Protein analysis by aqueous size exclusion chromatography

Application

Heinz Goetz

Abstract

Aqueous size exclusion chromatography (SEC), often referred to as gel filtration chromatography (GFC), is a entropically controlled separation technique in which molecules are separated based on hydrodynamic volume or size. In protein analysis, the technique is used for different applications such as:

- impurity testing,
- following progress of a reaction,
- separation of reaction components and products, (antibodies, fragments, and conjugates),
- folding studies,
- purification, and
- desalting and exchange of sample buffer.

In all these application areas, except for the last one, the protein molecular mass is typically determined in the same analysis. To obtain this important parameter a proper column calibration with well-characterized macromolecules or molecular mass sensitive detectors, such as mass spectrometry or light scattering, are used. This Application Note describes results in protein impurity testing and simultaneous molecular mass characterization with the Agilent 1100 Series GPC-SEC analysis system.

Agilent Technologies

Introduction

Using recombinant DNA techniques to obtain proteins through high-level expression in bacterial cells, rather than extraction and purification techniques from other biological resources, has become very popular lately and has brought forth a growing number of proteins of interest.
SEC is a mainstay technique for the analysis of proteins due to its excellent properties for impurity testing, quantification, molecular weight characterization and automation. The technique can handle a very wide range of molecular weights, from several hundred to about 10 million Dalton. A 50—100 % difference in the molecular weight is required. With the Agilent 1100 Series GPC-SEC analysis system the area/amount report, for example, on impurities, is obtained at the same time as the molecular weight report. Molecular weight accuracy and precision are typically better than 2 %. Another advantage is that the protein is not denatured and can easily be collected for further investigations.

Although protein analysis technologies are developing fast, many still rely on traditional methods such as sodium dodecylsulfate-polyacrylamide gel electrophoresis (SDS-PAGE), which include a number of laborious, time-consuming manual steps. It requires only a 10 % difference in molecular weight but offers limited repeatability in quantification.

Another technique is offered by the Agilent 2100 bioanalyzer with the Protein 200 LabChip®, which is a compact, electrophoresis based system for rapid and automated analysis of proteins on a chip. Here also, only a 10 % difference in molecular weight is necessary. The molecular weights can range from approximately 14000 to 200000 Dalton. The protein is denatured during the analysis and cannot be collected.

Excellent mass accuracy of 0.1 % is obtained by mass spectroscopy, for example, by MALDI-MS (Matrix-Assisted Laser Desorption Ionization MS), however, it cannot distinguish between the monomer and oligomers and is limited to molecular masses up to approximately 120000 Dalton. This technique also requires a significant financial investment.

Experimental

Chromatography was performed on an Agilent 1100 Series GPC-SEC analysis system equipped with a micro vacuum degasser, an isocratic pump with seal wash option, a thermostatted autosampler, a thermostatted column compartment, a diode array-detector and a fluorescence detector. The data was transferred to an HPLC 3D ChemStation with GPC-SEC data analysis software.
The columns were purchased from Agilent Technologies GmbH, Germany. These included TSK SW 2000xl, 7.8 × 300 mm, 5 µm, (Agilent part number 79912S2-597), TSK SW 3000xl, 7.8 x 300 mm, 5 µm (Agilent part number 79912S3-597) and TSK 4000 SW, 7.5 x 300 mm, 10 µm (Agilent part number 79912S4-197). All commercially available proteins of purest analytical grade were purchased from Sigma (Munich, Germany). The following mobile phases were prepared:

Eluent A: 0.1 M Na-Sulfate, 0.1 M Na-phosphate, pH 6.7
Eluent B: 0.40 M NaCl, 10 mM Na-phosphate, pH 7.4
Eluent C: 0.15 M NaCl, 20 mM Na-phosphate, 0.01 % Tween 20, pH 6.8

Results and Discussion

Calibration

SEC calibration of the chromatographic system is performed by injecting macromolecule standards with known molecular weight and then connecting the data points. The graph shows the retention time/elution volume on the x-axis and the logarithm of standard's molecular weight on the y-axis[2]. Commercially available proteins with the molecular weights as specified by the supplier were used. Depending on protein, type of column, buffer and buffer concentration, the hydrodynamic volume changes and therefore a mixed mode separation mechanism, for example, SEC mixed with IEC (ion exchange chromatography), occurs[3]. To ensure a widely SEC-based separation it is important to carefully determine correct conditions for each standard and sample protein.

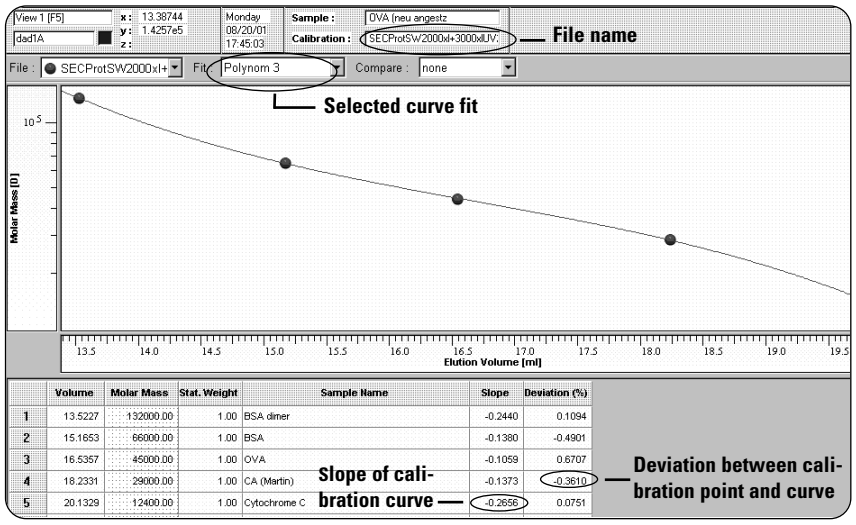
The proteins listed in table 1 fulfilled the criterion of a widely size based separation in the various mobile phases. The peptides Angiotensin II (molecular weight MW 1046 Da) and Leucin-Enkephalin (MW 555 Da) could not be used because they eluted with all three eluents with or even behind the amino acid tyrosine (MW 180). This means, they did not elute at all in the SEC mechanism. To determine the best calibration curve it is extremely useful that the Agilent ChemStation GPC-SEC data analysis software offers a wide variety of curve fits, such as linear, 2nd to 7th fits and three special ones – PSS 3, PSS 5 and PSS 7. The best curve fit for the data points was determined by the following criteria:

- Curve shape must be physically meaningful.
- Curve slope must be negative for all points (as predicted in theory).
- Small deviation between each measured data point and curve. Deciding on the the curve fit simply based on an average parameter such as the chi-square value or the coeffcient of regression is not recommended. Such a parameter is only an average value for all data points and does not provide detailed information at a certain elution volume.

Protein/amino acid	Molecular weight (Da)	Remark
Dimer of bovine serum albumin	132000	Present in serum albumin
Dimer of ovalbumin	86000	Present in ovalbumin
Bovine serum albumin	66000	
Ovalbumin	43000	
Carbonic anhydrase	29000	
Cytochrome C	12400	Strongly tailing with eluent C
Lysozyme	14300	
Tyrosine	180	

Table 1
Compounds used to calibrate the Agilent 1100 Series GPC-SEC analysis system

Figure 1 shows the calibration curve and table used for characterizing the erythropoietin. All important information, such as calibration file name, selected curve fit and calibration point data including their deviation from the curve itself is shown. Such clearly presented information makes the correct decision on curve fit easy and fast.



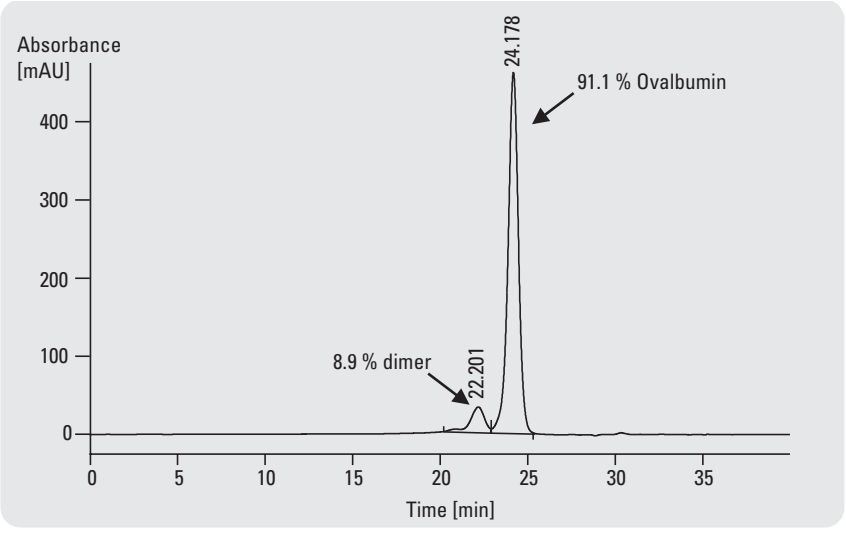
	Volume	Molar Mass	Stat. Weight	Sample Name	Slope	Deviation (%)
1	13.5227	132000.00	1.00	BSA dimer	-0.2440	0.1094
2	15.1653	66000.00	1.00	BSA	-0.1380	-0.4901
3	16.5357	45000.00	1.00	OVA	-0.1059	0.6707
4	18.2331	29000.00	1.00	CA (Martin)	-0.1373	-0.3610
5	20.1329	12400.00	1.00	Cytochrome C	-0.2656	0.0751

Figure 1
Typical calibration curve and table obtained with commercially available proteins

Columns	TSK SW 2000xl in series with TSK SW 3000xl
Eluent B	0.40 M NaCl, 10mM Na-phosphate, pH 7.4
Flow rate	0.8 ml/min
Column compartment temperature	42 °C
Injection volume	100 µl
Sample preparation	Proteins were dissolved in the mobile phase (concentration 0.1%)

Impurity testing and molecular weight characterization of commercially available proteins

Several commercially available proteins were analyzed by aqueous SEC for impurities and molecular weight. The advantage of aqueous SEC is that both types of information are obtained with good precision and accuracy from a single, fully automated analysis. Automation here includes the complete process from sample injection to chromatographic analysis, impurity identification and quantification, protein molecular weight calculation and even printing of the user-specified reports. Figure 2 shows the SEC chromatogram of an ovalbumin sample with the software calculated area % data. The impurity is the ovalbumin dimer, as identified easily by the molecular weight. Figure 3 shows the user-configured molecular weight report for the sample. The molecular weight data of ovalbumin monomer and -dimer match agree well with the reference molecular weights (figure 3 and table 1). Besides the molecular weight M_p, derived from the peak apex, the software also calculates the polydispersity D and other molecular weight data which are derived from synthetic polymer characterization. The polydispersity value D is also of interest in protein analysis. A D-value smaller than 1.01 typically indicates peak purity. Of course, this requires a minimum separation of the compounds often not visible from the peak shape. In the example here, the D value for the monomer is 1.0145, for the dimer it is 1.02. Polydispersity D is calculated by dividing the weight average molecular weight M_w by the number average molecular weight M_n.



Columns	TSK SW 3000xl in series with TSK SW 4000xl
Eluent A	0.1 M sulfate, 0.1 mM Na-phosphate, pH 6.7
Flow rate	0.8 ml/min
Column compartment temperature	22 °C
Injection volume	100 µl
Sample preparation	Proteins were dissolved in the mobile phase (concentration 0.1%)

Figure 2
SEC chromatogram of an ovalbumin sample with the software calculated area % values for ovalbumin and an impurity, ovalbumin dimer

Figure 4 shows the analysis of a 100 % pure commercial beta-lactoglobulin including the identity confirmation by the molecular weight. Peak purity was confirmed by:

- Visual inspection of the peak shape.
- A polydispersity D of 1.007, which is significantly smaller than 1.01.
- The overlay of the peak spectra acquired at inflection points and apex. The molecular weight (Mp = 35600) is in very good agreement with the reference value (M = 35000) from gel electrophoresis.

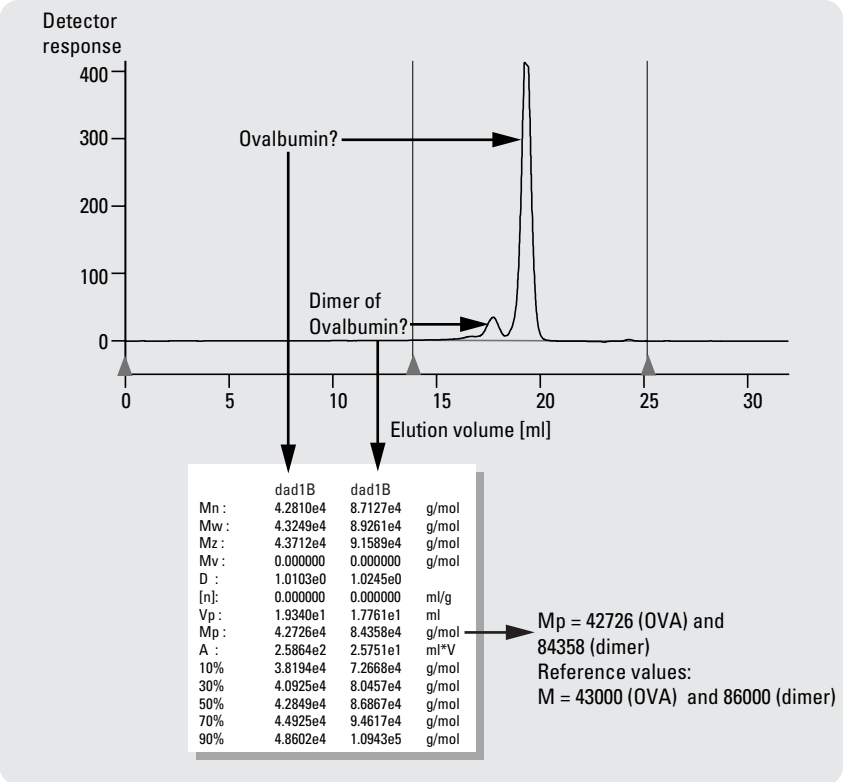


Figure 3
SEC report for ovalbumin and ovalbumin dimer

Columns	TSK SW 2000xl in series with TSK SW 3000xl
Eluent C	0.15 M NaCl, 20mM Na-phosphate, 0.01 % Tween 20, pH 6.8
Flow rate	0.8 ml/min
Column compartment temperature	42 °C
Injection volume	100 µl
Sample preparation	Proteins were dissolved in the mobile phase (concentration 0.1%)

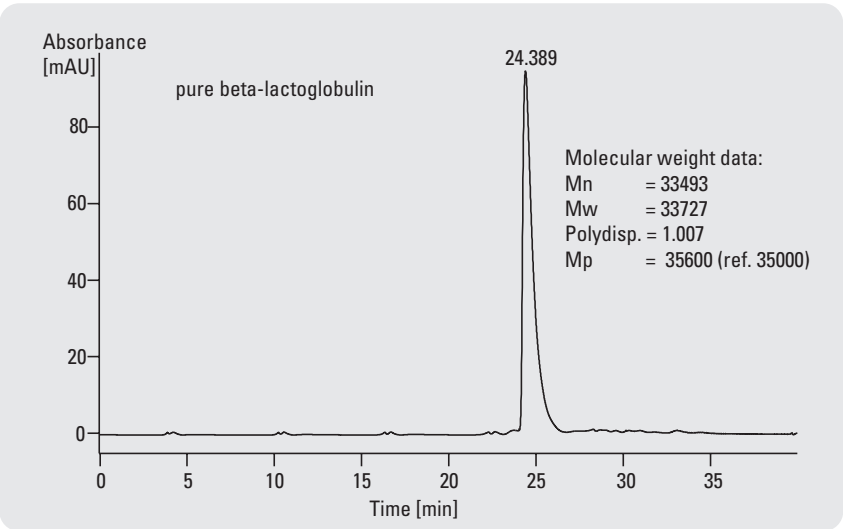


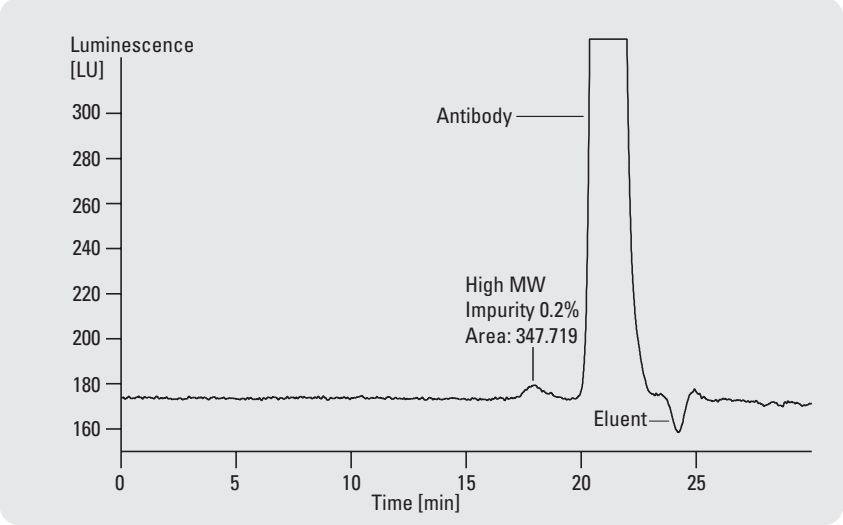
Figure 4
Protein impurity testing and identity confirmation of a beta-lactoglobulin sample

Protein analysis of research samples

1. Impurity testing and molecular weight characterization of a human antibody

UV detection at 214 or 278 nm is most commonly used for protein analysis. Fluorescence detection is also of interest due to improved selectivity and sensitivity. Protein fluorescence derives from the aromatic amino acids tryptophan and tyrosine, which are present in almost every protein. The UV detector used should be a diode-array detector (DAD) because it acquires several wavelengths and spectra at the same time. Figure 5 shows the SEC analysis of a human antibody sample with fluorescence detection at an excitation wavelength of 220 nm and an emission wavelength of 350 nm. The large antibody peak and a small impurity of only 0.2 % can be seen. The impurity could not be detected with the DAD at 214 nm due to the small concentration. A typical excitation and emission fluorescence protein spectrum, obtained with the Agilent 1100 Series fluorescence detector (FLD) without stopping the flow is shown in figure 8.

In addition, there was good agreement between molecular weight (M_p = 149000) calculated by SEC and the reference value (M = 157000) from the Agilent 2100 bioanalyzer. The difference of 5 % is mainly due to the fact that SEC separates under native conditions while the Agilent 2100 bioanalyzer uses the denaturing sodiumdodecylsulfate (SDS).



Columns	TSK SW 3000xl in series with TSK SW 4000xl
Eluent A	0.1 M Na-sulfate, 0.1 mM Na-phosphate, pH 6.7
Flow rate	0.8 ml/min
Column compartment temperature	22 °C
Injection volume	100 µl
Sample preparation	Proteins were dissolved in the mobile phase (concentration 0.1%)

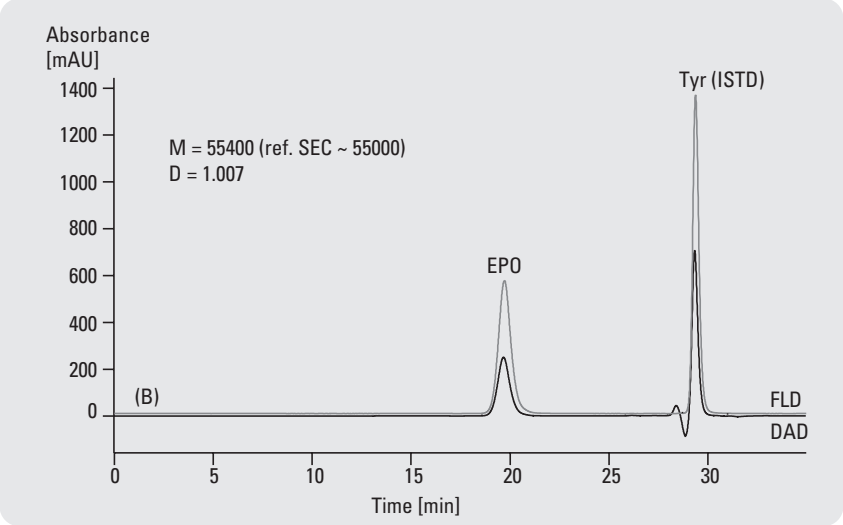
Figure 5
Sensitive fluorescence detection of an human antibody and an impurity

2. Impurity testing and molecular weight characterization of Erythropoietin (EPO)

Erythropoietin (EPO) is the main factor responsible for regulating red blood cell production during steady-state conditions and accelerating recovery of red blood cell mass following hemorrhage. EPO synthesis in adult organisms takes place in the kidney. It is a 165 amino acid glycoprotein with a molecular weight of approximately 30000 to 35000 Dalton.

EPO can nowadays be produced by a recombinant process and is used to treat anemia associated with chronic renal failure for patients on dialysis. It has become widely known in the context of misuse by athletes for blood-doping in endurance sports.

Figure 6 shows the overlay of a simultaneous diode array and fluorescence detector analysis of an EPO sample. For medical use it is mandatory that the sample is highly pure, that it does not form aggregates and that the molecular weight is correct. Therefore,by using two different detectors, additional confidence is gained.



Columns	TSK SW 2000xl in series with TSK SW 3000xl
Eluent B	0.40 M NaCl, 10 mM Na-phosphate, pH 7.4
Flow rate	0.8 ml/min
Column compartment temperature	42 °C
Injection volume	100 µl
Sample preparation	Proteins were dissolved in the mobile phase (concentration 0.1%)

Figure 6
Impurity testing of erythropoietin (EPO) with UV-diode-array- and fluorescence detection

Besides looking at the chromatographic signals, purity was also demonstrated by the overlay of the UV spectra acquired at inflection points and apex of the EPO peak (figure 7), and polydispersity D = 1.007 Figure 8 shows the excitation and emission spectra acquired with the Agilent 1100 Series FLD. The molecular weight M_p = 55400 agreed very well with the reference value M_p of 55000 which was determined in the supplier's laboratory, also using SEC. It should be noted that 55000 is not the "true" molecular weight of this EPO sample. Since it is glycosilated by 40 % and therefore larger and very hydrophilic it elutes earlier than an EPO without glycolization. With MALDI-MS a molecular weight of 30000 Dalton was determined. The disadvantage of MS-based techniques in protein analysis is that aggregates cannot be distinguished from the monomeric protein.

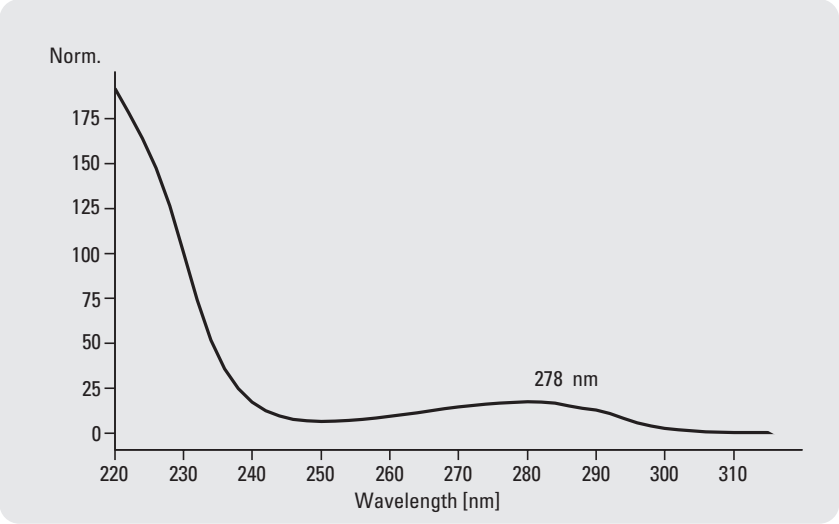


Figure 7
Overlay of UV spectra acquired at inflection points and apex of EPO peak which represents a further indication of peak purity

Columns	TSK SW 2000xl in series with TSK SW 3000xl
Eluent B	0.40 M NaCl, 10 mM Na-phosphate, pH 7.4
Flow rate	0.8 ml/min
Column compartment temperature	42 °C
Injection volume	100 µl
Sample preparation	Proteins were dissolved in the mobile phase (concentration 0.1%)

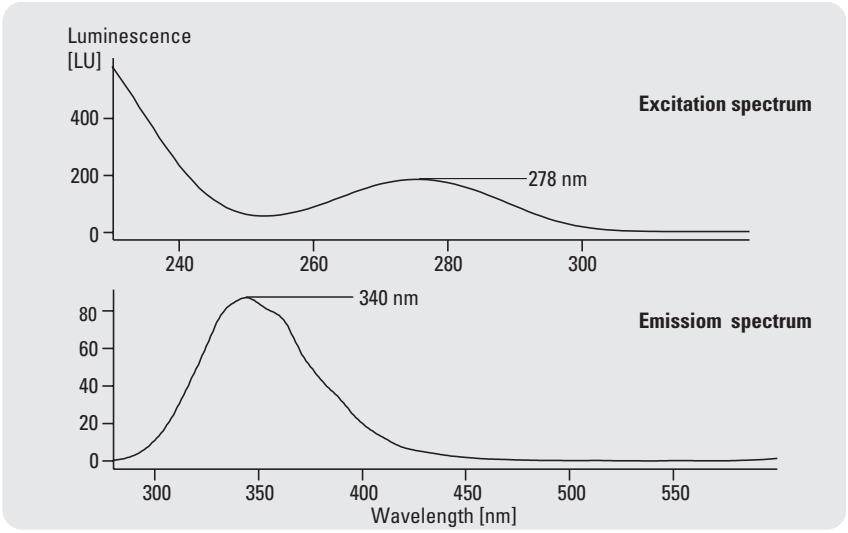


Figure 8
Excitation and emission spectra of EPO

Columns	TSK SW 2000xl in series with TSK SW 3000xl
Eluent B	0.40 M NaCl, 10 mM Na-phosphate, pH 7.4
Flow rate	0.8 ml/min
Column compartment temperature	42 °C
Injection volume	100 µl
Sample preparation	Proteins were dissolved in the mobile phase (concentration 0.1%)

3. Protein identification and molecular weight characterization of proteins in human plasma

Figure 9 shows the analysis of a human plasma sample using a DAD and FLD within one run. Eluent A and one TSK SW3000xl column were in series with one TSK SW 4000-column. Due to the many proteins, which are present in plasma, a full separation of all compounds was not expected. Despite the limited resolution the largest peak could be easily identified as human albumin by molecular weight comparison. The molecular weight of 65500 Dalton, determined by SEC, agrees very well with the reference value of 66000 Dalton. The first peak eluting is obviously an immunoglobuline with a molecular weight of 717000 Dalton. To obtain a better separation, especially in the area between the immunoglobuline and the albumin peak a further column, for example, a TSK SW 2000xl column should be installed.

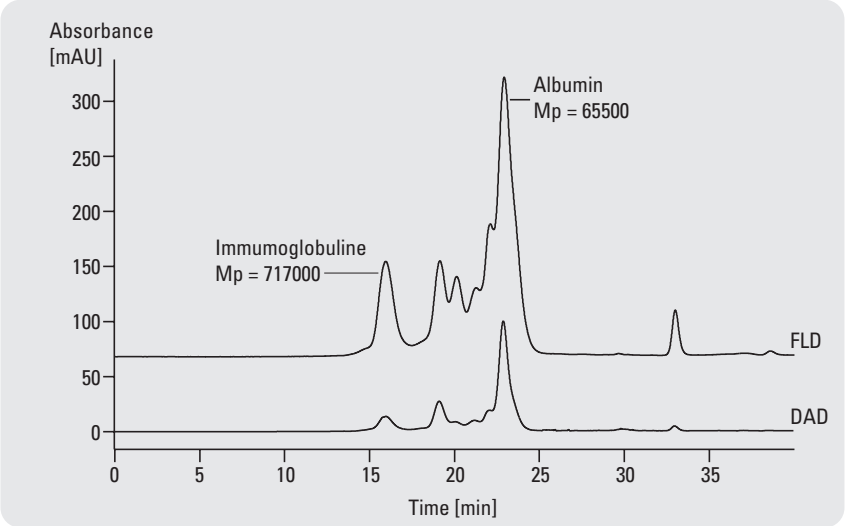


Figure 9
Protein identification in human plasma by molecular weight

Table 2 shows a comparison of the molecular weight results obtained by aqueous SEC and the reference methods. The difference between the methods is typically smaller than 5 %, in most cases even smaller than 2 %. The larger difference between the MALDI-MS and the SEC-value for EPO is due to the 40 % glycolization of the sample, which makes it larger and very hydrophilic. In SEC this results in smaller elution volumes but does not effect the MALDI experiment. There was a good agreement (only 1.8 % difference) between the EPO molecular weight determined in our lab and the reference value determined in the supplier's lab using SEC.

Columns	TSK SW 3000xl in series with TSK SW 4000xl
Eluent A	0.1 M Na-sulfatel, 0.1 mM Na-phosphate, pH 6.7
Flow rate	0.8 ml/min
Column compartment temperature	22 °C
Injection volume	100 µl
Sample preparation	Proteins were dissolved in the mobile phase (concentration 0.1%)

Protein	Reference MW	SEC MW (Mp)	Delta(%)
Myoglobin	17000 (PAGE)	17100/Eluent A	0.5
Ovalbumin	43000 (PAGE)	42700/Eluent A	0.7
Ovalbumin dimer	86000 (PAGE)	84400/Eluent A	0.7
Human albumin in plasma	66000 (PAGE)	65000/Eluent A	1.5
Antibody	157000 (2100 bioanalyzer)	149000/Eluent B	5.0
Ferritin	460000 (PAGE)	459600/Eluent A	0.1
EPO	approx. 55000 (SEC)	54000/Eluent B	1.8
	approx. 30000(MALDI)		80

Table 2
Comparison of molecular weight data obtained by aqueous SEC and reference methods

Conclusions

Aqueous SEC is an important technique for protein analysis. In a single run, quantitative information on the degree of dimerization/oligomerization, lower molecular weight impurities and the molecular weight is obtained. Molecular weight accuracy and precision are typically better than 2 %.

Information on purity is obtained not only by the chromatographic signals of the UV diode-array detector and the fluorescence detector but also by the polydispersity value calculated by the software and by overlaying spectra. Low concentration impurities can be quantitatively determined in the presence of the main compound. The molecular weight calibration of the Agilent 1100 Series GPC-SEC analysis system is performed with commercially available, well-characterized proteins. Since proteins are used for calibration, "absolute" detection such as cumbersome light scattering or viscosimetry, is not needed. The technique requires a minimum molecular weight difference of about 50—100%. The molecular weight range is very large, ranging from several hundert to about 10 million Dalton.

SEC is also very flexible. The column and the mobile phases can be optimized to ensure a size-based separation for a wide range of proteins. The purified proteins, mostly separated under native conditions can be easily fractionated for further investigations, for example, for sequencing.

References

1.
"Polymer analysis by GPC-SEC", Agilent Technologies Technical Note 5988 -0110, (**2000**)

2.
Mikes, O., "High performance liquid chromatography of biopolymers and biooligomers, Part B" *Elsevier*, *Science Publisher B.V.*, Amsterdam, Netherlands, **1988**

3.
Kopaciewicic, W., Regnier, F.E. "Nonideal size-exclusion chromatography of proteins: effects of pH at low ionic strength", Anal. Biochem. 126, 8-16, **1982**

Heinz Goetz is an application chemist at Agilent Technologies GmbH, Waldbronn, Germany.

The author would like to thank Dr. Rudolf Grimm, Hexal Biotech Research GmbH, Munich, Germany for providing the EPO sample and for helpful discussions.

www.agilent.com/chem

Published September 1, 2001
Publication Number 5988-4266EN

Agilent Technologies