Journal of Chromatography B 1264 (2025) 124721

Contents lists available at ScienceDirect

Journal of Chromatography B

journal homepage: www.elsevier.com/locate/jchromb

PeakClimber: A software tool for the accurate quantification of complex HPLC chromatograms

Joshua T. Derrick [a,b], Pragney Deme [c], Norman J. Haughey [c], Steven A. Farber [a,b,*], William B. Ludington [a,b,*]

[a] Department of Biology, Johns Hopkins University, Baltimore, MD, United States
[b] Department of Embryology, Carnegie Institute for Science, Baltimore, MD, United States
[c] Department of Neurology, JHMI, Baltimore, MD, United States

ARTICLE INFO

Keywords:
HPLC
Lipids
Bidirectional
Exponentially modified Gaussian
Fatty acids
drosophila
Algorithm
Software

ABSTRACT

High-performance liquid chromatography (HPLC) is a common medium-throughput technique to quantify the components of complex mixtures like those typically obtained from biological tissue extracts. However, analysis of HPLC data from multianalyte samples is hampered by a lack of tools to accurately determine the precise analyte quantities on a level of precision equivalent to mass spectrometry approaches. To address this problem, we developed a tool we call PeakClimber that uses a sum of bidirectional exponentially modified Gaussian (BEMG) functions to accurately deconvolve overlapping, multianalyte peaks in HPLC traces. Here we show that HPLC peaks are well-fit by a BEMG function, that PeakClimber more accurately quantifies known peak areas than standard industry software and other open-source software packages for HPLC, and that PeakClimber accurately quantifies differences in triglyceride abundances between colonized and germ-free fruit flies.

1. Introduction

Liquid chromatography is a series of techniques to separate individual analytes from a mixture of chemicals using a liquid-phase solvent [1]. Originally, liquid chromatography techniques relied on gravity for solvent flux, which meant that individual chromatographs took hours or days to run. In the 1960s, high-pressure (or high-performance) liquid chromatography (HPLC) was introduced, speeding up the flow rate by forcing the solvent through an extremely narrow column at high-pressure [2]. Despite improvements in column performance, trade-offs between mass transfer resistance and diffusive behaviors fundamentally limit peak resolution [3]. For many HPLC applications, peak integration is sufficient because these analyses principally are concerned with presence/absence of specific peaks or with quantification of relatively pure analytes with little peak overlap. For the quantification of more complicated chemical and biological samples with overlapping peaks, however, integration alone is inaccurate. Historically, this meant that the operator spent considerable efforts to develop protocols to fully separate analyte peaks of interest, something that is not always possible.

Various solutions have been proposed to this problem. Common industry software, such as ThermoFisher's Chromeleon and Waters' MassLynx utilize a method known as valley-to-valley [4], where a line is dropped from the lowest point between two peaks to the chromatograph's baseline, which is determined by the rolling-ball method [5]. The two peaks are then integrated on either side of the line. This method has the advantages of being neutral to the underlying peak shape, independent of the surrounding peaks, and having a fast runtime. However, most peaks map to some variation of the Gaussian distribution [3,6–11] and are not independent of neighboring peaks with which they overlap. Two more recent open-source software packages, HappyTools [12] and hplc.io [13], improve on the valley-to-valley method by fitting each chromatograph to a sum of Gaussian or skewed Gaussian curves, respectively. However, these theoretical peak shapes are not necessarily suited to the underlying data, and the shape of a single peak is not universally agreed upon. Early quantitative models of liquid chromatography showed that analytes unbind the column with an exponential decay that is convolved by Gaussian noise based on their distribution along the length of the column and their diffusion in the liquid phase before reaching the detector [6,7,14–16]. While the shape of a peak depends on the amount of sample loaded on the column, Langmuir surface binding kinetics usually lead to a Gaussian shaped peak with tailing [16,17].

* Corresponding authors at: Department of Biology, Johns Hopkins University, Baltimore, MD, United States.
E-mail addresses: sfarber3@jh.edu (S.A. Farber), will.ludington@gmail.com (W.B. Ludington).

https://doi.org/10.1016/j.jchromb.2025.124721
Received 24 April 2025; Received in revised form 27 June 2025; Accepted 30 June 2025
Available online 5 July 2025

In this manuscript, we show that HPLC analyte peaks are well-fit with a bidirectional exponentially modified Gaussian (BEMG) function. Our tool, PeakClimber, fits chromatographs to a sum of BEMG curves. We show that these curves are mathematically and empirically good fits for single analyte peaks, and are consistent with extensive literature suggesting that this approach empirically aligns with chromatography data [6,7,18–20]. PeakClimber also makes iterative improvements in denoising data, detrending data, and reducing the runtime of the analysis, as compared to other open-source software tools. To highlight the utility of PeakClimber, we compare its performance to other algorithms by analyzing co-injections of three fatty-acids. Finally, we use PeakClimber to quantify the differences in lipid composition between *Drosophila melanogaster* that were reared with and without bacteria.

2. Results

2.1. Theoretical results

2.1.1. Traditional chromatography analysis methods fail to accurately quantify complex peaks

Valley-to-valley integration methods produce a mismatch between the calculated and true peak shape (Fig. 1A). To quantify the error of this approach, we conducted a simulation with three-synthetic BEMG peaks with randomized parameters that overlapped significantly. Our simulations showed that the valley-to-valley method has significant error between the true peak shape and the valley-to-valley integration regions, but this error is especially marked for the first peak in the trace (Fig. 1B). This is likely due to the undercounting of the exponential tail region of the first peak by valley-to-valley analysis.

2.1.2. Analytical and computational evidence demonstrates that an exponentially modified Gaussian distribution fits single analyte HPLC peaks

We investigated which shape best fits individual peaks. There is extensive discussion of this question in the literature [1,3,4,7–11,14–18,21], but there is broad agreement as to a generally Gaussian peak shape with some amount of tailing. To this end, we developed analytical, computational, and empirical models, which supported the BEMG as the true shape of a chromatographic peak.

Consider a column of finite length, initially containing no solute. An injectant containing solute S is added to the column, and S is completely adsorbed by the column at a single location. Solvent U is then run over the column. Solute S has affinity k_1 for solvent U. We assume that only unbinding from the column occurs, and not rebinding, as unbound S flows away in the solvent much faster than it can rebind due to its high affinity to the solvent U. This behavior can be described by the differential equation:

$$\frac{dS}{dt} = -k_1 S \tag{1}$$

which we can solve analytically:

$$S(t) = M * k_1 e^{-k_1 t} \tag{2}$$

producing an exponential function, where M is the total number of molecules of S. Using an agent-based Monte-Carlo simulation with parameters for S (amount of analyte), k_1 (affinity for solvent U), column length, and flow rate that are relevant to common HPLC columns, we recapitulated the analytical solution almost exactly (Fig. 2A, red line on blue histogram).

However, this initial model contains several incorrect assumptions, most notably that column binding and unbinding is a single event. In reality there are many binding and unbinding steps [9,16,22], especially when the affinity of S for the solvent is low relative to its affinity to the column, which occurs during column loading. Thus, the distribution of analyte S will not be at a single site, but rather spread out across the column after many unbinding and binding events. We thus represent the probability of a single particle binding to location x on the column with the exponential probability distribution, with λ being the average

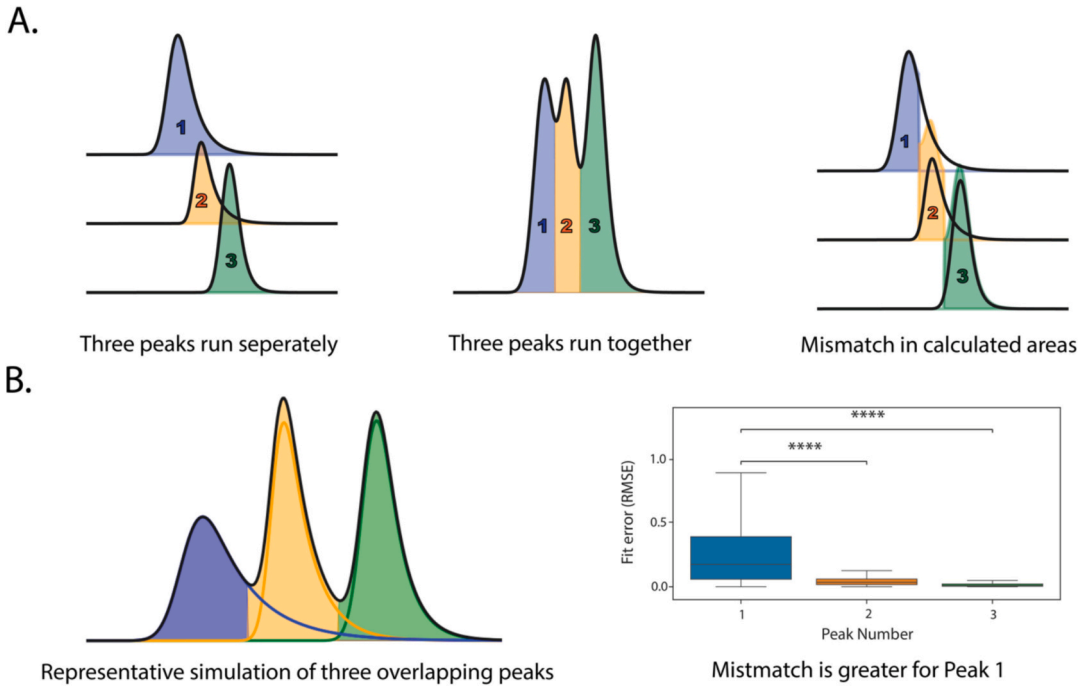


Fig. 1. The problem of peak quantification. (A) A cartoon depiction of a common inaccuracy in peak quantification. When three analytes are well-separated, their area is accurately calculated by peak integration. When three analytes have overlap in the trace, the valley-to-valley area calculation algorithm will not accurately determine peak areas due to overlap of the tails. (B) A simulation of three overlapping peaks. The shaded region represents the integration regions identified by the valley-to-valley algorithm; the solid lines represent the true peaks. The difference between the two is quantified by root mean square error (RMSE) ($n = 1000$, Mann-Whitney followed by Wilcoxon Ranked Test, $^{****}p < 1e\text{-}04$).

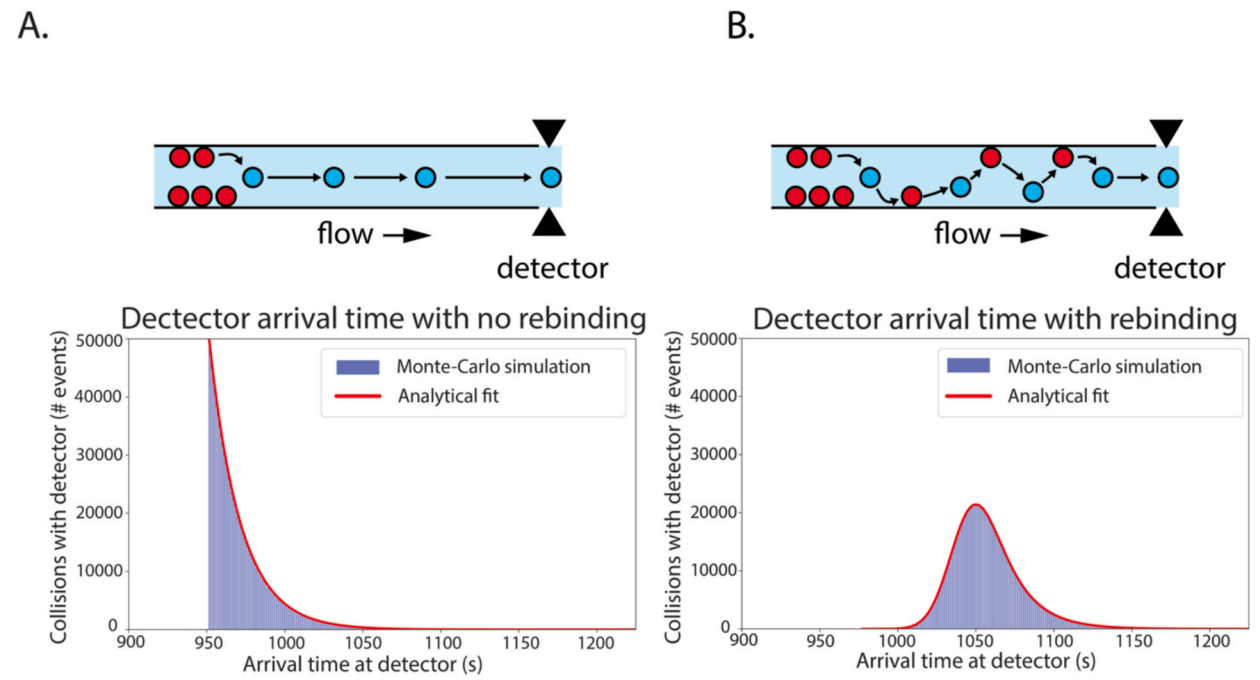


Fig. 2. Simulation of column dynamics fits an exponential Gaussian function. Monte Carlo simulation (blue histogram) of solute arrival time at the detector without (A) or with (B) rebinding, fit by the analytical functions (red histogram) for an exponential distribution (A) or bidirectional exponentially modified Gaussian distribution (B) (n = 10,000 simulations consisting of 100 analyte molecules each). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

distance a particle travels in solution before being adsorbed onto the column wall.

$$c(x) = \lambda e^{-\lambda x} \tag{3}$$

λ is directly dependent on the speed of the mobile phase (μ) and inversely proportional to the diffusion coefficient (D) and relative affinity for the column over the solute.

For a single particle, this event does not happen a single time, but many times over the course of column loading. To represent this for n binding/unbinding events, we can sum n exponential functions together, generating an Erlang distribution [23].

$$C(x) = \frac{\lambda^n x^{n-1}}{(n-1)!} e^{-\lambda x} \tag{4}$$

At large n, the Erlang distribution will converge to a normal distribution [23] with mean $n\lambda$ and variance $n\lambda^2$.

$$C(x) = \frac{1}{n\lambda^2\sqrt{2\pi}} e^{-\frac{(x-n\lambda)^2}{2(n\lambda^2)^2}} \tag{5}$$

This is the probability distribution for the location of a single particle along the column.

To convert this distribution to the arrival time domain, we divide the distance x from the column by the flow rate μ.

$$C(t) = \frac{1}{n\lambda^2\sqrt{2\pi}} e^{-\frac{(t\mu-n\lambda)^2}{2(n\lambda^2)^2}} \tag{6}$$

$$C(t) = \frac{1}{n\lambda^2\sqrt{2\pi}} e^{-\frac{(t-n\lambda/\mu)^2}{2(n\lambda^2)^2/\mu^2}} \tag{7}$$

To simplify the expression, we define two new variables $b = n\lambda^2/\mu$ and $c = n\lambda/\mu$. These variables are the spatial mean and variance from eq. 5 converted to the arrival time domain by dividing by the flow rate μ. This transformation yields the following equation:

$$C(t) = \frac{1/\mu}{b\sqrt{2\pi}} e^{-\frac{(t-c)^2}{2b^2}} \tag{8}$$

This is a Gaussian distribution, which is supported in the chromatography literature as the canonical distribution for peaks in isotonic elution conditions [7,8]. However, when performing elution over a gradient of solvents, the relative affinity of the analyte for the column and mobile phases shifts, encouraging single-step Langmuir kinetics at a critical point on the gradient near the retention time, which results in the exponential decay behavior with no rebinding at a discrete time point, which is observed in eq. 2. There are other additional sources [24] of this exponential "tailing function", such as the structure of the column bed [16], differences in velocity across the column diameter [25,26], and extra-column effects such as dead volume between the column and the detector. To simplify, we have lumped all these behaviors into a single exponential decay function. To combine this function with the previously derived Gaussian, we convolve the two functions.

$$Z = C^*S \tag{9}$$

$$Z(t) = \int_0^\infty C(t-\tau)S(\tau)d\tau \tag{10}$$

$$Z(t) = \frac{M * k_1/\mu}{b\sqrt{2\pi}} \int_0^\infty e^{-\frac{(t-\tau-c)^2}{b^2}} e^{-k_1{}^*\tau} d\tau \tag{11}$$

$$Z(t) = \frac{M * k_1/\mu}{2} e^{\frac{k_1}{2}(2c-2t+k_1b)} erfc\left(\frac{c+k_1b^2-t}{b\sqrt{2}}\right) \tag{12}$$

This matches the standard form of the exponentially modified Gaussian (EMG) function, with *erfc* representing the inverse error function $1 - erf(x)$, with $erf(x) = \frac{2}{\sqrt{\pi}}\int_x^\infty e^{-t^2}dt$. This equation gives us several insights into the factors that influence the shape of the resulting function. The amplitude, or height of the EMG function is directly proportional to the number of analyte molecules, and inversely proportional to the flow rate. The center c of the distribution is dependent on

the average travel distance of the particle during column loading and the number of binding/unbinding events, whose dependency has been previously described. The width or $\sigma\,(b)$ is dependent on the same parameters but can also be affected by other minor parameters such as longitudinal diffusion and column inhomogeneities and, thus, is not directly proportional to the distribution center. Finally, k_1 is the unbinding coefficient of the analyte from the column and represents the gamma variable, γ, of the EMG. This determines how large the tails of the function are. Although our model makes several simplifying assumptions, such as a constant λ during column loading and no longitudinal diffusion, it provides a sound biophysical justification for use of the EMG, which has been utilized in previous chromatography studies [7,19].

One downside of the EMG is the fact that the tail region always lies to the right of the peak center. This is not reflective of all peaks because fronting, in addition to tailing, can also result from column overloading or competitive interactions between Langmuir isotherms. To preserve the properties of the EMG, but also to allow peak fronting to be modeled, we decided to use the BEMG [27], whose closed form function is merely the reflection of the function $Z(t)$ over the peak center. To simplify, we have adopted the form from Misra et al. [28] in eq. Eq 13 below.

$$Z(t) = \frac{M}{2{*}k_1} exp\left(\frac{b^2}{2{k_1}^2} + \frac{c-t}{k_1}\right) * \left[erf\left(\frac{t-c}{\sqrt{2}{*}b} - \frac{b}{\sqrt{2}\,c}\right) + sign(k_1)\right] \qquad (13)$$

To verify this analytical equation, we conducted a Monte-Carlo simulation recapitulating the assumptions of an initial period of unbinding/rebinding to the column followed by a kinetic phase in which the analyte has strong affinity for the solvent. This simulation fit a BEMG equation almost exactly (Fig. 2B, red line on blue histogram), further supporting the BEMG as a good distribution to model HPLC peaks.

2.1.3. *PeakClimber: A software package to rapidly and accurately quantify chromatography peak areas*

We implemented an algorithm in a Python package that identifies and quantifies individual peaks on a chromatographic trace by fitting a sum of BEMG functions to the underlying HPLC trace (Fig. 3). We named this open-source software PeakClimber.

Taking a text file of the trace as input, PeakClimber first denoises and detrends the data. Denoising is accomplished by first filtering the data with a low-pass Fast Fourier Transform (FFT) filter using the Kaiser window function with a low beta value [29–31], as well as a time-averaged convolution. Detrending is accomplished with a high-pass Bohlmann-Whittaker [32,33] baseline subtraction algorithm that was developed for chromatography, called the peaked signal's asymmetric least squares algorithm (psalsa) [34], which estimates a more accurate baseline in the presence of large peaks as compared to the original asymmetric least squares chromatography algorithm [35]. Exact parameters for these detrending algorithms are input by the user. We chose our default values by fitting single peaks of real HPLC data (Fig. 3–1). Peaks are then identified on the denoised data using SciPy's peak finding algorithm, relying on prominence cutoffs to determine if peaks are real or noise [36,37]. The prominence cutoff is also user-defined in PeakClimber. In this paper, we use a value of 0.05, meaning peaks must be 5 % above the contour trough of surrounding peaks to be analyzed. Additional peaks that form shoulders on more prominent peaks can be optionally identified by identifying local minima and maxima in the first derivative of the HPLC trace that are close to 0 (Fig. 3–2).

2.1.4. *Parameter estimation and algorithm function*

For each identified peak, a BEMG function is fit using lmfit [38], a non-linear least squares curve fitting package for Python. Individual peaks are identified using the SciPy algorithm find_peaks. This function returns a list of peak centers and corresponding heights. These correspond to the initial parameters for peak center (c) and amplitude (M). Peak height significantly overestimates amplitude: PeakClimber divides

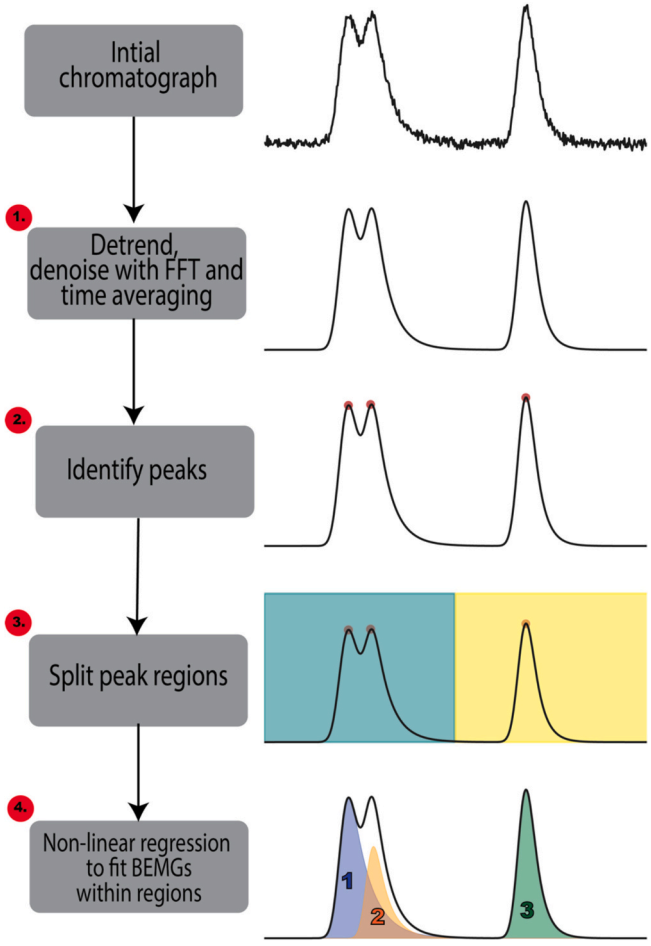


Fig. 3. The PeakClimber workflow. (1) PeakClimber first denoises (FFT) and detrends chromatography data before (2) identifying peaks using prominence cutoffs. (3) To decrease runtime, peaks are split into regions based on intersections of the trace with the x-axis. (4) Within each region, peaks are fit to a bidirectional exponentially modified Gaussian distribution.

by a constant of 2 to correct for this. Each peak is then fed into another SciPy algorithm peak_widths, which calculates the full width at half max (FWHM). This is empirically equivalent to approximately 2.35*b, so $FWHM/2.35$ is used as an initial guess for b. Finally, k_1 is estimated using the first statistical moment of the distribution.

$$Sample\ mean = parameter\ mean + 1/k_1 \qquad (14)$$

The bounds for this distribution are the same used to calculate FWHM. Acceptable ranges for each of the four parameters are hard-coded based on ranges we have seen from peak fitting, but these can be modified by the user from within the program.

Boundaries between discrete peak regions are set where the background-subtracted trace hits zero (Fig. 3–3). The discrete peak regions of the graph are effectively independent of each other, meaning fits can be performed independently on each region without loss of accuracy. Each group of Gaussians is fit to the trace in the appropriate region using a non-linear regression to minimize the least-squared distance between the generated sum of functions and the underlying trace using the initial parameter estimates described in the preceding paragraph (Fig. 3–4). The algorithm recombines the fits for the different windows and returns a summary graph of the resulting trace, overlaid with individual peaks, as well as a table with peak number, runtime, and peak area. Readers can find a user guide for practical implementation of PeakClimber on their own systems in Supplementary Document 1.

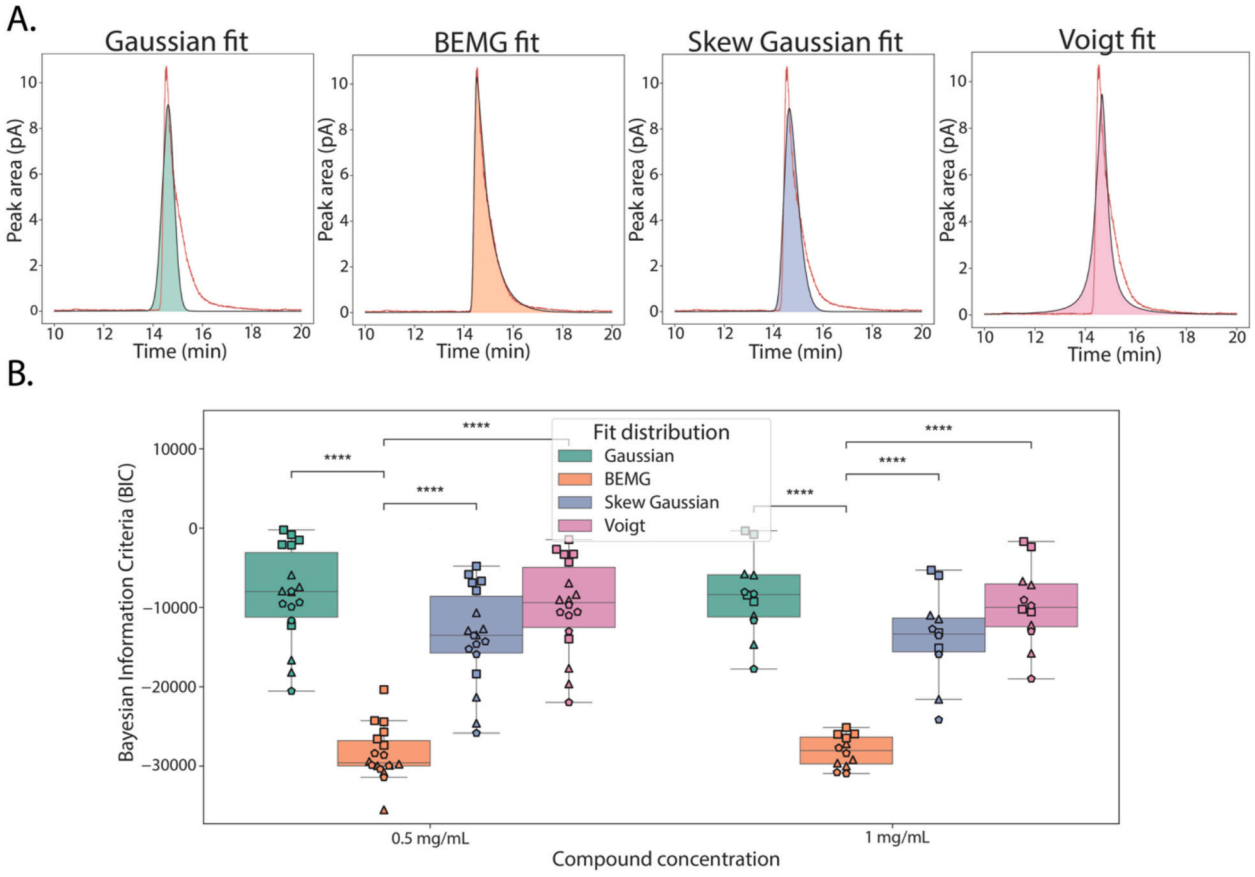


Fig. 4. Individual HPLC analyte peaks correspond to an exponential Gaussian distribution. (A) Empirical fits of (left to right) Gaussian, BEMG, skew Gaussian, and Voigt distributions to a single linoleic acid peak. (B) Bayesian Information Criteria (BIC) of the fit of above distributions on pooled arachidonic acid (triangle), docosahexaenoic acid (circle), and linoleic acid (square) single peaks. Analytes are grouped by injection volume. ($n = 12$ [4 experimental replicates of each of the 3 fatty-acids], Mann-Whitney U test with Bonferroni correction, ***: $p <$ 1e-03; ****: p < 1e-04).

2.2. Empirical results

2.2.1. Injections of HPLC standards confirm that the exponentially modified Gaussian fits real peaks

To empirically test our theoretical BEMG distribution on real data, we injected single, pure fatty acid analytes (linoleic acid, arachidonic acid, and docosahexaenoic acid) onto a C18 column, at individual concentrations of either 0.5 mg/mL or 1 mg/mL. Analytes were eluted from the column on a 3:1 methanol water to acetonitrile gradient (see methods) [39]. We then used the Python package lmfit [38] to fit one of four functions commonly used in chromatography to each of the fatty acid peaks. Fig. 4A depicts a representative chromatograph of linoleic acid fit to (*i*) a Gaussian distribution [12], (*ii*) a BEMG distribution [27,28], (*iii*) a skewed Gaussian distribution [13], and (*iv*) a Voigt distribution [40]. The goodness of fit was calculated using the Bayesian Information Criteria (BIC), which scores models based on both their

Table 1
Commercial and open source chromatography software. Software name (column 1), publisher (column 2), availability, where to download and cost (column 3), how peak fits are conducted (column 4), computer operating systems (column 5) and usability, how the user interacts with the software (column 6) of common chromatography analysis packages.

Software name	Publisher	Availability	Peak fitting function	OS	Usability
PeakClimber	Joshua Derrick/ Johns Hopkins University	Open source on GitHub	Bidirectional EMG	Windows/Mac/ Linux	GUI (Python-based)
Hplc.io [13]	Griffin Chure/Stanford	Open source on GitHub	Skew Gaussian	Windows/Mac/ Linux	Python package
HappyTools [12]	Bas Jansen/Thermofisher	Open source on GitHub	Gaussian	Windows/Mac/ Linux	GUI (Python-based)
Chromeleon	Thermofisher	~$1000 license	Valley-to-valley	Windows	GUI
OpenLab	Agilent	~$1000 license	Valley-to-valley	Windows	GUI
MassLynx	Waters	~$1000 license	Valley-to-valley	Windows	GUI
PeakLab	AIST	$449 annual license	General HVL (and many other functions)	Windows	GUI
PeakFit	Grafiti	$499 license	Bidirectional EMG (and 83 other models)	Windows	GUI
Clarity	Dataapex	Quote	Valley-to-valley	Windows	GUI
CHROMuLAN	Pikron	Free download from website	Valley-to-valley	Windows	GUI

residual function and the number of parameters (Fig. 4B). The skewed Gaussian and BEMG functions both have an additional parameter compared to the Gaussian and Voigt functions, making this comparison necessary. The BEMG had by far the lowest BIC for both concentrations of analytes (Fig. 4B).

2.2.2. Comparison of PeakClimber to other common HPLC quantification algorithms

To test the utility of PeakClimber, we compared its performance to publicly available software and commercial software with free trial versions. Other commercial software that was not tested was summarized in Table 1. To generate a test dataset with known standards, we injected three fatty acids with overlapping retention times: docosahexaenoic acid (12.3 min), arachidonic acid (12.5 min), and linoleic acid (12.9 min). We ran the analytes at concentrations of either 0.5 or 1 mg/mL. Thus, in each injection, the analytes were either of equal concentration or one analyte was double the concentration of the other two (Fig. 5A). This was done to test the dynamic range of PeakClimber as compared to other algorithms. The raw HPLC trace was then smoothed and normalized before being fit to three peaks by the four following algorithms: PeakClimber is the algorithm described in this paper (Fig. 5B). hplc.io [13] is a free, Python-based, chromatographic fitting software that uses skewed Gaussian functions as representative of single peaks (Fig. 5C). HappyTools is a free, standalone software package that uses Gaussian functions to fit single peaks [12] (Fig. 5D). Finally, valley-to-valley is an abstraction of algorithms [4,5] used by common HPLC-software such as Thermofisher's Chromeleon, Agilent's OpenLab CDS, or Waters' MassLynx that integrates the area under the curve of the trace between the lowest points between two identified peaks (Fig. 5E). Fits (black line in Fig. 5B-E) were performed on the entire trace (red line in Fig. 5B-E). Error comparisons are reported for each individual peak for each of the three analytes (lower panel Fig. 5B-E; analytes are DHA, ARA, and LA from left to right) using root mean square error (RMSE). Fit peaks were recentered on the canonical single analyte peaks because run times shifted to later elution times in the co-injections. PeakClimber outperformed all other software regardless of peak position (Fig. 5F). Particularly for the first peak in the co-injection, PeakClimber has lower error than the other algorithms due to the correct fitting of the tail of the peaks (Fig. 5G). PeakClimber also performed better for the second and third peaks (Supplemental Fig. 1). When error is calculated through percent error of the peak area, rather than RMSE, this pattern still holds (Supplement Fig. 1). We also conducted an additional test comparing the percent error of the peak area between the generalized Haarhoff-Van der Linde (HVL) [41] function utilized by the commercial software PeakLab, and found that there was no statistical difference between PeakClimber and PeakLab implementations (Supplementary Fig. 1).

2.2.3. Testing the limits of peak climber

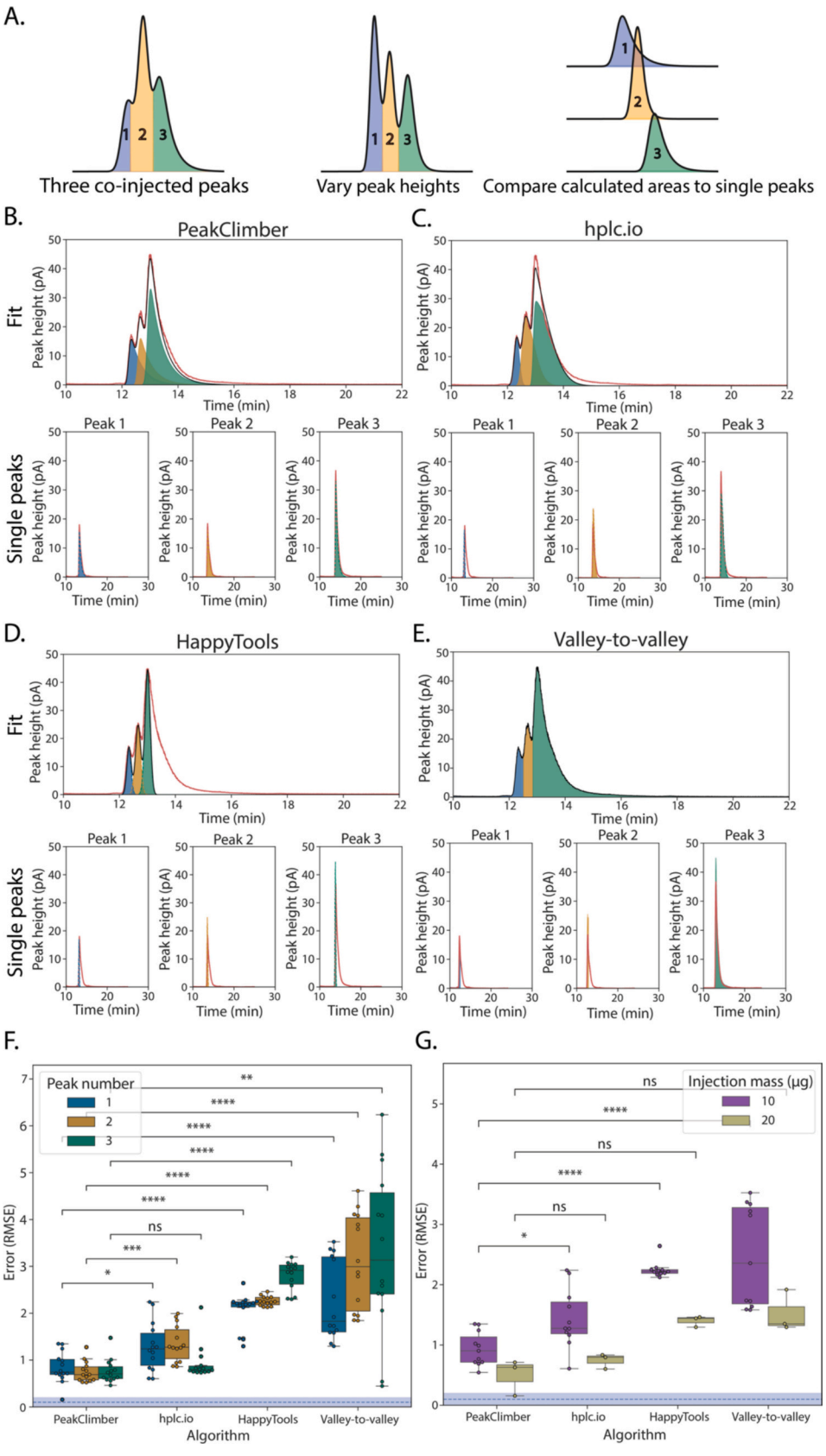
All algorithms, including PeakClimber, have reduced accuracy for groups of peaks under three separate circumstances: low signal-to-noise ratios, small distances between peaks, and uneven ratios between small and large peaks. To test the bounds specifically for PeakClimber, we computationally created traces of partially overlapping peaks using the real fatty acid traces that we generated in Fig. 2, with different levels of noise added on top of each trace. The first and second peak overlap, while the third peak is functionally independent, serving as a negative control (Supplemental Fig. 2). To test the fitting ability of each algorithm, rather than peak finding, which is already well-tested in other works [36], we provided the peak locations to each algorithm. The percent error for each of these cases is shown in the respective subpanels of Fig. 6.

For noise on single peaks, PeakClimber outperforms manual integration for added noise at a level of 0.3 times or greater than the true peak size (Fig. 6A). This is likely because PeakClimber better captures the shape of the underlying peak. For the distance between peaks, PeakClimber accuracy begins to drop off when the distance between peaks is less than 0.25 min. The valley-to-valley method is similarly sensitive to peak overlap only at a threshold distance of 1.5 min (Fig. 6B). For the ratio between peaks, we held peaks one and two a fixed distance of 1.5 min apart. Varying the ratio between these peaks did not change the error rate for the larger first peak, although PeakClimber outperformed valley-to-valley at every peak ratio. For the second peak, both algorithms have large error rates at ratios below 10:1 large peak: small peak. However, PeakClimber's error drops rapidly to 0 by a ratio of 4:1, whereas the valley-to-valley method drops in error more slowly and converges to a steady error rate of about 85 % (Fig. 6C). This error rate is the lower bound for the valley-to-valley method for peaks with this interpeak distance (Fig. 6B).

2.2.4. PeakClimber can be used to accurately quantify lipid differences between biological samples

To test the utility of PeakClimber on real biological data, we raised female *Drosophila melanogaster* from the larval stage on two different microbial conditions (germ-free or colonized by a defined, 7-species bacterial community from wild *D. melanogaster* with lactobacilli and *Acetobacters*) on a standard diet to ten days into adulthood. We then performed a lipid extraction and ran the isolated lipids on HPLC, using a two-step gradient (first methanol:water to acetonitrile, then acetonitrile to isopropanol) to separate lipid species by polarity and size, as adapted from [39]. Significant differences are observed by eye between germ-free and colonized animals (Fig. 7A), especially in the triglyceride region running from 60 to 70 min (Fig. 7A, inset). Individual peaks were quantified using either PeakClimber (Fig. 7B, left panel) or Thermofisher Chromleon (Fig. 7B, right panel). The two algorithms identified the same peaks but produced differences in the magnitude and statistical significance between peaks in colonized and germ-free animals (Fig. 7C). Chromeleon identifies all peaks in this region as significantly differentially abundant between samples, whereas PeakClimber only identifies some of these peaks as differentially abundant. This is not due to sample variance: PeakClimber and Chromeleon both capture biological sample variance equally. This discrepancy is likely because the tail of the first peak contributes to the area counted as the second peak by Chromeleon, causing a false positive when the area is counted this way. This does not occur with PeakClimber, which is able to deconvolve the tail of the first peak from the rest of the second peak. This suggests that PeakClimber has more utility in identifying real differentially abundant peaks as compared to standard industry software.

To identify the lipids contained in these peaks, we first performed a lipidomic-mass spectrometry analysis of whole male and female flies to establish a dataset for canonical fly lipid compounds (Supplementary Table 1). Then, we isolated the 8 sample peaks identified in Fig. 7B and ran them through a liquid chromatography-mass spectrometry (LC-MS) system to determine their identities. We used the lipidomic data to verify the LC-MS results from individual peaks. Many m/z numbers from the individual peak analyses were not found in the lipidomic-mass spec dataset. One possible reason for this could be due to difference in sample preparation: lipids were directly extracted from *D. melanogaster* to generate the lipidomics dataset, whereas individual peaks had to been run through a chromatography system and dried down during vacuum centrifugation in order to be isolated. Since water was present in the column mobile phase, and the vacuum centrifugation took 18 h, we hypothesized that significant saturation of unsaturated carbon-carbon bonds could occur. Thus, we also considered compounds in the lipidomics database with m/z values that were multiples of 2 less than the measured m/z value of the individual peaks. Hydrogen has an m/z value of 1, and each conversion of an unsaturated to saturated carbon bond adds two hydrogens to the molecule. In support of this hypothesis, we observed that only lipids containing unsaturated bonds changed mass between the lipidomics dataset and the analysis of individual peaks, which were composed of triglycerides containing unsaturated fatty acid tails, excepting the phospholipid peak at 60.1 min (Table 2). These peaks were also relatively rich in medium-chain triglycerides, which is in



(caption on next page)

Fig. 5. PeakClimber is more accurate and precise than industry software. (A) Schematic of the experiment performed in this figure. Three overlapping fatty-acid peaks were injected at either a ratio of 1:1:1, 1:2:1, 2:1:1, or 1:1:2. The calculated areas (using the algorithms listed below) were compared to the real injected areas of the individual peaks. (B) PeakClimber fit to a chromatograph of a co-injection of C18:1, C20:4 and C22:6 mixed in a 1:2:1 ratio. Red trace: raw data, black trace: predicted sum of peak areas, blue, orange, green shaded regions: predicted individual peak areas. (C-E) Fits to the same data in B by (C) hplc.io, (D) HappyTools, and (E) valley-to-valley. Bottom subpanels show the fits to each individual peak. (F) Quantification of error rates by RMSE of pooled co-injections of C18:1, C20:4 and C22:6 depending on peak position by above algorithms. Blue dotted line represents minimum RMSE error obtained from the single-peak fits. (G) Quantification of error rates for peak 1 alone comparing 10 μg and 20 μg injections. (Kruskal-Wallis Test with Bonferroni correction for subpanels F and G, n = 12, [3 biological replicates each with 4 experimental replicates], *: $p < 5e\text{-}02$**: $p < 1e\text{-}02$, ***: $p < 1e\text{-}03$, ****: $p < 1e\text{-}04$). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

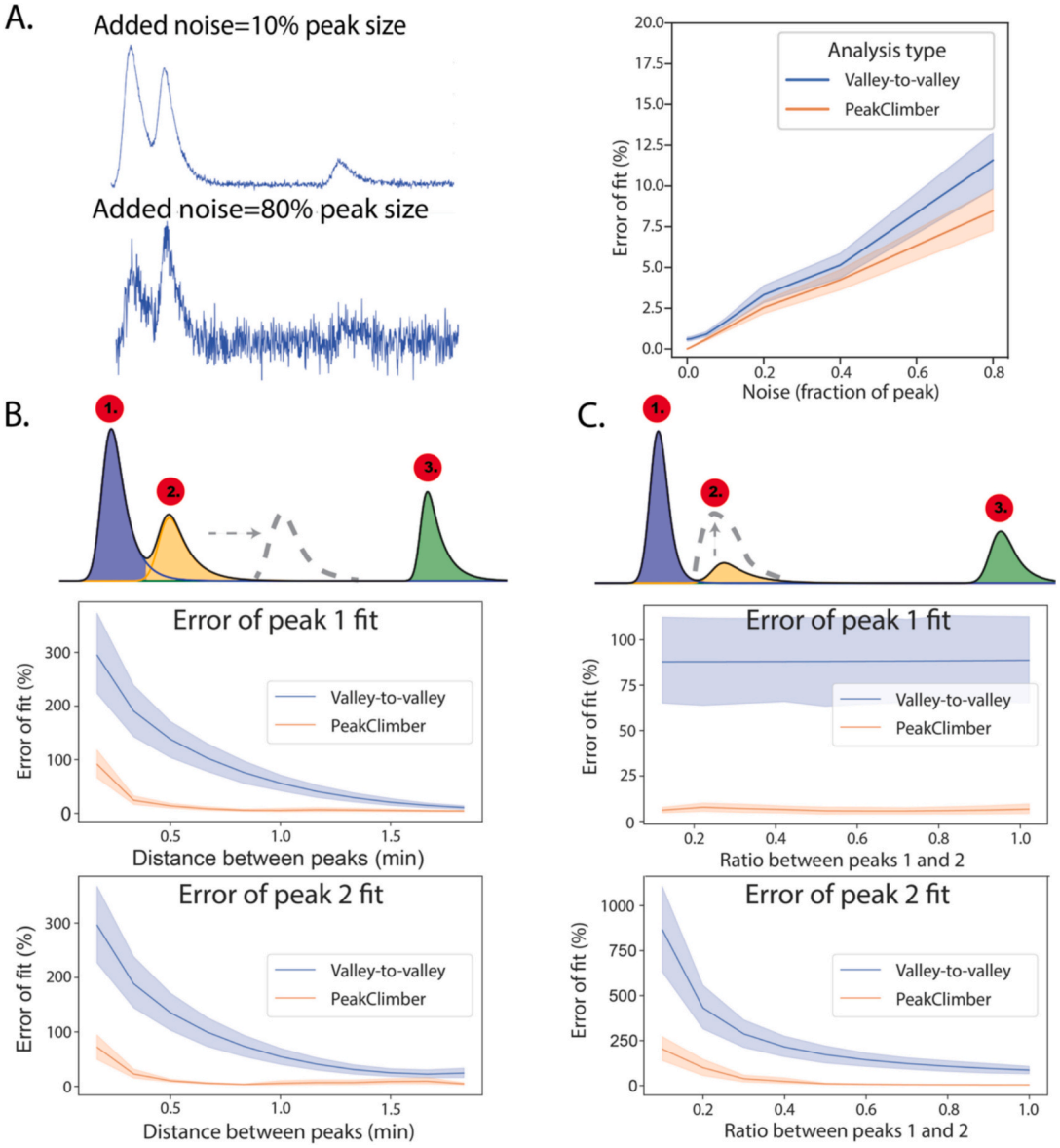


Fig. 6. PeakClimber is more robust to noise and peak overlap than the valley-to-valley method. (A) Schematic: Gaussian noise as a fraction of the peak amplitude is added to a single synthetically generated peak with amplitude between 1 and 2, σ between 0.1 and 0.2, γ between 2.9 and 3 and center at 1. This noise is detrended and removed using the PeakClimber algorithm and then the resulting peak area is either found by fitting with PeakClimber or integration with the valley-to-valley method. The calculated area is compared to the known area. (B) Three analyte curves are superimposed and shifted 0.1–2 min (peak 2) or 10 min later (peak 3). Curves are generated from real traces of arachidonic acid, docosahexaenoic acid, and linoleic acid. (C) Using the same parameters for BEMGs as in A, but with a fixed distance of 0.75 min between peaks 1 and 2, and 10 min between peaks 1 and 3, the ratio between peak 1 and 2 was varied between 0.1 and 1 ($n = 24$ [4 experimental replicates of each of the 3 fatty-acids at 2 concentrations]).

agreement with other literature on *Drosophila* lipids [42]. The peak at 63.1 min, which only contains saturated bonds, was identified without modification in the lipidomics dataset. Additionally, the specific elution time of these triglyceride peaks agrees with prior HPLC data of zebrafish lipid extracts that were also subject to mass spec confirmation [39].

Three out of the four significantly enriched peaks in germ-free animals contained long-chain polyunsaturated fatty acids (63.6,66,67 min). None of the non-significant peaks contained any polyunsaturated fatty acid tails, perhaps suggesting that colonized animals more readily metabolize these fatty-acids, or that they are preferentially absorbed by

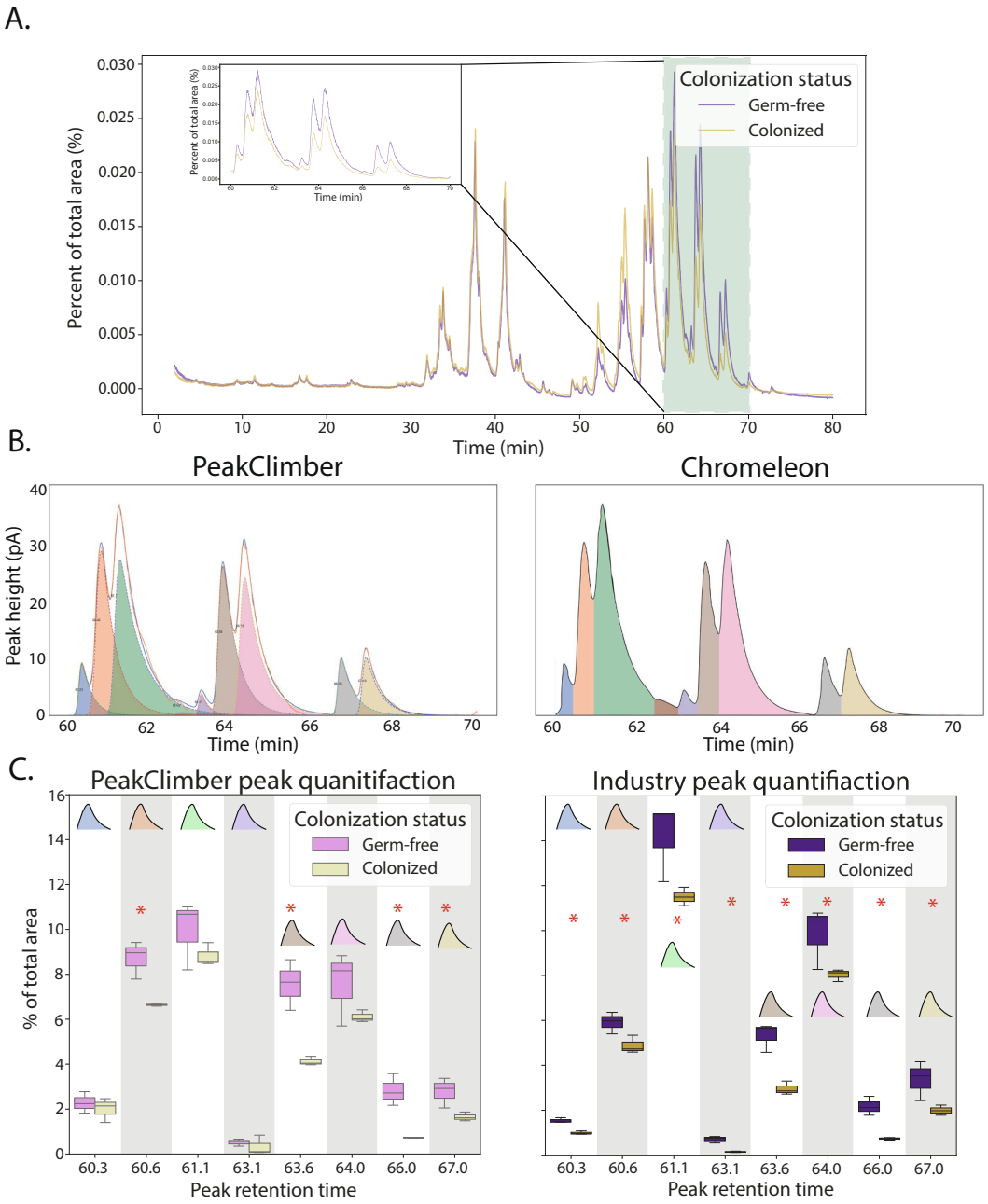


Fig. 7. PeakClimber more accurately quantifies biological differences between germ-free and colonized flies. (A) A lipid profile for germ free (purple) or colonized (yellow) female fruit flies normalized to total chromatograph area. Highlighted regions in green are shown in higher resolution in the inset are quantified below in panel B. (B) Algorithmic fit of the highlighted regions above using PeakClimber (left) and Thermofisher Chromeleon (right) on the germ-free trace. (C). Quantification of peak areas for selected peaks in the highlighted region (Kruskal-Wallis Test: *: $p < 0.05$ $n = 7$, 8 flies per sample). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

microbes, and are thus lost through feces.

2.2.5. *Uniqueness of peakclimber solution*

PeakClimber identifies peak areas by fitting BEMG functions to the underlying chromatography trace using non-linear regression [38]. We can define the error of this function as the sum of residuals between the y_i and the sum of n BEMGs $f_n(x_i)$

$$r_i = y_i - \sum_{1}^{n} f_n(x_i, \mu_n, A_n, \sigma_n, \gamma_n) \tag{15}$$

With $\mu_n, A_n, \sigma_n, \gamma_n$ being the center, amplitude, width, and decay function of each BEMG respectively. This residual function will have a single solution if the second derivatives of the function r are all positive, i.e., if the function is convex. When the shape of y is equivalent to the sum of BEMGs, this function will simplify to 0, which is trivially convex, meaning there is only a single solution.

Additionally, we can empirically restrict the sample space of parameters by observing real behavior of single HPLC peaks. For example, peak centers do not vary from their locations in identified traces, meaning that we can effectively reduce the parameter space down to 3 parameters for each BEMG. Kinetics and diffusion-to-flow ratios also place biophysical limits on the upper and lower bounds for the γ_n (tail, k_1 in equation eq. 11), and σ (width, b in eq. 11) parameters. In this reduced parameter space, we find a single optimum for two overlapping BEMGs fit to a region of the lipid profile of *D. melanogaster* thought to

Table 2
Mass spectrometry analysis and identification of *Drosophila* lipids eluted from the triglyceride region of the total lipid chromatograph. The retention time (column 1), measured m/z value (column 2), predicted unsaturated m/z value (column 3), and compound identity (column 4), and PeakClimber significance (Kruskal-Wallis rank-sum: *: $p < 0.05$ $n = 7$, 8 flies per sample) for each peak in the 60–70 min region of the fly lipid profile (TG = Triglyceride, PC = phosphatidylcholine).

Retention Time	Measured m/z	Predicted unsaturated m/z	Compound ID	Significant
60.3	654.27/ 846.67	654.56/834.6 (−12H)	TG(36:1) + NH4, PC (40:6) + H	No
60.6	820.66	818.72 (−2H)	TG(16:1,16:1,16:1) + NH4	Yes
61.1	794.65	790.69 (−4H)	TG(14:1,16:1,16:1) + NH4	No
63.1	768.64	768.7	TG(14:0,14:0,16:0) + NH4	No
63.6	848.69	846.75 (−2H)	TG(18:3,16:0,16:0) + H	Yes
64.0	822.68	820.73 (−2H)	TG(16:0,16:1,16:1) + NH4	No
66.0	876.72	869.75 (−7H)	TG(21:4,16:0,16:0) + NH4	Yes
67.0	850.70	848.77 (−2H)	TG(14:0,14:0,18:2) + NH4	Yes

contain only two peaks. Since the space is mapped by 6 parameters (not including the fixed centers), we used dimensional reduction to visualize the result as a PCA, which has a single minimum of the residual χ^2 function (Supplementary Table 2), indicating a unique solution.

3. Discussion

In this paper we have shown three principal findings. First, the BEMG function is a good fit for HPLC peaks. We showed this both analytically, computationally with Monte-Carlo simulations, and empirically by calculating the error of the fit for various common distributions used in chromatography to fit single analyte peaks. Many previous works from the 1970s and 1980s attempt to analytically work out these solutions, and their models also approximated a BEMG distribution [7,11,14,15,21,43,44]. HPLC peaks often do not represent single compounds, but groups of compounds. This means that a single peak is often a sum of individual compounds, all with behavior as described in Fig. 2. Due to the central limit theorem, this would suggest that the chromatographic traces that we observe should have more of a Gaussian character, but this is not what we observe empirically.

Second, we demonstrated the effectiveness of PeakClimber as compared to other commercially and freely available software tools to quantitatively analyze chromatography data with overlapping peaks. This is due to the ability of our algorithm to capture the tail region of the first peak in a group of peaks, which prevents undercounting and reduces distortion by larger surrounding peaks. Other commercial software packages that were not extensively tested in this paper, such as PeakLab or PeakFit have functions with higher statistical moments that may be able to more accurately fit peaks that the BEMG distribution used here. However, we found no statistical difference between the areas calculated by PeakClimber and PeakLab's generalized Haarhoff-Van der Linde (HVL) function [41], suggesting that this more complicated function is not necessary to model the chromatographs analyzed in this paper (Supplementary Fig. 1). For other columns and analytes, these more powerful commercial software packages may perform better peak fits. We also show that, given biophysical assumptions that limit the parameter space, there is only a single best fit solution for the underlying trace. This is vital for accurately quantifying peaks.

The package and documentation for PeakClimber are freely available on GitHub with an easy-to-use graphic user interface (GUI). A user manual is also included in the supplementary data of this paper (Supplementary Document 1), as well as on GitHub.

Third, we demonstrated the utility of our algorithm for the analysis of biological data. While mass spectrometry will always be the gold standard for metabolic analysis [45], HPLC represents a lower-cost, higher-throughput option than mass spectrometry [46–48]. Consider an experiment similar to one that we set up in Fig. 7 with multiple replicates of different dietary, genetic, or microbial conditions. Rather than analyze each replicate by mass spectrometry, one replicate from each group could be run through mass spectrometry, and the rest on HPLC, where relative changes in the compounds identified by mass spectrometry could be much more accurately quantified with PeakClimber. The recognition of HPLC as a medium-throughput bridge between mass spectrometry and high-throughput methods such as colorimetric kits could be one reason for the recent interest in development of tools to better analyze this type of data [45,47–49].

The reduction in triglycerides containing long-chain and polyunsaturated fatty acids in flies colonized with *Lactobacillus* and *Acetobacter* is an additional interesting finding from this work. Previous work in mice [50,51] has shown that various *Lactobacillus* species can protect against obesity by acting as a sponge for fatty acids, which are then excreted in the feces. These results also agree with work in the fly that shows that colonization can reduce triglyceride accumulation [52,53]. Why these bacteria reduce the presence of polyunsaturated fats in particular is unclear but could be due to a preference of *Lactobacillus* for these lipids, as their membranes are largely composed of unsaturated fatty acids [54].

Although we did not observe it in our dataset, neighboring peaks in HPLC are often composed of extremely similar compounds that are part of biochemical pathways such as fatty acid elongation or conversion between different phospholipid compounds [55,56]. PeakClimber could be used to find the precise step in these pathways that is affected by the genetic mutation, diet, or colonization condition of interest. This method could provide an additional advantage over alternative methods such as transcriptomic or proteomic analysis due to the ability to measure actual metabolite levels rather than the proxies of mRNA or protein levels, the activity of which can both be affected by downstream processing such as translation (in the case of mRNA), or post-translational modifications and confirmational changes (in the case of protein).

3.1.1. Limitations and comparison to other algorithms

Our mathematical model makes several simplifying assumptions about the geometry and flow rate of common HPLC systems. Based on complexities of experimental conditions that influence the quality of the data, more complex analysis programs, such as those that incorporate functions with higher statistical moments, including the aforementioned HVL [41], could be needed in specific cases. Future updates of PeakClimber could include these functions, or simpler functions such as a Gaussian. Due to the open-source nature of the software, these modifications can easily be made by users with some knowledge of Python and the relevant distributions.

One limitation of the test data used to validate PeakClimber is that it was only used to test HPLC data from lipid chromatography. Theoretically other biomolecules should have the same kinetic and diffusive behaviors as lipids, and many chromatographic traces present in the literature show single peaks that appear to be similar to BEMG functions [6,7,57–60].

However, despite these limitations, we want to highlight the performance advantages of PeakClimber compared to other available software. There is an ongoing debate on the utility computational separation to quantify overlapping peaks in the chromatography community [61]. We hope by providing a free, reliable, and powerful software tool we can facilitate more widespread use of signal processing for

the analysis of this type of data. Here with PeakClimber we show that we can in fact extract highly quantitative data from HPLC traces that contain overlapping peaks. Even compared to other open-source software that attempts to tackle this problem in a similar manner, PeakClimber much more accurately quantifies areas of overlapping peaks, due to its ability to consider long peak tails. For the analysis of biological data that contain many overlapping and non-overloaded peaks, we believe that PeakClimber will prove to be valuable.

4. Materials and methods

4.1. Monte-Carlo simulations

Exponential decay simulation: A column 1000 units long with 100 analyte particles at position 1 bound to the column is instantiated. At each time step the analyte has a 5 % chance of unbinding from the column (representing a k_1 value of 0.05). Once unbound the particle arrives at the detector a fixed time later, in this case 900 time-steps. The simulation was performed 10,000 times and results were pooled.

Multi-step reaction simulation: A column 1000 units long with 100 analyte particles at position 1 is instantiated. Particles are allowed to bind to the column with a probability of 0.5 and unbind with the same probability for 100 steps. When not bound to the column, particles move at the flow rate (1 binding site per step). After 100 steps, the probability of unbinding is reduced to 0.05, and the probability of rebinding is reduced to 0. Once unbound, these particles arrive at the detector a fixed time later, in this case 900 time-steps. The simulation was performed 10,000 times and results were pooled.

4.2. Fatty-acid chromatography

Fatty acid aliquots were obtained from Cayman Chemicals: linoleic acid (LA) (CAS 60–33-3), arachidonic acid (ARA) (CAS 506–32-1), and docosahexaenoic acid (DHA) (CAS 6217-54-5). The fatty acids were suspended in HPLC-grade isopropanol in stock concentrations of 10 mg/mL. Aliquots were then further diluted to either 0.5 mg/mL or 1 mg/mL as individual analytes or as part of one of the four mixtures analyzed (0.5:0.5:0.5, 1:0.5:0.5, 0.5:1:0.5, 0.5:0.5:1 mg/mL of DHA: ARA: LA respectively. 20 μL of each individual analyte or mixture was injected onto the HPLC system. The components of each sample were separated and detected by an HPLC system using a LPG-3400RS quaternary pump, WPS-3000TRS autosampler (maintained at 20 °C), TCC-3000RS column oven (maintained at 40 °C), Accucore C18 column (150 × 3.0 mm, 2.6 μm particle size; 8 μL flow cell maintained at 45 °C), FLD-3100 fluorescence detector and a Dionex Corona Veo charged aerosol detector (all from Thermo Fisher Scientific). Component peaks were resolved over a 30 min time range in a multistep mobile phase gradient as follows: 0–5 min = 0.8 mL/min in 98 % mobile phase A (methanol-water-acetic acid, 750:250:4) and 2 % mobile phase B (acetonitrile-acetic acid, 1000:4); 5–30 min = 0.8–1.0 mL/min, 98–30 % A, 2–44 % B, and 0–3.3 % mobile phase C (2-propanol) [42]. HPLC-grade acetic acid and 2-propanol were purchased from Fisher Scientific and HPLC-grade methanol and acetonitrile were purchased from Sigma-Aldrich.

4.3. Error tolerance simulations

Noise simulation: A single BEMG peak was initialized with the following parameters: amplitude between 1 and 5, γ (skew) between 2.9 and 3, and sigma between 0.1 and 0.2. Noise was added to the peak between 0 and 80 % of its amplitude. Peak area was calculated using PeakClimber or manual integration after denoising and compared to the known area of the generated peak.

Proximity and ratio simulations: Individual analyte traces of either linoleic acid (CAS 60–33-3), arachidonic acid (CAS 506–32-1), or docosahexaenoic acid (CAS 6217-54-5) were smoothed, and background subtracted as described in the body of the paper. Three copies of the corrected trace were superimposed on top of each other, and the resulting three peaks were computational separated by 0.1 to 2 min, or 10 min, respectively. The area of each of the three peaks was calculated either using PeakClimber, or the valley-to-valley algorithm, and the compared to the known underlying peak. Error was calculated by dividing the chi-square function of the residual error by the total peak area. For peak ratio, the second peak was held at a constant distance of 0.75, but the relative size of the peak was scaled between 0.1 and 1 of the size of the first peak.

4.4. Fly husbandry

Drosophila melanogaster Canton-S flies were initially isolated from long term germ-free stocks kept in lab. The parental generation of flies waseither maintained germ-free or inoculated at 5 days after eclosion with a 7-species microbiome mixture, consisting of *Lactiplantibacillus plantarum, Levilactobacillus brevis, A. pomorum, A. orientalis, A. cerevisiae, A. sicerae*, and *A. tropicalis* that recapitulates the microbiome found in a wild fruit fly. Parental flies were fed a diet consisting of 10 % glucose (v/v), 0.42 % propionic acid (v/v), 1.2 % (w/v) agar, and 5 % yeast (w/v). These flies were allowed to lay eggs on their food for 3 days. The resulting offspring were raised until 10 days post-eclosion before lipid extractions.

4.5. Lipid extractions and chromatography for Drosophila melanogaster

Groups of 8 flies were macerated using a bead beater in 500 μL of lipid extraction buffer (10 mM Tris, 1 mM EDTA, 7.8 pH). 400 μL of extract was mixed with 1.5 mL 2:1 chloroform:methanol (with 1 ng/mL of TopFluor cholesterol as an internal standard) and allowed to sit for 10 min. Then 500 μL of chloroform followed by 500 μL of extraction buffer was added to the mixture. The mixture was centrifuged at 2300 rcf for 5 min and the organic (bottom) phase was harvested. This was evaporated to dryness under vacuum centrifugation and then resuspended in 100 μL of HPLC grade isopropanol.

20 μL of the sample was injected onto the HPLC system as described earlier. Component peaks were resolved over an 80 min time range in a multistep mobile phase gradient as follows: 0–5 min = 0.8 mL/min in 98 % mobile phase A (methanol-water-acetic acid, 750:250:4) and 2 % mobile phase B (acetonitrile-acetic acid, 1000:4); 5–35 min = 0.8–1.0 mL/min, 98–30 % A, 2–65 % B, and 0–5 % mobile phase C (2-propanol); 35–45 min = 1.0 mL/min, 30–0 % A, 65–95 % B, and 5 % C; 45–73 min = 1.0 mL/min, 95–60 % B and 5–40 % C; and 73–80 min = 1.0 mL/min, 60 % B, and 40 % C. HPLC-grade acetic acid and 2-propanol were purchased from Fisher Scientific and HPLC-grade methanol and acetonitrile were purchased from Sigma-Aldrich.

For downstream LC-MS analysis, peaks were collected in the following intervals via fraction collector 60–60.4 min (1), 60.4–60.9 min (2), 60.9–61.5 min (3), 61.5–63 min (4) 63.2–64 min (5),64–66 min (6),66–67 min (7), 67–69 min (8). Peaks were then evaporated to dryness and resuspended in 100 μL methanol and DCM (50/50 v/v) with a concentration of 5 mM ammonium acetate in the final solution.

4.6. Liquid chromatography-high resolution (q-tof) mass spectrometry analysis

Ammonium acetate, methanol, and dichloromethane (DCM) were purchased from Thermo Fisher Scientific Inc. (Waltham, MA). Full scan mass spectral analyses of isolated peaks were conducted using an AB Sciex Quadrupole Time of Flight Mass Spectrometer controlled by Analyst 1.8 (5600 Q-TOF) (Framingham, MA). The mass spectrometer was coupled to a Shimadzu ultrafast liquid chromatographic system (UFLC, Kyoto, Japan), which consisted of a degasser, a quaternary pump, an autosampler, and a temperature-controlled column compartment. Each individual peak fraction (20 μL injection volume) was directly infused into the mass spectrometer's electrospray ionization (ESI) source

chamber through the UFLC autosampler. The mobile phase comprised methanol and DCM (50/50 v/v) spiked with 5 mM ammonium acetate, with a flow rate set to 100 µL/min. ESI parameters were as follows: source gases were set to 20 for Gas 1 and 30 for Gas 2, while the curtain gas was set to 30. The source temperature was set to 250 °C, and the Ion Spray Voltage Floating (ISVF) was set to 5.5 kV. The compound decluttering potential was set to 80. The mass spectrometer operated in TOF high-resolution full scan mode within an m/z scan range of 100 to 1200, with an accumulation time of 0.25 s for one minute for each sample run. The high-resolution mass spectrometer was calibrated with manufacturer solvent (PI: 4460131) to maintain mass accuracy.

Canonical lipid peaks were obtained with the following method. 8 frozen *Drosophila melanogaster* females (10 days post-eclosion) were resuspended in MTBE (1 mL), vortexed and then transferred to an Eppendorf tube. 300 µL of methanol with internal standard was added and samples were shaken for 10 min. 200 µL of water was added to facilitate phase separation. The extracts were centrifuged at 2000 rcf for 10 min. The top layer was removed, dried down, and reconstituted in 100 µL of IPA for analysis. Avanti's deuterated lipid mix, Equisplash, was used as an internal standard. This was spiked into the methanol at 1.5 µg/mL and used for extraction. Analysis was performed using a Thermo Q Exactive Plus coupled to a Waters Acquity H-Class liquid chromatograph. A 100 mm × 2.1 mm, 2.1 µm Waters BEH C18 column was used for separations. The following mobile phases were used: A- 60/40 ACN/H20 B- 90/10 IPA/ACN; both mobile phases had 10 mM ammonium formate and 0.1 % formic acid.

A flow rate of 0.2 mL/min was used. Starting composition was 32 % B, which increased to 40 % B at 1 min (held until 1.5 min) then 45 % B at 4 min. This was increased to 50 % B at 5 min, 60 % B at 8 min, 70 % B at 11 min, and 80 % B at 14 min (held until 16 min). At 16 min the composition switched back to starting conditions (32 % B) and was held for 4 min to re-equilibrate the column.

CRediT authorship contribution statement

Joshua T. Derrick: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Pragney Deme:** Methodology, Investigation. **Norman J. Haughey:** Supervision, Resources, Methodology. **Steven A. Farber:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Conceptualization. **William B. Ludington:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Conceptualization.

Code and data availability

All data and Jupyter Notebooks used to generate figures in this manuscript can be found at github.com/ATiredVegan/PeakClimberManuscriptRepository. The PeakClimber package and user instructions can be found at github.com/ATiredVegan/PeakClimber as well as in Supplementary Document 1.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We acknowledge Dr. Brandie Ehrmann of UNC for helping with the generation of the canonical fly lipid profile. We thank Dr. Huiqiao Pan, Dr. Darby Sweeney, and Dr. McKenna Feltes for testing the PeakClimber GUI. Dr. Darby Sweeney also helped write the PeakClimber user's guide. We acknowledge the Farber and Ludington labs for their helpful feedback conceptually and on the actual body of the manuscript. Finally, we thank an anonymous reviewer for their constructive criticism that significantly improved the work.

This work was funded by NIH grant R01DK093399 (SAF), NIH grant R01DK128454 (WBL), NSF Grant IOS 2144342 (WBL), the Carnegie Institution for Science Endowment (WBL, SAF), and the Mathers Foundation (SAF).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jchromb.2025.124721.

References

[1] R. Morley, M. Minceva, Liquid–Liquid Chromatography: Current Design Approaches and Future Pathways, Annu. Rev. Chem. Biomol. Eng. 12 (2021) 495–518, https://doi.org/10.1146/annurev-chembioeng-101420-033548.
[2] Y. Ito, M. Weinstein, I. Aoki, R. Harada, E. Kimura, K. Nunogaki, The coil planet centrifuge, Nature 212 (1966) 985–987, https://doi.org/10.1038/212985a0.
[3] J.J. Van Deemter, F.J. Zuiderweg, A. Klinkenberg, Longitudinal diffusion and resistance to mass transfer as causes of nonideality in chromatography, Chem. Eng. Sci. 5 (1956) 271–289, https://doi.org/10.1016/0009-2509(56)80003-1.
[4] N.A. Dyson, Chromatographic Integration Methods, Royal Society of Chemistry, 1998.
[5] Sternberg, Biomedical Image Processing, Computer 16, 1983, pp. 22–34, https://doi.org/10.1109/MC.1983.1654163.
[6] B. Steffen, K.P. Müller, M. Komenda, R. Koppmann, A. Schaub, A new mathematical procedure to evaluate peaks in complex chromatograms, J. Chromatogr. A 1071 (2005) 239–246, https://doi.org/10.1016/j.chroma.2004.11.073.
[7] A. Felinger, Data Analysis and Signal Processing in Chromatography, Elsevier, Amsterdam Lausanne New York [etc.], 1998.
[8] A.J.P. Martin, R.L.M. Synge, A new form of chromatogram employing two liquid phases, Biochem. J. 35 (1941) 1358–1368, https://doi.org/10.1042/bj0351358.
[9] J.C. Giddings, H. Eyring, A molecular dynamic theory of chromatography, J. Phys. Chem. 59 (1955) 416–421, https://doi.org/10.1021/j150527a009.
[10] L.C. Craig, C. Columbic, Identification of small amounts of organic compounds by distribution studies; use of a solid phase, Science 103 (1946) 587–589.
[11] G. Guiochon, Fundamentals of Preparative and Nonlinear Chromatography, 2nd ed, Elsevier Science & Technology, Chantilly, 2006.
[12] B.C. Jansen, L. Hafkenscheid, A. Bondt, R.A. Gardner, J.L. Hendel, M. Wuhrer, D.I. R. Spencer, HappyTools: A software for high-throughput HPLC data processing and quantitation, PLoS One 13 (2018) e0200280, https://doi.org/10.1371/journal.pone.0200280.
[13] G. Chure, J. Cremer, Hplc-py: A Python utility for rapid quantification ofComplex chemical chromatograms, J. Open Source Softw. 9 (2024) 6270, https://doi.org/10.21105/joss.06270.
[14] N.R. Amundson, The Mathematics of Adsorption in Beds. III. Radial Flow. Leon, Lapidus, J. Phys. Colloid Chem 54 (1950) 821–829, https://doi.org/10.1021/j150480a011.
[15] G.L. Frey, E. Grushka, Numerical solution of the complete mass balance equation in chromatography, Anal. Chem. 68 (1996) 2147–2154, https://doi.org/10.1021/ac960220o.
[16] G. Gotmar, T. Fornstedt, G. Guiochon, Peak tailing and mass transfer kinetics in linear chromatography, J. Chromatogr. A 831 (1999) 17–35, https://doi.org/10.1016/S0021-9673(98)00648-7.
[17] I. Langmuir, The constitution and fundamental properties of solids and liquids. Part i. Solids, J. Am. Chem. Soc. 38 (1916) 2221–2295, https://doi.org/10.1021/ja02268a002.
[18] P.J. Naish, S. Hartwell, Exponentially modified Gaussian functions—A good model for chromatographic peaks in isocratic HPLC? Chromatographia 26 (1988) 285–296, https://doi.org/10.1007/BF02268168.
[19] Eli Grushka, Characterization of exponentially modified Gaussian peaks in chromatography, Anal. Chem. 44 (1972) 1733–1738, https://doi.org/10.1021/ac60319a011.
[20] K. Lan, J.W. Jorgenson, A hybrid of exponential and gaussian functions as a simple model of asymmetric chromatographic peaks, J. Chromatogr. A 915 (2001) 1–13, https://doi.org/10.1016/S0021-9673(01)00594-5.
[21] Sadroddin Golshan-Shirazi, Georges Guiochon, Analytical solution for the ideal model of chromatography in the case of a Langmuir isotherm, Anal. Chem. 60 (1988) 2364–2374, https://doi.org/10.1021/ac00172a010.
[22] D.A. McQuarrie, On the stochastic theory of chromatography, J. Chem. Phys. 38 (1963) 437–445, https://doi.org/10.1063/1.1733677.
[23] L.M. Leemis, J.T. McQueston, Univariate distribution relationships, Am. Stat. 62 (2008) 45–53, https://doi.org/10.1198/000313008X270448.
[24] M.F. Wahab, D.C. Patel, R.M. Wimalasinghe, D.W. Armstrong, Fundamental and practical insights on the packing of modern high-efficiency analytical and capillary columns, Anal. Chem. 89 (2017) 8177–8191, https://doi.org/10.1021/acs.analchem.7b00931.

[25] K. Miyabe, G. Guiochon, Peak tailing and column radial heterogeneity in linear chromatography, J. Chromatogr. A 830 (1999) 263–274, https://doi.org/10.1016/S0021-9673(98)00852-8.
[26] T. Farkas, G. Guiochon, Contribution of the radial distribution of the flow velocity to band broadening in HPLC columns, Anal. Chem. 69 (1997) 4592–4600, https://doi.org/10.1021/ac970530m.
[27] SeaSolve Software, PeakFit; User's Manual, Sea Solve Software, Framingham, MA, 2003.
[28] S. Misra, M.F. Wahab, D.C. Patel, D.W. Armstrong, The utility of statistical moments in chromatography using trapezoidal and Simpson's rules of peak integration, J. Sep. Sci. 42 (2019) 1644–1657, https://doi.org/10.1002/jssc.201801131.
[29] F.J. Harris, On the use of windows for harmonic analysis with the discrete Fourier transform, Proc. IEEE 66 (1978) 51–83, https://doi.org/10.1109/PROC.1978.10837.
[30] J.F. Kaiser, Ch 7 digital filters, in: Syst. Anal. Digit, Comput., John Wiley and Sons, New York, 2025, pp. 218–255.
[31] J. Kaiser, R. Schafer, On the use of the Lsinh window for spectrum analysis, IEEE Trans. Acoust. Speech Signal Process. 28 (1980) 105–107, https://doi.org/10.1109/TASSP.1980.1163349.
[32] G. Bohlmann, Ein ausgleichungsproblem, Nachrichten Von Ges, Wiss. Zu Gött. Math.-Phys. Kl. 1899 (1899) 260–271.
[33] E.T. Whittaker, On a new method of graduation, Proc. Edinb. Math. Soc. 41 (1922) 63–75.
[34] S. Oller-Moreno, A. Pardo, J.M. Jimenez-Soto, J. Samitier, S. Marco, Adaptive Asymmetric Least Squares baseline estimation for analytical instruments, 2014 IEEE 11th Int. Multi-Conf. Syst. Signals Devices SSD14, 2014, pp. 1–5, https://doi.org/10.1109/SSD.2014.6808837.
[35] P. Eilers, H. Boelens, Baseline Correction with Asymmetric Least Squares Smoothing, Unpubl Manuscr, 2005.
[36] P. Virtanen, R. Gommers, T.E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S.J. Van Der Walt, M. Brett, J. Wilson, K.J. Millman, N. Mayorov, A.R.J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E.W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E.A. Quintero, C.R. Harris, A.M. Archibald, A.H. Ribeiro, F. Pedregosa, P. Van Mulbregt, SciPy 1.0 Contributors, A. Vijaykumar, A. P. Bardelli, A. Rothberg, A. Hilboll, A. Kloeckner, A. Scopatz, A. Lee, A. Rokem, C. N. Woods, C. Fulton, C. Masson, C. Häggström, C. Fitzgerald, D.A. Nicholson, D. R. Hagen, D.V. Pasechnik, E. Olivetti, E. Martin, E. Wieser, F. Silva, F. Lenders, F. Wilhelm, G. Young, G.A. Price, G.-L. Ingold, G.E. Allen, G.R. Lee, H. Audren, I. Probst, J.P. Dietrich, J. Silterra, J.T. Webber, J. Slavič, J. Nothman, J. Buchner, J. Kulick, J.L. Schönberger, J.V. De Miranda Cardoso, J. Reimer, J. Harrington, J.L. C. Rodríguez, J. Nunez-Iglesias, J. Kuczynski, K. Tritz, M. Thoma, M. Newville, M. Kümmerer, M. Bolingbroke, M. Tartre, M. Pak, N.J. Smith, N. Nowaczyk, N. Shebanov, O. Pavlyk, P.A. Brodtkorb, P. Lee, R.T. McGibbon, R. Feldbauer, S. Lewis, S. Tygier, S. Sievert, S. Vigna, S. Peterson, S. More, T. Pudlik, T. Oshima, T.J. Pingel, T.P. Robitaille, T. Spura, T.R. Jones, T. Cera, T. Leslie, T. Zito, T. Krauss, U. Upadhyay, Y.O. Halchenko, Y. Vázquez-Baeza, SciPy 1.0: fundamental algorithms for scientific computing in Python, Nat. Methods 17 (2020) 261–272, https://doi.org/10.1038/s41592-019-0686-2.
[37] A. Kirmse, J. De Ferranti, Calculating the prominence and isolation of every mountain in the world, Prog. Phys. Geogr. Earth Environ. 41 (2017) 788–802, https://doi.org/10.1177/0309133317738163.
[38] M. Newville, T. Stensitzki, D.B. Allen, A., Ingargiola, Non-Linear Least-Square Minimization and Curve-Fitting for Python, LMFIT, 2014, https://doi.org/10.5281/ZENODO.11813.
[39] V.H. Quinlivan, M.H. Wilson, J. Ruzicka, S.A. Farber, An HPLC-CAD/fluorescence lipidomics platform using fluorescent fatty acids as metabolic tracers, J. Lipid Res. 58 (2017) 1008–1020, https://doi.org/10.1194/jlr.D072918.
[40] J.T. Reilly, J.M. Walsh, M.L. Greenfield, M.D. Donohue, Analysis of FT-IR spectroscopic data: the Voigt profile, Spectrochim. Acta part Mol Spectrosc. 48 (1992) 1459–1479, https://doi.org/10.1016/0584-8539(92)80154-O.
[41] P.C. Haarhoff, H.J. Van der Linde, Concentration dependence of elution curves in non-ideal gas chromatography, Anal. Chem. 38 (1966) 573–582, https://doi.org/10.1021/ac60236a013.
[42] W. Palm, J.L. Sampaio, M. Brankatschk, M. Carvalho, A. Mahmoud, A. Shevchenko, S. Eaton, Lipoproteins in Drosophila melanogaster—assembly, function, and influence on tissue lipid composition, PLoS Genet. 8 (2012) e1002828, https://doi.org/10.1371/journal.pgen.1002828.
[43] R.E. Pauls, L.B. Rogers, Band broadening studies using parameters for an exponentially modified Gaussian, Anal. Chem. 49 (1977) 625–628, https://doi.org/10.1021/ac50012a030.
[44] Y. Kalambet, Y. Kozmin, K. Mikhailova, I. Nagaev, P. Tikhonov, Reconstruction of chromatographic peaks using the exponentially modified Gaussian function, J. Chemom. 25 (2011) 352–356, https://doi.org/10.1002/cem.1343.
[45] A.C. Schrimpe-Rutledge, S.G. Codreanu, S.D. Sherrod, J.A. McLean, Untargeted metabolomics strategies—challenges and emerging directions, J. Am. Soc. Mass Spectrom. 27 (2016) 1897–1905, https://doi.org/10.1007/s13361-016-1469-y.
[46] D.J. Kao, J.M. Lanis, E. Alexeev, D.J. Kominsky, HPLC-based Metabolomic analysis of Normal and inflamed gut, in: A.I. Ivanov (Ed.), Gastrointest. Physiol. Dis, Springer, New York, New York, NY, 2016, pp. 63–75, https://doi.org/10.1007/978-1-4939-3603-8_7.
[47] L. Perez De Souza, S. Alseekh, F. Scossa, A.R. Fernie, Ultra-high-performance liquid chromatography high-resolution mass spectrometry variants for metabolomics research, Nat. Methods 18 (2021) 733–746, https://doi.org/10.1038/s41592-021-01116-4.
[48] M. Jacob, A.L. Lopata, M. Dasouki, A.M. Abdel Rahman, Metabolomics toward personalized medicine, Mass Spectrom. Rev. 38 (2019) 221–238, https://doi.org/10.1002/mas.21548.
[49] S.T. Ovbude, S. Sharmeen, I. Kyei, H. Olupathage, J. Jones, R.J. Bell, R. Powers, D. S. Hage, Applications of chromatographic methods in metabolomics: A review, J. Chromatogr. B 1239 (2024) 124124, https://doi.org/10.1016/j.jchromb.2024.124124.
[50] H.R. Jang, H.-J. Park, D. Kang, H. Chung, M.H. Nam, Y. Lee, J.-H. Park, H.-Y. Lee, A protective mechanism of probiotic Lactobacillus against hepatic steatosis via reducing host intestinal fatty acid absorption, Exp. Mol. Med. 51 (2019) 1–14, https://doi.org/10.1038/s12276-019-0293-4.
[51] H. Chung, J.G. Yu, I. Lee, M. Liu, Y. Shen, S.P. Sharma, M.A.H.M. Jamal, J. Yoo, H. Kim, S. Hong, Intestinal removal of free fatty acids from hosts by *lactobacilli* for the treatment of obesity, FEBS Open Bio 6 (2016) 64–76, https://doi.org/10.1002/2211-5463.12024.
[52] P.D. Newell, A.E. Douglas, Interspecies interactions determine the impact of the gut microbiota on nutrient allocation in Drosophila melanogaster, Appl. Environ. Microbiol. 80 (2014) 788–796, https://doi.org/10.1128/AEM.02742-13.
[53] J.G. McMullen, G. Peters-Schulze, J. Cai, A.D. Patterson, A.E. Douglas, How gut microbiome interactions affect nutritional traits of Drosophila melanogaster, J. Exp. Biol. 223 (2020) jeb227843, https://doi.org/10.1242/jeb.227843.
[54] J.H. Veerkamp, Fatty acid composition of *Bifidobacterium* and *Lactobacillus* strains, J. Bacteriol. 108 (1971) 861–867, https://doi.org/10.1128/jb.108.2.861-867.1971.
[55] W. Bernhard, M. Linck, H. Creutzburg, A.D. Postle, A. Arning, I. Martincarrera, K. F. Sewing, High-performance liquid chromatographic analysis of phospholipids from different sources with combined fluorescence and ultraviolet detection, Anal. Biochem. 220 (1994) 172–180, https://doi.org/10.1006/abio.1994.1315.
[56] X. Huang, X.-F. Guo, H. Wang, H.-S. Zhang, Analysis of catecholamines and related compounds in one whole metabolic pathway with high performance liquid chromatography based on derivatization, Arab. J. Chem. 12 (2019) 1159–1167, https://doi.org/10.1016/j.arabjc.2014.11.038.
[57] Z. Huang, W.P. Fish, Development of simple isocratic HPLC methods for siRNA quantitation in lipid-based nanoparticles, J. Pharm. Biomed. Anal. 172 (2019) 253–258, https://doi.org/10.1016/j.jpba.2019.04.026.
[58] B. Ksas, M. Havaux, Determination of ROS-induced lipid peroxidation by HPLC-based quantification of Hydroxy polyunsaturated fatty acids, in: A. Mhamdi (Ed.), React. Oxyg, Species Plants, Springer, US, New York, NY, 2022, pp. 181–189, https://doi.org/10.1007/978-1-0716-2469-2_13.
[59] C.T. Mant, Y. Chen, Z. Yan, T.V. Popa, J.M. Kovacs, J.B. Mills, B.P. Tripet, R. S. Hodges, HPLC analysis and purification of peptides, in: G.B. Fields (Ed.), Pept. Charact. Appl. Protoc, Humana Press, Totowa, NJ, 2007, pp. 3–55, https://doi.org/10.1007/978-1-59745-430-8_1.
[60] Y. Zhang, M. Wu, J. Xi, C. Pan, Z. Xu, W. Xia, W. Zhang, Multiple-fingerprint analysis of Poria cocos polysaccharide by HPLC combined with chemometrics methods, J. Pharm. Biomed. Anal. 198 (2021) 114012, https://doi.org/10.1016/j.jpba.2021.114012.
[61] S.C. Moldoveanu, V. David, Modern Sample Preparation for Chromatography, Second edition, Elsevier, Amsterdam, 2021.