Supporting information

Probing protein denaturation during size exclusion chromatography using native mass spectrometry

Iro K. Ventouri[1,2,3,4], Daniel B. A. Malheiro[1,3], Robert L. C. Voeten[1,2,3], Sander Kok[5], Maarten Honing[1,5], Govert W. Somsen[1,2], Rob Haselberg[1,2]*

1. Division of Bioanalytical Chemistry, AIMMS Amsterdam Institute of Molecular and Life Sciences, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands
2. Centre for Analytical Sciences Amsterdam, The Netherlands
3. TI-COAST, Amsterdam, the Netherlands
4. Analytical Chemistry Group, van't Hoff Institute for Molecular Sciences, University of Amsterdam, Amsterdam, the Netherlands
5. DSM Materials Science Center, Geleen, the Netherlands

Current affiliation Daniel B. A. Malheiro

MS-Omics, Vedbæk, Denmark

Current affiliation Sander Kok

ChemConnection B.V., Pivot Park, Oss, the Netherlands

Current affiliation Maarten Honing

M4I institute, Faculty Health, Medicine & Life Sciences, University Maastricht, the Netherlands.

* **Corresponding author**

Division of BioAnalytical Chemistry

AIMMS Amsterdam Institute of Molecular and Life SciencesVrije Universiteit Amsterdam

de Boelelaan 1085

1081 HV Amsterdam

The Netherlands.

E-mail: r.haselberg@vu.nl

Automated analysis of protein mass spectra. MS data evaluation has been performed with an in-house developed script in Matlab (version 2015b). The purpose of this script was the automated estimation of the fraction of folded myoglobin based on the charge state distribution of the DI-ESI-MS and SEC-ESI-MS experiments. The data obtained with DI-ESI-MS and SEC-ESI-MS were exported as ASCII files. To allow most accurate data evaluation first a baseline correction was performed. The latter was based on the asymmetric truncated cost function due to the sole presence of positive peaks which have shown to provide minimal mean square error values as compared to similar methods. The approach assumes $\mathbf{y}=\mathbf{b}+\boldsymbol{\varepsilon}$ where $\mathbf{b}$ and $\boldsymbol{\varepsilon}$ represent the background and the residuals, respectively. In addition, $\mathbf{b}$ is defined as a product of the Vandermonde matrix and the corresponding vector of the polynomial coefficients i.e. $\mathbf{b}=\mathbf{Vp}$. Instead of the least-squares approach, minimization of equation S1 is the considered criteria.

$$\boldsymbol{\delta}(\mathbf{p}) = \sum_{i=1}^{n} \boldsymbol{\phi}(\mathbf{y}_i - (\mathbf{Vp})_i) \qquad \textit{Equation S1}$$

Where φ represents the asymmetrical cost function and is defined as:

$$\forall \mathbf{x} \in \mathbf{R}, \quad \boldsymbol{\phi}(\mathbf{x}) = \begin{cases} \mathbf{x}^2 & \textbf{if } \mathbf{x} < \mathbf{t} \\ \mathbf{t}^2 & \textbf{otherwise} \end{cases} \qquad \textit{Equation S2}$$

The truncating value t ($1*10^{-5}$) ensured minimal influence of the peak heights with respect to the baseline model. Subsequently, each intensity value at a measured m/z is compared to a 3σ threshold in addition to the mean of the background. Every value below this threshold is regarded as background and the remaining values are regarded as signals. These signals were automatically tunemix (ESI-L, Agilent Technologies, Waldbronn, Germany) corrected, to ensure highest mass accuracy. The corresponding baseline-substracted and mass-corrected signals were considered true responses and related to the protein. Depending on the m/z value at which the signal is observed it was assigned to the denatured protein (below m/z 1700) or native protein (above m/z 1700). The ratio folded was subsequently calculated as the fraction of summed intensity caused by signals with an m/z above 1700 and the sum of all signals (see main text).

Figures & Description

Figure S1 shows the calibration curve of the TOSOH TSKgel® G2000SWXL column by plotting log M_w vs. the elution volume derived from a set of known analytes.

Figures S2 shows some relevant chromatographic descriptors obtained during SEC-UV analysis of myoglobin using a variety of conditions. Clearly, low ionic strength eluents lead to increased peak width and consequently lower plate numbers for most volatile eluents. Also, peak tailing increases with lowering of the ionic strength. Ammonium acetate shows the most similar behavior compared to the golden standard, sodium phosphate.

Figure S3 displays the chromatograms and subsequent myoglobin mass spectra as obtained using ammonium formate and bicarbonate at both low and high ionic strength. Clearly, under all conditions protein denaturation can be observed (in the mass spectra), as well as extensive protein-column interaction (peak tailing at low ionic strength).

Figure S4 represents the fraction folded myoglobin as obtained throughout the peaks displayed in Figure 5 (main text). The average value over ~0.1 min is plotted against the elution volume. Under high ionic strength conditions the fraction folded is by approximation 1, indicating that the system does not induce any structural changes. The opposite is true at low ionic strength. Here interaction between protein and column material (represented by the tailing peak and shifted peak maximum) leads to protein denaturation (shown as a decrease in fraction folded over the peak width).

Figure S5 displays the results of the SEC-MS experiments analyzing carbonic anhydrase and cytochrome C using the two examined columns in combination with ammonium acetate eluents of varying ionic strength. The chromatograms as well as the mass spectra of the two proteins obtained in the peak apex are presented under the various conditions. To study protein denaturation, the average charge state of each protein is plotted against the various ionic strength conditions. Higher values indicate that the protein is less compact and thus undergoes some degree of denaturation.

Figure S6 shows the chromatograms of myoglobin and the respective mass spectra obtained when the protein is analyzed using column 2 in combination with ammonium acetate eluents of various ionic strength.

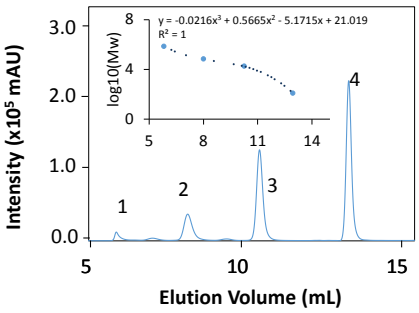

Figure S1. Calibration curve obtained on the TOSOH TSKgel® G2000SWXL column using thyroglobulin (1), bovine serum albumin (2), myoglobin (3) and uracil (4). Eluent: sodium phosphate (0.1 M, pH 6.9) containing sodium sulfate (0.1 M), and sodium azide (0.05%).

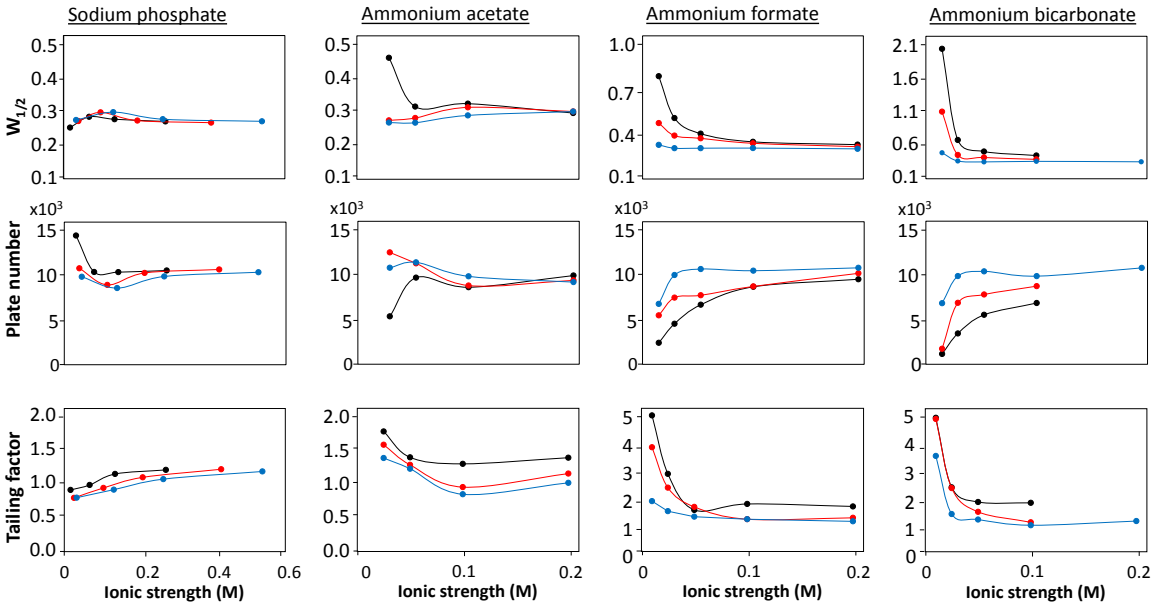

Figure S2. Peak width at half maximum (in min), plate number and tailing factor for myoglobin analyzed by SEC-UV using the conditions (salt type; pH and ionic strength of eluent) as provided in Figure 1. Eluent pH, 5.9 (black), 6.9 (red), and 7.5 (blue). The connecting lines between the points highlight trends and are not obtained by fitting.

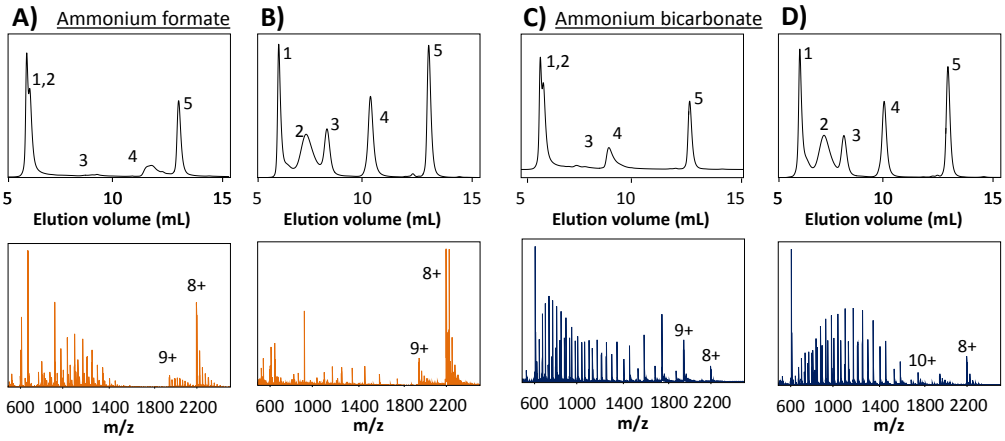

Figure S3. SEC-ESI-MS of the protein test mixture containing thyroglobulin (1), γ-globulin (2), ovalbumin (3), myoglobin (4), and vitamin B12 (5) using an eluent of (A) 0.01 M ammonium formate (pH 7.5), (B) 0.2 M ammonium formate (pH 7.5), (C) 0.01 M ammonium bicarbonate (pH 7.5), and (D) 0.2 M ammonium bicarbonate (pH 7.5). Average mass spectra of the myoglobin peaks are provided.

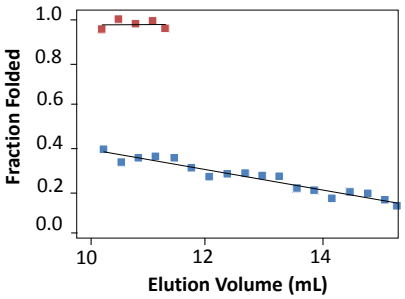

Figure S4. Fraction folded myoglobin measured over the entire peak width as obtained during SEC-ESI-MS using 0.01 M (blue trace) and 0.2 M (red trace) ammonium acetate (pH 6.9) as eluent. The corresponding chromatograms are in Figure 5. Data points represent the average fraction of approximately every 0.1 min. The lines indicate the trend of the data points and is not obtained through fitting.

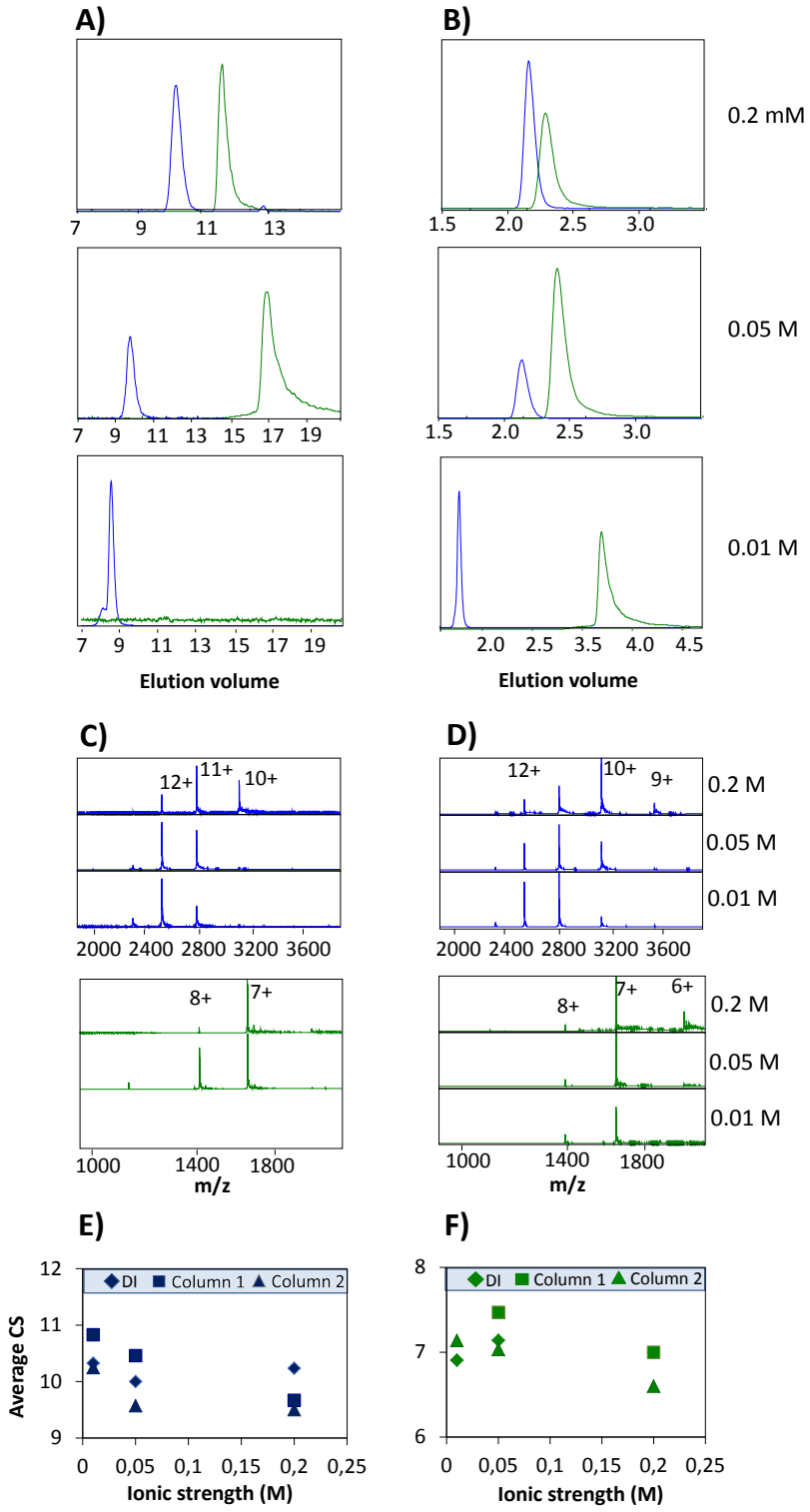

Figure S5. SEC-ESI-MS of the proteins carbonic anhydrase (blue traces) and cytochrome c (green traces). Chromatograms (A, B), mass spectra (C, D) and average charge states (E, F) obtained for the two proteins using the TOSOH TSKgel® G2000SWXL column (A, C; column 1) and AdvanceBioSEC column (B, D; column 2) in combination with ammonium acetate eluents of varying ionic strength are displayed. Average charge state as obtained with direct infusion (DI) are added to panels E and F for comparison.

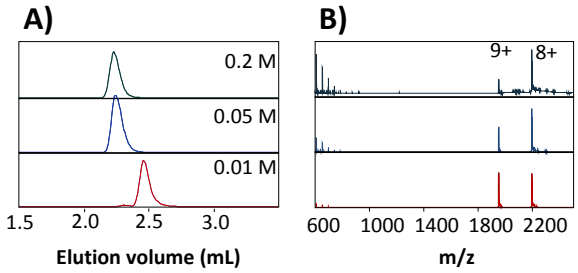

Figure S6. SEC-ESI-MS of myoglobin using the AdvanceBioSEC column (column 2) in combination with ammonium acetate eluents of varying ionic strength. Chromatograms (A) and mass spectra (B) are displayed.