Fingerprint Multiplicity in MALDI-TOF Mass Spectrometry of Copolymers

Jiří Horský, Zuzana Walterová*

Summary: Due to the limited resolution of matrix-assisted laser desorption/ionization time-of-flight mass spectrometry, the assignment of binary-copolymer composition to the peak molecular weight observed in the spectrum is not unique. Consequently, several clusters, called fingerprints, appear on calculated 2-D distribution, but usually only one of them corresponds to the sample composition. The ambiguity in the assignment of the composition to peaks in mass spectra of binary copolymers was explained in the framework of number theory and several criteria for identification of the correct fingerprint were proposed. The results were applied in analysis of poly(propylene glycol)-*b*-poly(ethylene glycol)-*block*-poly-(propylene glycol) triblock copolymer Reverse Pluronic 17R4.

Keywords: composition; copolymers; distribution; matrix-assisted laser desorption/ionization mass spectrometry time-of-flight (MALDI-TOF MS); molecular weight

Introduction

The resolution of matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) is so high that the molecular weight of the individual components of an investigated mixture can be determined.[1] For copolymers, the molecular weight is given by the content of each of the monomeric units in the molecule and thus conveys the information on the composition.[2–6] The transformation of MALDI-TOF mass spectra to compositional distributions is complicated by the fact that even though the relation between molecular weight and composition is unique in principle, the relation is one-to-several rather than one-to-one because of finite MALDI-TOF MS resolution. More than one composition can be assigned to the molecular weight observed in the spectrum.[3] As a result, several copies of virtually the same pattern, called fingerprints, can be found in the 2-D compositional distribution obtained from a MALDI-TOF mass spectrum of binary copolymer but only one of them corresponds to the composition of the measured sample, with the exception of some special and/or deliberately prepared copolymer mixtures.[7]

In this paper, we review the process of transformation of the MALDI-TOF mass spectrum of a binary copolymer into the 2-D compositional distribution, analyze the origin of isotopic multiplets' overlap and of "copolymer fingerprints" multiplicity in the scope of number theory, and derive the rules for elimination of spurious fingerprints and test a method for elimination of the partial isotopic multiplets' overlap. The results are used in the characterization of the commercial poly(propylene glycol)-*block*-poly(ethylene glycol)-*block*-poly-(propylene glycol) triblock copolymer Reverse Pluronic 17R4.

Experimental Part

Materials. Poly(propylene glycol)-*block*-poly(ethylene glycol)-*block*-poly(propylene glycol) copolymer (Reverse Pluronics) 17R4, dimethylformamide, sodium

Institute of Macromolecular Chemistry, Academy of Sciences of the Czech Republic, Heyrovsky Sq. 2, 162 06 Prague 6, Czech Republic
E-mail: horsky@imc.cas.cz

Table 1.
Molecular weight and composition of Reverse Pluronic 17R4.

w_e [a)]	w_e [b)]	M_n [c)]	M_n [d)]	M_w [e)]
0.40	0.38	2700	2740	2760

[a)] mass fraction of PEG given by the vendor, verified by the ^{1}H-NMR spectroscopy.
[b)] mass fraction of PEG from MS.
[c)] number-average molecular weight by the vendor.
[d)] number-average molecular weight from MS.
[e)] weight-average molecular weight from MS.

trifluoroacetate ($NaAcF_3$), and dithranol (anthracene-1,8,9-triol) were obtained from Aldrich and were used as received. Number-average molecular weight and the content of poly(ethylene glycol) (PEG) provided by the vendor are given in Table 1; ; the content of PEG was verified by the ^{1}H-NMR spectroscopy.[8]

MALDI-TOF Mass Spectrometry. The samples were prepared by the dried droplet method: 1 μL of a solution of the sample (2 mg/mL), matrix (dithranol, 16 mg/mL), and ionization agent (sodium trifluoroacetate, 0.4 mg/mL) in dimethylformamide was deposited on the ground-steel target plate and allowed to dry. MALDI-TOF mass spectra were obtained with a Biflex III mass spectrometer (Bruker Daltonics) in the positive ion reflectron mode, using delayed extraction with a N_2 laser emitting at 337 nm. The spectra consisted of peaks corresponding to adducts with a single Na^+ ion. Four spectra from four depositions were acquired for the sample.

Results and Discussion

The separation power and the resolution of MALDI-TOF mass spectrometry (MS) are so high that the signals corresponding to individual species are detected, provided that the molecular weight is not too high.[1] Figure 1 shows the MALDI-TOF mass spectrum of the Reverse Pluronic 17R4. It can be seen that the spectra of copolymers, even binary ones, are rather complex.

The number of peaks is much higher for copolymers then for homopolymers of the same non-uniformity with the respect to molecular weight because various combination of monomer units are possible. For a binary copolymer with the molecules composed of $n1$ monomeric units of type 1 with molecular weight M_1 and $n2$ monomeric units of type 2 with molecular weight M_2, the observed molecular weight M is related to the composition[3]

$$M = n1\,M_1 + n2\,M_2 + M_{e1} + M_{e2} + M_{ia} \quad (1)$$

where M_{e1}, $M_{e2,}$ and M_{ia} are the molecular weights of the end groups and the ionization agent, respectively. Thus MALDI-TOF

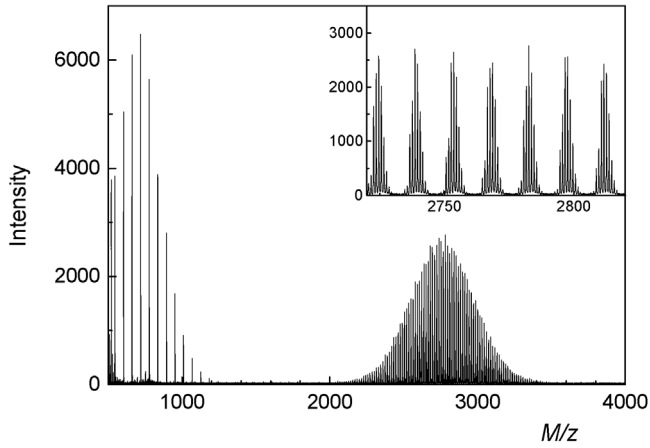

Figure 1.
MALDI-TOF mass spectrum of the Reverse Pluronic 17R4 with a portion of the spectrum expanded in the inset.

mass spectra of copolymers convey not only the information on molecular weight distribution, but also on the composition and the compositional uniformity. The composition $n1$, $n2$ can be assigned to the peak in the spectrum if the difference between the theoretical molecular weight (M_{th}) given by Equation 1 and the experimental value found in the spectrum (M_{ex}) is small, or more precisely if

$$|M_{th} - M_{ex}| \leq \frac{\Delta M}{2}, \tag{2}$$

where ΔM represents the accuracy.[7]

Several signals, reflecting various isotopic compositions, can be observed for each compound in the low molecular weight region of a high-resolution spectrum. The transformation of a MALDI-TOF mass spectrum to the compositional distribution is complicated - among other things - by the fact that isotopic patterns corresponding to different combinations of n_1 and n_2 can overlap. Precise coincidence of two such isotopic patterns is possible only if both monomeric units consist of the same elements and the ratio of the number of atoms in monomeric units is identical for each element (e.g., in the case of the ethylene-propylene copolymer). In practice, two signals corresponding to slightly different molecular weights merge to one peak because of the finite resolution of MALDI-TOF MS. Thus the molecular weight of a $n1$, $n2$ copolymer can be considered to be equal to that of a $n1'$, $n2'$ copolymer if

$$\left|M_{n1,n2} - M_{n1',n2'}\right| \leq \frac{\Delta M}{2}, \tag{3}$$

where ΔM is now the full width of the peaks at half of their maximum height.[9] One of the consequences of the peak overlap is that the peak height is not proportional to the analyte content. For that reason, peak area should be used instead.[10]

Due to the validity of Equation 3, general analysis of peak overlap can be done taking M_1 and M_2 as integers. In this case, the left-hand side of Equation 3 should equal zero and

$$(n1 - n1')\,M_1 = (n2' - n2)\,M_2. \tag{4}$$

In other words, the substitution of $\Delta n1$ monomer units **1** for $\Delta n2$ monomer units **2** does not change the copolymer molecular weight if [2]

$$\Delta n1\,M_1 = \Delta n2\,M_2. \tag{5}$$

While searching for the lowest values of $\Delta n1$ and $\Delta n2$ satisfying Equation 5 (denoted $\Delta n1_{min}$ and $\Delta n2_{min}$), we are in fact looking for the least common multiple of M_1 and M_2, LCM(M_1, M_2)

$$\begin{aligned}\Delta n1_{min}\,M_1 = \Delta n2_{min}\,M_2 &= \mathrm{LCM}\,(M_1, M_2)\\ &= \frac{M_1 M_2}{\mathrm{GCD}\,(M_1, M_2)},\end{aligned} \tag{6}$$

where GCD(M_1, M_2) is the greatest common divisor of M_1 and M_2. From Equation 6 we get

$$\begin{aligned}\Delta n1_{min} &= \frac{M_2}{\mathrm{GCD}\,(M_1, M_2)} \text{ and}\\ \Delta n2_{min} &= \frac{M_1}{\mathrm{GCD}\,(M_1, M_2)}.\end{aligned} \tag{7}$$

M_1 and M_2 are relatively small numbers and so their GCD can easily be found by the prime factorization. In more complicated cases, Euclid's algorithm is used for several thousand years.[11] Nevertheless, the explicit formula for GCD was found recently.[12] For a copolymer with $n1 < \Delta n1_{min}$ and $n2 < \Delta n2_{min}$, the assignment of peaks is unique because no substitution is possible while keeping the same molecular weight. This condition is met for all copolymers with $M < \mathrm{LCM}(M_1, M_2) + M_{E1} + M_{E2} + M_{ia}$; whereas no such copolymer can be found above $M = 2\ \mathrm{LCM}(M_1, M_2) - M_1 - M_2 + M_{E1} + M_{E2} + M_{ia}$. Only copolymers of low molecular weight can be analyzed by MALDI-TOF MS and thus $n1$ and $n2$ are usually relatively small numbers and can be directly used as independent coordinates in

a three-dimensional plot (3-D plot) of the results corresponding to the two-dimensional number compositional distribution (2-D compositional distribution). An example for 17R4 is given in Figure 2. Note that the dominant peaks with $M < 1000$ observable in the 17R4 spectrum (Figure 1) did not find their way onto the plot. Their regular distance 58 corresponds to poly(propylene glycol) (PPG) but there is a downshift of 18 from theoretical PPG peaks, indicating loss of water i.e. cyclization.

The base of the 3-D plot can be divided by a rectangular grid with the mesh size $\Delta n1_{\min} \times \Delta n2_{\min}$ (22×29 for PG/PE copolymers.) Denoting rectangles by their position along the $n1$ and $n2$ axes in the chess-like manner, we can see that the relation between composition and molecular weight is unique only in the rectangle a1. The rectangles a2 and b1 are equivalent; the same molecular weight (as identified in the spectrum) corresponds to two different compositions: one from rectangle a2 and one from b1, in which 29 PE units were replaced with 22 PG units. Similarly triple assignment exists in rectangles a3, b2, and c1, and so on. In the plot several copies of the virtually same pattern (fingerprints) may be seen.[7] Under some conditions, we can identify the correct copy reflecting composition of the sample if these patterns are isolated and do not overlap.

Thus, the fingerprint extending to rectangle a1 would be the correct one. If there is no such fingerprint, some additional information can be of help. For example, a known molecular weight distribution of one block of a diblock copolymer will specify the position of the correct fingerprint along one of the axes[7] and thus pinpoint the

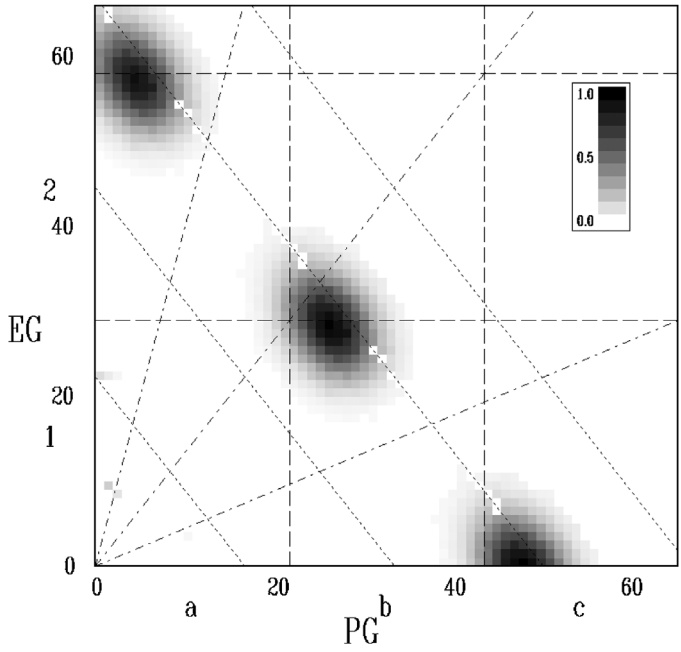

Figure 2.
The monomer-units distribution in the Reverse Pluronic 17R4. PG is the number of propylene glycol monomeric units in the copolymer molecule (i.e., the total length of the outer blocks); EG is the number of ethylene glycol monomeric units (i.e. the length of the inner block). The relative experimental MS signal intensity is denoted using a gray scale map. Rectangular areas with the same peak-assignment multiplicity are outlined with dash lines and identified by a chess-like notation, with labels at the EG and PG axis. Dotted lines indicate the copolymer molecular weight – 1000, 2000, 3000, and 4000, respectively; dashed-dotted lines indicate copolymers with mass fraction of PEG 0.25, 0.50, and 0.75.

correct fingertip. Similarly, a known averaged composition will fix the angular position of the correct fingerprint. On the other hand, a known average molecular weight is not useful; all fingerprints in the distribution plot give an identical molecular weight distribution. The correct 17R4 fingerprint is identified from the partial overlap with rectangle a1 and known averaged content of PEG (40% w/w) as the one mostly located in rectangles b1 and b2. With this knowledge and using standard relations for average molecular weight and composition,[13] we obtained the average molecular weight and composition of 17R4, which were in fair agreement with the vendor's values (Table 1).

This finding may seem surprising for several reasons.

(i) The peaks corresponding probably to cyclic poly(propylene glycol) (cPPG) were not considered in calculations. The cPPG degree of polymerization is about half of that for PG component of 17R4 or similar to that of one PG block. The content of cPPG is estimated to be 10% (w/w) from the area under the curve. Omission of cPPG from the calculation should result in higher molecular weight and higher content of EG than for the correct values. One of the reasons why the discrepancies with the vendor's values was not observed can be the overestimation of cPPG content due to molecular weight and compositional bias of MALDI-TOF MS.[14]

(ii) The peak height was used as a primary quantity proportional to the concentration in our distribution calculation even though the quantity directly related to molar concentration of a compound in the sample is the area under the curve, specifically the area under the curve in a time domain.[14] One of the reasons why the height of a peak or of an isotopic multiplet is not proportional to the concentration in a sample with a proportional constant valid over the whole spectrum is that the number of peaks in the isotopic multiplet increases and their proportion changes with increasing molecular weight.[1] However, this was accounted for in our calculations using a correction factor which related the area of the whole multiplet to the area of the first peak in the multiplet (so called monoisotopic) and which was obtained from the theoretical isotopic distribution calculated by the procedure of Yergey.[15]

As stated above, the height of a peak is not a suitable measure of concentration for overlapping peaks[9,10] because area is an additive quantity, not the height. However, there are no significantly overlapping monoisotopic peaks in the spectrum of 17R4. The preceding analysis dealt with the potential overlap of the first - i.e. the monoisotopic - peaks of isotopic multiplets (i.e. full overlap of isotopic multiplets), which leads to the appearance of several fingerprints in the 2-D distribution obtained from MALDI-TOF mass spectrum of a binary copolymer. With the exception of some special and/or deliberately prepared mixtures, only one of those fingerprints is the real one and corresponds to the composition of the sample. The remaining fingerprints are artifacts and have no relevance to the actual sample composition. The actual full overlap of isotopic multiplets can occur only when the real fingerprint overlap with a spurious one; the fingerprints, however, are isolated for 17R4.

(iii) So far, only the full overlap of isotopic multiplets (i.e. the coincidence of isotopic patterns) was considered and dealt with. However, a partial overlap of isotopic patterns is also possible and was not accounted for in the analysis of 17R4.

We will now analyze the phenomenon of the partial overlap of isotopic patterns in terms of number theory.[9] Substitution of

$\Delta n1^a$ monomeric units **1** in the copolymer molecule by $\Delta n2^a$ monomeric units **2** results in the shift of the isotopic pattern by a if

$$\Delta n1^a M_1 = \Delta n2^a M_2 + a. \quad (8)$$

If M_1 and M_2 are relatively prime, i.e., $\mathrm{GCD}(M_1, M_2) = 1$, any integer value of a is acceptable. Otherwise, a has to be a multiple of $\mathrm{GCD}(M_1, M_2)$, e.g., an odd a cannot be obtained if both M_1 and M_2 are even. The determination of $\Delta n1^a$ and $\Delta n2^a$ is related to the so-called Chinese remainder problem.[11] Without loss of generality, M_1 and M_2 can be assumed to be relatively prime, as can be seen by dividing Equation 8 by $\mathrm{GCD}(M_1, M_2)$. Under this condition

$$\Delta n1^a = a\,\Delta n1^1, \quad (9)$$

where $\Delta n1^1$ is a multiplicative inverse M_1 mod M_2, which can be found by direct search or by the extended Euclid's algorithm.[11] In the present context, the direct search will be sufficient in most cases because $\Delta n1^a{}_{\min} < \Delta n1_{\min}$. The shift a of isotopic patterns of an ethylene glycol and propylene glycol copolymer after replacing $\Delta n_{PG}{}^a$ propylene glycol units by $\Delta n_{EG}{}^a$ ethylene glycol units are given in Table 2 for some relevant values of a.

As mentioned above, the total overlap of isotopic patterns ($a = 0$) can be resolved only under special circumstances allowing elimination of spurious fingertips. On the other hand, a partial overlap ($a > 0$) can be accounted for *in principle* by procedure called de-isotoping.[11] Starting with the occurrence of the isotopic pattern overlap of the lowest molecular weight, we identify the copolymer composition from the position of the left-most peak; calculate the theoretical isotopic pattern[15] using a correct intensity for its leftmost (monoisotopic) peak; and substract these intensities except the intensity for the monoisotopic peak from the spectrum. We move to the next peak corresponding to the monoisotopic molecular weight of some copolymer and repeat the procedure. After traversing the whole molecular weight span in this way, the spectrum is reduced to monoisotopic peaks with corrected intensities.[16] However, as the molecular weight increases, the use of this procedure becomes questionable because the relative intensity of the monoisotopic peak decreases and the position of the maximum in the isotopic pattern moves to a higher molecular weight. As the intensity of the monoisotopic peak decreases, the relative error of its intensity increases. Moreover, the resolution power of the method decreases with increasing molecular weight.

Table 2.
The shift a of the isotopic multiplet of an ethylene glycol and propylene glycol copolymer after replacing $\Delta n_{PG}{}^a$ propylene glycol units by $\Delta n_{EG}{}^a$ ethylene glycol units.

$\Delta n_{PG}{}^a$	$\Delta n_{PE}{}^a$	a
16	21	−4
19	25	−2
22	29	0
3	4	2
6	8	4
9	12	6

Fair agreement of the MALDI-TOF MS analysis of 17R4 with the vendor's data was achieved even without de-isotoping the spectrum (see Table 1.) One of the possible reasons why the partial overlap is not significant in the present case is that both M_{EG} and M_{PG} are even and thus only even values of a are permissible. Therefore, the monoisotopic peak may overlap with monoisotopic + 2, monoisotopic + 4, and so on. For low-molecular-weight copolymers, the intensity of the monoisotopic + 2 peak is lower than that of the monoisotopic peak and the contribution of the monoisotopic + 4 is almost negligible. There is no effect of the partial overlap on the values of molecular weight averages calculated directly, i.e. in the same way as for homopolymers.

Not only the de-isotoping seems unnecessary in the present case, but it in fact worsens the results - the fingerprints become distorted (see Figure 3) and the average PEG content decreases even further from the value given by the vendor and confirmed by ^{1}H-NMR spectroscopy.

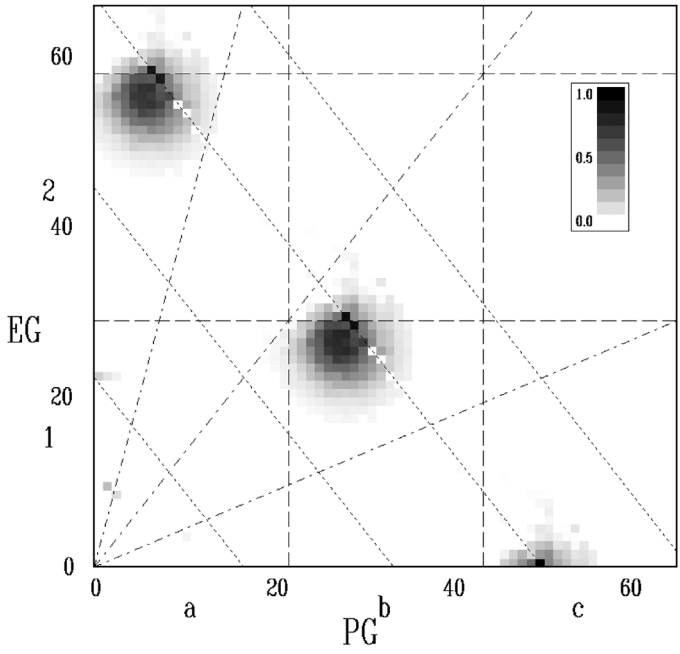

Figure 3.
The monomer units' distribution in the Reverse Pluronic 17R4 obtained using the partial-overlap rectification procedure. The detailed description of the graph given at the Figure 2 also applies here.

The decrease of the calculated ethylene glycol (EG) content upon de-isotoping is expected: de-isotoping decreases the apparent abundance of copolymers which has higher content of EG because the contribution of EG_nPG_m is substracted from signal of $EG_{n+4}PG_{m-3}$ according to Table 2, PG being propylene glycol. Thus de-isotoping cannot be recommended for EG/PG copolymers without further analysis at this point in time. Nevertheless, the situation may be different for another combination of monomers.

Our analysis of fingerprint multiplicity predicts that spurious and real fingerprints are identical, just shifted. Nevertheless, the differences can occur in practice because the assignment is done on the basis of unrounded molecular weights. Thus monoisotopic molecular weight read in the spectrum for EG_nPG_m copolymer is compared with calculated molecular weights of EG_nPG_m and $EG_{n-29}PG_{m+22}$ which differ by 0.16 (replacement error). The value of ΔM in Equation 2 and the accuracy of experimental molecular weight determine which composition or compositions are assigned to the experimental value. In our experiment, identical fingertips were obtained. This indicates that the accuracy of measurement was acceptable but resolution was somewhat limited. The departure from integral values and replacement error may be higher for other monomers. A higher replacement error seems to be advantageous because the multiplicity predicted by integer-based analysis need not to occur. However, coincidence may occur for other monomer combinations instead. In principle, such coincidence can be also found by the procedure outlined above after multiplying monomer molecular weights by a suitable constant because the numerical value of molecular weight depends on chosen units.

Conclusion

We have demonstrated the power of MALDI-TOF MS in analysis of copolymers.

The wide-spread use of MALDI-TOF MS in this area is limited by the same two factors as the use of MALDI-TOF MS with polymers in general;[1] namely, (i) discrimination due to molecular weight and composition and (ii) relatively low molecular weights for which isotopic resolution can be achieved. Limitation (i) is of less concern in applications where changes in compositional distribution rather than absolute values are sought.

Driven mostly by the needs of biomedicine, the development of MALDI-TOF MS instrumentation results in higher and higher resolution and thus in the easing of limitation (ii). The application of the results presented here for copolymer fingerprints multiplicity, which were obtained in the scope of number theory, to copolymers of mid and high molecular weight will however require careful re-examination. The way in which the molecular weight of monomeric units is approximated by integers and the accumulation of errors caused by such approximation must be considered.

[1] H. Pasch, W. Schrepp, "*MALDI-TOF Mass Spectrometry of Synthetic Polymers*", Springer, Berlin Heidelberg **2003**.
[2] G. Wilczek-Vera, Y. Yu, K. Waddell, P. O. Danis, A. Eisenberg, *Rapid Commun. Mass Spectrom.* **1999**, *13*, 764.
[3] G. Wilczek-Vera, P. O. Danis, A. Eisenberg, *Macromolecules* **1996**, *29*, 4036.
[4] P. O. Danis, F. J. Huby, *J. Am. Soc. Mass Spectrom.* **1995**, *6*, 1112.
[5] J. Horský, J. Mikešová, O. Quadrat, J. Šňupárek, *J. Rheol.* **2004**, *48*, 23.
[6] A. Krupková, J. Čermák, Z. Walterová, J. Horský, *Anal. Chem.* **2007**, *79*, 1639.
[7] R. X. E. Willemse, B. B. P. Staal, E. H. D. Donkers, A. M. van Herk, *Macromolecules* **2004**, *37*, 5717.
[8] U. Holzgrabe, B. W. K. Diehl, C. Schollmayer, I. Waver, *J. Pharm. Biomed. Anal.* **1998**, *17*, 557.
[9] A. Felinger, *Anal. Chem.* **1997**, *69*, 2976–2979.
[10] G. J. van Rooij, M. C. Duursma, C. G. de Koster, R. M. A. Heeren, J. J. Boon, *Anal. Chem.* **1998**, *70*, 843.
[11] G. A. Jones, J. M. Jones, *Elementary Number Theory*, Springer undergraduate mathematic series. Springer, London **2005**.
[12] M. Polezzi, *Am. Mat. Mon.* **1997**, *104*, 445.
[13] M. S. Montaudo, *Mass Spectrom. Rev.* **2002**, *21*, 108.
[14] Z. Walterová, J. Horský, *Anal. Chim. Acta* **2011**, *693*, 82.
[15] J. A. Yergey, *Int. J. Mass Spectrom. Ion Phys.* **1983**, *52*, 337.
[16] G. J. W. M. van Alebeek, H. A. Schols, A. G. J. Voragen, *Carbohydr. Polym.* **2001**, *46*, 311.

