macromolecular and Organic Chemistry, Department of Biomedical Engineering, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

Received 30 September 2007; accepted 24 December 2007
Available online 16 January 2008

Abstract

HPLC–MS analysis of tryptic protein digests in combination with fluorescence detection is presented as a convenient and quantitative method to gain insight into the relative reactivity of lysine side chains. In this scheme (tandem) mass spectrometry was used for identification of the modified residue, whereas fluorescence detection allowed determination of their relative abundance. Our method identified 'labeling hot-spots' at two flexible parts of the collagen-binding protein CNA35, positions that were consistent with all available structural and biochemical data on the collagen-binding properties of CNA35.

Keywords: Fluorescence; Labeling; LC–MS; Collagen-binding protein; Tryptic digestion

1. Introduction

Proteins are attractive ligands for applications that require highly specific targeting. Using the classical concept of the "magic bullet" proposed almost 100 years ago by Paul Ehrlich, such a targeting protein needs to be conjugated to either a spectroscopic probe or a drug [1]. Commonly, the amine groups present in lysines are chosen for the conjugation of synthetic groups, since lysines are widely available in proteins and the positively charged groups are preferentially located at protein surfaces [2,3]. However, lysine modification often results in conjugation at multiple sites on the protein, which sometimes results in protein inactivation. Insights into factors that determine the relative reactivities of lysine side chains are still limited, in part because of a lack of convenient analytical approaches. Besides providing fundamental knowledge about lysine reactivity, such a method should allow detailed characterization of proteins after bioconjugation with fluorescent labels.

Mass spectrometric identification of peptide fragments after proteolysis is a common method to map post-translational modifications in proteins [4–6]. Several studies have previously used this method to assess the amine reactivities of lysine residues [7–9]. Przybylski and coworkers used amino-acylation to probe the surface accessibility of lysine side chains in four model proteins [7]. More recently, the use of amidination was reported as a general method to probe protein tertiary structure [8]. The latter modification has the advantage that it retains the positive charge of the original amine. However, because the signal intensity in mass spectrometric detection depends strongly on the ionization efficiency of the peptide fragments, quantitative comparison of the relative reactivities of various amine groups in the same protein remains difficult using this method.

In the present study we have combined mass spectrometry with fluorescence detection to analyze tryptic digests of the fluorescently labeled collagen-binding protein CNA35. In this scheme fluorescence detection is used to identify and quantify the relative abundance of fluorescently labeled peptide fragments, whereas mass spectrometry allows identification of the specific lysine residues that were modified (Fig. 1). Our method identified 'labeling hot-spots' at two flexible parts of

Abbreviations: CNA35, collagen-binding protein (PDB ID: 2F68); OG488, Oregon Green 488; CNA35–OG488, Oregon Green 488-labeled CNA35 protein.

* Corresponding author. Tel.: +31 40 2474728; fax: +31 40 2451036.
E-mail address: m.merkx@tue.nl (M. Merkx).

1570-0232/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jchromb.2007.12.026

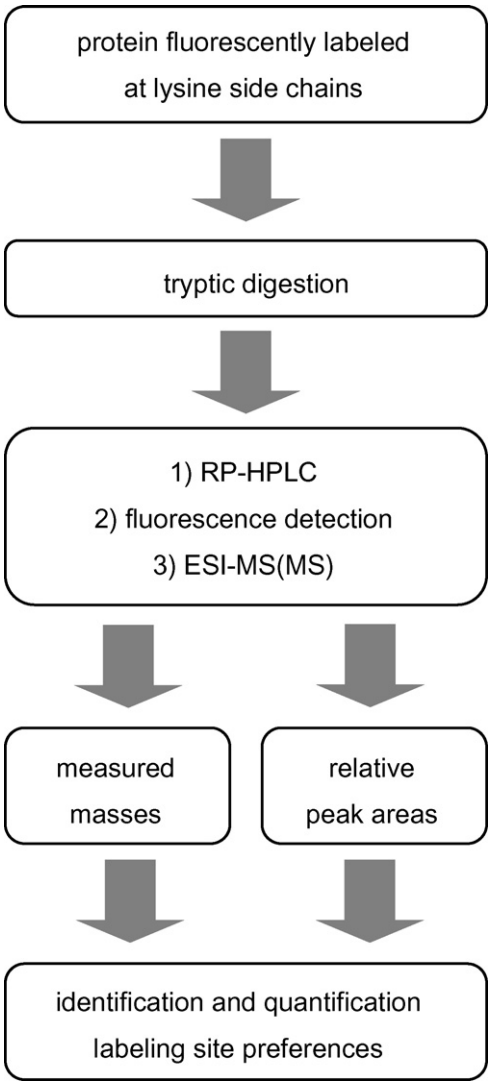

Fig. 1. Flow diagram showing each step of the analytical procedure to identify fluorescently labeled residues and determine their relative abundance.

CNA35, positions that were consistent with all structural and biochemical data available on the collagen-binding properties of CNA35.

2. Experimental

2.1. Preparation of CNA35 and CNA35–OG488

The collagen-binding protein CNA35 was expressed, purified and labeled with a 10-fold molar excess of the succinimidyl ester of Oregon Green 488 carboxylic acid (Invitrogen, Leiden, The Netherlands) as previously reported [10]. The labeling ratio was determined by measuring the absorbance (MultiSpec-1501, Shimadzu, Kyoto, Japan) at 496 nm for OG488 ($\varepsilon_{496} = 72{,}000\,\mathrm{cm^{-1}\,M^{-1}}$) and at 280 nm for the protein ($\varepsilon_{280} = 33{,}167\,\mathrm{cm^{-1}\,M^{-1}}$). The absorbance at 280 nm was corrected for the contribution by the dye by subtraction of $0.12 \times A_{496}$ nm (Invitrogen, Leiden, The Netherlands). Afterwards the protein solutions were stored at −80 °C.

2.2. Tryptic digestion of CNA35 and CNA35–OG488

CNA35 and CNA35–OG488 solutions were diluted to a final concentration of 0.5 mg/mL and 1.0 mg/mL, respectively, in 100 mM Tris–HCl buffer, pH 8.5 and 20% (v/v) acetonitrile. Proteomics grade trypsin (Sigma, Steinheim, Germany) was added from a 1.0 mg/mL stock solution in 1.0 mM HCl to a final trypsin-to-protein ratio of 1:50 (w/w) and incubated at 37 °C. Digestion was followed in time for at least 1 day and was stopped by the addition of 1.0 M HCl to obtain pH 1–2.

2.3. LC–MS characterization

LC–MS characterization of intact and digested CNA35 and CNA35–OG488 was performed on a LCQ Deca XP MAX (Thermo Finnigan, USA), equipped with a Surveyor autosampler, a RP-HPLC Polaris C18-A column (Varian, Torrance, CA, USA): 100 mm × 2.0 mm, 200 Å, 3 μm, and a Surveyor PDA detector connected to an ESI-ion trap mass spectrometer. Two LC-10AD VP pumps and a DGU-14A degasser (all from Shimadzu, Kyoto, Japan) were also connected, one pump for ultrapure water and one for HPLC grade acetonitrile (Biosolve, Valkenswaard, The Netherlands), which both contained 0.1% (v/v) formic acid (98–100%, Merck Darmstadt, Germany). The LC–MS was controlled by Xcalibur Software (Version 1.3). The column temperature was set to 35 °C, the flow rate was 0.2 mL/min, and 2.0–4.0 μL of sample solution was loaded on the column. The CNA35–OG488 digestion mixture was analyzed with an RF-10AXL fluorescence detector (Shimadzu, Kyoto, Japan) inserted between the end of the column and the inlet of the ESI-MS system. Gradient elution was optimized for separation of the fluorescent peptides, where a gradient was applied from 15% to 35% acetonitrile in 40 min. The settings for ionization, for the ion optics and the ion trap were default, except that mass analysis was adjusted to two microscans and the mass range was minimized to m/z 300–1500. Data-dependent tandem MS analysis was performed via two scan events, the first in TurboScan MS mode and the second in data-dependent tandem MS mode. MS data was displayed in Xcalibur Qual Browser, where smoothing (Gaussian, 7 points) and baseline subtraction were applied (<10% abundance, two degrees of freedom).

The mass spectra belonging to the intact proteins were deconvoluted using MagTran (Version 1.02) [11]. The fluorescent peaks were analyzed using DAx software (Version 8.0, Van Mierlo Software Consultancy, The Netherlands). The expected fragments for tryptic-digested CNA35 were simulated using the BioLynx Protein/Peptide Editor supported by MassLynx (Version 4.0, Micromass, UK). Trypsin was defined to cleave at the C-terminal site of all lysines and arginines, except if a proline was located next to them. Complete digestion would result in 36 tryptic peptides, denoted from T1 to T36, where T1 originated from the N-terminal site of CNA35 and T36 from the C-terminal site. Up to two missed cleavage-sites were taken into account when analyzing the LC–MS data. For the digestion of labeled CNA35–OG488, up to five missed cleavages were taken into account. The number of ligated OG488 labels was varied from one to the total amount of lysines present in the peptide, where

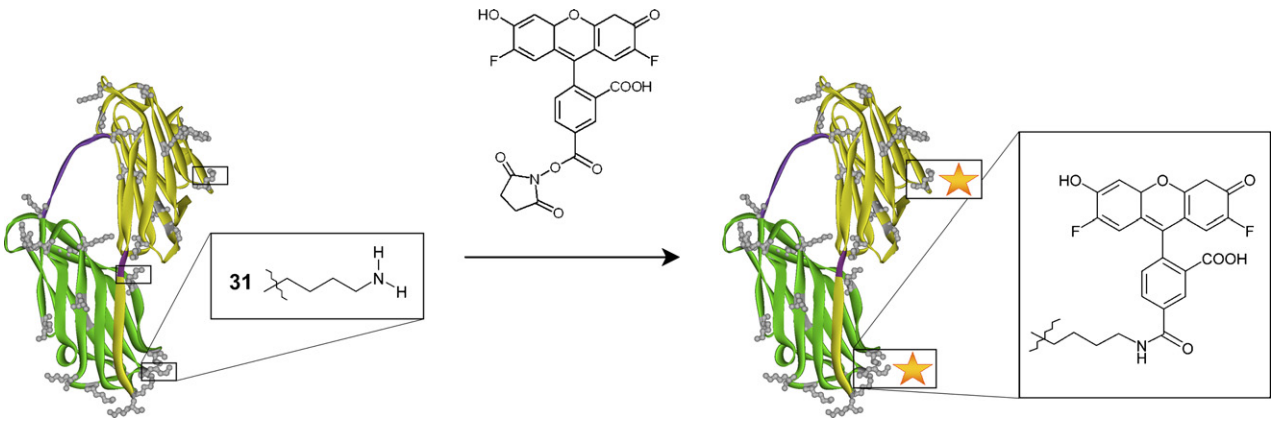

Fig. 2. Fluorescent labeling of the collagen-binding protein CNA35 (PDB ID: 2F68) with Oregon Green 488 succinimidyl ester results into the collagen-binding probe CNA35–OG488.

the peptide mass increased with 394 Da per ligated OG488 label. Tandem MS data were analyzed using Bioworks Browser 3.0 (Xcalibur, ThermoFinnigan), allowing for modification of both lysine and hydroxyl containing amino acids.

3. Results and discussion

We recently reported the application of the fluorescently labeled collagen-binding protein CNA35 as an excellent probe for the visualization of collagen in tissues and life cell cultures [10,12,13]. Treatment of CNA35 with 10 equiv. of Oregon Green 488 succinimidyl ester was shown to result in the conjugation of on average three fluorescent labels, despite the presence of 31 surface accessible lysine residues (Fig. 2). Since the collagen-binding properties of CNA35 were not affected by fluorescent labeling [10], the lysine residues present at the collagen-binding site were probably not reactive. These properties and the availability of a high resolution X-ray structure [14,15] make CNA35 an ideal protein to validate our method of using a combination of fluorescence and MS analysis of tryptic peptide digests. Fig. 3 shows the ESI-MS spectra of CNA35 obtained before (A) and after labeling with OG488 succinimidyl ester (B). Deconvolution of the mass spectrum of OG488-labeled CNA35 showed the presence of protein with one to four OG488 labels per protein, which is consistent with the average labeling ratio of three per CNA35 obtained using optical spectroscopy [10].

Because trypsin cleaves specifically at lysines and arginines, tryptic digestion is ideally suited to probe lysine modifications. Complete digestion of all unmodified lysine residues would result in fluorescent peptide fragments without any non-labeled lysine residues, and in this way remove a possible source of ambiguity. Optimal digestion conditions were first established for unlabeled CNA35 protein (Supporting information, Fig. S1). Tryptic digestion for 1 day at 37 °C in 20% (v/v) acetonitrile and pH 8.5 generated peptide fragments covering most of the CNA35 protein sequence (84%). Extension of the digestion time to up to 5 days did not further increase the amount of protein coverage. The same digestion conditions were then used for the CNA35–OG488 protein and the peptide mixture was analyzed using a fluorescence detector connected to the LC–MS system. The chromatogram obtained using fluorescence detection (Fig. 4A) was very different from the total ion current chromatogram (Fig. 4B). This finding illustrates the importance of using fluorescence detection for identifying labeled peptide fragments. Unlike mass spectrometry [7,8], the peak areas in fluorescence can be used to quantify the relative amount of each peptide, and therefore the reactivity of each lysine present in the protein. To determine the peak areas complete peak separation

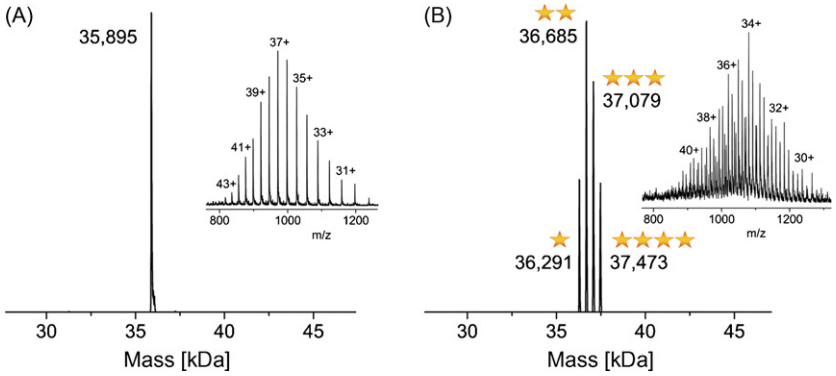

Fig. 3. (A) Mean protein mass ($n = 5$) of intact CNA35 protein (expected MW 35,905 Da) and (B) labeled CNA35–OG488 as measured with ESI-MS. The ligation of OG488 succinimide (MW 509 Da) is illustrated by the stars, where each star corresponds to an additional mass of 394 Da. The insets show the corresponding mass spectra with indicated some of the charges.

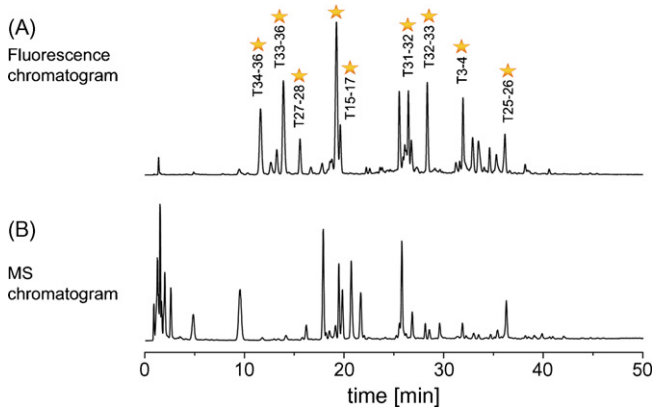

Fig. 4. (A) Fluorescence chromatogram (excitation at 496 nm and emission at 524 nm) and (B) MS chromatogram of CNA35–OG488 fragmented by tryptic digestion (1:50 (w/w)) for 21 h at 37 °C in 100 mM Tris–HCl buffer, pH 8.5, 20% (v/v) acetonitrile. After LC separation, the peptide mixture was analyzed by the fluorescence detector immediately followed by ESI-MS. Peak separation was optimized for the fluorescent peptides specifically using a gradient from 15% to 35% acetonitrile in 40 min. The assignment of the most intense fluorescent peaks is shown in Table 1 and relates to the marked peaks in the fluorescence chromatogram.

is required. Compared to the tryptic digest of unlabeled CNA35 (Supporting information, Fig. S1), however, poorer peak separation was noticed for the labeled peptides and the gradient needed to be adjusted to less than 1% acetonitrile per minute (Fig. 4).

In the assignment of the fluorescent peaks we focused on the peaks with at least 1.5% of the total fluorescence intensity (Table 1). Several of the small peaks with <1.5% intensity could be identified too, but they were not considered further in our analysis of labeling preferences. Most of the peaks in Table 1 could be assigned to the modification of a specific lysine residue as their mass corresponded to a unique peptide fragment containing only a single OG488 label. Despite the relatively high cleavage coverage for trypsin, several peptide fragments were identified with missed cleavage-sites. In that case tandem MS analysis was used to further fragment the specific peptide fragment to allow assignment of the lysine residue that was modified (Table 1: peaks 1–3, 12, and 14; see Fig. S2 for an example). Tandem MS analysis was also used to confirm that OG488 labeling had exclusively occurred on lysine residues for all peptides listed in Table 1, no peptides were identified that contained modification of, e.g., hydroxyl side chains. Only for peaks 6 and 7 the modified residue was not identified. The large size of these peptides, 3369 Da and 7768 Da, respectively, yielded very complex product ion spectra that prevented unique assignments.

Fig. 5 shows the overall structure of the CNA35 protein with all lysine side chains depicted in ball-and-stick representation. Color-coding was used to illustrate the relative reactivity of each lysine based on the intensity of their fluorescent peaks (Table 1). Although lysine modification was detected for 11 lysine residues distributed over the entire protein, two 'hot-spots' for fluorescent labeling can be identified on CNA35. As shown in Fig. 5, one site is located in a flexible loop in the N-terminal subdomain (in green) around lysine K30. The most prominent site of fluorescent labeling are the five-lysine residues present at the C-terminus of CNA35. The absence of this C-terminal part in the X-ray structure indicates a high flexibility [15], suggesting

Table 1
Identification of the labeled lysines in trypsin-digested CNA35–OG488 using fluorescence and LC–MS

Peak#	Fluorescence		MS			Assignment			
	t_R[a] (min)	A[a,b] (%)	t_R[a,c] (min)	z measure	M measure[d] (Da)	Peptide[e]	M calc[f] (Da)	Total labels	Labeled K#[g]
1	11.64	5.01	11.74	1, 2, 3	1199	T34–36	1198	1	323
2	13.24	2.22	13.36	2, 3	1468	T33–35	1469	1	320
3	13.90	6.65	14.08	2, 3	1539	T33–36	1539	1	320
4	15.56	2.86	15.74	2, 3	1483	T27–28	1484	1	265
5	19.21	15.2	19.04	1	515	OG488	509		
6	19.56	3.73	19.63	3, 4, 5	3369	T15–17	3370	1	130 or 139[h]
7	25.57	5.66	25.57	6, 7, 8, 9	7768	T9–14	7766	3	Unknown[h]
						T14–19	7766	1	Unknown[h]
8	26.51	5.26	26.82	3, 4	2945	T31–32	2945	1	313
9	26.80	2.56	26.84	2, 3	2502	T15–16	2503	1	130
10	28.40	6.46	28.55	1, 2, 3	1179	T32–33	1180	1	317
11	31.97	6.15	31.87	2, 3	2657	T3–4	2658	1	30
12	32.94	3.40	32.99	2, 3	1934	T33–36	1933	2	320 and 326
13	33.51	3.58	33.49	2, 3, 4	2912	T12–13	2912	1	101
14	34.63	1.65	34.67	2, 3, 4	2361	T32–36	2361	2	317 and 320
15	35.31	2.15	35.36	2, 3	2339	T8–9	2340	1	57
16	36.15	3.58	36.26	3, 4	3647	T25–26	3648	1	241

[a] Values are the average over four measurements.
[b] Relative peak area measured by the fluorescence detector, where total peak area of all fluorescent peaks corresponds to 100%.
[c] The retention time measured by MS was determined by plotting the mass chromatogram for each peptide using the z values shown in the adjacent column.
[d] The measured mass was determined by deconvolution of the mass spectra.
[e] Tryptic peptides were denoted from T1 to T36, where T1 originated from the most N-terminal site of CNA35 and T36 from the most C-terminal site.
[f] The calculated average mass determined in the digest simulation, which included up to five missed cleavage-sites and a varying amount of OG488 labels from one to five, equal to the number of present lysines.
[g] Residue number of OG488-labeled lysines.
[h] Product ion spectra obtained with tandem MS analysis were too complex to identify the modified lysines.

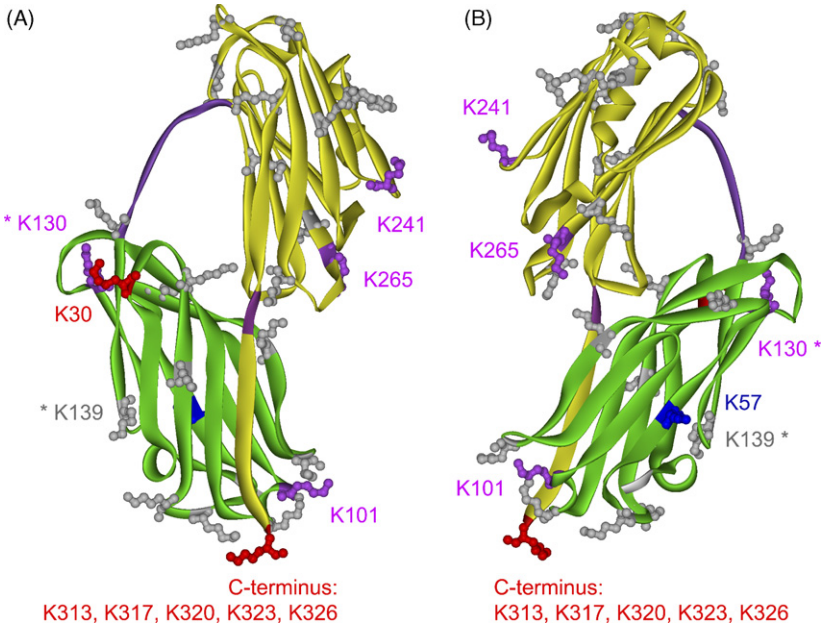

Fig. 5. (A) Frontview and (B) backview of the CNA35 X-ray structure with the lysine side chains shown in ball-and-stick representation (PDB ID: 2F68). The color of the lysines indicates their relative peak area in the fluorescence chromatogram (Table 1). Red, >5.0%; purple, 2.5–5.0%; blue, 1.5–2.5%; grey, <1.5%. Residues 313–327 at the C-terminus are not defined in the X-ray structure, but they do contain five lysine residues that were highly reactive: K313 (5.3%), K317 (7.3%) K320 (11.4%), K323 (5.0%), K326 (1.7%). The * indicates that either K130 or K139 is more favored for labeling than implied by their colors, as peak 6 is not included in the assignment.

that flexible regions in a protein structure are highly reactive towards reaction with succinimide ester dyes. Remarkably, none of the lysines located around the collagen-binding cavity were modified, which is consistent with our previous finding that the affinity for collagen was not affected after fluorescent labeling [10].

4. Conclusion

ESI-MS analysis of the fluorescently labeled collagen-binding probe CNA35 showed a heterogeneous mixture of CNA35 proteins with one to four Oregon Green 488 labels per protein. To identify the exact locations of the fluorescent labels, the modified protein was treated with trypsin and characterized with LC–MS directly connected to a fluorescence detector. MS was used for the identification of the labeled lysine residues in the probe and fluorescence detection allowed their quantification, as the peak area of the fluorescent peaks gives a good estimation of the relative amount of each labeled lysine in the digestion mixture. Combining both techniques revealed two preferred sites for fluorescent labeling in the CNA35 probe, where the 'labeling hot-spots' were primarily located in flexible regions of the protein. The approach presented here is a convenient way to validate newly designed protein-based fluorescent sensors, especially when the conjugates, like the fluorescein-based OG488, suppress the ionization process.

Acknowledgements

The authors thank M. Breurken for the expression, purification and labeling of the CNA35 protein and E.W. Meijer for general support (Eindhoven University of Technology, The Netherlands).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jchromb.2007.12.026.

References

[1] F. Winau, O. Westphal, R. Winau, Microbes Infect. 6 (2004) 786.
[2] F. Baud, S. Karlin, Proc. Natl. Acad. Sci. U.S.A. 96 (1999) 12494.
[3] L. Lins, A. Thomas, R. Brasseur, Protein Sci. 12 (2003) 1406.
[4] R. Aebersold, M. Mann, Nature 422 (2003) 198.
[5] B. Domon, R. Aebersold, Science 312 (2006) 212.
[6] E.P. Romijn, J. Krijgsveld, A.J.R. Heck, J. Chromatogr. A 1000 (2003) 589.
[7] M.O. Glocker, C. Borchers, W. Fiedler, S. Suckau, M. Przybylski, Bioconjugate Chem. 5 (1994) 583.
[8] D.J. Janecki, R.L. Beardsley, J.P. Reilly, Anal. Chem. 77 (2005) 7274.
[9] A. Scholten, N.F.C. Visser, R.H.H. van den Heuvel, A.J.R. Heck, J. Am. Soc. Mass Spectrom. 17 (2006) 983.
[10] K. Nash Krahn, C.V.C. Bouten, S. van Tuijl, M.A.M.J. van Zandvoort, M. Merkx, Anal. Biochem. 350 (2006) 177.
[11] Z. Zhang, A.G. Marshall, J. Am. Soc. Mass Spectrom. 9 (1998) 225.
[12] R.A. Boerboom, K. Nash Krahn, R.T. Megens, M.A. van Zandvoort, M. Merkx, C.V. Bouten, J. Struct. Biol. 392 (2007) 392.
[13] R.T.A. Megens, M.G.A. Oude Egbrink, J. Cleutjens, M.J.E. Kuijpers, P.H.M. Schiffers, M. Merkx, D.W. Slaaf, M.A.M.J. van Zandvoort, Mol. Imaging 6 (2007) 247.
[14] J. Symersky, J.M. Patti, M. Carson, K. House-Pompeo, M. Teale, D. Moore, L. Jin, A. Schneider, L.J. DeLucas, M. Höök, S.V.L. Narayana, Nat. Struct. Biol. 4 (1997) 833.
[15] Y. Zong, Y. Xu, X. Liang, D.R. Keene, A. Höök, S. Gurusiddappa, M. Höök, S.V.L. Narayana, EMBO J. 24 (2005) 4224.