# SUPPORTING INFORMATION

## Optimum Cell Pathlength or Volume for Absorbance Detection in Liquid Chromatography. Transforming Longer Cell Results to Virtual Shorter Cells.

Akinde F. Kadjo, Purnendu K. Dasgupta* and Charles Phillip Shelor

Department of Chemistry and Biochemistry, University of Texas at Arlington, Arlington, Texas 76019-0065, United States

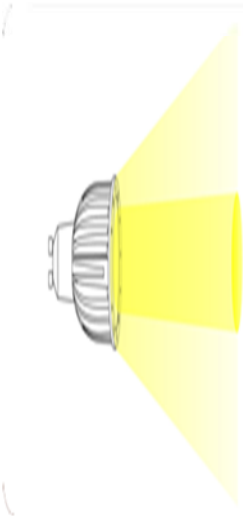
### Table of Content



Table of contents continued on next page

*Corresponding Author

Email: Dasgupta@uta.edu Fax: (817)-272-3808



| Flow → | | | | | What The Detector Sees | Step Difference |
|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 0 | 0 | 0 | 1 | 1 |
| 3 | 1 | 0 | 0 | 0 | 4 | 3 |
| 7 | 3 | 1 | 0 | 0 | 11 | 7 |
| 11 | 7 | 3 | 1 | 0 | 22 | 11 |
| 14 | 11 | 7 | 3 | 1 | 36 | 14 |
| 16 | 14 | 11 | 7 | 3 | 51 | 16 |
| 17 | 16 | 14 | 11 | 7 | 65 | 17 |

**Figure S1.** The cell is imagined as a series of 5 box cars. The row-wise depiction is 8 sequential snapshots of the cell in time. Flow goes from left to right. Initially all five boxes are filled with non-absorbing eluent (0 mAU, say). The total absorbance read (green column) is zero. During the observation of the second data point, the first box is filled with 1 mAU worth of absorption, the overall reading is the same and by subtracting the previous total from it, one is able to ascertain that the first box has been filled with 1 mAU worth of material (red column). By the next snapshot, the contents of the leftmost box has moved over to the next one and the leftmost box now represents 3 mAU worth of absorption, the detector reading 4. But subtracting the previous reading we are able to determine the contents of each box. This goes on until the $7^{th}$ row where and whenceforth we must also take note that the contents of the rightmost box is now no longer in our view.

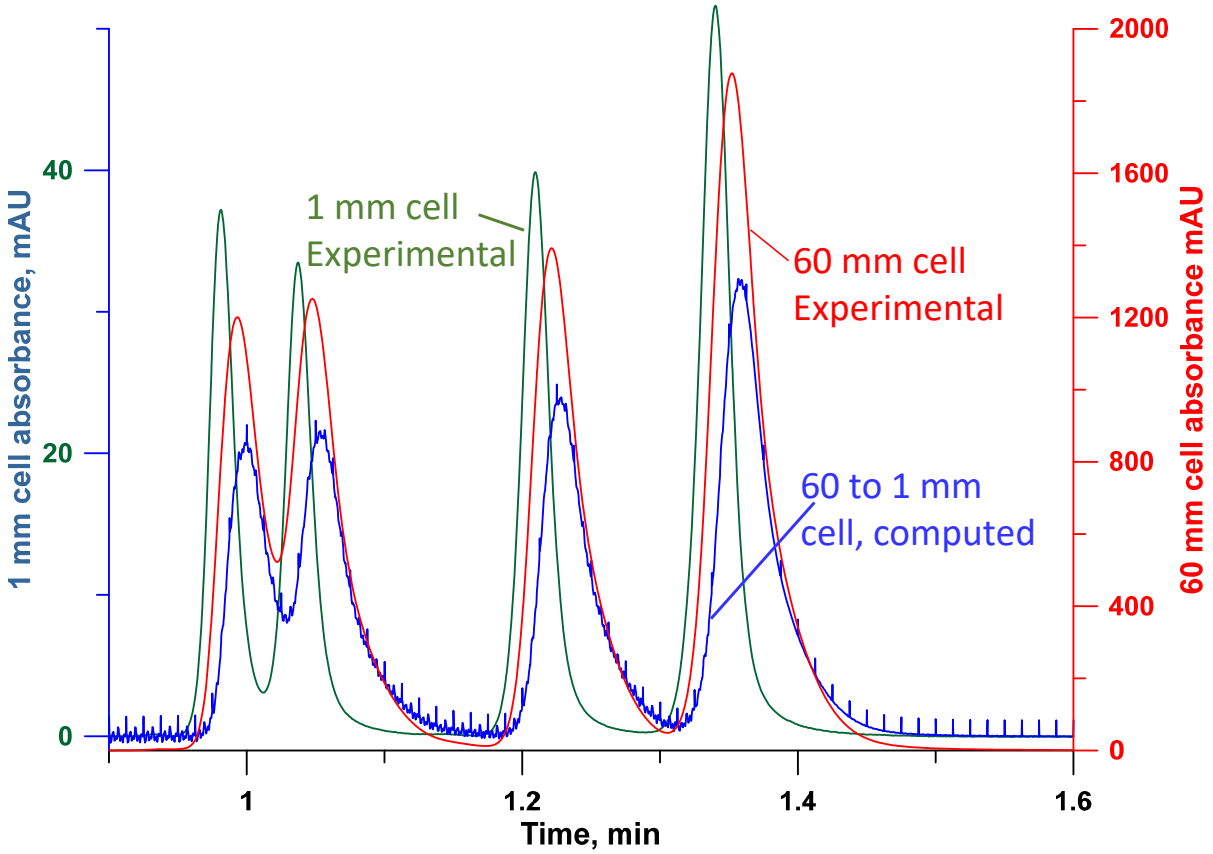



**Figure S2.** Applying a reverse box car transform to 60 mm path data to generate data for the 1 mm cell does not match what is observed in a real experimental 1 mm cell. Peaks are much wider. Note also that at a fixed interval (equaling the assumed overall box length in the time domain, i.e., the average residence time assumed for a given fluid element in the cell), spikes are observed.

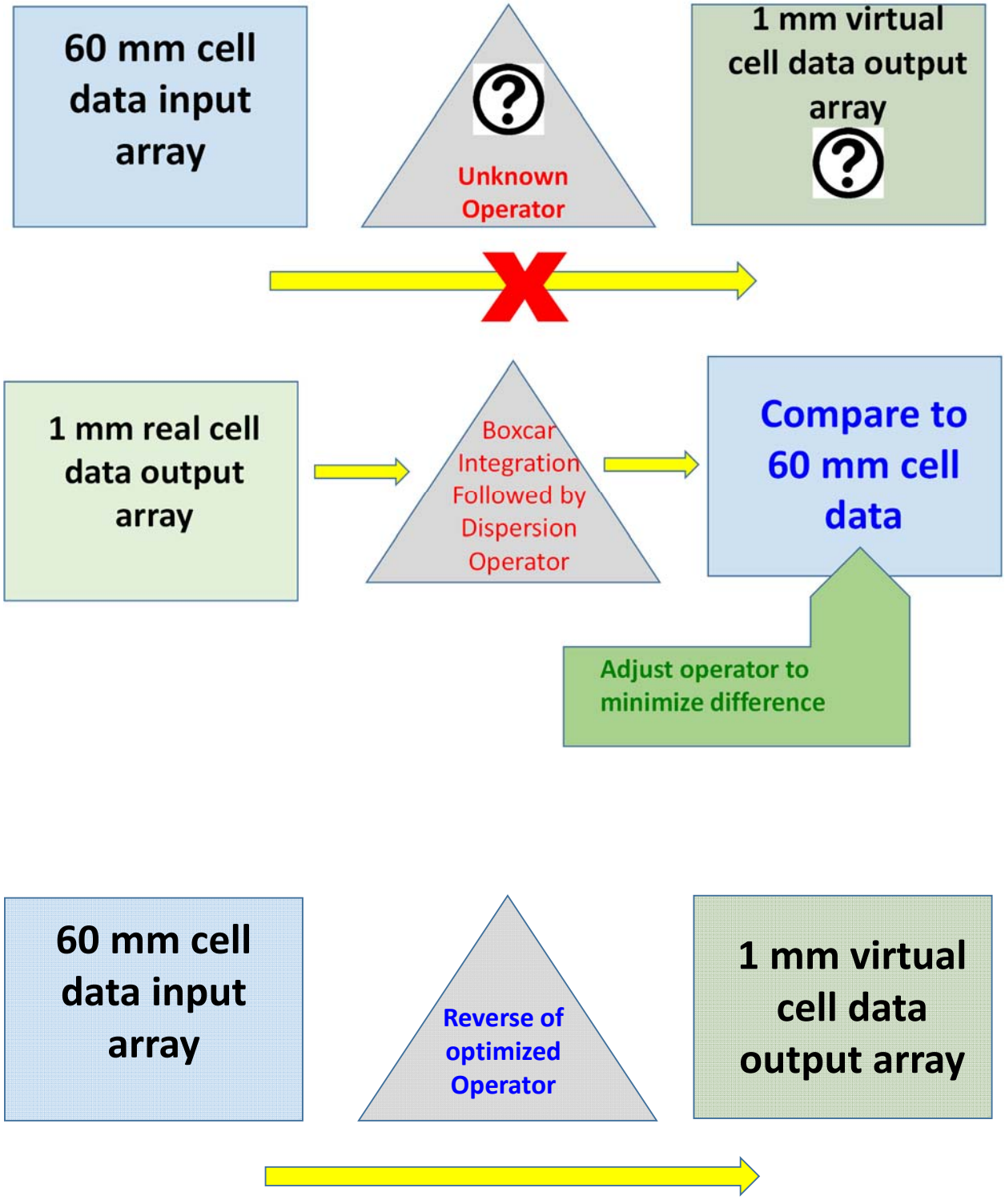
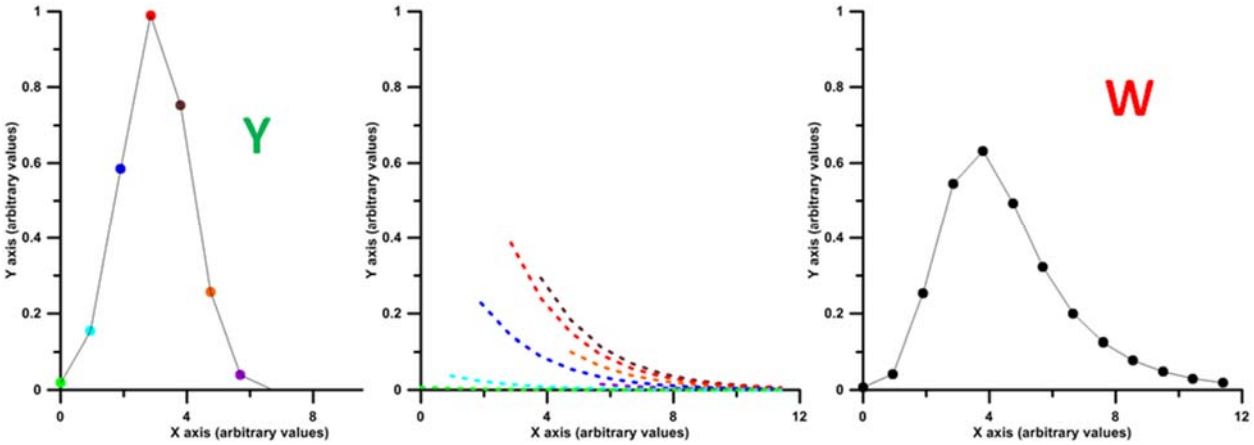



**Figure S3.** The strategy involves finding a dispersion operator to act on the boxcar-integrated 1 mm cell data to best match the observed 60 mm cell data. The operator will then be reversed to act on the 60 mm cell data and divided as needed to produce the data for the desired cell length.

**The Dispersion Model**




Y data points each decay as $\beta e^{-\alpha t}$
The dispersed signal becomes:
$W_n = W_{n-1} * e^{-\alpha} + Y_n * \beta$, $W_0$ being given by $W_0 = Y_0 * \beta$

**Figure S4.** Consider for example our peak is represented by 7 points ($t$ = 0, 1, ...6). The original peak therefore consists of the 7 points $y_0$, $y_1$, $y_2$, ...$y_6$, etc. Each point is then dispersed forward, creating 28 member two-dimensional array (See Table S1). $W_0$ through $W_6$ as shown overleaf are then simply the sum of the corresponding time columns.

**Conservation principle**

For the conservation of peak areas, we have:

$$Y = \sum_{t=0}^{\infty} Y\beta e^{-\alpha t} \iff 1 = \sum_{t=0}^{\infty} \beta e^{-\alpha t} \iff 1 = -\frac{\beta}{-1+e^{-\alpha}} \iff \beta = 1 - e^{-\alpha} \quad \ldots(S1)$$

**Table S1. Illustration of the Dispersion Operation (see Figure S4).**

| $y_{i,t}'$ | $t=0$ | $t=1$ | $t=2$ | $t=3$ | $t=4$ | $t=5$ | $t=6$ |
|---|---|---|---|---|---|---|---|
| $y_{0,t}'$ | $\beta y_0$ | $\beta y_0 e^{-\alpha}$ | $\beta y_0 e^{-2\alpha}$ | $\beta y_0 e^{-3\alpha}$ | $\beta y_0 e^{-4\alpha}$ | $\beta y_0 e^{-5\alpha}$ | $\beta y_0 e^{-6\alpha}$ |
| $y_{1,t}'$ | | $\beta y_1$ | $\beta y_1 e^{-\alpha}$ | $\beta y_1 e^{-2\alpha}$ | $\beta y_1 e^{-3\alpha}$ | $\beta y_1 e^{-4\alpha}$ | $\beta y_1 e^{-5\alpha}$ |
| $y_{2,t}'$ | | | $\beta y_2$ | $\beta y_2 e^{-\alpha}$ | $\beta y_2 e^{-2\alpha}$ | $\beta y_2 e^{-3\alpha}$ | $\beta y_2 e^{-4\alpha}$ |
| $y_{3,t}'$ | | | | $\beta y_3$ | $\beta y_3 e^{-\alpha}$ | $\beta y_3 e^{-2\alpha}$ | $\beta y_3 e^{-3\alpha}$ |
| $y_{4,t}'$ | | | | | $\beta y_4$ | $\beta y_4 e^{-\alpha}$ | $\beta y_4 e^{-2\alpha}$ |
| $y_{5,t}'$ | | | | | | $\beta y_5$ | $\beta y_5 e^{-\alpha}$ |
| $y_{6,t}'$ | | | | | | | $\beta y_6$ |

$$W_0 = \beta y_0 \quad \ldots(S2)$$
$$W_1 = \beta y_0 e^{-\alpha} + \beta y_1 \quad \ldots(S3)$$
$$W_2 = \beta y_0 e^{-2\alpha} + \beta y_1 e^{-\alpha} + \beta y_2 \quad \ldots(S4)$$
$$W_3 = \beta y_0 e^{-3\alpha} + \beta y_1 e^{-2\alpha} + \beta y_2 e^{-\alpha} + \beta y_3 \quad \ldots(S5)$$
$$W_4 = \beta y_0 e^{-4\alpha} + \beta y_1 e^{-3\alpha} + \beta y_2 e^{-2\alpha} + \beta y_3 e^{-\alpha} + \beta y_4 \quad \ldots(S6)$$
$$W_5 = \beta y_0 e^{-5\alpha} + \beta y_1 e^{-4\alpha} + \beta y_2 e^{-3\alpha} + \beta y_3 e^{-2\alpha} + \beta y_4 e^{-\alpha} + \beta y_5 \quad \ldots(S7)$$
$$W_6 = \beta y_0 e^{-6\alpha} + \beta y_1 e^{-5\alpha} + \beta y_2 e^{-4\alpha} + \beta y_3 e^{-3\alpha} + \beta y_4 e^{-2\alpha} + \beta y_5 e^{-\alpha} + \beta y_6 \quad \ldots(S8)$$

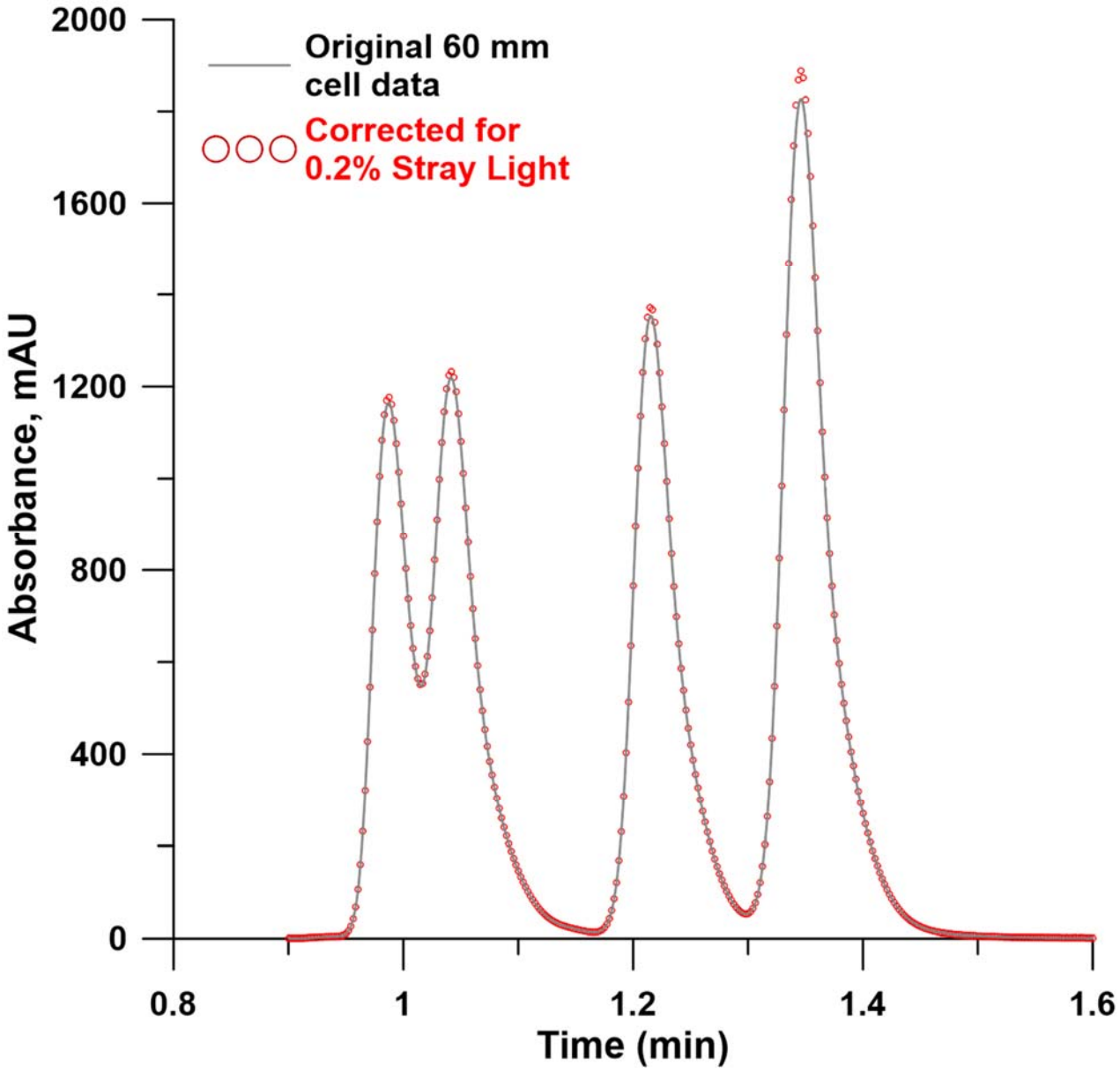



**Figure S5.** Stray light correction (0.2%) on the 60 mm cell data produces a minor but perceptible difference.

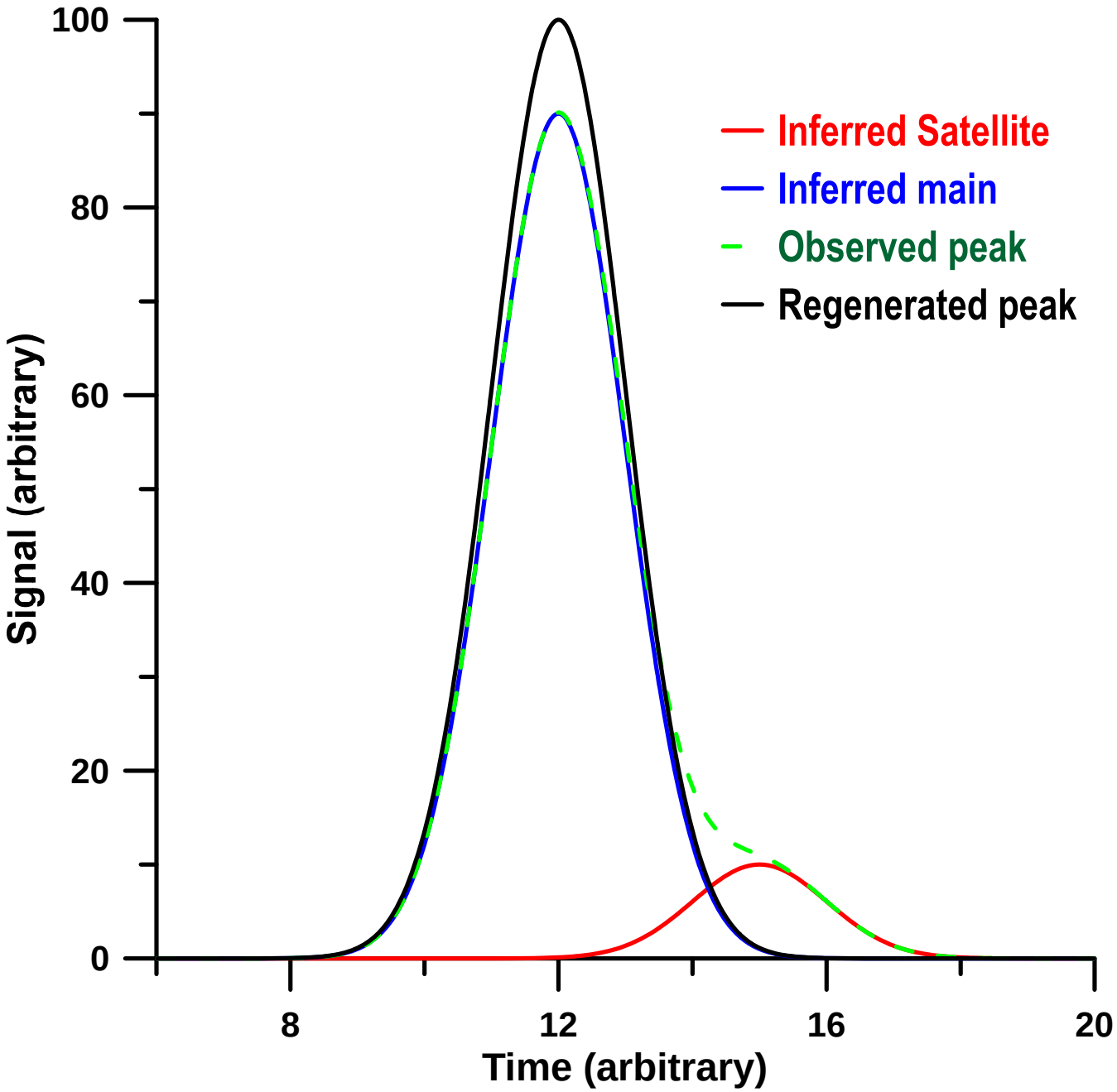
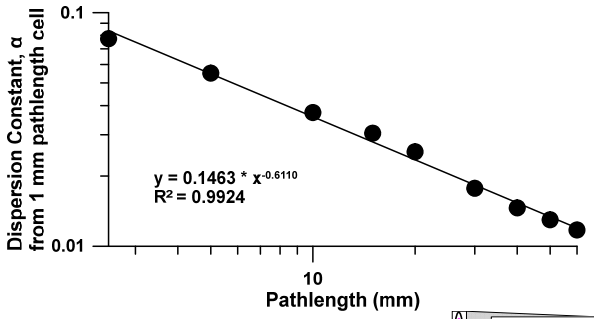
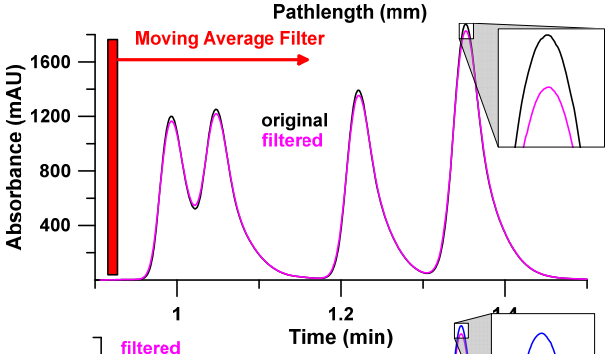



**Figure S6.** The green trace simulates the signal as it appears upon dispersion reversal. of the satellite peaks on reversal. While this is illustrated with Gaussian peaks, the approach to correction for these peaks is independent of peak shapes. The green trace can be seen to be composed of a main peak (blue trace) and a reduced version thereof (red trace, satellite) appearing a finite time after. Equations 18-19 removes the satellite and generate the black trace (with the same area as the green trace).

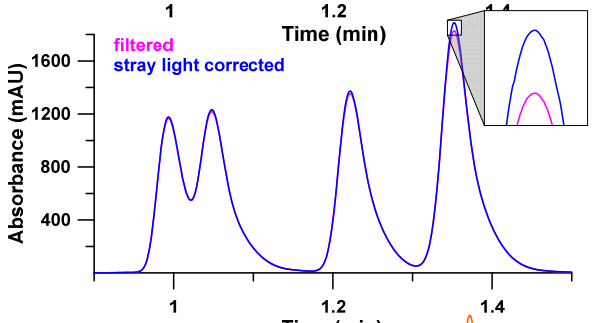
**Step 1**. Determine the dispersion constants, α and β necessary to reverse dispersion. α is a physical property of a cell and only needs determined once and may be extrapolated from known lengths.
EQ 14: β = 1- exp(-α)



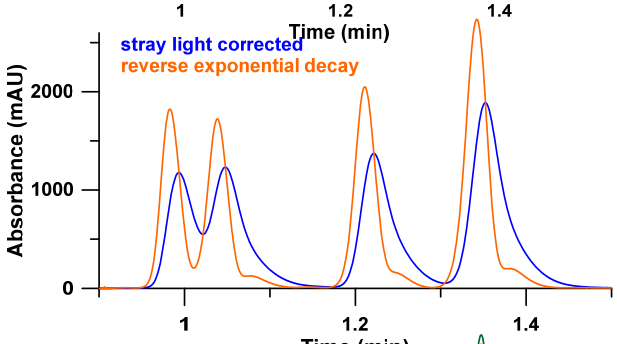

**Step 2**. A simple 61 point moving average filter is applied centered at the same time.




**Step 3.** Stray light correction.
Eq 16: $A_{corr} = -\log \frac{(100 + s)\ 10^{-A_{obs}} - s}{100}$




**Step 4.** Reverse the exponential decay (dispersion)
Eq 17: $Y_i = (W_i - W_{i-1} * e^{-a})/\beta$




**Step 5.** Satellite Peak Removal
Eq 18: $X_i = 1/(1-a)Z_i - aX_{i-n}$

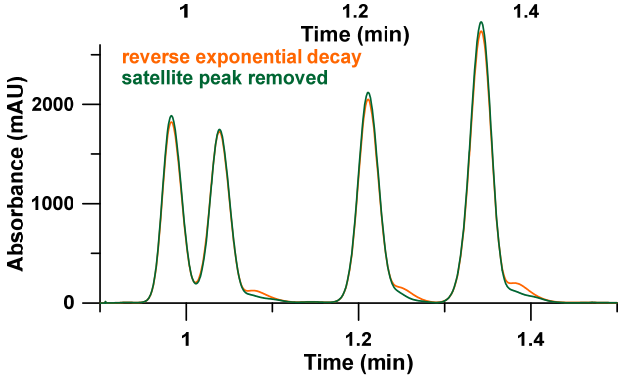



**Figure S7.** Flow chart depicting the sequential steps and formulae used to correct dispersion from a long to short path (60 mm to 1 mm shown). The relevant equations from the main text are provided. The final step not shown is simply dividing the signal by the ratio of the long path to the short path lengths to achieve the simulated shorter path chromatogram.

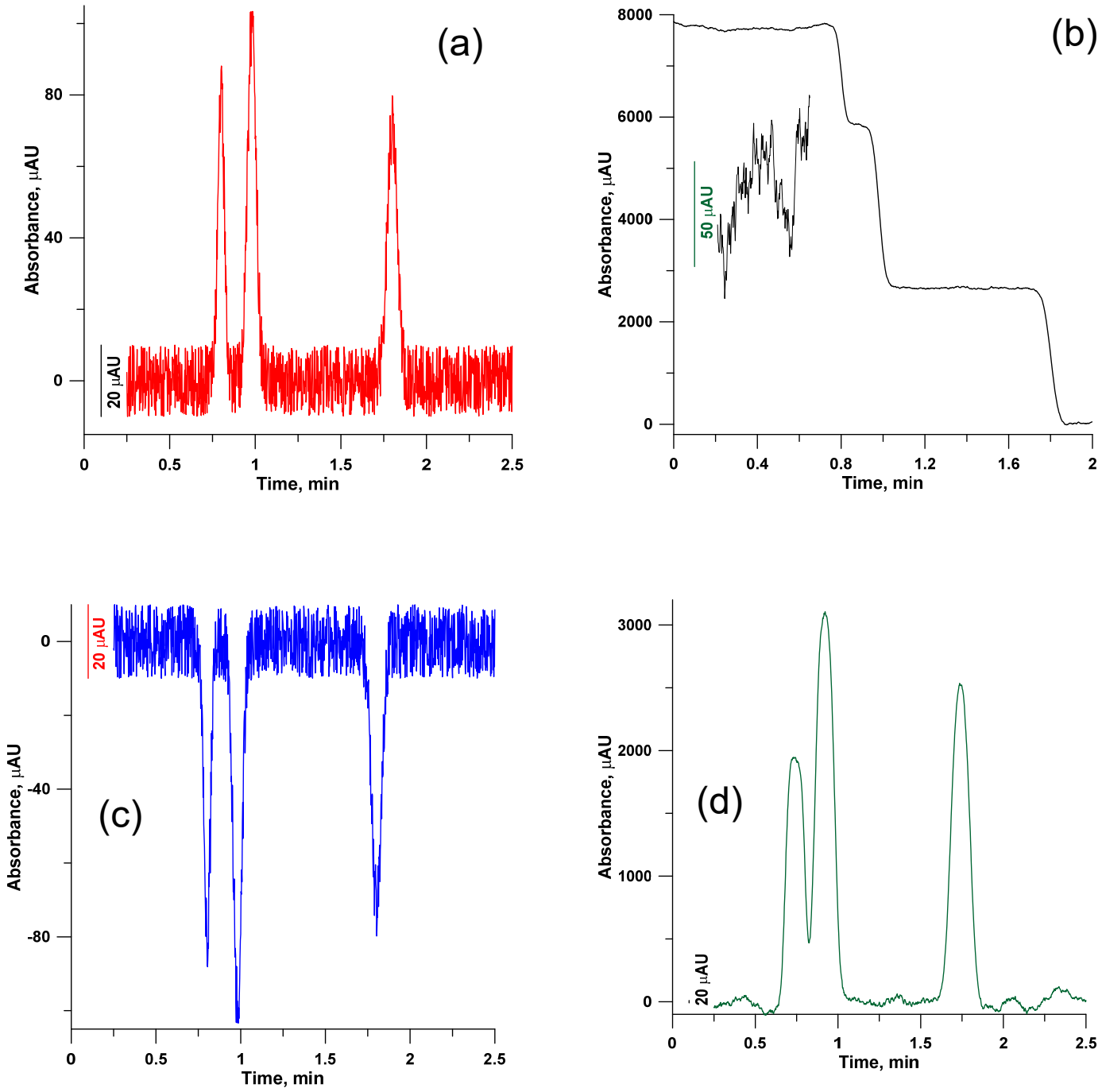



**Figure S8.** (a) A simulated chromatogram containing a significant amount of random noise, as it may look through a cell of unit path length at the end of a 1250-unit long column. The flow rate is such that the fluid is moving through a unit path length per data slice. (b) The signal as it would appear (we do not consider here the additional shot noise from vastly decreased light throughput) if the sample was introduced at time zero and the entire column served as the detector. Note the magnified inset for a portion of the trace - the short term noise has not increased but the accumulation of the random noise contributes to drift; (c) the differential of the signal in b is a mirror image of the chromatogram in a. (d) chromatogram that would be obtained with an end-column cell with a 60 unit path length, Differentiation of d will yield the same chromatogram as a. The generation of figures b - d from the data in a did not take longitudinal dispersion into account.

**Table S2. Table 3 Data (including uncertainties).**

| | S/N | | | $W_{0.5}$ (s) | | | $\Delta W_{0.5}$ (ms) | | | S/N Gain | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | | |
| 1-mm unfiltered | 408 | ± | 67 | 2.783 | ± | 0.007 | | | | | | |
| 1-mm SMA-filtered | 647.3 | ± | 147 | 2.850 | ± | 0.000 | 66.7 | ± | 7.2 | 1.6 | ± | 0.1 |
| 1-mm GKMA filtered | 617 | ± | 142 | 2.825 | ± | 0.000 | 41.7 | ± | 7.2 | 1.5 | ± | 0.1 |
| | 0 | | | | | | | | | | | |
| 60-mm unfiltered | 9793 | ± | 2930 | 3.546 | ± | 0.019 | | | | | | |
| 60-mm SMA filtered | 12969 | ± | 5760 | 3.608 | ± | 0.019 | 62.5 | ± | 27.0 | 1.3 | ± | 0.2 |
| 60-mm Gaussian filtered | 12713 | ± | 5467 | 3.583 | ± | 0.019 | 37.5 | ± | 27.0 | 1.3 | ± | 0.2 |
| | 0 | | | | | | | | | | | |
| 60 mm → 1 mm unfiltered | 618.3 | ± | 134 | 2.692 | ± | 0.007 | | | | | | |
| 60 mm → 1 mm SMA filtered | 6347 | ± | 1837 | 2.758 | ± | 0.007 | 66.7 | ± | 10.2 | 10.3 | ± | 0.9 |
| 60 mm → 1 mm GKMA filtered | 6770 | ± | 1970 | 2.733 | ± | 0.007 | 41.7 | ± | 10.2 | 10.9 | ± | 1 |

[a]S/N defined as peak amplitude/(4*baseline standard deviation)
[b]standard (rectangular kernel) moving average filter, 60 points (750 ms) wide
[c]Gaussian kernel moving average filter standard deviation of 14 points (175 ms)

The same filter applied to any particular cell results essentially in the same added broadening as listed in the column $\Delta W_{0.5}$. However, the S/N improvement is >10 times in the dispersion-reversed long cell. This is simply due to the filtering of high frequency noise and the feedback into the system caused by it. In any case, one gets an order of magnitude improvement even relative to the same filtration in the short cell. The noise is the average from 8 15 s samples, 3 samples from before elution of peak 1, 2 samples from the middle (between peaks 4 and 5), and 3 samples from the end (after peak 5).

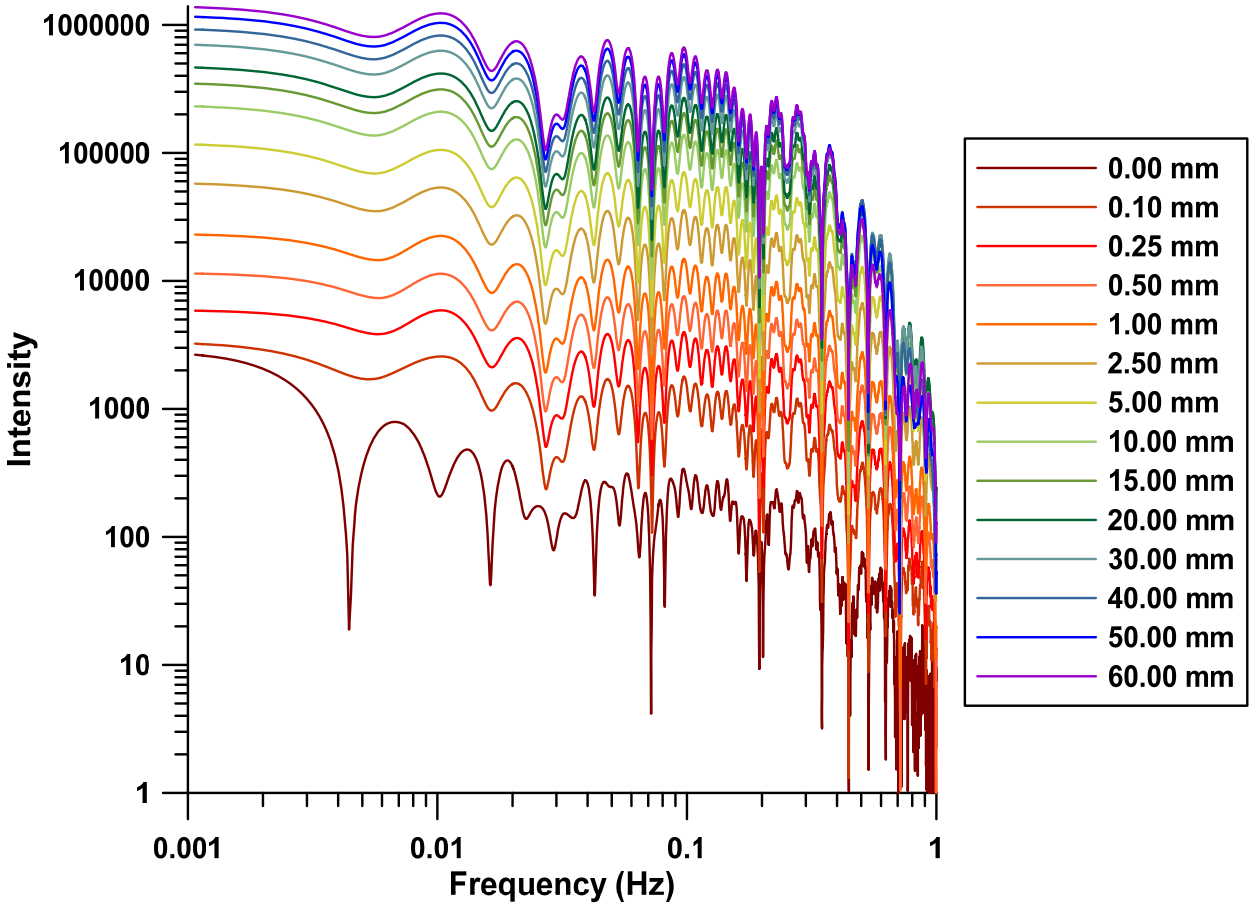



**Figure S9**. Fourier Transformations of chromatograms (frequency domain spectrum, $L(\omega)_{av}$ for different cell lengths *L*. The DC component is the integrated peak area and increases with the increasing pathlength. Note the similarities in fine structure. Data begin to converge at frequencies higher than that shown. Zero path length has the optical fibers touching each other but apparently a liquid layer is still present between them. Zero padding was used to increase available data points to $2^{20}$, see Figure S11 and main text.

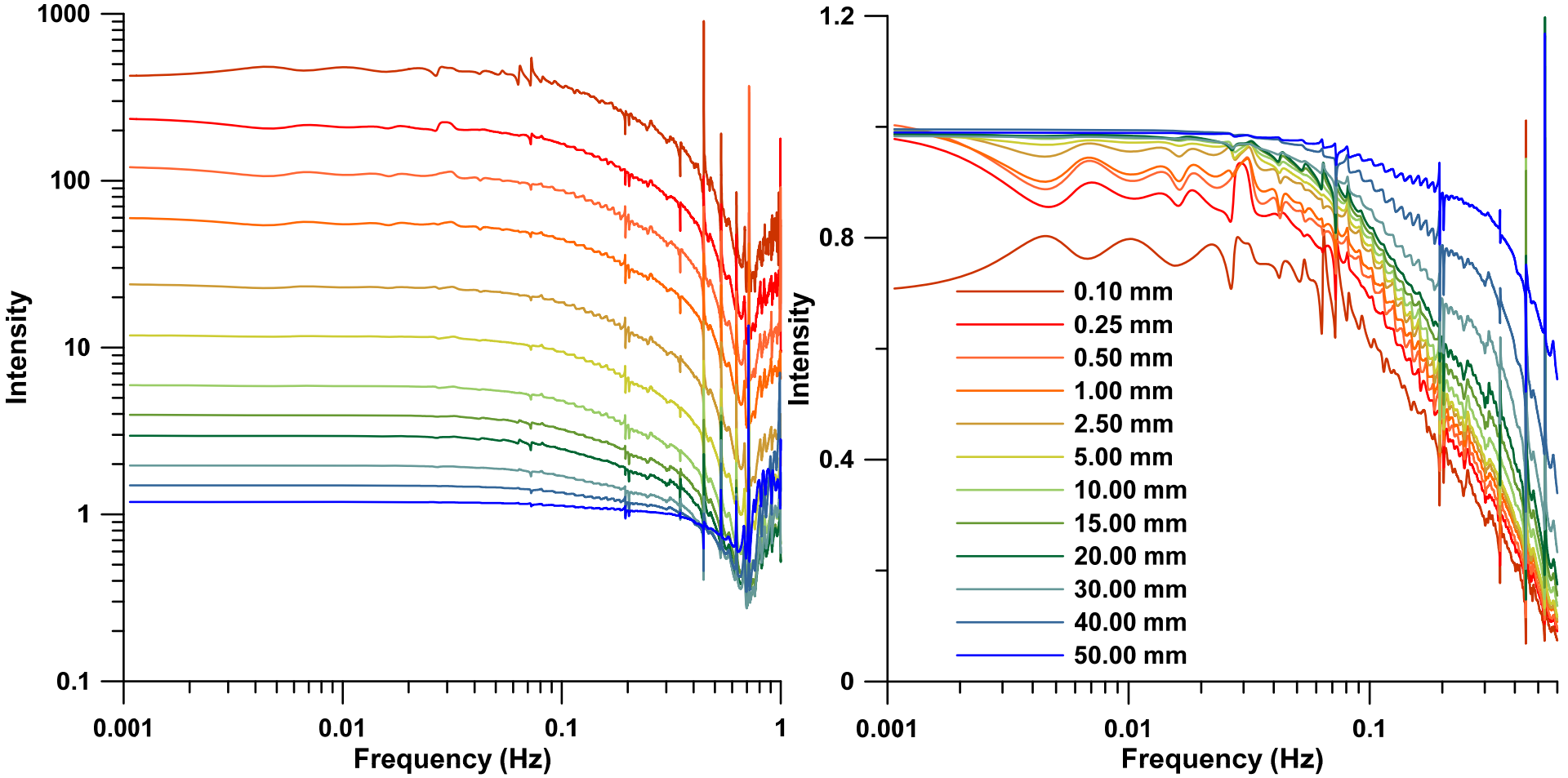



**Figure S10**. Left: Deconvolution Function $D(\omega)$, the longer path *L* was fixed at 60 mm, the different color-coded traces (see right panel) pertain to different short path lengths *S*. The Right panel shows the same function normalized by the pathlength ratio. Note that the DC contributions of the normalized function in the right panel converge to unity. The shape of the high frequency portion of the curve controls how the peak is reshaped upon deconvolution. The steeper the slope, the greater the required dispersion. Minor inaccuracies in pathlength or non-linearity due to uncorrected stray light may account for the observed differences for the short path cells in the low frequency region. Note that the left figure has log scaling of the ordinate and the right does not. Zero padding was used to increase available data points to $2^{20}$, see Figure S11 and main text.

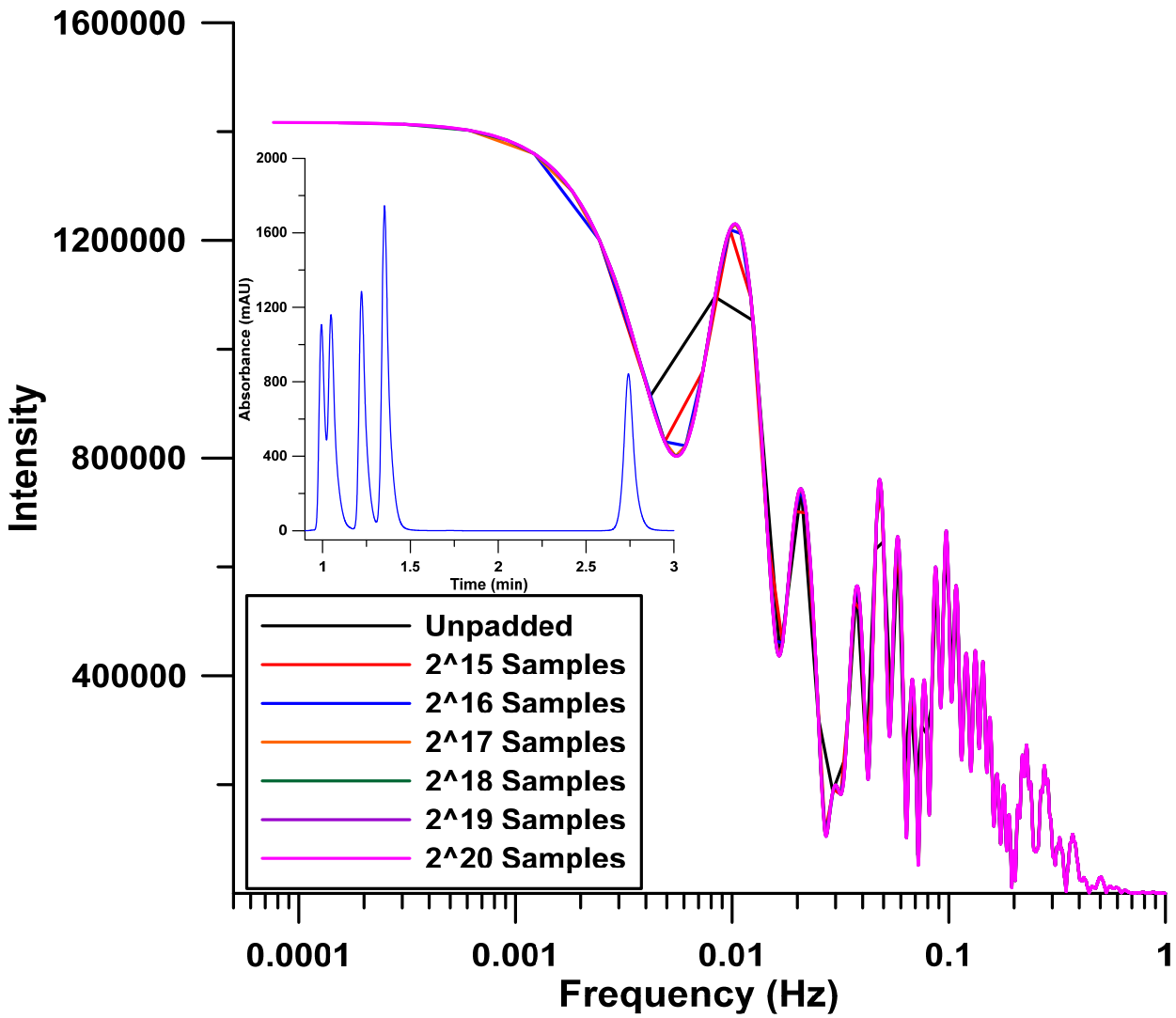



**Figure S11.** Zero Padding Effect on Fourier Transformation of chromatograms. Shown is the 60 mm data truncated to 1 Hz with the time domain chromatogram in the inset. Resolution following FT is *f/N* where *f* is the sampling frequency (80 Hz) and *N* the number of data points; the output is centrosymmetric around 0 up to the Nyquist Frequency (*f/2*). Data has been truncated to 1 Hz, note log scaling of the abscissa. The low frequency/DC portions of the spectrum are due to the peaks while the higher frequency low intensity portions are mostly due to the system noise. The low resolution of the unpadded data particularly at low frequencies (see 0.004-0.02 Hz) may cause improper binning of data that negatively effects the deconvolution noise (*vide infra*). The raw chromatogram had 19201 data points (~$2^{14.23}$). Padding data to a high integer power of 2 is computationally more efficient to perform FFT. Unless stated otherwise, $2^{20}$ data points were used.

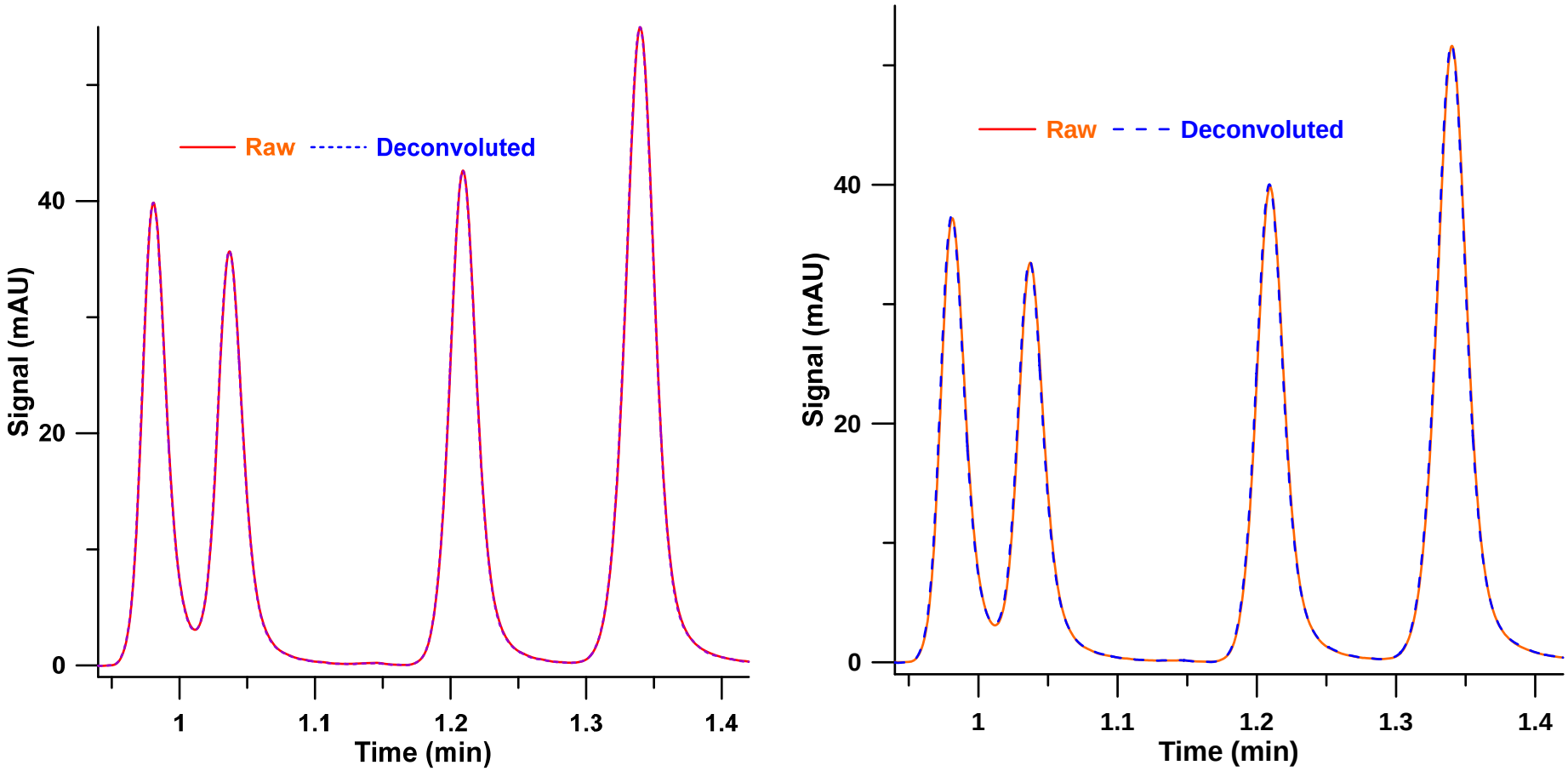



**Figure S12.** Two other case comparisons for the same case as Figure 7. In all three cases the divisor array was the average of the three raw 1-mm path chromatograms, these are shown as individual traces in Figure S14. Three individual 1-mm path chromatograms were generated from three individual 60-mm path chromatograms. In both of the above cases and in Figure 7, the deconvoluted data is shown with the best matching raw data, The match in the left panel is as good as that in Figure 7 and both of these are slightly better than that in the right panel. if they are randomly chosen as in Figure S13, the match is much less perfect. However, on the average, the difference in no greater than the variability of the replicate raw chromatograms in Figure S14.

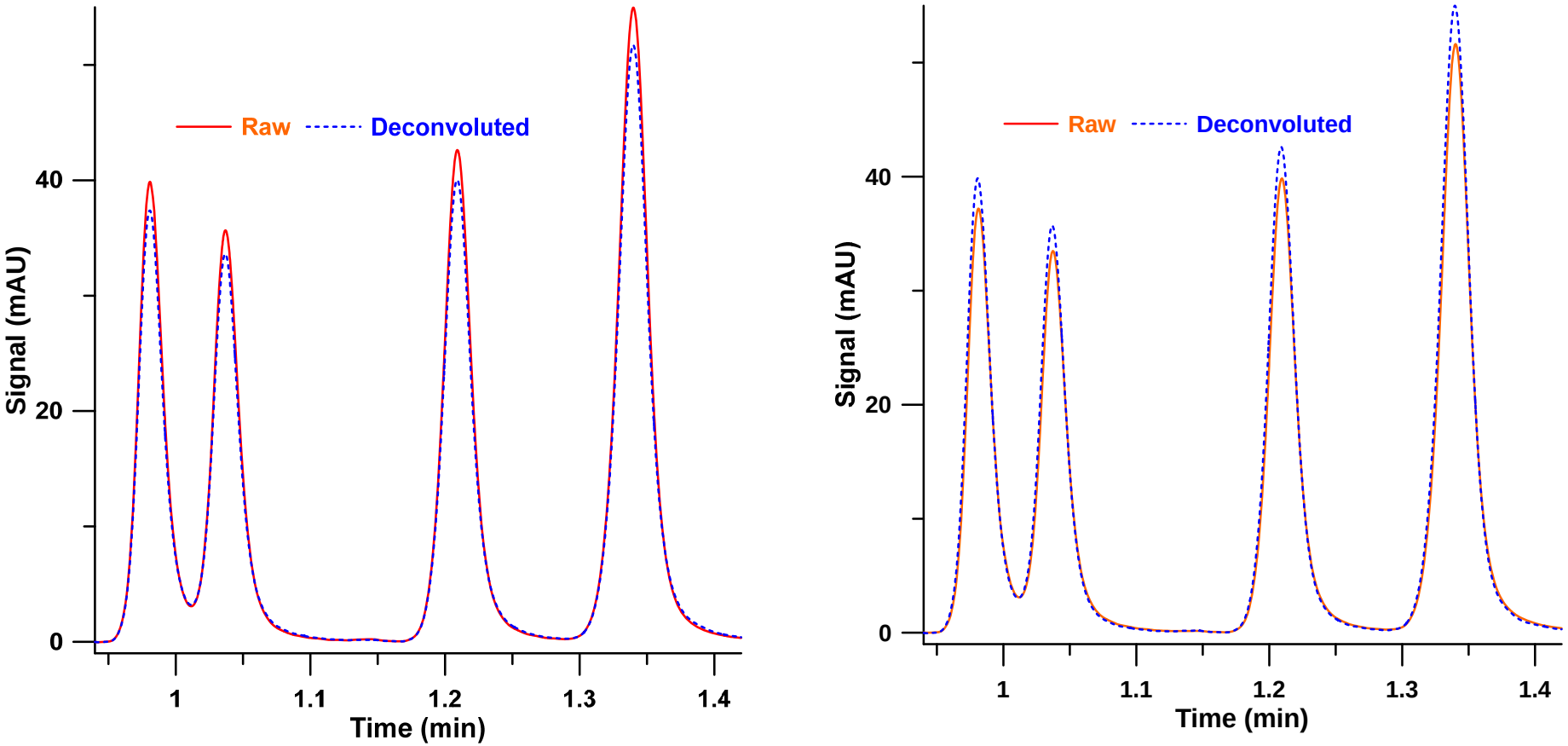



**Figure S13.** Two other case comparisons for the same case as Figure 7 and Figure S12. Here the raw and deconvoluted data pairs were randomly chosen. While the match is less perfect than in Figure 7 or Figure S12, the variability of the raw chromatograms in Figure S14 is comparable.

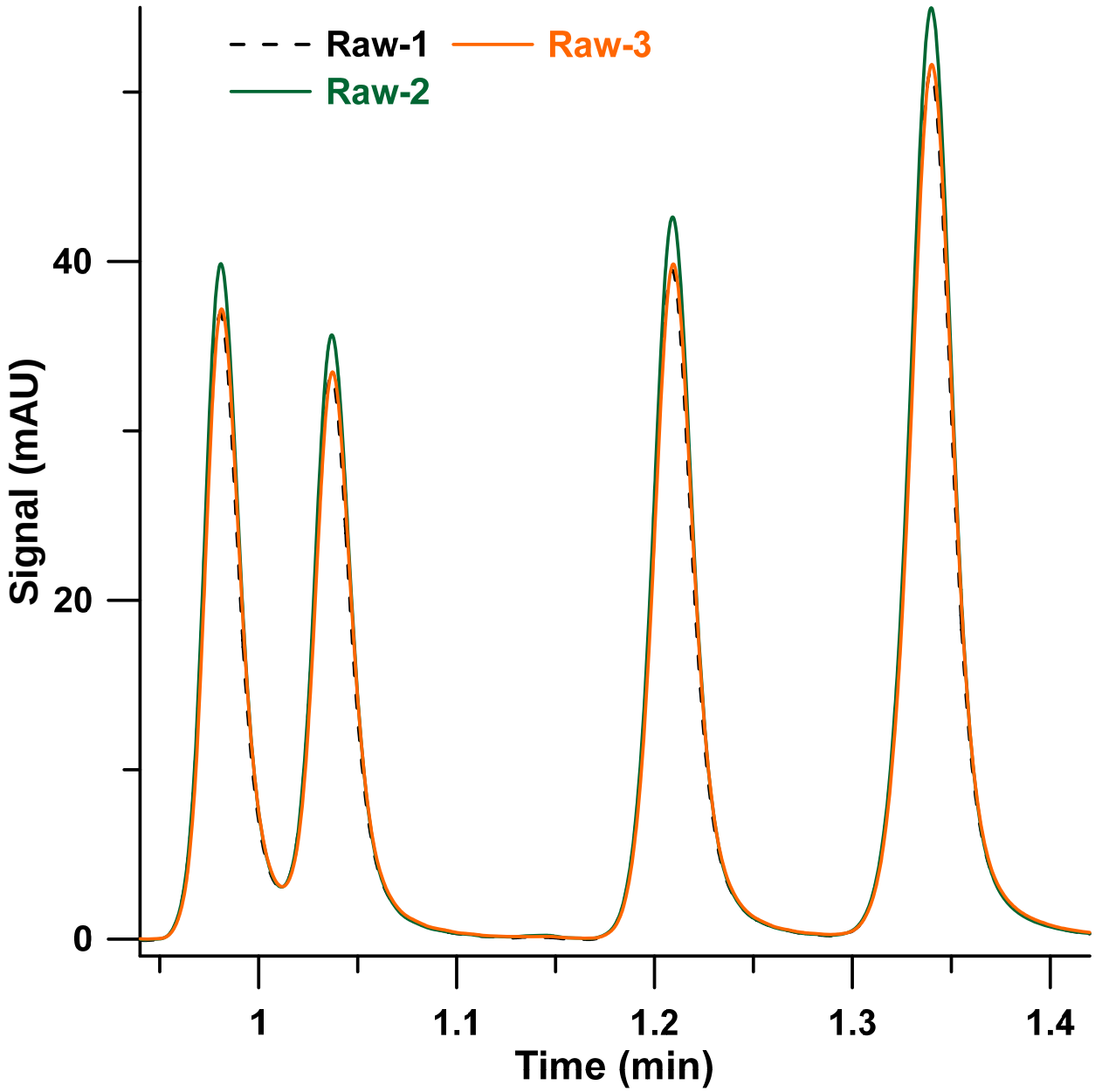



**Figure S14.** Triplicate runs of 1 -mm cell path chromatograms whose average was used in the deconvolution divisor. Figure 7 compares “Raw-1” data to the best matching set of data derived from a 60-mm cell path. Figure S12 in contrast compares “Raw-2” and “Raw-3” with the best matching transforms.

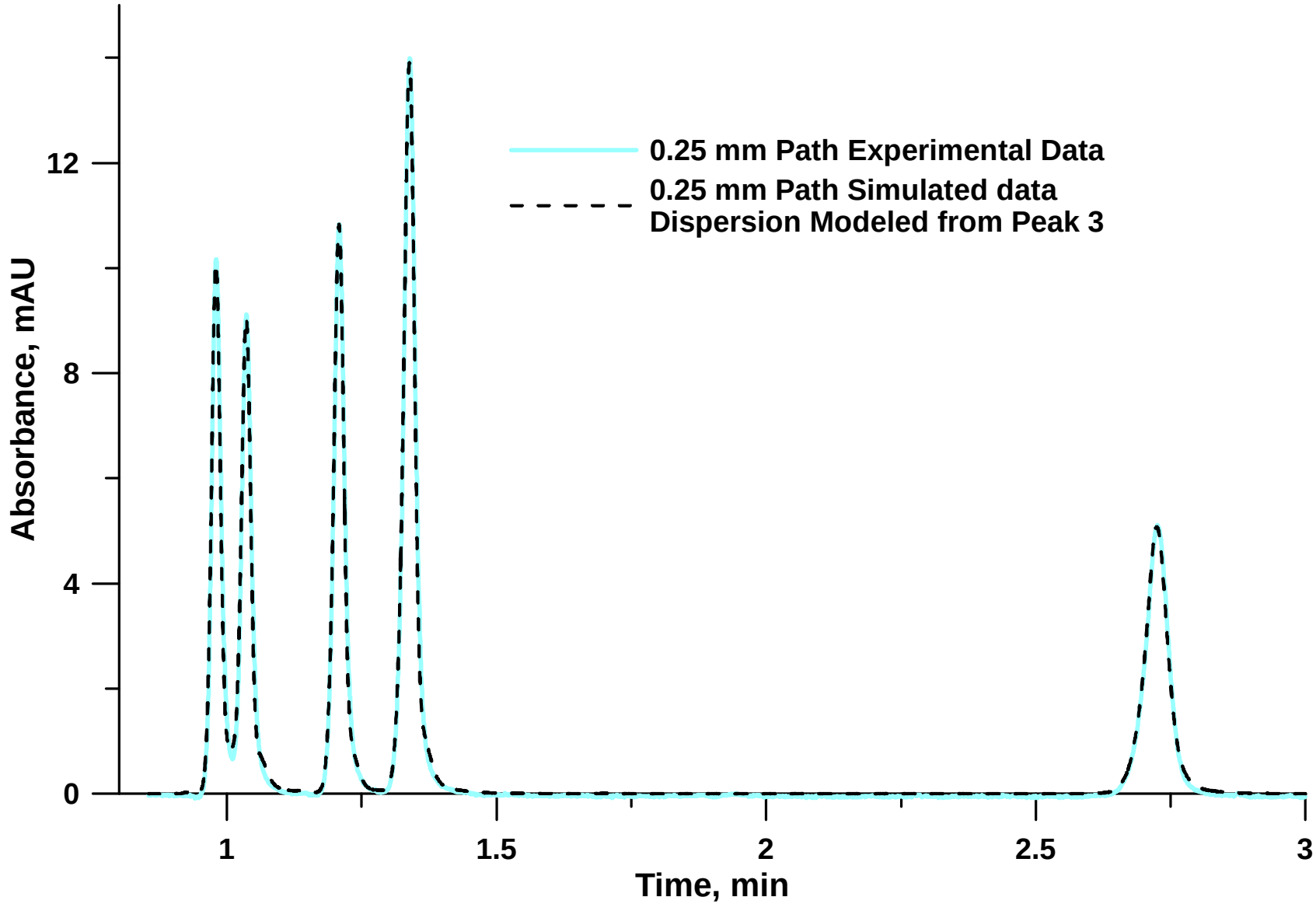



**Figure S15.** Similar to Figure 5 except entire chromatogram for a 0.25 mm path is regenerated from the data from a 30 mm path cell using only the region around peak 3 for modeling dispersion.

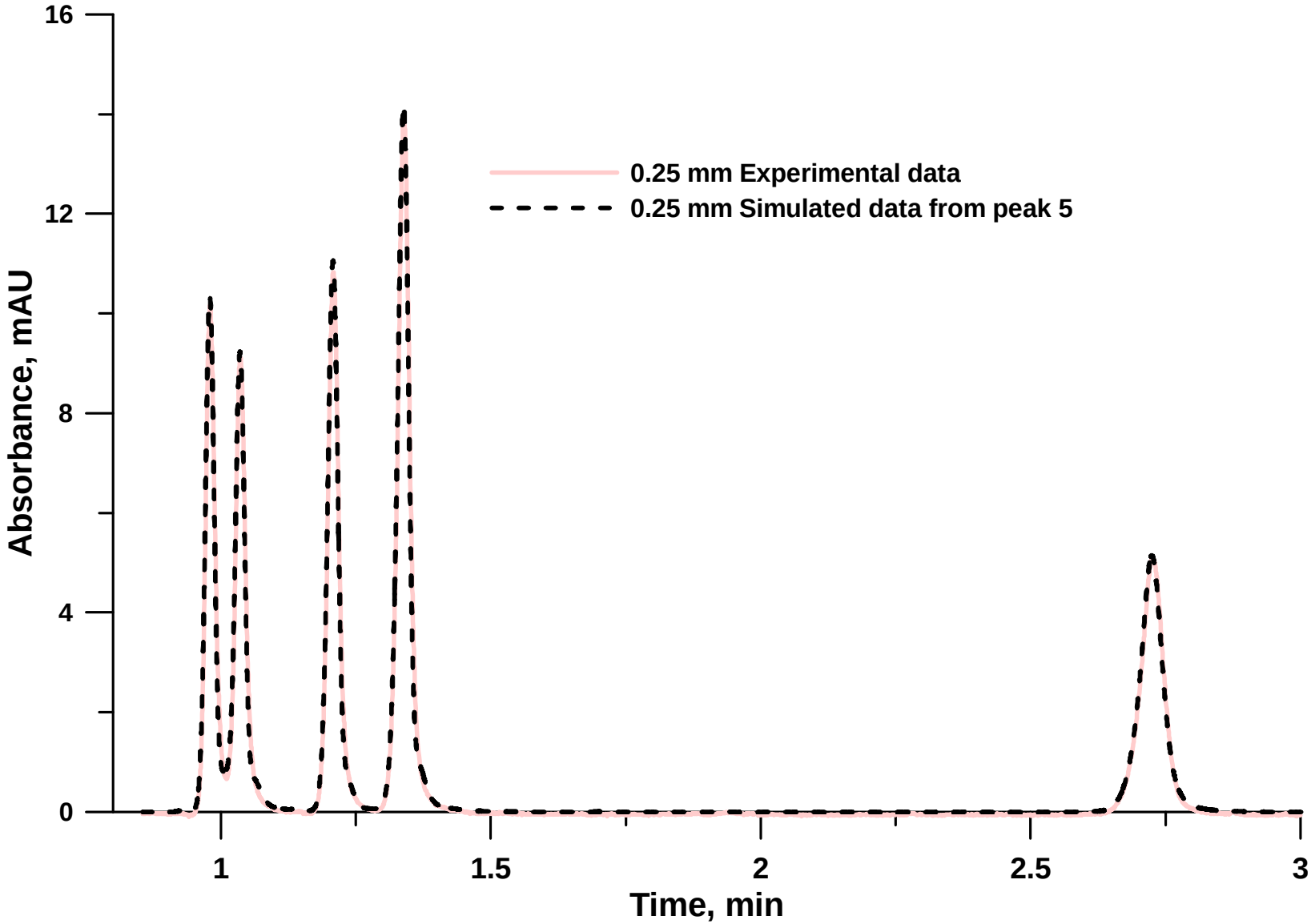



**Figure S16.** Similar to Figure S15 except entire chromatogram for a 0.25 mm path is regenerated from the data from a 30 mm path cell using only the region around peak 5 for modeling dispersion.

## Simplified Dispersion Reversal Calculations for Figures S15 and S16

For the simulation of the 30 mm to 0.25 mm cell data, photometric correction is not necessary as the absorbance values are within the linear range. Additionally, no satellite peak was discernible without the step used to eliminate satellite peaks and no separate steps for satellite removal were therefore taken. The following two steps were taken. First, the peak area ratio of the 30-mm path data for to the 0.25-mm path was computed and all the ordinate values for the 30-mm data are divided by this value to make this new peak area is equal to that of the experimental 0.25 mm data. The reverse of the exponential decay (Equation 17) is now applied, the nominal alpha value is adjusted for the best fit.

### Table S3. Peak 4 S/N characteristics upon 30 mm to 0.25 mm reversal

| Peak 4 Characteristics | | S/N | Half-width, s |
|---|---|---|---|
| Experimental data | 30 mm | 9830 | 1.806 |
| | 0.25 mm | 170 | 1.437 |
| Experimental 0.25 mm data | No filter | 170 | 1.437 |
| | SMA (30 pts, 375 ms) | 242 | 1.471 |
| | GKMA (stdev 62.5 ms) | 205 | 1.448 |
| Simulated 0.25 mm data | No filter | 1200 | 1.395 |
| | SMA (30 pts, 375 ms) | 8300 | 1.412 |
| | GKMA (stdev 62.5 ms) | 4140 | 1.400 |

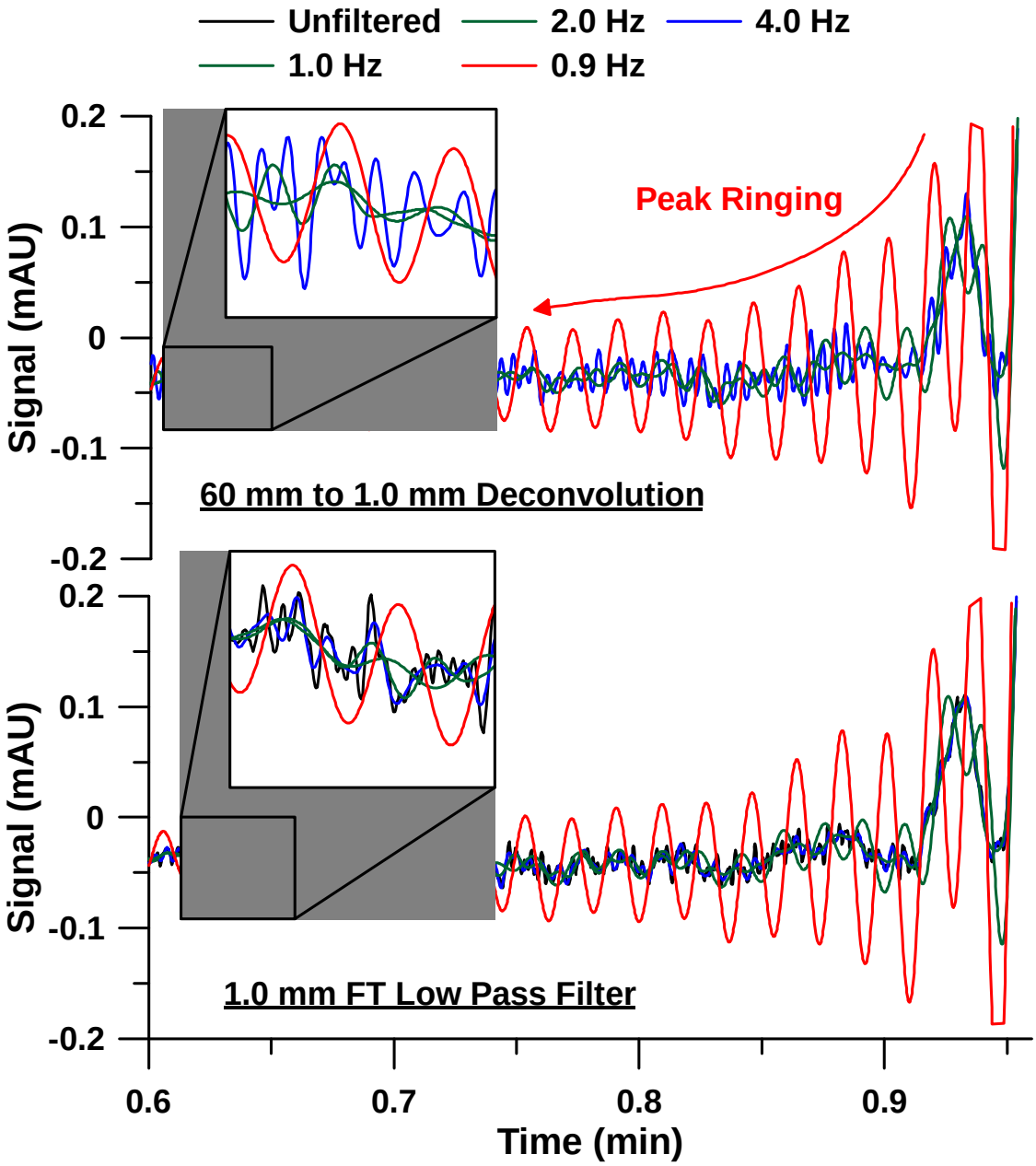



**Figure S17**. Magnified view of baseline immediately before Peak 1 begins. Effect of Cutoff filter upon baseline noise. As the cutoff frequency is lowered below 1 Hz, peak ringing appears. Clear similarities are seen with the temporal noise profile in the regards to the filtered 1 mm chromatogram and the deconvoluted 1 mm chromatogram indicating the source is the shorter pathlength data.

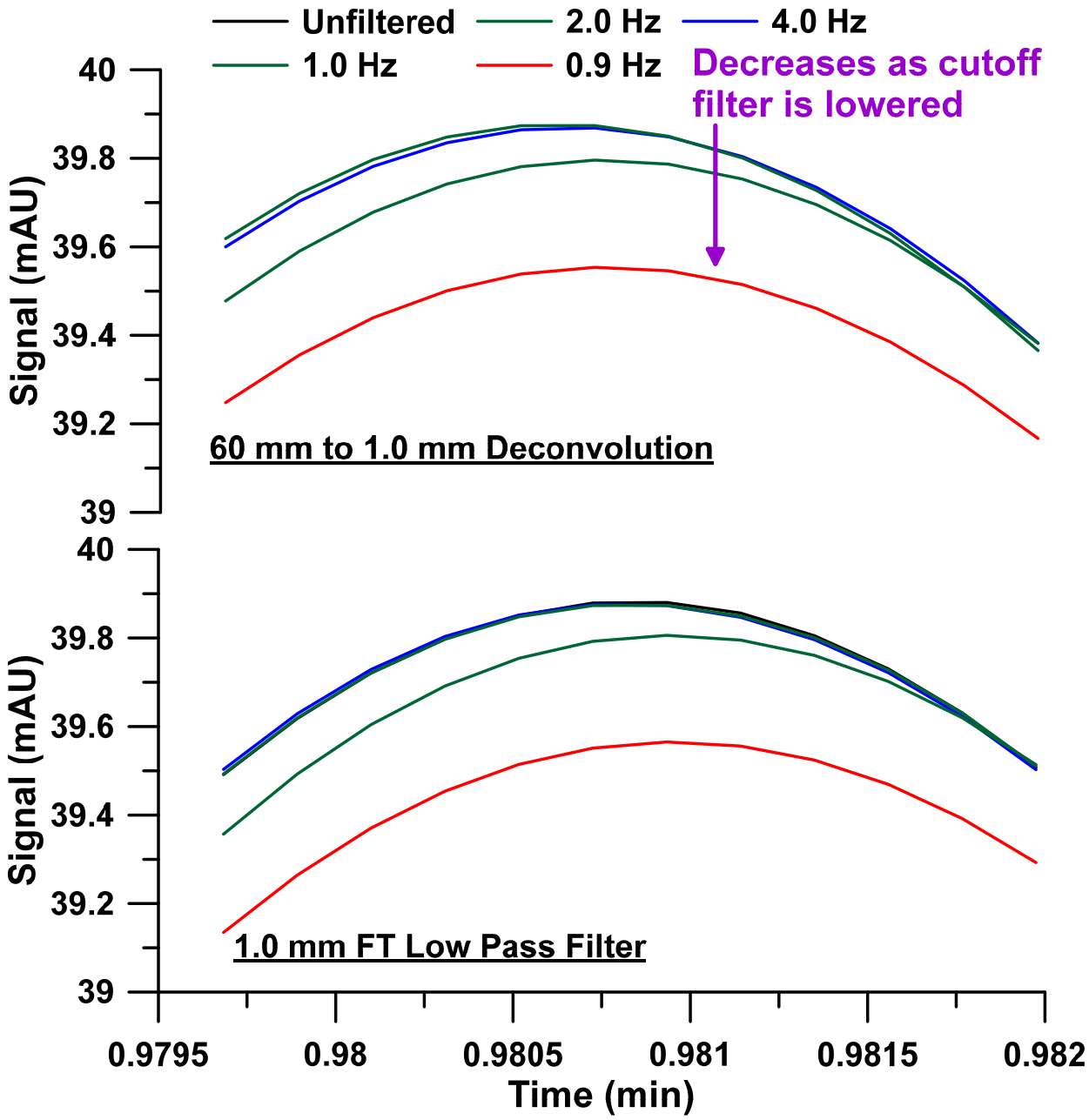



**Figure S18**. Magnified view of apex of peak 1. Choice of LPF cutoff frequency on the peak height. Peak height is affected below an LPF cutoff frequency of 2 Hz.

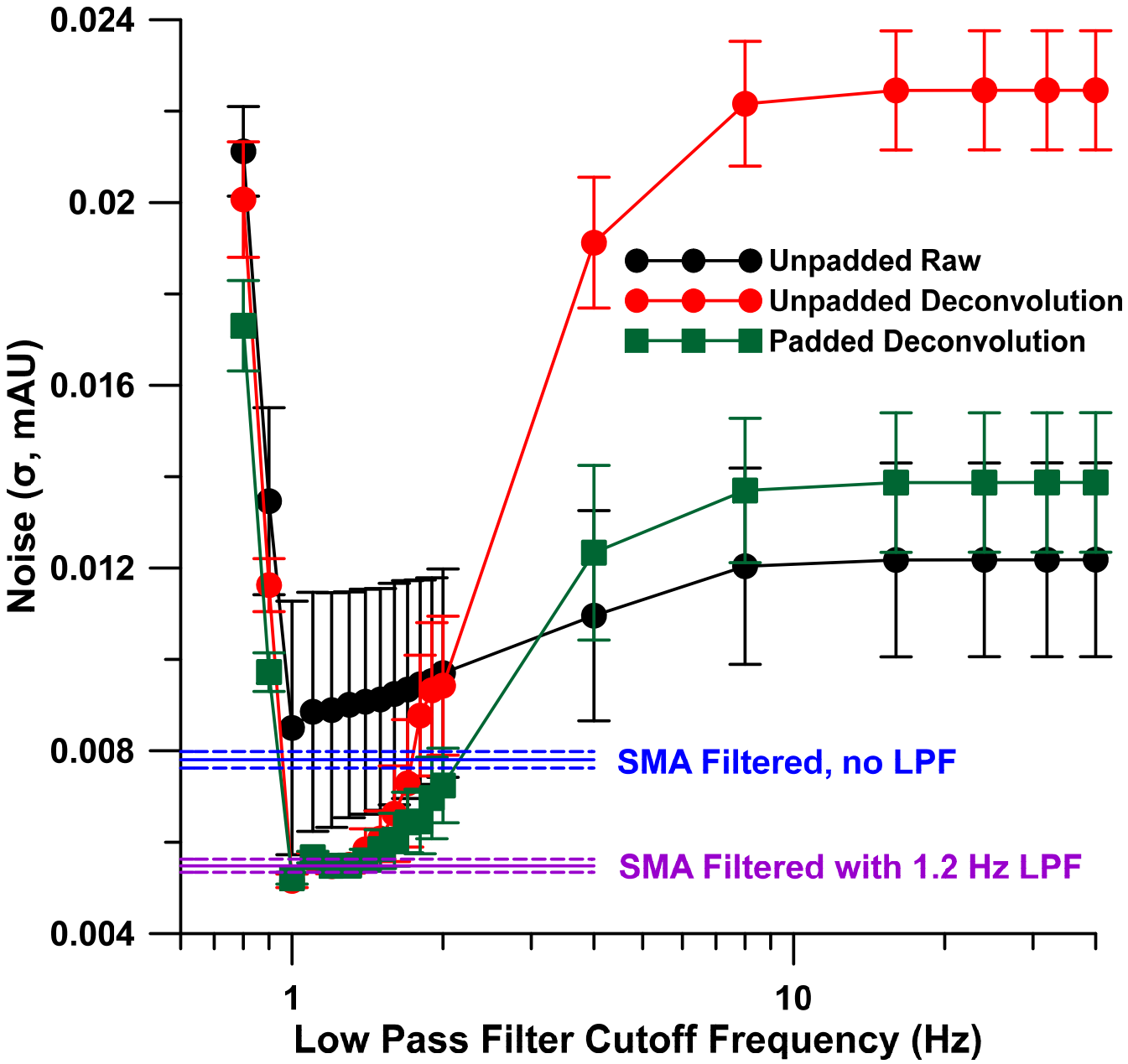



**Figure S19**. Effect of the choice of cutoff frequency of the low pass filter on noise in FT deconvolution. The error bars indicate ± 1 standard deviation. The base case (black trace) is the raw 1-mm cell data, subjected to FFT filtered with an LPF at the indicated cut off frequencies. When the cutoff frequency is >10 Hz, the LPF has no effect. Upon deconvolution of the unpadded data (red trace), the noise increases dramatically with LPF frequency cutoffs ≥ 2 Hz. This is likely due to the poor data density and improper binning. The deconvoluted padded data ($2^{20}$ points, green trace) has significantly lower noise than the deconvoluted unpadded data (red trace) at any LPF cutoff frequencies ≤ 1.5 Hz. At higher LPF cutoff frequencies ($f \geq 4$ Hz), within experimental error, noise is essentially the same for the padded deconvoluted data (green) as the original 1-mm cell (black). If the LPF cutoff frequency is in the optimal window of 1-2 Hz, low-pass filtration provides ~60% less noise in the padded deconvoluted data (green vs. black). Compared to the base case with no LPF at all (e.g., black trace at $f > 10$ Hz), this is a 2.2x improvement. The blue line depicts the noise level, with ± 1 σ error bars if a 60-point SMA filter is applied to the deconvoluted data to which an LPF, with a cutoff at 1.2 Hz has been applied. In this case, no benefit is seen, there is no benefit to applying both.

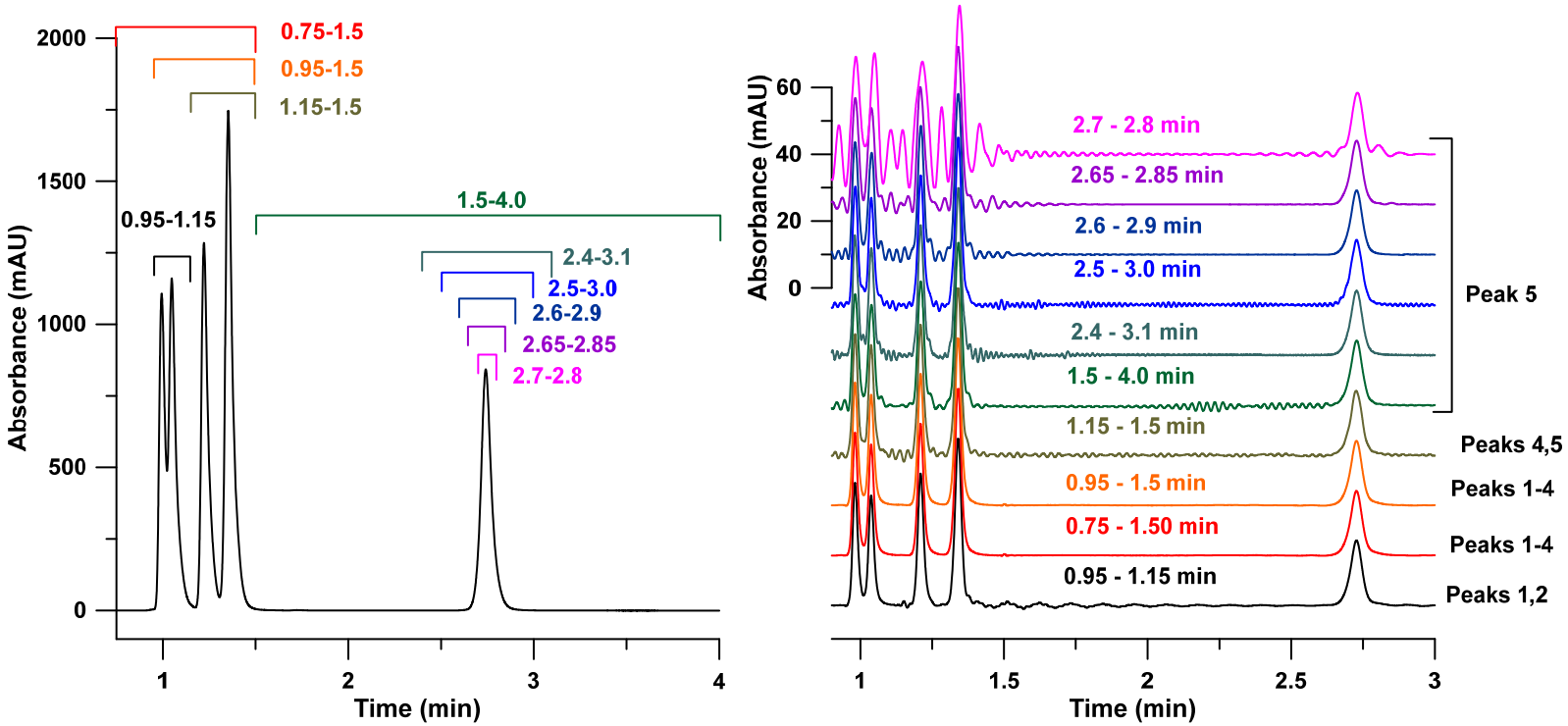



**Figure S20.** Left Panel: Region selected as reference for Fourier Transform deconvolution. Right Panel: Deconvoluted chromatograms. Text at right indicates the peaks included in the deconvolution. Expanding the windows on the 5$^{th}$ peak reduces the noise/ringing but inclusion of more peaks provides the best deconvolution. Even using 2 peaks may be "adequate", depending on how much ringing/noise is considered acceptable. Figure S19 shows the traces in more detail when differing spans around peak 5 are used as the reference.

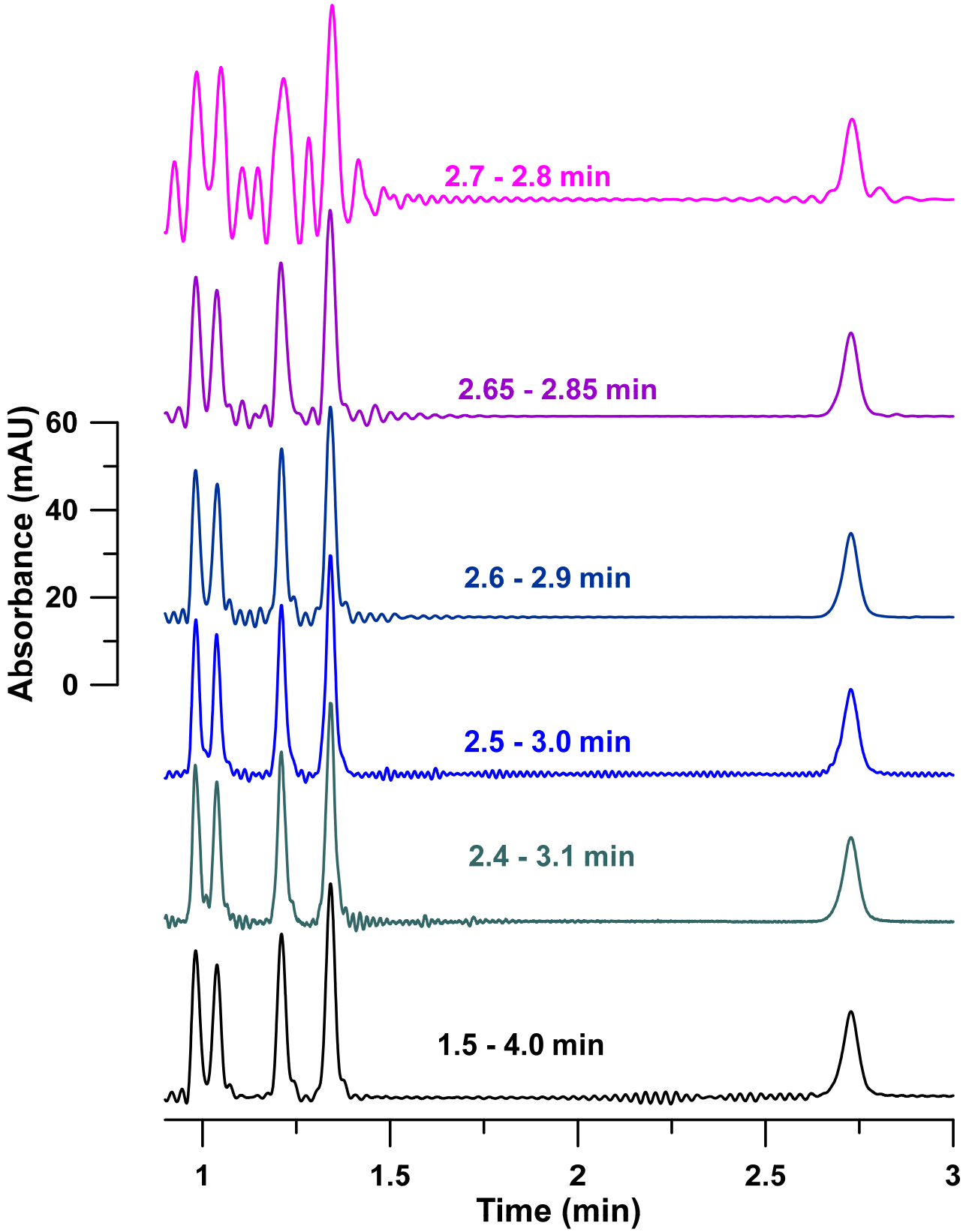



**Figure S21.** Different spans around peak 5 but excluding peaks 1-4 are used for FT deconvolution. While extending the span may be helpful, some degree of oscillation/noise was impossible to eliminate.

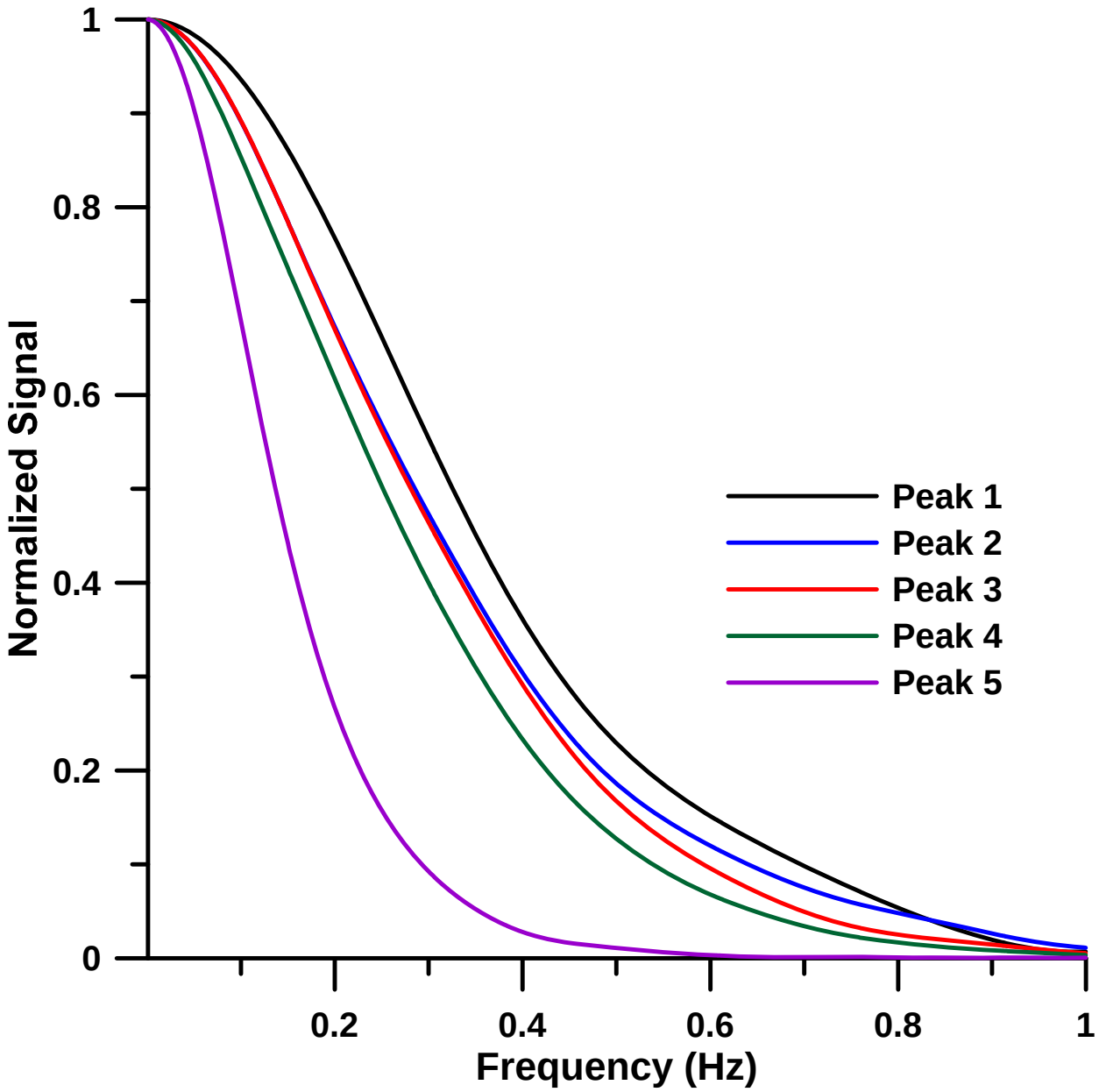



**Figure S22.** The frequency domain spectra for each peak (peaks 1 and 2 cannot be completely separated, the demarcation point is taken to be the lowest point in the valley between them) in the 1 mm cell i.e., the *S(ω)* spectra.

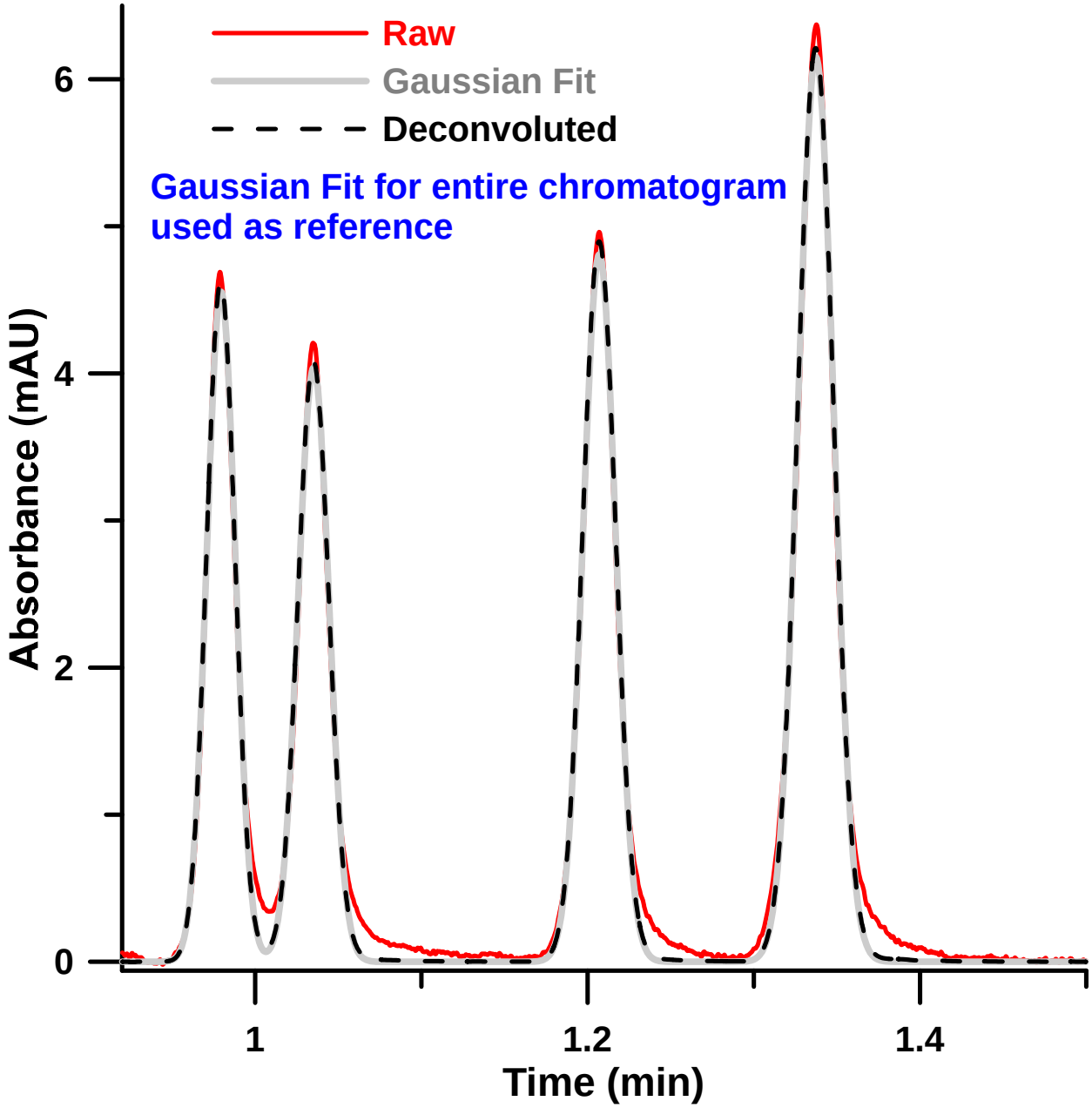



**Figure S23.** A model Gaussian fit (gray) for the entire chromatogram is used as the reference. The deconvoluted chromatogram (dashed black lines) closely follow the provided model and generally ignore the imperfections (tailing, etc.) of the in the actual chromatogram. The fifth peak wasn't included in this particular analysis but also fits the provided model very well if it is included.

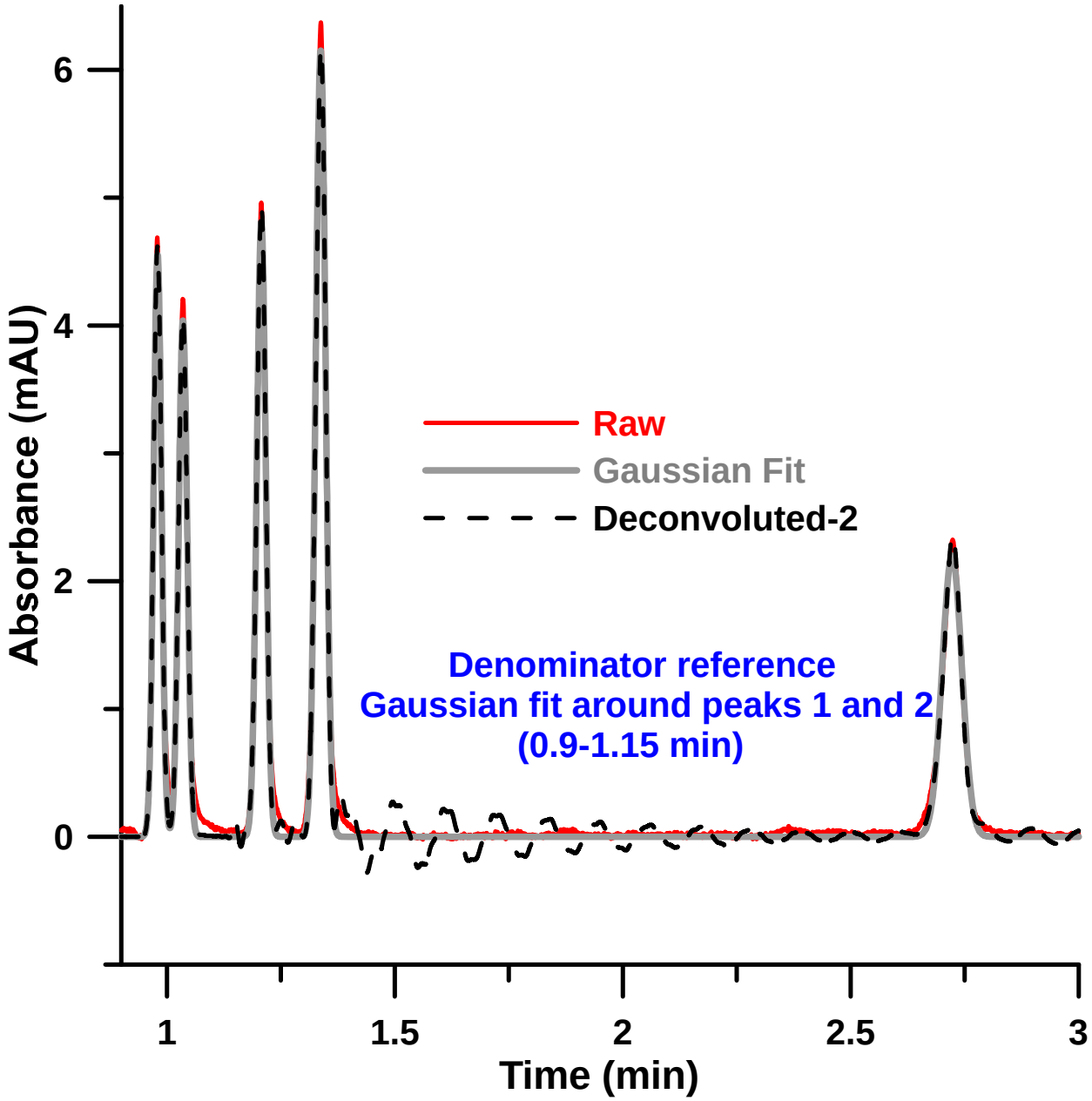



**Figure S24.** Only the Gaussian fit for the region around the first two peaks is used as the reference. The deconvoluted chromatogram (dashed black lines) closely follow the provided model and generally ignore the imperfections (tailing, etc.) but a lot of oscillations are observed between the peaks.

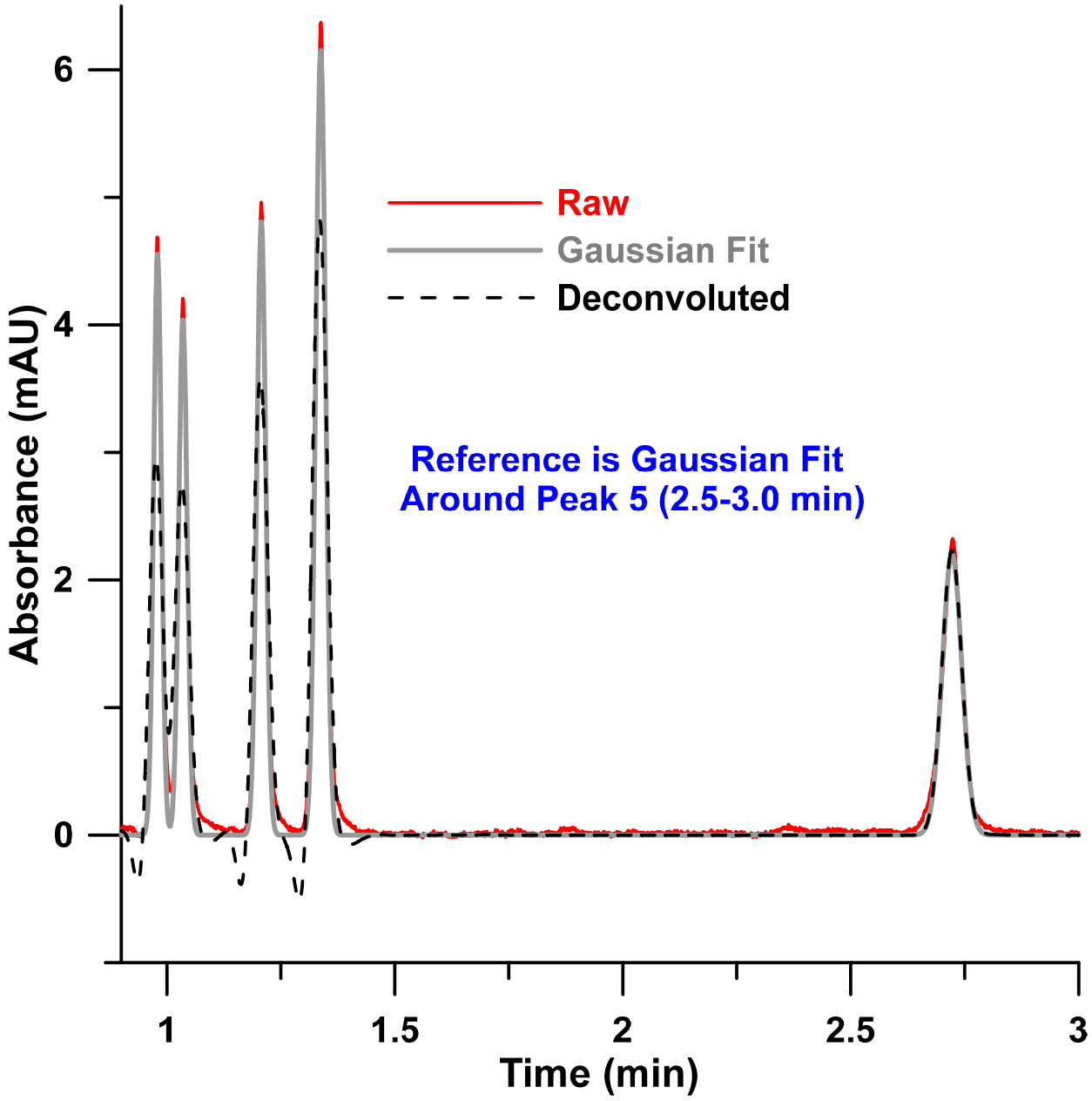



**Figure S25.** Only the Gaussian fit for the region around peak 5 was used as the reference. The deconvoluted chromatogram (dashed black lines) closely follow the provided model and generally ignore the imperfections (tailing, etc.) but severe dips precede most peaks other than 5. The model invokes a better resolution between peaks 1 and 2 than the real situation while it is considerably worse than the real situation in the model. First four peak heights decrease substantially.

**Computing Time Cost.** Although our aforedescribed dispersion reversal work was initially carried out in Microsoft Excel,™ In the interest of fair comparison, the same algorithm was written in LabView and thus compared to FT deconvolution on the same platform. However, FFT and IFFT routines have been optimized in LabView, this was not the case for our algorithm. Using a 2.8 GHz quad-core desktop CPU with a standard amount of RAM, analysis time was negligible for either method compared to the time to acquire chromatographic data. Our approach and the FT deconvolution approach (no zero padding) respectively required 6.2 ± 0.7 and 1.51 ± 0.09 ms (in 1000 trials, $\alpha$ and $D(\omega)$ were considered already determined and available). To zero-pad to $2^{20}$ points and then compute increased the FT approach computing time to ~100 ms but even this is trivial to the time needed for chromatography.