University of Delaware Mass Spectrometry Facility

Polymerix Data Analysis Guide

Author: Katherine Martin Date: 04/06/2023

Table of Contents

Background

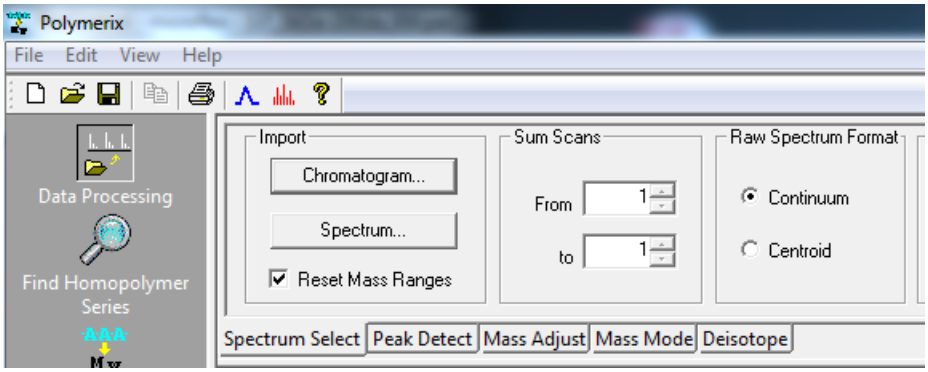
A license for **Polymerix (ver. 3.01)** by Sierra Analytics was included with the MALDI-TOF with an install formatted for total ion chromatograph (TIC) import of Bruker files. It appears that this may be locked to exclude other vendor formats, but other files can be brought in as peak lists in ASCII mass/intensity format. See the first example in this document for importing a Bruker file and the second example for bringing in an outside peak list. See the very detailed manual under *Polymerix>Help>Help Topics*, which was used to make this overview.

Polymerix has more features than PolyTools, including spectral processing features that can account for resolution and peak width changes across a spectrum. It can handle low, high, and ultrahigh mass resolution data and can also handle negative and multiply charged data. It can be used for copolymer analysis and also allows for double bond equivalent (DBE) and element ratio restrictions. It calculates number average molecular weight (M_n), weight average molecular weight (M_w), Z-average (third order) molecular weight (M_z), polydispersity index (PD), number average degree of polymerization (DP_n), weight average degree of polymerization (DP_w), third order degree of polymerization (DP_z), percentage of series ($\%s_i$), percentage of spectrum ($\%S_i$), weight percentage of series ($\%w_i$), and weight percentage of spectrum ($\%W_i$). The math behind these parameters is available in the Polymerix manual on the MALDI-TOF computer.

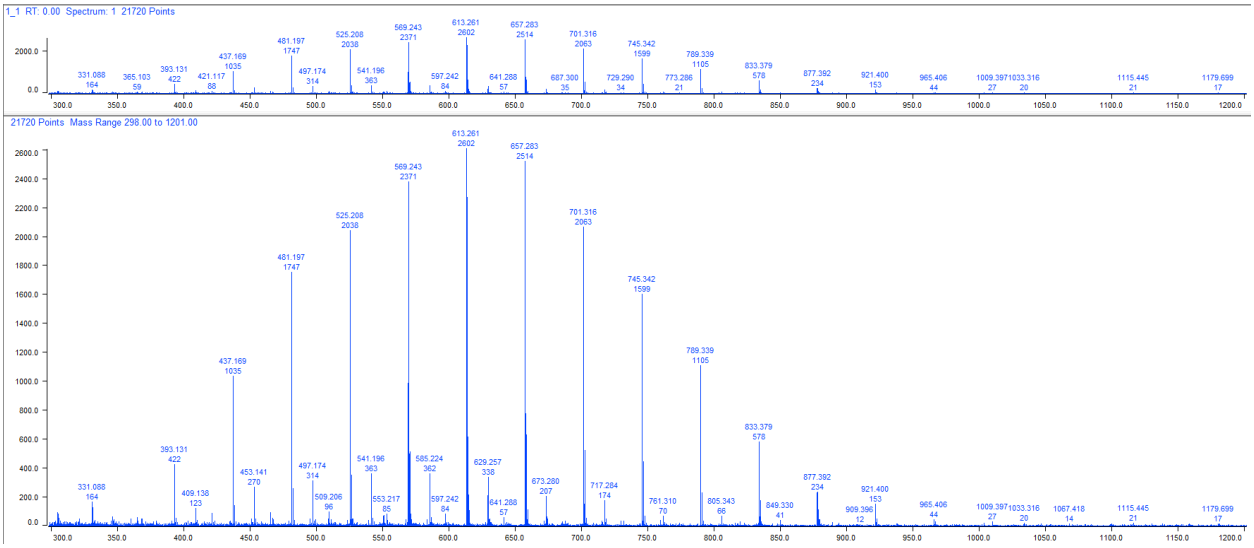
PEG700 (Homopolymer Example)

Data Processing

This example is with the mass spectrum of a polyethylene glycol (PEG700) standard. This is a relatively simple spectrum where end groups are already known. Open Polymerix and then click on the "*Data Processing*" button and "*Spectrum Select*" tab. To import Bruker files go to *File>Import Chomatogram* and navigate to the Bruker spectrum to import. Alternatively, click the *Chromatogram* button under *Import*.

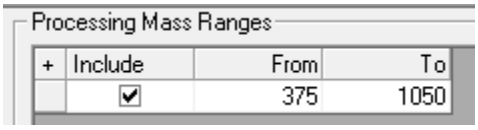


This imports the raw spectrum and ignores previously generated peak lists. *"Raw Spectrum Format"* should be selected based on the imported data. Direct import of Bruker spectra would be in continuum (or profile) spectrum format, while ASCII peak list imports are assumed to be centroid data. The spectrum can be zoomed in on by left clicking and dragging, and zoomed back out by left clicking. The middle panel is the original raw spectrum, while the bottom panel is the processed spectrum that will update throughout data processing.



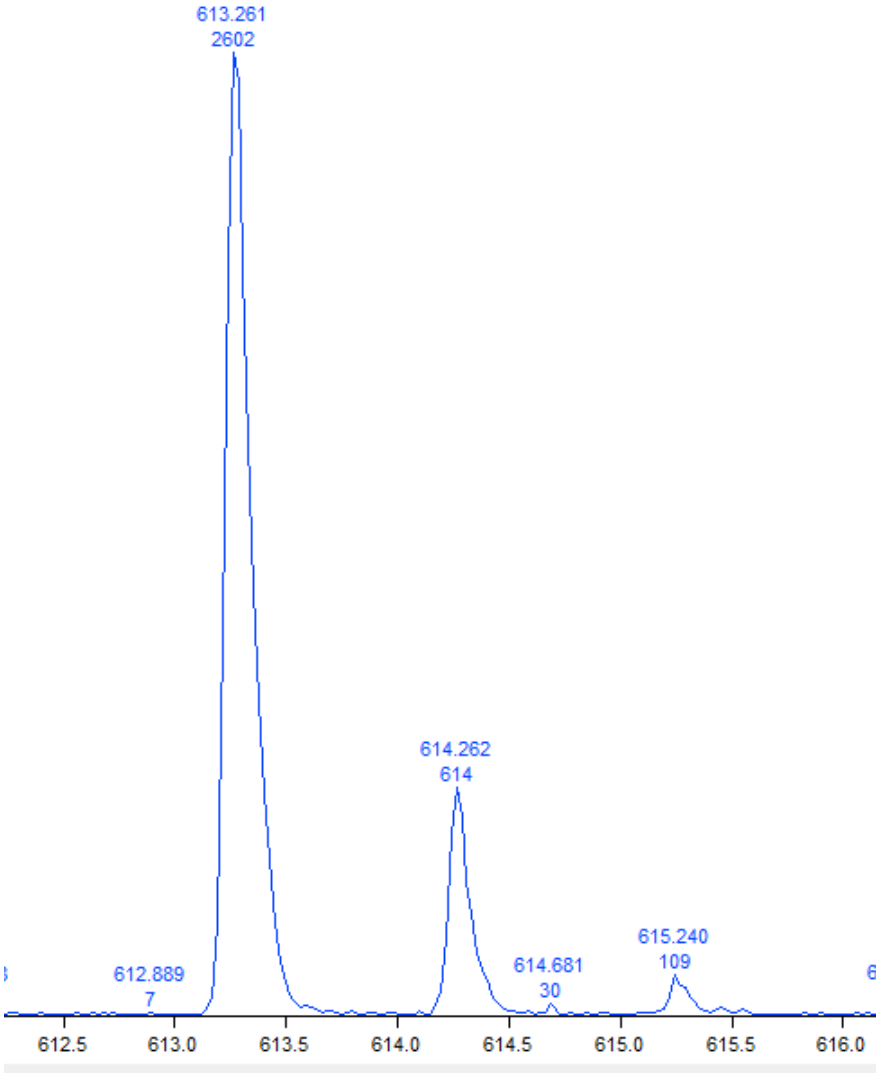
The original spectrum was created from 298-1201 *m/z*, but the processing range will be cut to 375 to 1050 *m/z* because the regions beyond this appear to just feature noise and peaks that are clearly not part of the repeating unit pattern of interest. To do this, add the desired

range to the *Processing Mass Ranges* table, then click apply. You can also hit plus in the table for multiple processing ranges, such as if there are multiple distributions in the spectrum or an impurity in the middle of the spectrum.

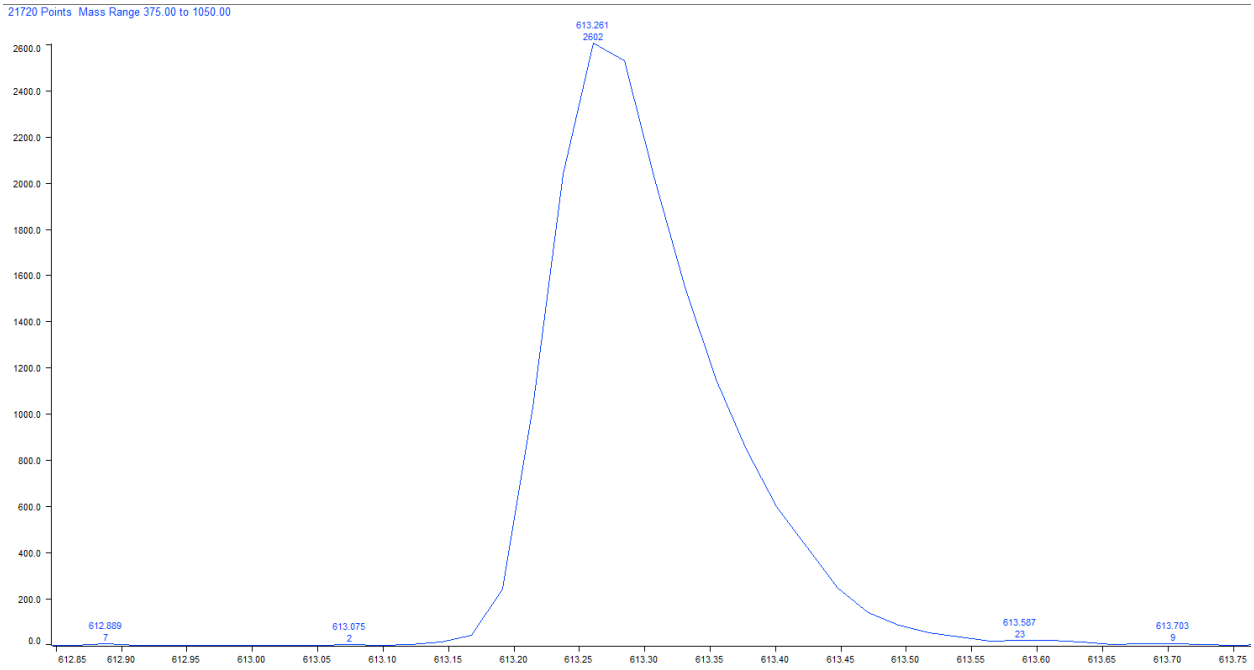


Ranges can also be made by highlighting the line of the table, then right clicking and dragging across the region of interest. To delete a range, click the *"-"*, or a range can be unchecked if only temporarily excluding. When done click apply.

If data has isotopic resolution, such as from reflectron mode, it is recommended to process in peak detection mode (centroided during peak detection). If the data does not have isotopic resolution or if peaks aren't fully resolved, it can be processed in profile mode. Here, the PEG700 data has isotopic resolution, so it will be processed in peak detection mode.



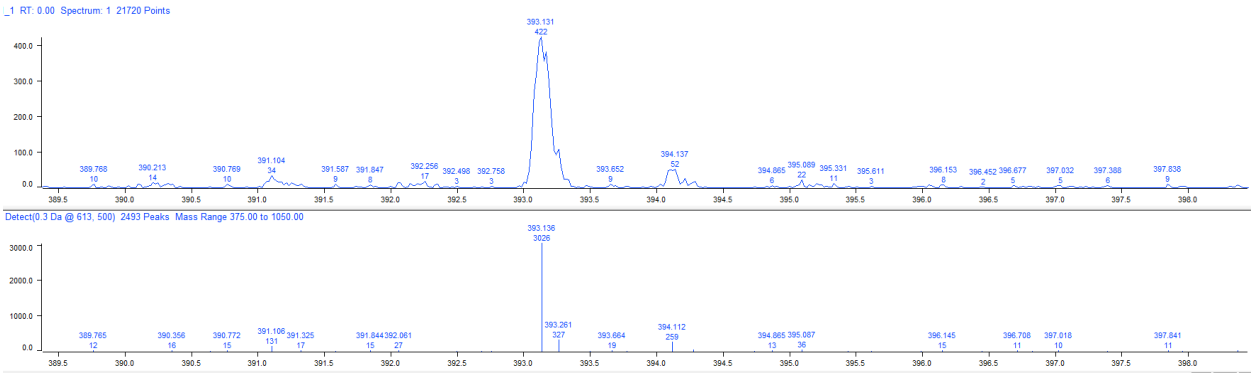
Click the *"Peak Detection"* tab, and under *"Peak Width"* select TOF and select an approximate resolution that reflects the data. (*"Constant"* or almost constant peak width data is the type usually produced by quadrupole, FTICR, or Orbitrap instruments, while *"proportional"* peak width data is produced by magnetic sector instruments.) Peak width here needs to be an estimate of full width at 5% peak height or 10% valley. At ~613 *m/z* there is a ~0.3 *m/z* peak width at 5% peak height (~130 intensity in arbitrary units), and most peaks across the example spectrum have essentially the same peak widths.



Checking the *"Peak detect and centroid"* button in the Peak Detection/Baseline section, automatically also selects *"Subtract baseline from spectrum"*. Once done click *"Apply"* and compare the profile spectrum to the centroided peak detection now in the bottom panel, checking for accuracy. If single peaks are appearing as multiple peaks in the centroided data, then the peak width parameter was likely set too small. If peaks are being missed or noise is being assigned, then the peak width was likely set too large

"Subtract baseline from spectrum" is automatically applied with peak detection, but can also be used without peak detection to correct a sloping or elevated baseline in data that will be interpreted fully in profile mode. Baseline width is the width of the spectrum that will be averaged to compute baseline offset, and the default 500 Da is usually okay. A smaller value may be needed to account for more narrow offsets in the baseline. "*Merge to peak width*" is used for noisy data where small satellite peak artifacts are appearing around the true centroid. It merges these artifacts based on the peak width parameter. "*MS Threshold*" is for setting a peak relative intensity threshold filter based on the base peak to remove small peaks from the peak detection spectrum.

Merge to peak width was not selected and a 0.6% MS threshold was applied. The 0.6% threshold is still including a lot of noise in the detection, but a higher cut-off led to the final ~1009 *m/z* peak not being detected, which appears to be a real peak in the main repeating unit series. A lower MS threshold led to even more background noise being peak detected. An initial peak width of 0.3 Da at 613 *m/z* was selected based on the above observations; however, this was observed to give a peak width issue at the low end of the spectrum. The ~393 *m/z* peak is being incorrectly detected as 2 centroids due to the peak being broader and having a more jagged peak shape than the higher *m/z* peaks.



Therefore, the final peak width settings were switched to be more based on this peak, 0.35 *m/z* at 393 *m/z*, as this corrected this, while still centroiding other peaks correctly.

Under the "*Mass Adjust*" tab are options for recalibrating the *m/z* of the spectrum. This is likely unnecessary as the instrument method should have been calibrated during data generation, but may be needed if post-hoc calibration is necessary or for outside data brought in as an ASCII peak list. To use this, the *"Adjust spectrum masses"* box would be clicked and then *Offset* or *Slope* adjustment, or both, selected and set as needed, shifting the spectrum. Where *offset* shifts the spectrum the same value across the entire *m/z* range, while positive *slope* would be an increasing adjustment with increasing *m/z*, or a negative *slope* a decreasing adjustment with increasing *m/z*. Mass adjustment can also be applied later in the analysis process.

The *"Mass Mode"* tab can be used to account for the resolution of the spectrum. Data with isotopic resolution would be processed in *"Monoisotopic"* mode. Polymerix also allows for processing in "*Chemical*" aka profile mode for unresolved peaks, or "*Nominal*" mode for data only at a nominal mass or rounded to a nominal mass. Polymerix can also be used to account for loss of resolution across a spectrum, such as if the lower end of the spectrum was at isotopic resolution (monoisotopic), but isotopic peaks become unresolved at higher *m/z* in the spectrum (chemical).

Note: The de-isotoping feature has been largely superseded and the homopolymer analysis feature will only work on data that has not been de-isotoped.

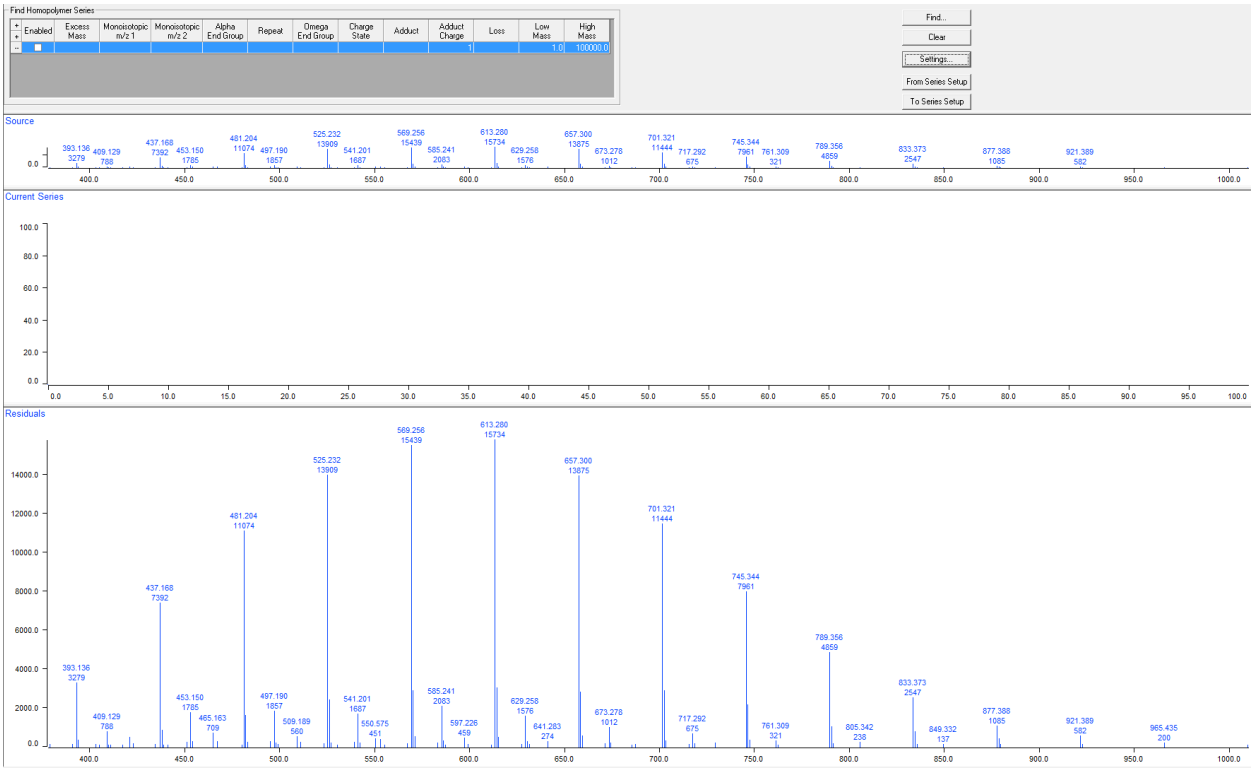
The de-isotoping feature is available after peak detection, and is to account for higher isotopic weighting at higher *m/z* for polydisperse data, as this would lead to intensity bias against higher *m/z* peaks if considering only the first peak (usually the monoisotopic peak). It shifts predicted intensity in the isotopic peaks back into the monoisotopic. Using this requires a model chemical formula input including at least a repeating unit, but should also include end groups and adduct identity if they are predicted to have a significant contribution to the higher isotope pattern.

Processing template files may be saved to save time processing the same spectrum or similar spectra. Go to *File>Save Template* and the template can be saved as a **.pmt** file, it can then be opened and applied through *File>Open Template*.

If the identity of the polymer is already known, skip to the Homopolymer Analysis section.

Find Homopolymer Series

Polymerix can also help with identifying composition of a polymer spectrum. However, polymerix is best designed for handling polymer spectra with known or partially known identities, and has limited use in a situation of a true unknown in terms of all compositional aspects. Though this is an unlikely scenario as usually some info, like repeat unit, end group, adducts, or possible elemental composition, is already known or predicted. To use this, **data should be peak detected and NOT de-isotoped**. This is for homopolymer analysis only. Click the "*Find Homopolymer Series*" button on the left-most panel.

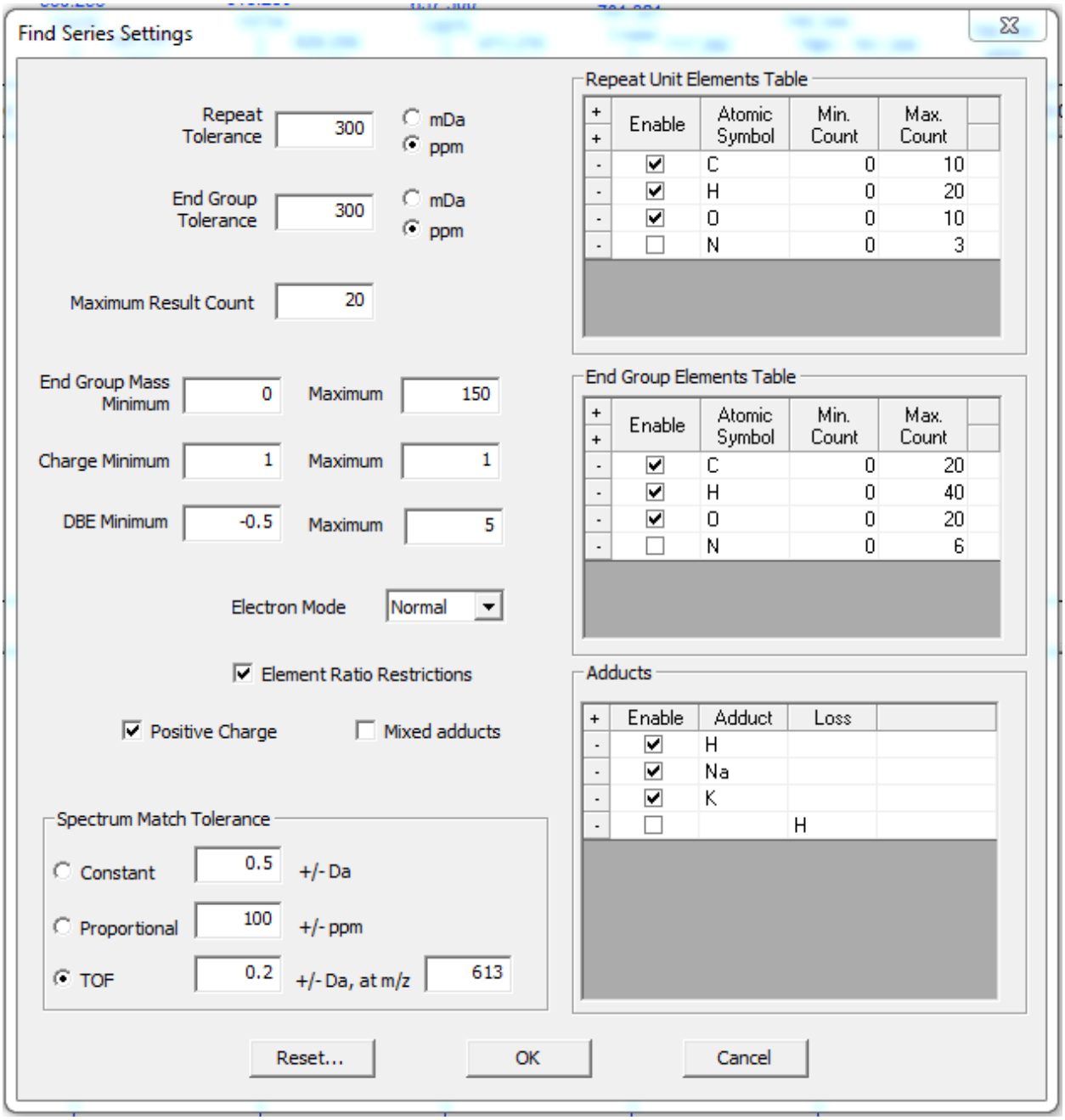


The window should look similar to above, where the top panel shows the original spectrum and the middle panel will show the current reference peaks, once they are selected. The bottom panel is the residuals, or the peaks still not accounted for in the identified polymer series. No series have been identified yet, so all peaks are still considered residuals. Clicking "*Find*" will start the process of Polymerix identifying "*reference peaks*".

In comparison to PolyTools which works by matching end group and repeating measurements to tables provided by the user, Polymerix works by using a heuristic algorithm to select 2 reference peaks believed to be from a repeating unit series. These reference peaks are then used to calculate a repeating unit *m/z* value and do elemental ID matches to the repeating unit according to settings under the "*Settings…*" button. Residual *m/z* after a modulo with the repeating unit *m/z* is then matched to possible adduct and combined end group formula to account for the remainder, following settings.

Choosing proper settings will likely take some trial and error. Here it is known that the PEG sample should not contain elements beyond C, H, or O in the repeat unit or end group. No cationizing agent was added, so adducts are likely to only involve common background elements. Na and K are ubiquitous and can come from contact with glass or other materials of construction, be background in reagents, or already present on the MALDI plate surface.

This data has higher error partially inherent to MALDI-TOF. It was also not internally re-calibrated. Setting ***m/z* tolerances** at 300 ppm (an error of ~±0.113-0.315 *m/z* across this *m/z* range) appears to give reasonable suggestions. This is low resolution data, so **Maximum Results Count** was set higher than default to catch more reasonable IDs. High resolution data like from higher resolution TOFs, Orbitrap, or FTICR would use settings with much lower tolerances.

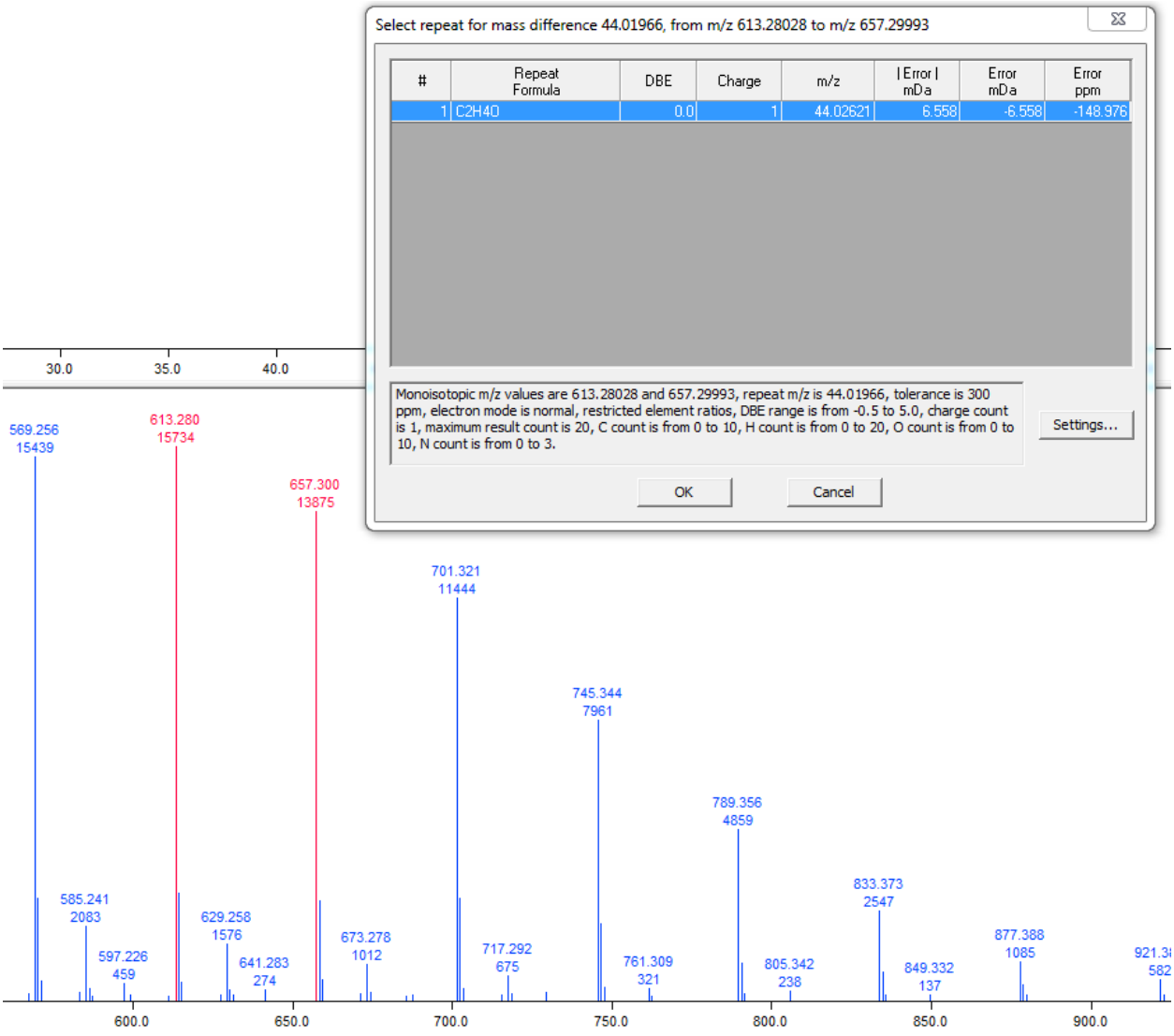


End group mass minimum and maximum was left as default, but this could be used to narrow results as there is usually some end group identity expectation. **Charge** settings were left at 1, as MALDI almost exclusively produces singly charged ions. **Double Bond Equivalent** (DBE) restrictions were left at default, but would have to be changed depending on the repeat unit (e.g. PEG has 0, polystyrene would need a maximum setting >4). DBE could also be used to better restrict results based on DBE expectations for the repeating unit and possibly end group.

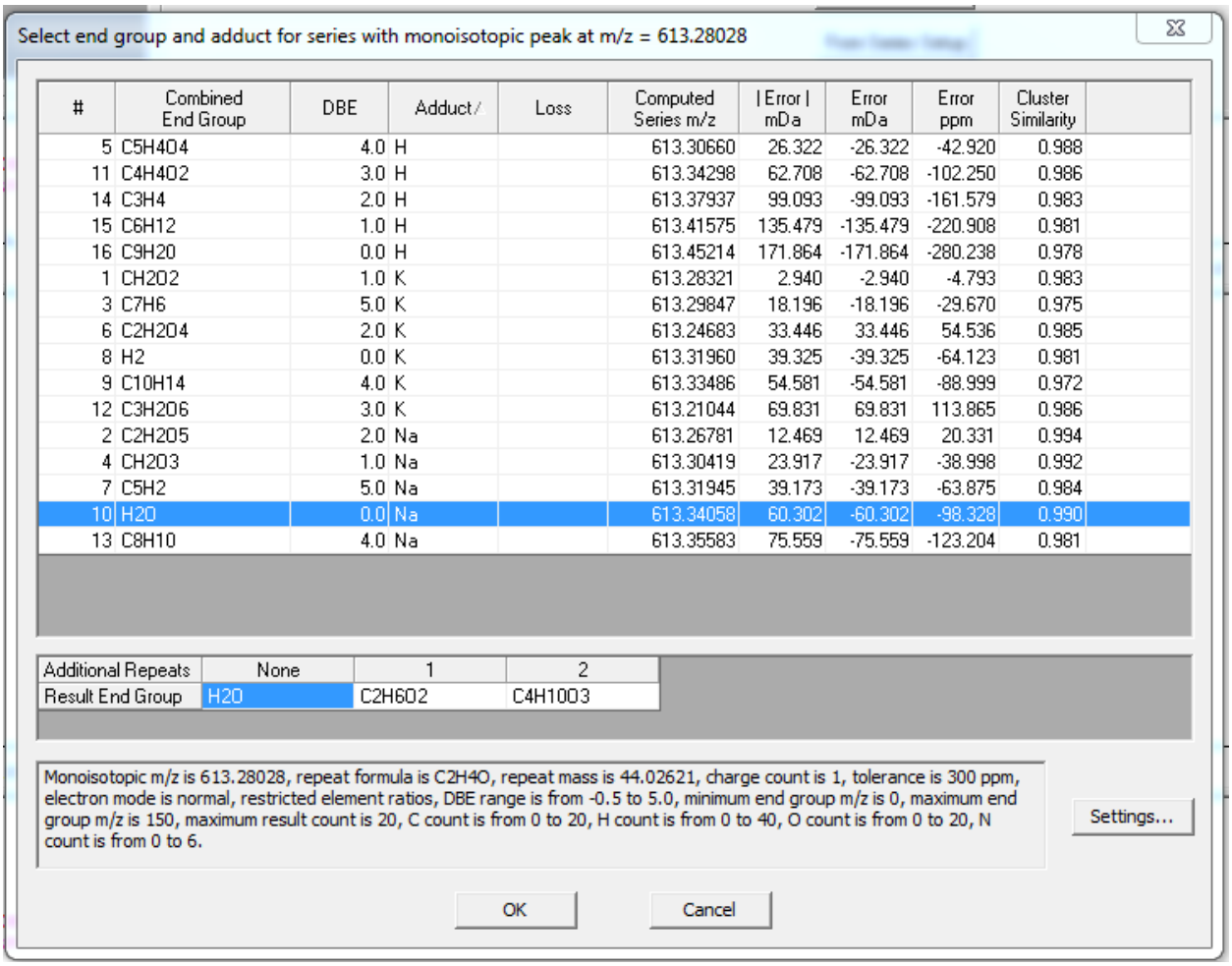
Electron Mode set at "*Normal*" restricts DBE to integer values, "*Reversed*" restricts to non-integer values, and "*None*" removes restrictions. Sum of end group DBE is integral under

"*Normal*" and vice-versa. **Element Ratio Restrictions** restrict repeat unit and end group formulas to what are deemed "reasonable". These actual restrictions are not made clear but are likely based on H/C, O/C, other heteroatom ratios with C. **Positive Charge** is ion polarity, but is also used to account for electron mass which is important for high resolution data. In Polymerix, **Mixed Adducts** only works for multiply charged data (e.g. $[M + H^+ + NH4^+]^{2+}$).

Spectrum Match Tolerance is as described previously. TOF is most appropriate for this data, and it was set to ±0.2 at 613 *m/z*.

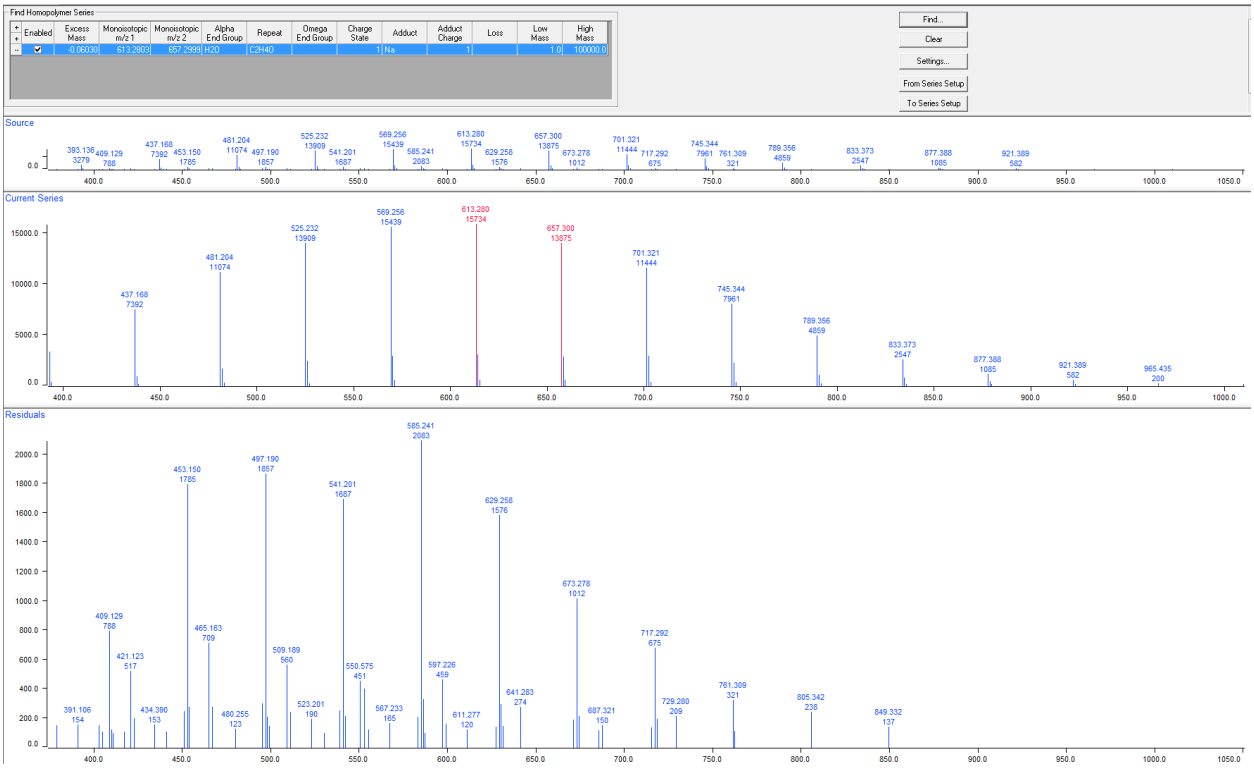


This shows the selected initial reference peaks in red and a list of the suggested repeat unit formula for these. Only one possibility was a suggestion here based on the applied settings, which here is a correct ID for a PEG ($O—CH_2—CH_2$) repeating unit.

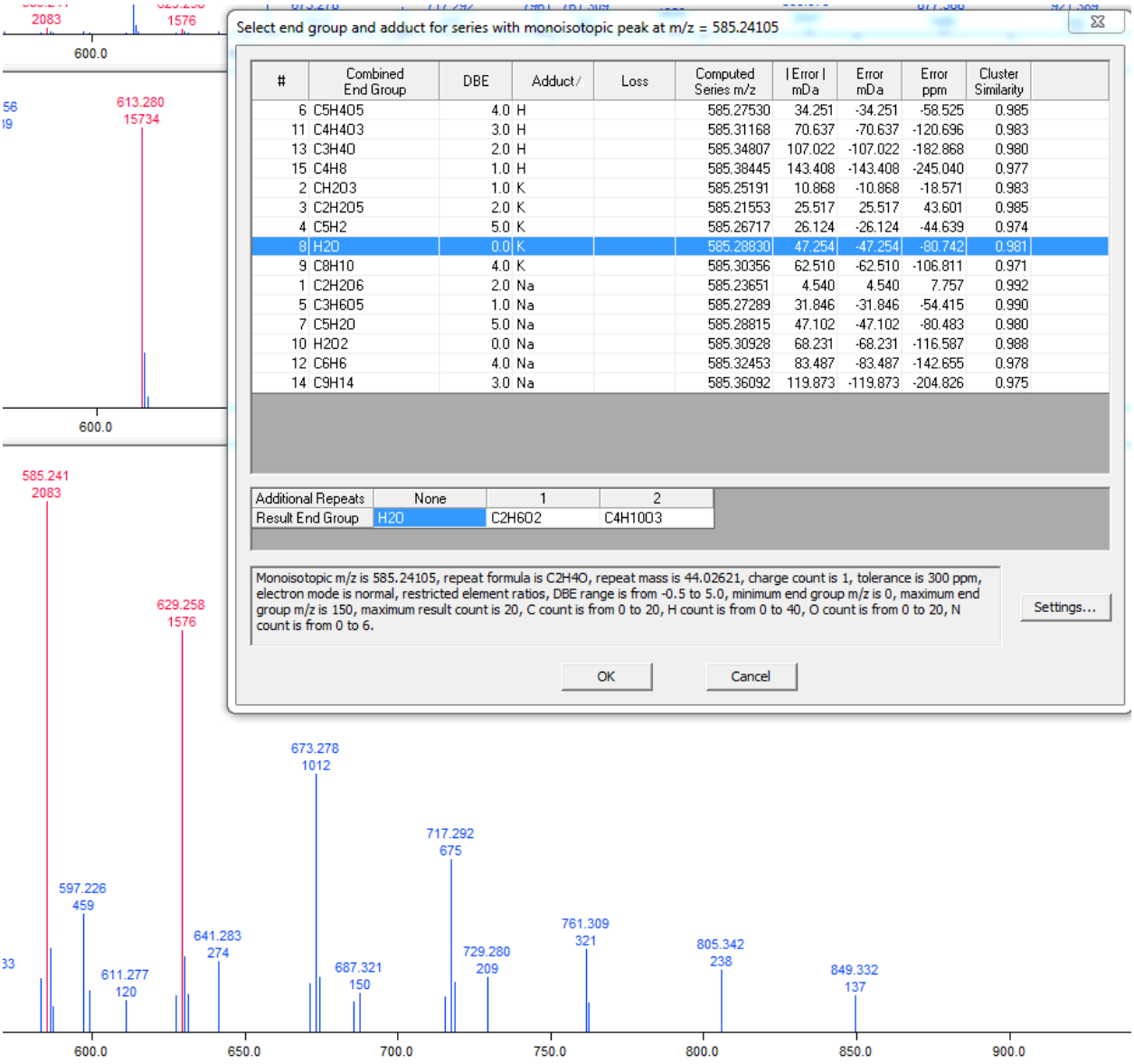


The next window after clicking "*OK*" is the combined end group suggestions table. It was chosen to sort by adduct because PEGs strongly favor cationization with alkali metal adducts, with a sodium adduct more predicted as the major series. Of these, H2O is most reasonable as a combined end group and is a correct ID for the standard. H and OH are common end groups for PEGs, it is a simpler ID with reasonable error for the spectrum, has a higher isotopic cluster similarity, and has reasonable DBE. The repeating unit would be expected to have 0 DBE and end groups here are expected to be small with no to very minimal degrees of unsaturation.

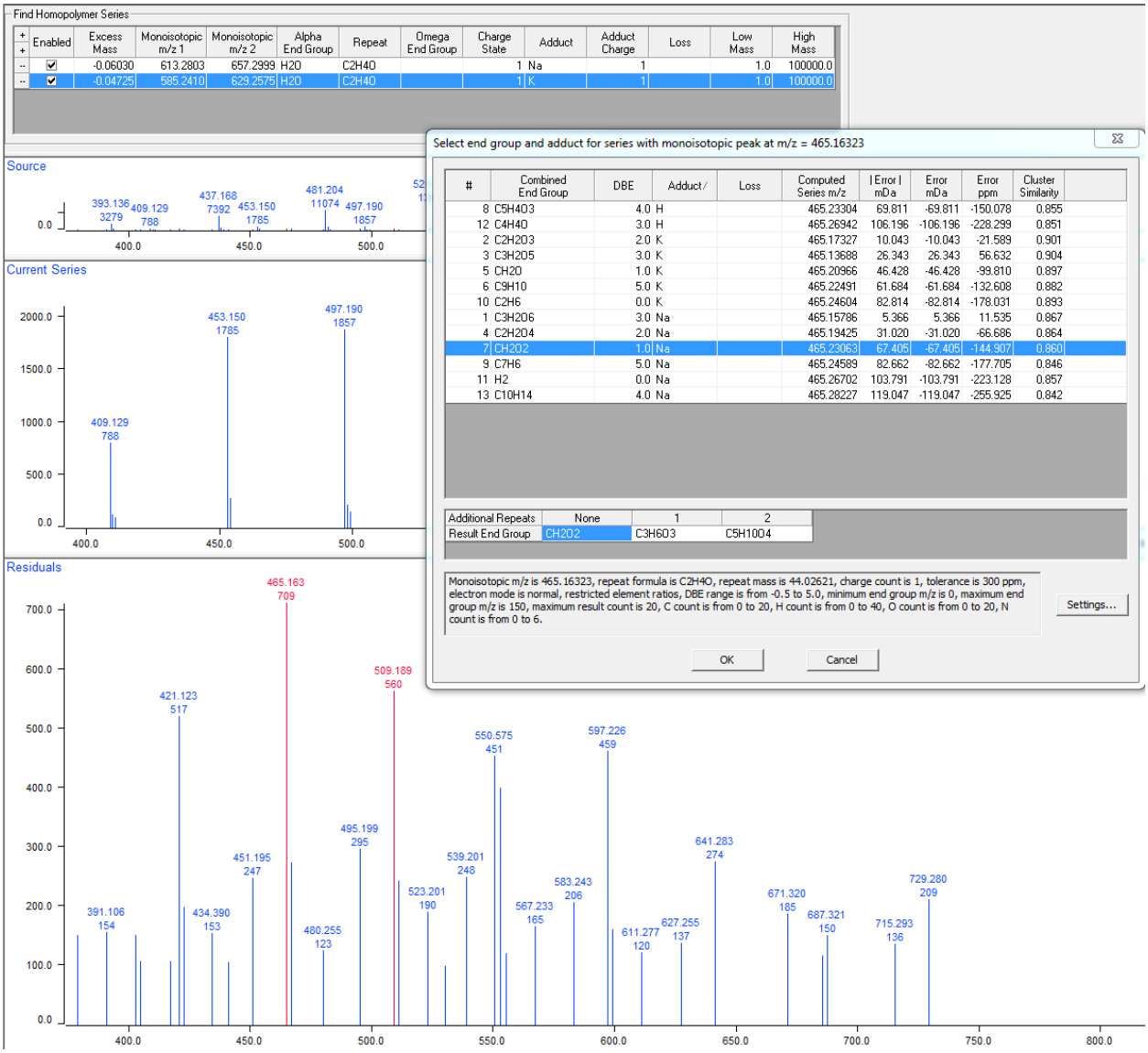
For low resolution data, picking the more likely correct identity from possibilities generally relies on previous knowledge and expectations for the sample and an understanding of more reasonable IDs. It is important to understand the type of ionization that is more reasonable for different sources and instruments (e.g. different expected polymer adduct formation and multicharging by ESI). Analysis of higher resolution data would be much more restricted by error and relying on *m/z* match and cluster similarity.



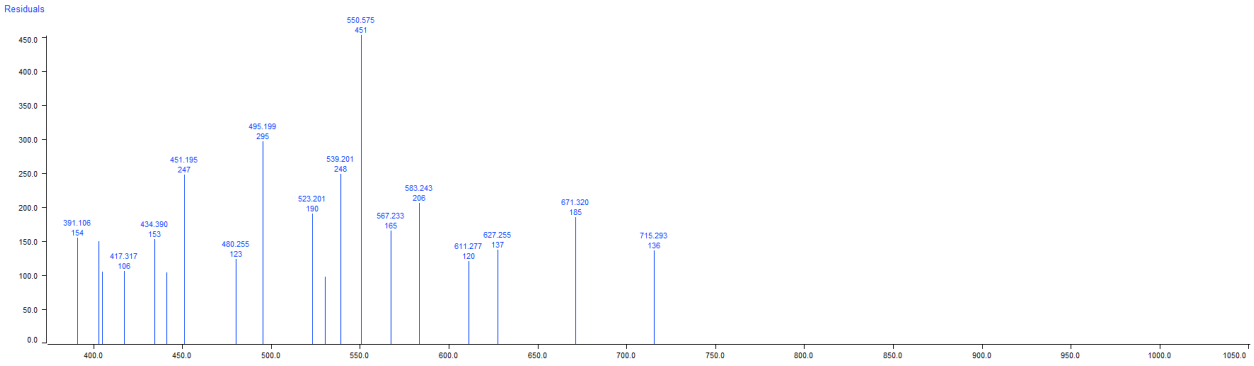
This first selection is now added to the list in the top panel. Continuing, a next potential series is identified with the reference peaks in red. Residuals at this point are shown in the bottom panel. This is the second most intense obvious series in the initial spectrum. The previously selected repeating unit is still a good fit, but if it was no longer a reasonable ID for the reference peaks, the original repeating unit suggestion table would reappear and a new selection would have to be made.



Of the combined end group suggestions, the equivalent identity as the largest series, but with a K adduct instead of Na makes the most sense and is expected. It is also of similar and reasonable error as the previous ID and has high cluster similarity. In the previous image of the residuals another series was obvious. So, continuing on, a third pair of reference peaks is identified. The PEG repeating unit is still a good fit.



This is a minor series closer to the baseline and noise, and would require more analysis for a more certain ID. It's +28 *m/z* from the main sodium adduct series, and might reflect something like a carboxyl end group (COOH vs OH), $[M + Na]^+$.

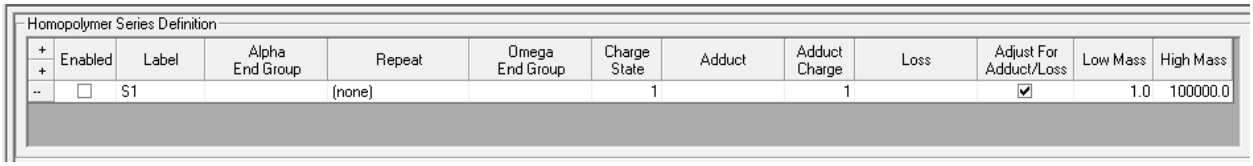


There were signs of a fourth PEG series in the spectrum at +2 *m/z* from the previous series. This had an even lower S/N and greater associated error and has already been cut from the residuals by the relative intensity cut during processing. The current residuals do not show another series.

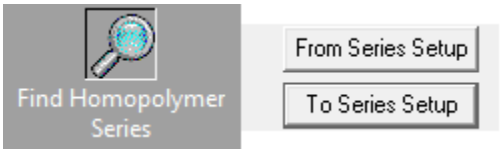
The Find Homopolymer Series tools may need troubleshooting. A common problem is in reference peak selection. Error across the spectrum is unlikely to be linear, so the algorithm may grab on to peaks that are monoisotopic peaks in a series, but with slightly higher error, giving no possibilities for matches or otherwise making matching to repeat unit or end groups not possible. Manual selection of different reference peaks in the series may address this, by right clicking and dragging across selected peaks in the source or residuals panels. Alternatively, temporarily restricting the *m/z* range in data processing just for series finding, then reverting back for actual analysis. A lack of identified possibilities may also be due to settings, which should be carefully selected.

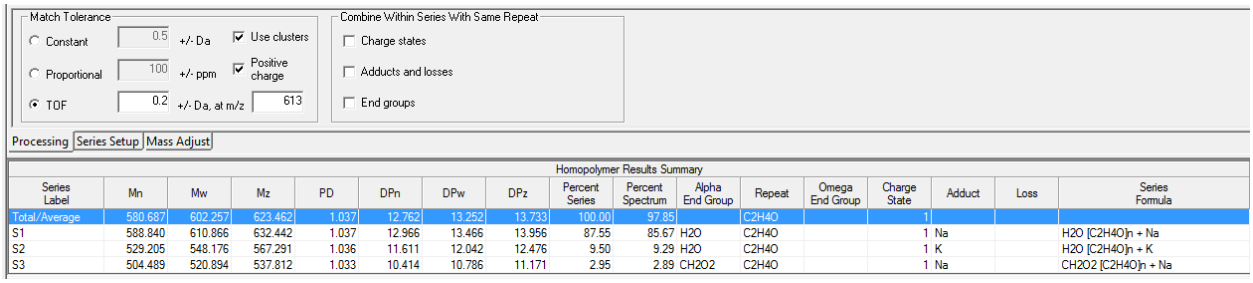
Homopolymer Analysis

Click the *Homopolymer Analysis* button and then click the "*Series Setup*" tab. This table can be manually filled out if the polymer identity is already known or has been identified.



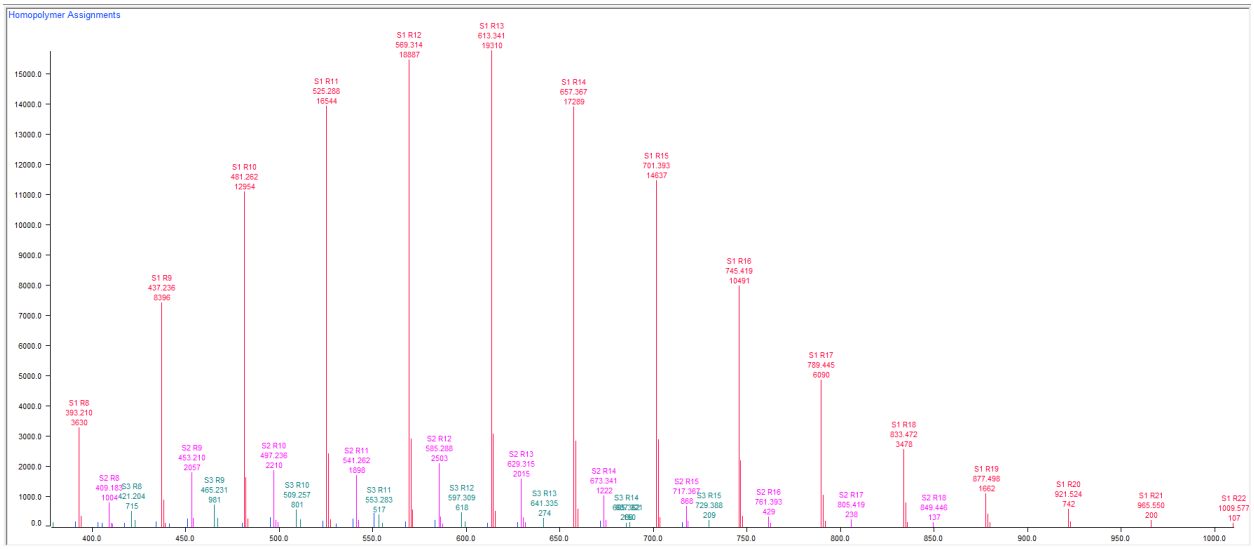
Alternatively, series selected in *Find Homopolymer Series* can be automatically brought to this table by going back to the button and clicking the "*To Series Setup*" button.



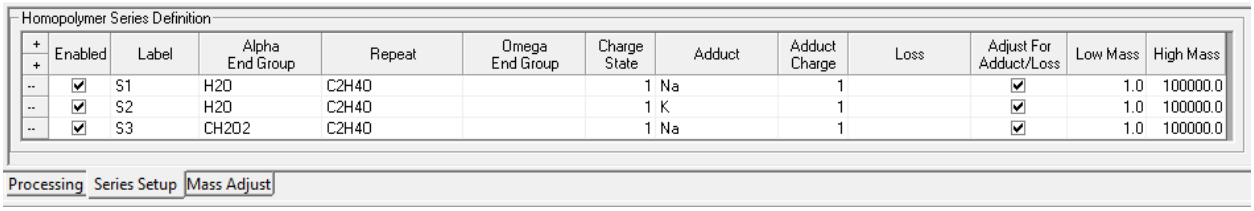


Click the "*Processing*" tab. *Match Tolerance* is the same as previously described. Settings in *Homopolymer Analysis* can be saved as a template **.pmt** file. Click the "*Assignments*" tab at the bottom display tabs. Check that isotopic clusters of the series are being properly matched in the series spectrum (or monoisotopic peaks if not using clusters). Here, they are not being correctly detected for S3, where +2 *m/z* peaks are being incorrectly considered part of the

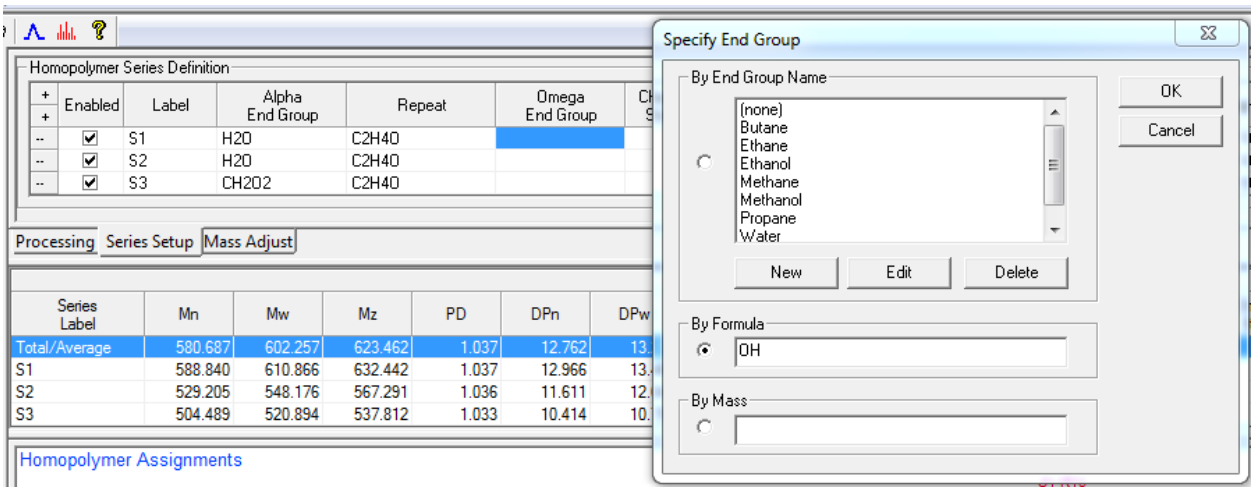
isotopic pattern. How to address this is described later. Residuals, peaks left after series assignment, are plotted in blue. The S1 series is plotted in red, S2 in pink, and S3 in dark green. These colors can be customized as described later. In the peak labels, the first line is the series label and number of monomers, the second line is the *m/z* value, and the third line is the intensity/"abundance" in arbitrary units.



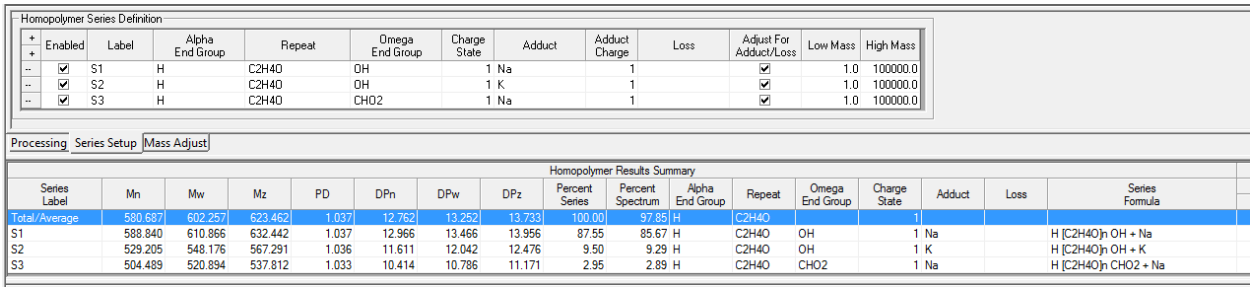
"*Combine Within Series With Same Repeat*" mathematically merges series that share features but are different only in the charge state, adducts and losses, or end groups. This will be revisited. Click on the "*Series Setup*" tab again. Polymerix identification is done as a combined end group identity, so if the part of this identity that belongs to the alpha versus omega end group is known, these can be separated.



To do this, click on the boxes in the *Definition Table* and type in new entries. Default lists of common end groups and repeating units are available to choose from for quicker filling of the table. This list can be edited or new options can be added for quick selection. Table edits can be done by formula. Table edits can also be only mass if the formula is unknown, to account for mass for proper calculations (e.g. know remainder is only the end group, but don't know identity or don't know identity of repeating unit). Series names can be changed by clicking in the Label column.



By typing in the formula box, alpha end group is manually set to H and omega end group is manually set to OH to reflect the PEG standard. The "*Adjust for Adduct/Loss*" box is selected to make sure parameters exclude the mass shift from the mass of a Na or K, as these are only adducts and are not related to the mass distribution of the polymer standard.



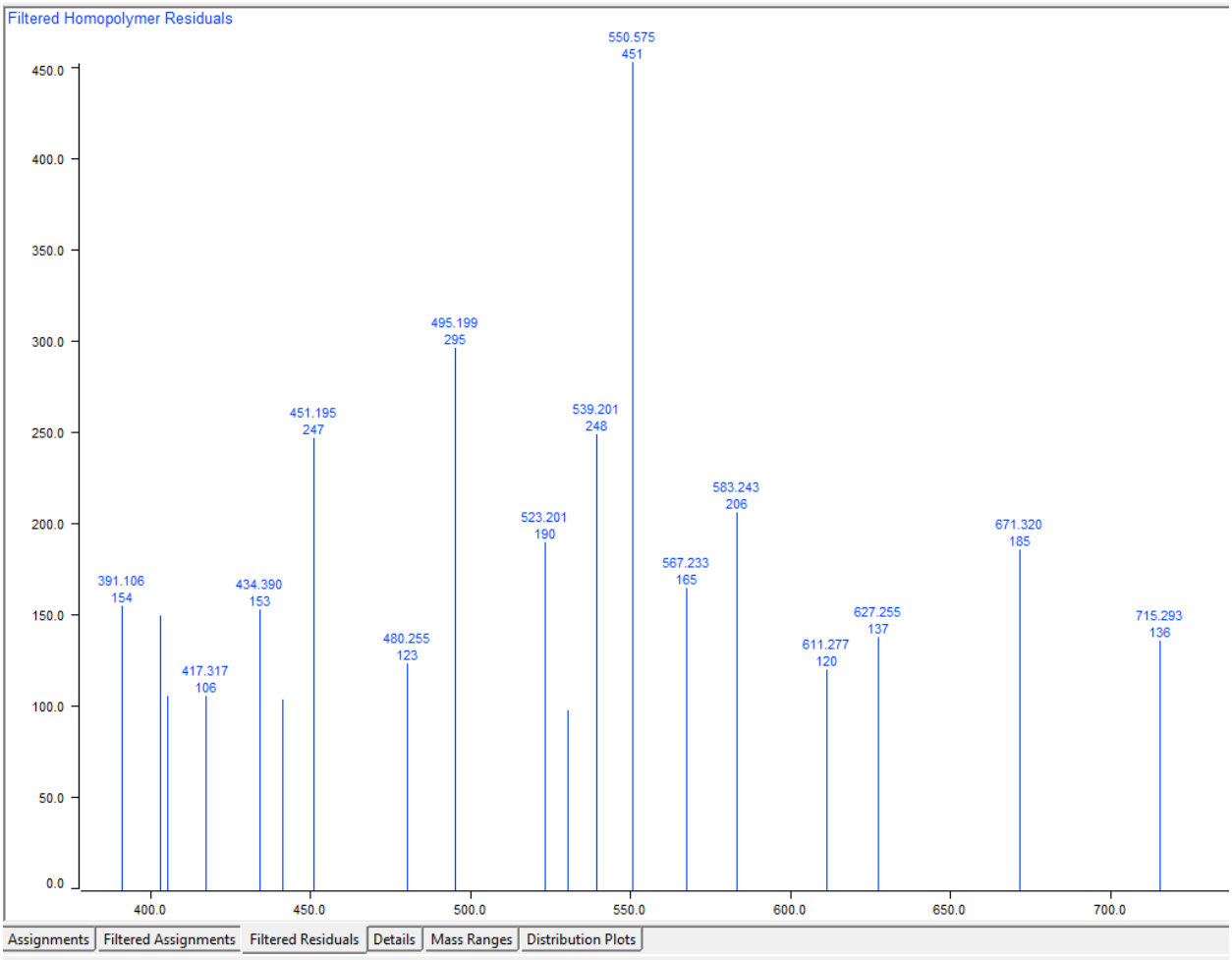
This is the final series identity table used for characteristic parameter calculation. The characteristic parameters for S1 are very similar to the PolyTools results for this series (see the PolyTools Guidance), though lower on M_n and M_w, as these were calculated including isotopic cluster peaks compared to the PolyTools example. The S2 series is more significantly off, due to the relative abundance cut-off causing the loss of 2 peaks from the second series. However, a lower relative abundance cut-off led to incorrect assignment of isotopic cluster peaks, which would also lead to some bias.

This analysis prioritized the main Na adduct series (S1) and incorporating isotopic clusters in parameter calculation. In S2 and S3, peaks were not correctly detected and these parameters are currently incorrect. If a data goal was a characterization of all series, this would need to be corrected. In S2, 2 peaks in the series are missing. In S3, peaks in the series are missing and +2 peaks that don't match the expected isotopic pattern are being considered part of the cluster. There are multiple ways to address this issue with S2 and S3.

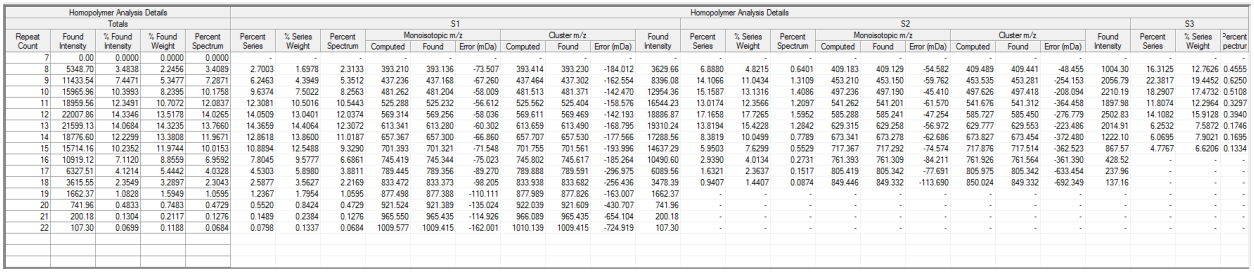
One way to address this would be to lower the processing cut-off with the acknowledgment of some minor bias from background noise incorrectly incorporated into isotopic clusters. This bias would be much lower than the bias caused by incorrectly excluding

peaks in the series. A second way to address this would be to lower the processing cut-off and then switch to monoisotopic series analysis by unchecking the use clusters box. There is also bias associated with this due to how isotopic pattern distribution shifts across the series range, though it would be minor here. The third, and probably best, way to address this and still use cluster incorporation would be to lower the relative abundance cut-off, then export the processed spectrum as a peak list by clicking *Edit>Copy Spectrum Data* and then pasting the peak list in Excel. This could then be manually edited to remove the background peaks from the peak list. The peak list would then be re-imported. Alternatively, processing could be done to begin with outside of Polymerix with data imported as a peak list (see the PolyTools Guidance). Cluster incorporation can have its own bias concerns, as the visibility of isotopic peaks can be affected by the localized baseline and background interference or overlap.

Results and produced figures can be further explored using the tabs at the bottom. Clicking the "*Filtered Assignments*" tab shows the series spectrum without the residuals. Clicking the "*Filtered Residuals*" tab shows a plot of only the residuals from the processed spectrum.



The "*Details*" tab shows detailed data tables for the current series.

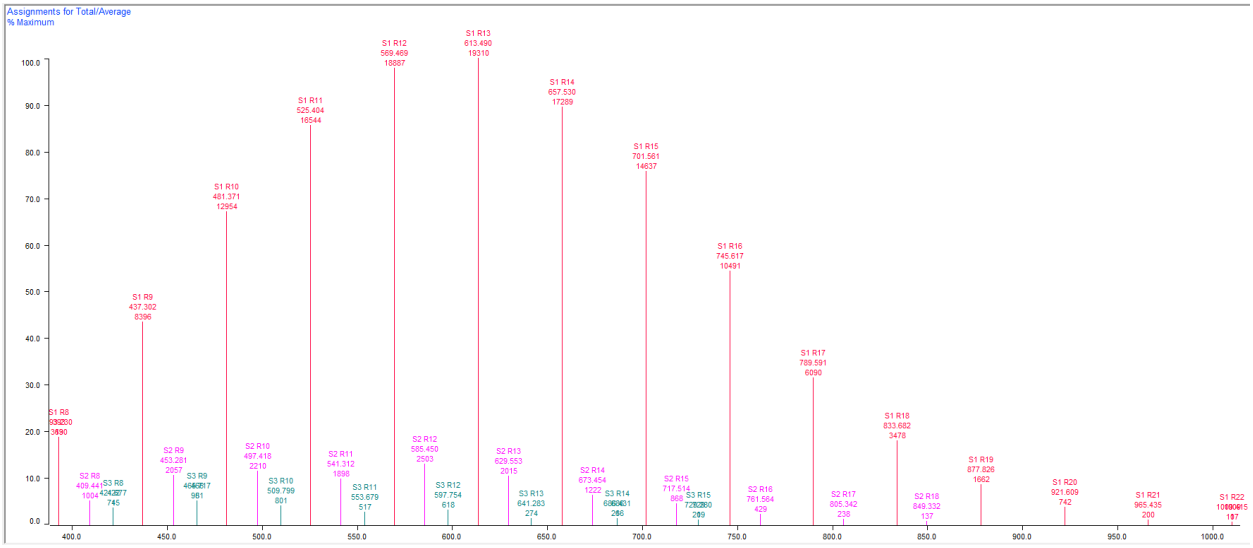


Homopolymer Analysis Details					Homopolymer Analysis Details																							
Totals					S1											S2										S3		
Repeat Count	Found Intensity	% Found Intensity	% Found Weight	Percent Spectrum	Percent Series	% Series Weight	Percent Spectrum	Monoisotopic m/z			Cluster m/z			Found Intensity	Percent Series	% Series Weight	Percent Spectrum	Monoisotopic m/z			Cluster m/z			Found Intensity	Percent Series	% Series Weight	Percent pectrur
								Computed	Found	Error (mDa)	Computed	Found	Error (mDa)					Computed	Found	Error (mDa)	Computed	Found	Error (mDa)				
7	0.00	0.0000	0.0000	0.0000	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
8	5348.70	3.4838	2.2456	3.4089	2.7003	1.6978	2.3133	393.210	393.136	-73.507	393.414	393.230	-184.012	3629.66	6.8880	4.8215	0.6401	409.183	409.129	-54.582	409.489	409.441	-48.455	1004.30	16.3125	12.7626	0.4555
9	11433.54	7.4471	5.3477	7.2871	6.2463	4.3949	5.3512	437.236	437.168	-67.260	437.464	437.302	-162.554	8396.08	14.1066	11.0434	1.3109	453.210	453.150	-59.762	453.535	453.281	-254.153	2056.79	22.3817	19.4452	0.6250
10	15965.96	10.3993	8.2395	10.1758	9.6374	7.5022	8.2563	481.262	481.204	-58.009	481.513	481.371	-142.470	12954.36	15.1587	13.1316	1.4086	497.236	497.190	-45.410	497.626	497.418	-208.094	2210.19	18.2907	17.4732	0.5108
11	18959.56	12.3491	10.7072	12.0837	12.3081	10.5016	10.5443	525.288	525.232	-56.612	525.562	525.404	-158.576	16544.23	13.0174	12.3566	1.2097	541.262	541.201	-61.570	541.676	541.312	-364.458	1897.98	11.8074	12.2964	0.3297
12	22007.86	14.3346	13.5178	14.0265	14.0509	13.0401	12.0374	569.314	569.256	-58.036	569.611	569.469	-142.193	18886.87	17.1658	17.7265	1.5952	585.288	585.241	-47.254	585.727	585.450	-276.779	2502.83	14.1082	15.9128	0.3940
13	21599.13	14.0684	14.3235	13.7660	14.3659	14.4064	12.3072	613.341	613.280	-60.302	613.659	613.490	-168.795	19310.24	13.8194	15.4228	1.2842	629.315	629.258	-56.972	629.777	629.553	-223.486	2014.91	6.2532	7.5872	0.1746
14	18776.60	12.2299	13.3808	11.9671	12.8618	13.8600	11.0187	657.367	657.300	-66.860	657.707	657.530	-177.566	17288.56	8.3819	10.0499	0.7789	673.341	673.278	-62.686	673.827	673.454	-372.480	1222.10	6.0695	7.9021	0.1695
15	15714.16	10.2352	11.9744	10.0153	10.8894	12.5488	9.3290	701.393	701.321	-71.548	701.755	701.561	-193.996	14637.29	5.9503	7.6299	0.5529	717.367	717.292	-74.574	717.876	717.514	-362.523	867.57	4.7767	6.6206	0.1334
16	10919.12	7.1120	8.8559	6.9592	7.8045	9.5777	6.6861	745.419	745.344	-75.023	745.802	745.617	-185.264	10490.60	2.9390	4.0134	0.2731	761.393	761.309	-84.211	761.926	761.564	-361.390	428.52	-	-	-
17	6327.51	4.1214	5.4442	4.0328	4.5303	5.8980	3.8811	789.445	789.356	-89.270	789.888	789.591	-296.975	6089.56	1.6321	2.3637	0.1517	805.419	805.342	-77.691	805.975	805.342	-633.454	237.96	-	-	-
18	3615.55	2.3549	3.2897	2.3043	2.5877	3.5627	2.2169	833.472	833.373	-98.205	833.938	833.682	-256.436	3478.39	0.9407	1.4407	0.0874	849.446	849.332	-113.690	850.024	849.332	-692.349	137.16	-	-	-
19	1662.37	1.0828	1.5949	1.0595	1.2367	1.7954	1.0595	877.498	877.388	-110.111	877.989	877.826	-163.007	1662.37	-	-	-	-	-	-	-	-	-	-	-	-	-
20	741.96	0.4833	0.7483	0.4729	0.5520	0.8424	0.4729	921.524	921.389	-135.024	922.039	921.609	-430.707	741.96	-	-	-	-	-	-	-	-	-	-	-	-	-
21	200.18	0.1304	0.2117	0.1276	0.1489	0.2384	0.1276	965.550	965.435	-114.926	966.089	965.435	-654.104	200.18	-	-	-	-	-	-	-	-	-	-	-	-	-
22	107.30	0.0699	0.1188	0.0684	0.0798	0.1337	0.0684	1009.577	1009.415	-162.001	1010.139	1009.415	-724.919	107.30	-	-	-	-	-	-	-	-	-	-	-	-	-

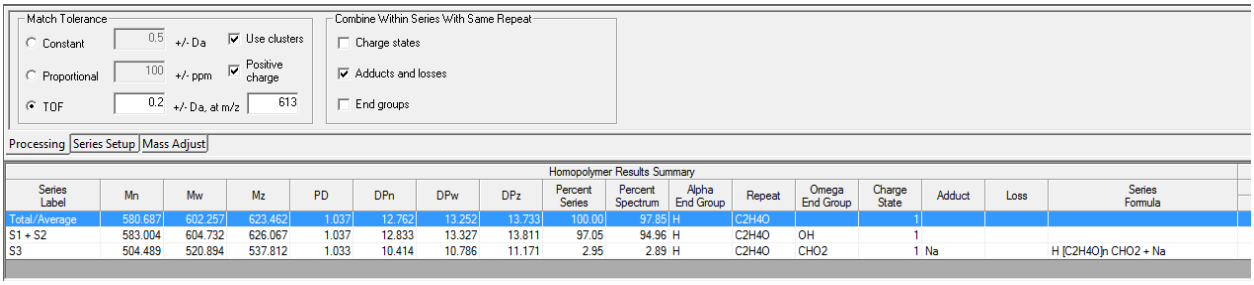
The "*Mass Ranges*" tab is a list of the *m/z* ranges used for determining series, defined by the matching settings. It also includes "collisions" or isobaric overlaps of series.

Low Mass	High Mass	Series	Repeat	Collisions
377.0213	379.3409	S3	7	
393.0493	395.3755	S1	8	
409.0200	411.3477	S2	8	
421.0386	423.3760	S3	8	
437.0668	439.4105	S1	9	
453.0377	455.3828	S2	9	
465.0564	467.4108	S3	9	
481.0847	483.4451	S1	10	
497.0557	500.4201	S2	10	
509.0746	511.4451	S3	10	
525.1030	527.4793	S1	11	
541.0742	544.4543	S2	11	
553.0931	555.4791	S3	11	
569.1216	571.5132	S1	12	
585.0929	588.4881	S2	12	
597.1119	599.5128	S3	12	
613.1405	615.5468	S1	13	
629.1119	632.5217	S2	13	
641.1309	643.5462	S3	13	
657.1597	659.5801	S1	14	
673.1311	676.5551	S2	14	
685.1502	687.5793	S3	14	
701.1791	703.6131	S1	15	
717.1506	720.5882	S2	15	
729.1698	731.6123	S3	15	
745.1987	747.6460	S1	16	
761.1703	764.6211	S2	16	
773.1895	776.6479	S3	16	
789.2185	792.6815	S1	17	
805.1901	808.6538	S2	17	
817.2094	820.6804	S3	17	
833.2384	836.7140	S1	18	
849.2102	852.6864	S2	18	
861.2295	864.7128	S3	18	
877.2586	880.7463	S1	19	
893.2303	896.7187	S2	19	
905.2497	908.7450	S3	19	
921.2789	924.7785	S1	20	
937.2507	940.7510	S2	20	
949.2701	952.7771	S3	20	
965.2993	968.8105	S1	21	
981.2712	984.7831	S2	21	
993.2906	996.8090	S3	21	
1009.3198	1012.8424	S1	22	

The "*Distribution Plots*" tab allows for individual series to be plotted with single peaks as well as *m/z* and intensity now representing the cluster. The y-axis is % of maximum. Here, all series are selected, but highlighting individual series in the table would only display that series at a time.

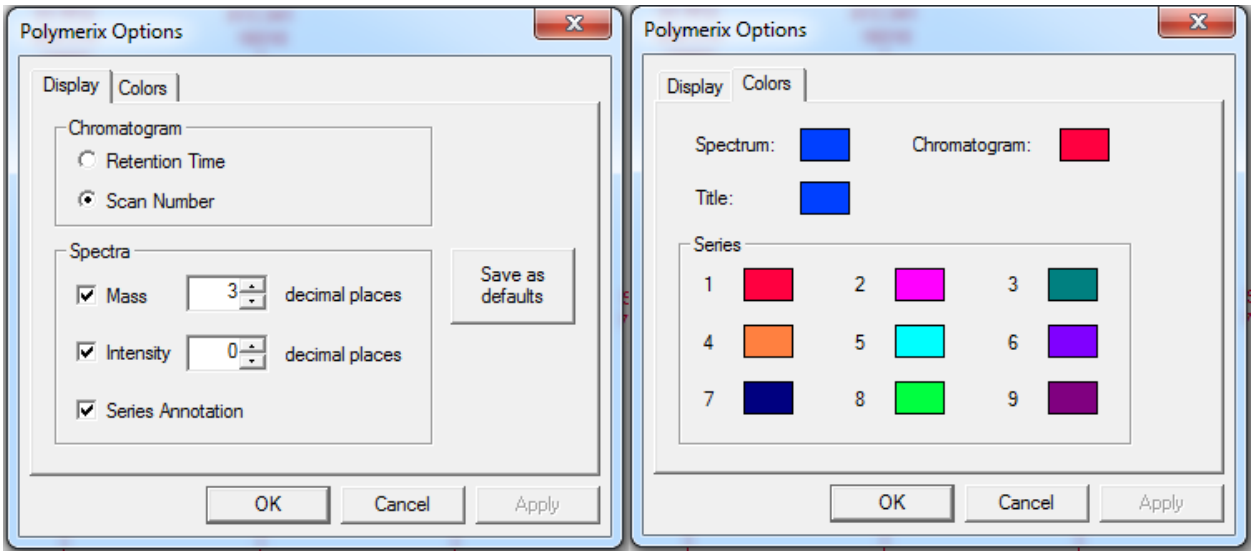


Going back, if for example, S1 and S2, should be combined because they represent the same polymer molecular identities, just as different adduct ions. This would be done by clicking the "*Adducts and losses*" box in "*Combine Within Series With Same Repeat*". The series summary table changes to reflect this.



Series Label	Mn	Mw	Mz	PD	DPn	DPw	DPz	Percent Series	Percent Spectrum	Alpha End Group	Repeat	Omega End Group	Charge State	Adduct	Loss	Series Formula
Total/Average	580.687	602.257	623.462	1.037	12.762	13.252	13.733	100.00	97.85	H	C2H4O		1			
S1 + S2	583.004	604.732	626.067	1.037	12.833	13.327	13.811	97.05	94.96	H	C2H4O	OH	1			
S3	504.489	520.894	537.812	1.033	10.414	10.786	11.171	2.95	2.89	H	C2H4O	CHO2	1	Na		H [C2H4O]n CHO2 + Na

Figure display settings can be edited by clicking the *Edit>Options…* at the top of Polymerix, and colors can be changed by clicking the "*Colors*" tab.



Polymerix allows any graphic to be copied to the clipboard as currently displayed by first left clicking it and then going to *Edit*>*Copy Spectrum Plot*. However, these can have resolution issues or appear as unexpected dimensions, so it is often better to just screenshot images using the Snipping Tool on the computer or by doing a print screen (**ctrl** + **print screen** key) and cropping in a graphics program like MS Paint. Spectra can be copied as peak lists through *Edit*>*Copy Spectrum Data*. Tables can be copied through *Edit*>*Copy Results* into Excel or Word. The original unprocessed spectrum can be exported as a peak list through *File*>*Export Spectrum*. Summary tables of results can be exported through *File*>*Export Results*. Result reports can be generated through *File*>*Print*, selecting what to include, and making sure print to pdf is selected under setup to create a pdf report.

Relative Area

Clicking the "*Relative Area*" button at the side panel brings up options for calculating relative intensity for selected ranges of the spectrum. Ranges can be made by typing in the table, adding or removing lines as needed, or by right clicking and dragging across the region of interest in the spectrum.

Mass Ranges to Analyze

+	Range	From	To
--	1	378.1	533.4

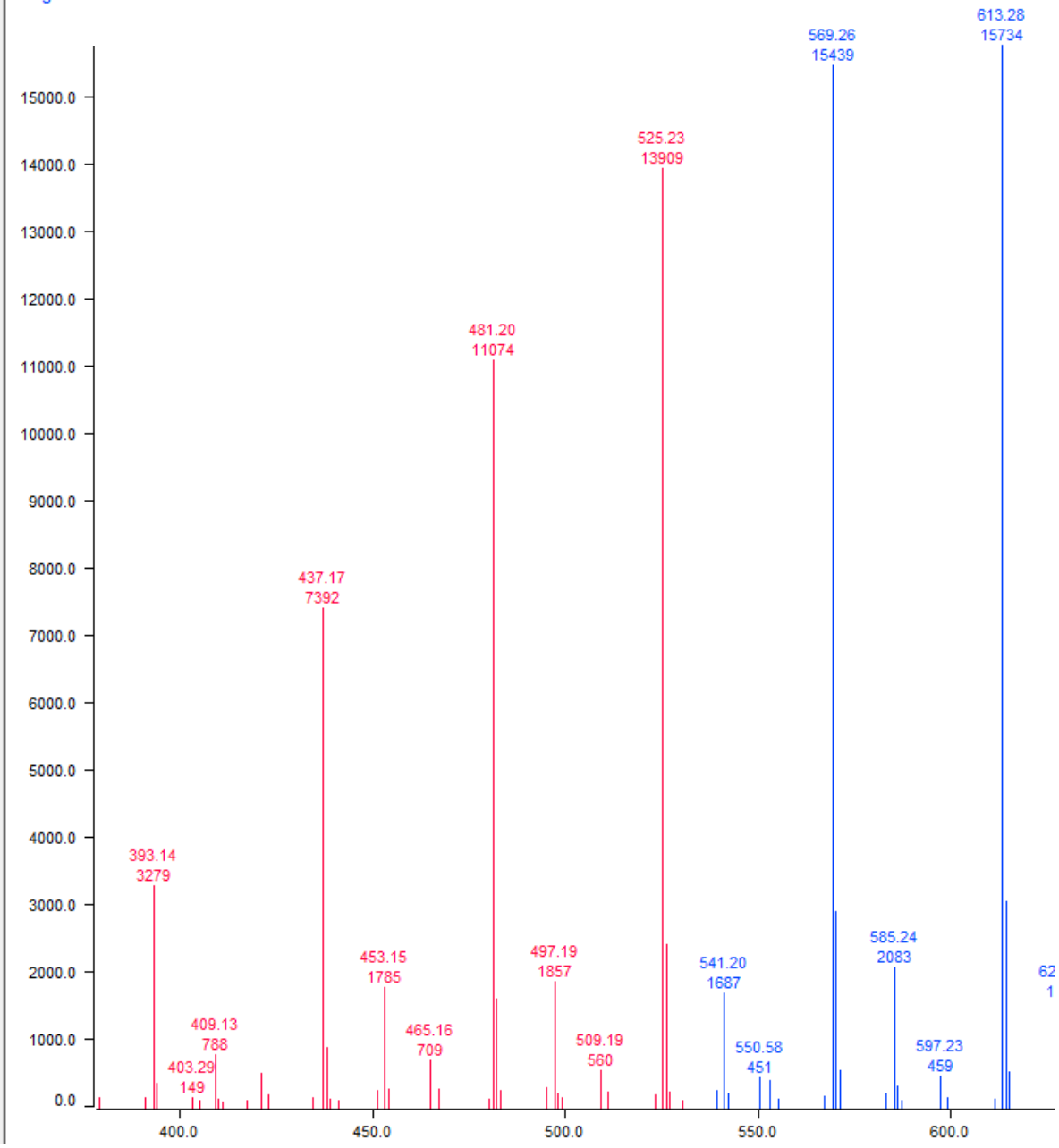
Apply

Cancel

Spectrum Mass Ranges Area Analysis Results

Range	From	To	Intensity	% Ranges	% Spectrum
Total			51166.7	100.0	32.6
1	378.1	533.4	51166.7	100.0	32.6

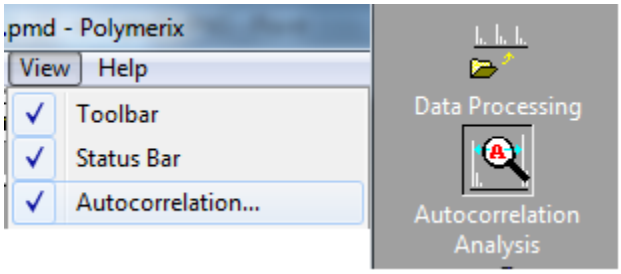
Ranges selected for area measurement



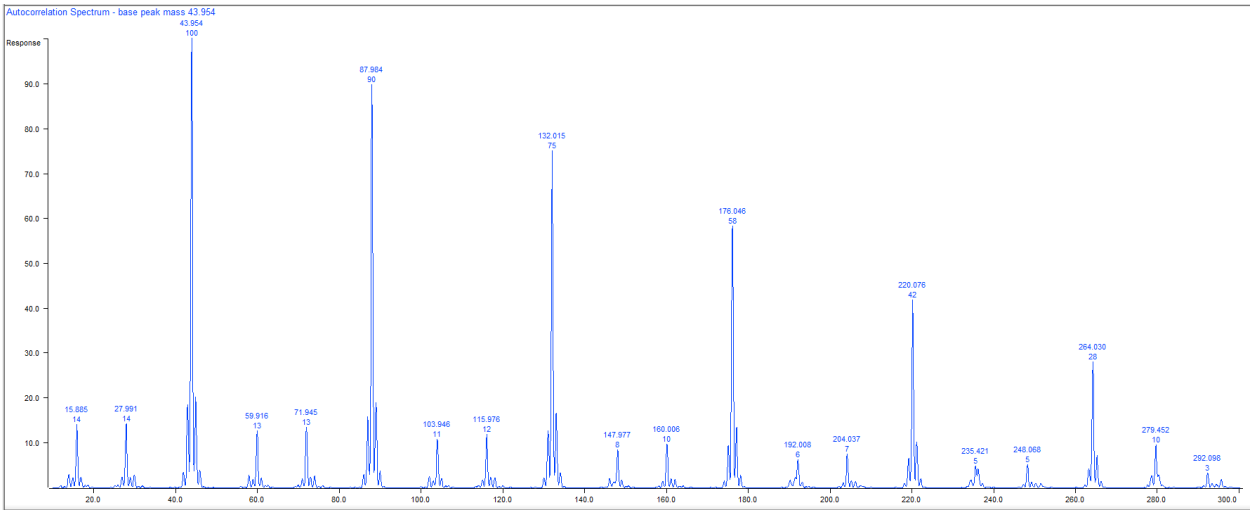
Autocorrelation Analysis

The *Autocorrelation Analysis* feature is largely replaced by *Find Homopolymer Series* options, but it still could be useful in all types of homologous series analysis and closely and intuitively follows more "manual" polymer spectral interpretation.

Note: The autocorrelation feature has been depreciated and hidden from initial startup options. If it is not currently visible in the sidebar, it can be brought back by going to *View>Autocorrelation…* and then closing and reopening the program.



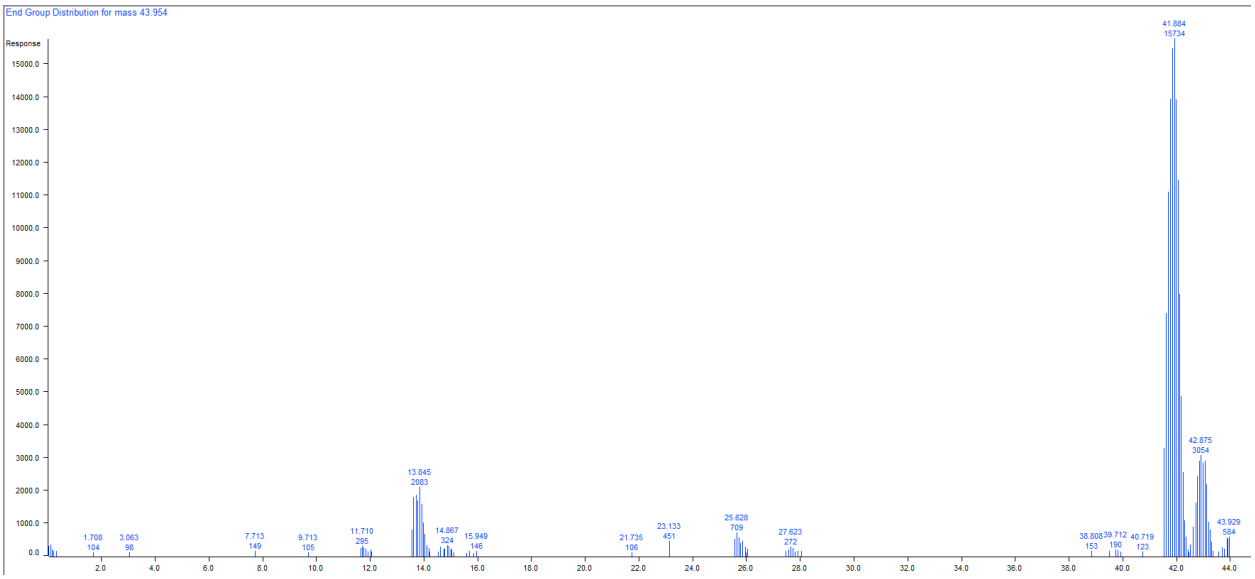
This will bring up data panels where the higher panel is an autocorrelation spectrum. This is done through a Fourier transform on the original spectral data.



This shows overtones at ~44 (the base peak), 28, and 16 *m/z*, as well as decreasing abundance at their higher order multiples. These provide a hint for the identities of the major repeating pattern series. The largest peak is likely the monomer unit, and 44 *m/z* suggests an ethylene oxide (C_2H_4O) unit, aka a polyethylene glycol (PEG) mass spacing. The 16 *m/z* spacing suggests possible presence of both a Na and corresponding K adduct series. The major isotopes of Na (~23 Da) and K (~39 Da) have a ~16 *m/z* mass difference. The 28 *m/z* overtone is less obvious, but could represent something like a carboxyl end group (COOH vs OH).

This requires some understanding of typical mass differences in mass spectrometry. Formula assignment tools like mentioned in the *MALDI Data Analysis* guidance document may help with interpretation of overtones, along with search engine and literature searches for observed

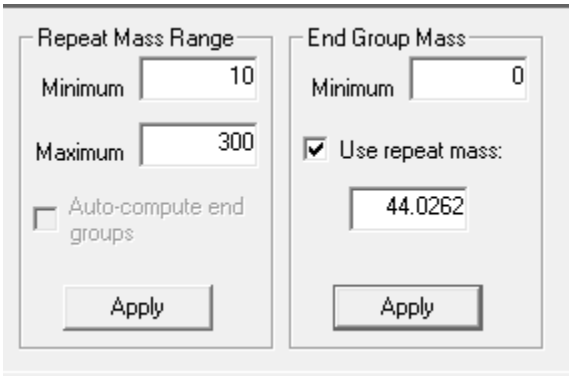
mass differences and spacings. There are also many lists of common repeating units and mass differences in mass spectrometry, such as the Excel list from University of Washington's Proteomics Resource (https://proteomicsresource.washington.edu/protocols05/esi_background_ions.php) or the MilliporeSigma document (https://www.emdmillipore.com/Web-PR-Site/en_CA/-/USD/ShowDocument-Pronet?id=201604.110).



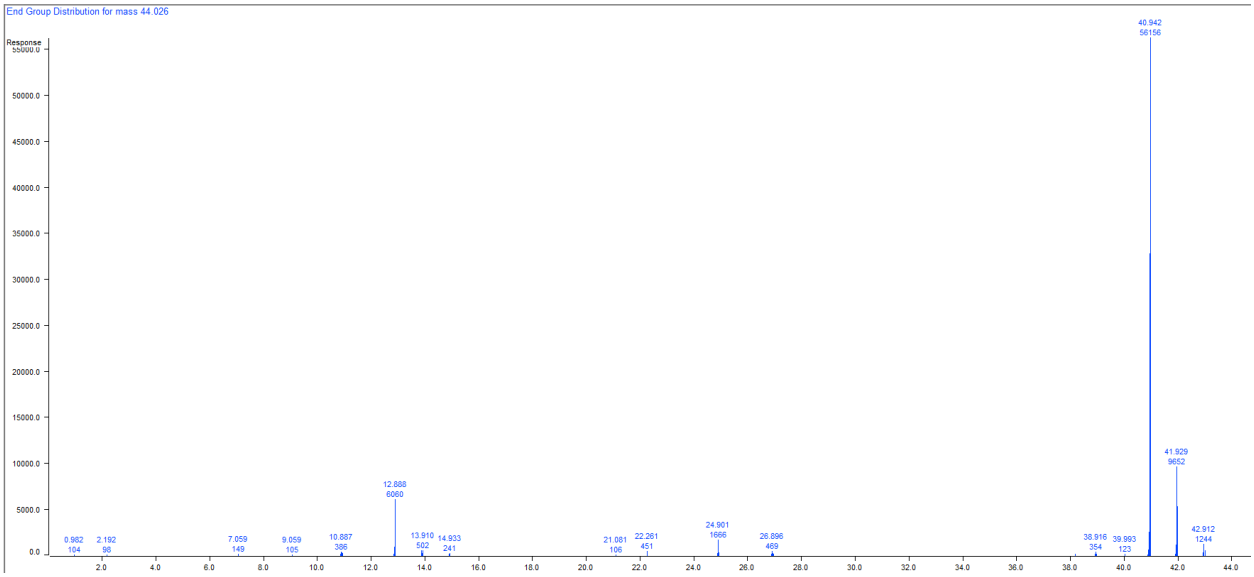
The autocorrelation spectrum is used to make the "End Group Distribution" spectrum. Polymerix takes the base peak from the autocorrelation spectrum and considers it the monomer unit. It then does a modulo on the *m/z* values of the original spectrum and plots remainders. There is a main cluster at ~41.8 and +1 and +2 *m/z* from this and other clusters visible; however, this spectrum is fairly noisy.

The "43.954" *m/z* monomer, does not clearly mean anything, though it appears to be very close to ethylene oxide (44.0262). This is due to the inherent inaccuracy of the initial sample spectrum, particularly for TOF spectra, where detector dead time causes peak tailing, impacting the accuracy of centroiding.

So, going through this process again with optimized settings. The first box of the top settings controls the autocorrelation spectrum settings.



The repeat mass range for identifying the monomer can be changed. Here, the default 10 to 300 *m/z* works well and is expected to cover the repeat unit. However, due to mass error, the default monomer was slightly off, so the end group spectrum was noisy. Unchecking "*Auto-compute end-groups*" and clicking *Apply*, allows this to be manually set to the actual mass of ethylene oxide, 44.0262 *m/z*, then click *Apply*. The end group spectrum is now much cleaner.

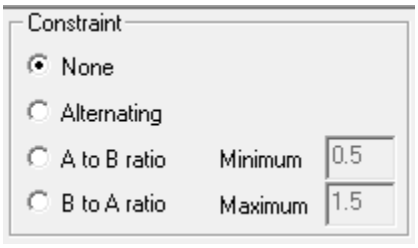


It is now clear that the peak is actually at ~40.9 and +1, +2 *m/z* isotopic peaks, with the second biggest at ~12.9 and +1, +2 *m/z* isotopic peaks. There are also smaller peaks at ~24.9 and 26.9 *m/z*. These remainder masses represent end groups and any adduct mass. For example, using what's been identified from the autocorrelation spectrum, a reasonable guess for the largest series is that it is a sodium adduct series. So, 40.942 - 22.990 (mass of Na) = 17.952 Da or ~18 which could suggest H_2O as a combined end group, which is correct for a H and OH terminated PEG.

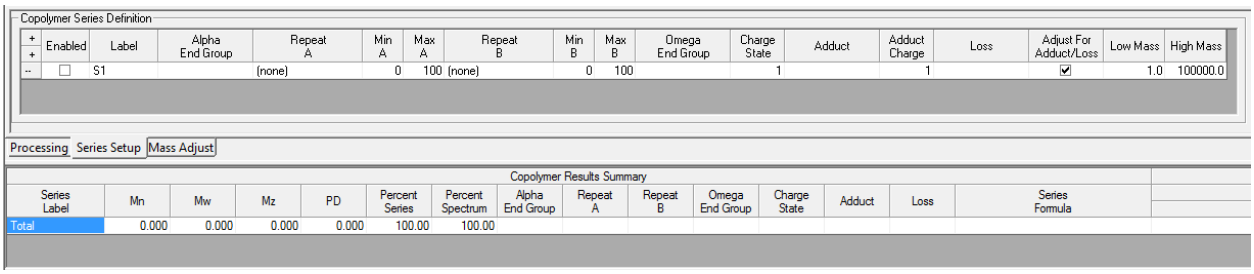
The next largest is at ~ 12.9, and from what is known is suspected of being a K adduct. If the expected remainder mass is larger than the repeating unit, then an extra repeating unit will be subtracted when doing the modulo operation. A larger ^{39}K (38.964 Da) and this remainder being fairly small, suggests this might have happened. So, 12.888 + 44.026(ethylene glycol) = 56.914 Da and 56.914 – 38.964 = 17.950 Da, once again suggesting a combined end group of H_2O. Continuing with the 3rd series, 24.901 + 44.026 = 68.927 Da and 68.927 - 22.990 (mass of Na) = 45.937, which could be something like CH_2O_2 in consideration against the H_2O. The 26.9 is the unclear very minor 4th series previously mentioned that only appears at the lower end of the spectrum. This may be some kind of fragmentation product (the 3rd series may also be a fragmentation product, other transformation product, or some kind of standard impurity). More certain identification of these would require further experimentation.

Copolymer Analysis

Click the "*Copolymer Analysis*" button in the side panel. In Polymerix, copolymer refers to a polymer with two different repeat units, A and B of n and m counts. It could be a block, alternate, or random co-polymer, but this is indistinguishable in the mass spectrum and all would be treated the same regarding parameter characterization. Most settings are similar to Homopolymer Analysis and will not be described here. Constraints can be set for the ratio of A and B.



Alternating is A = B±1, so allows for one extra A or B. If A = B is needed, this can be done through the ratio option. Click "*Apply*" after choosing settings.



The "*Series Setup*" tab now has options to fill in info for two repeating units. Click "*Apply*" to assign the spectrum. There are now additional display tabs for exploring the A and B distributions.