# ВОССТАНОВЛЕНИЕ ХРОМАТОГРАФИЧЕСКИХ ПИКОВ ПРИ ПОМОЩИ ЭКСПОНЕНЦИАЛЬНО-МОДИФИЦИРОВАННОЙ ГАУССИАНЫ

**Ю.Каламбет, к.ф.-м.н., К.Михайлова,к.ф.-м.н.,** ЗАО "Амперсенд", kalambet@ampersand.ru **Ю.Козьмин,** Институт биоорганической химии им. Шемякина РАН **И.Нагаев, к.х.н.,** Институт молекулярной генетики РАН

> Экспоненциально-модифицированная гауссиана широко используется для аппроксимации формы хроматографического пика. Найден надежный способ вычисления формы ЭМГ и определения положения вершины ее пика по одной из двух альтернативных формул. Предложен алгоритм разложения перекрывающихся хроматографических пиков на ЭМГ-пики методом линейной оптимизации. Процедура учитывает максимум линейного диапазона детектора для работы с зашкаленными пиками, когда из-за перегрузки детектора вершина пика обрезана. Рассмотрена задача восстановления формы одиночного зашкаленного пика.

Экспоненциально-модифицированная гауссиана (ЭМГ), вероятно, лучшая формула для описания формы хроматографического пика. Одно из основных преимуществ этой формулы – соответствующая ей физическая модель. Такая форма получится, если после идеальной хроматографической колонки, дающей гауссову форму пика, поставить камеру смешения, моделирующую неидеальность колонки.

вычисление эмг

۲

История развития теории ЭМГ насчитывает более 45 лет, ей посвящено много публикаций [1–14]. Исследователи предложили несколько формул, описывающих ЭМГ. Первая из них используется наиболее широко [1–14]:

$$F(t) = \frac{h\sigma}{\tau} \sqrt{\frac{\pi}{2}} e^{\left(\frac{\mu-t}{\tau} + \frac{\sigma^2}{2\tau^2}\right)} \operatorname{erfc}\left(\frac{1}{\sqrt{2}}\left(\frac{\mu-t}{\sigma} + \frac{\sigma}{\tau}\right)\right) = H(t)K(t), (1)$$

где t - время; h - высота гауссианы; σ<sup>2</sup> - дисперсия; μ - позиция немодифицированной гауссианы; т – время релаксации; параметр экспоненты, используемый для модификации гауссианы,

$$\operatorname{erfc}(z) = 1 - \operatorname{erf}(z); \quad \operatorname{erf}(z) = \frac{2}{\sqrt{-\pi}} \int_{0}^{z} e^{-t^{2}} dt; \quad \operatorname{H}(t) = e^{\left(\frac{\mu-t}{\tau} + \frac{\sigma^{2}}{2\tau^{2}}\right)}$$
  

$$\operatorname{K}(t) = \frac{h\sigma}{\tau} \sqrt{\frac{\pi}{2}} \quad \operatorname{erfc}\left(\frac{1}{\sqrt{2}}\left(\frac{\mu-t}{\sigma} + \frac{\sigma}{\tau}\right)\right).$$

Альтернативная формула ЭМГ была выведена Delley [7]:

$$F(t) = h e^{\frac{-(\mu-t)^2}{2\sigma^2}} \frac{\sigma}{\tau} \sqrt{\frac{\pi}{2}} \operatorname{erfcx}\left(\frac{1}{\sqrt{2}}\left(\frac{\mu-t}{\sigma^+}\frac{\sigma}{\tau}\right)\right) = G(t)E(t), \quad (2)$$

где 
$$G(t) = he^{\frac{-(\mu-t)^2}{2\sigma^2}}$$
 - исходная гауссиана  
и  $E(t) = \frac{\sigma}{\tau} \sqrt{\frac{\pi}{2}} \operatorname{erfcx} \left( \frac{1}{\sqrt{2}} \left( \frac{\mu-t}{\sigma} + \frac{\sigma}{\tau} \right) \right)$  - функция-множи-

тель, преобразующая гауссиану в ЭМГ.

www.j-analytics.ru

۲

۲

В формуле (1) при малых значениях т экспонента H(t) может быть очень большой и легко принимает большее значение, чем самое большое действительное число в формате двойной точности в стандарте IEEE-754 [15], тогда как K(t) для того же самого т будет практически равна нулю. Хотя произведение имеет приемлемое значение, оно не может быть вычислено с помощью компьютера. Будет получен неопределенный результат произведения, где, по крайней мере, значение одной из функций выпадает из вычислительного диапазона.

۲

При вычислении по альтернативной формуле (2) не возникает проблем при малых значениях т, поскольку для больших аргументов z

erfcx(z) 
$$\approx \frac{1}{z \sqrt{\pi}}$$
. В нашем случае  $z = \frac{1}{\sqrt{2}} \left( \frac{\mu - t}{\sigma} + \frac{\sigma}{\tau} \right)$ .

Однако функция erfcx(z) неудобна при отрицательных z, так как она очень быстро растет, и тогда неправильно использовать формулу (2) для вычислений ЭМГ. Данная ситуация проиллюстрирована рис.1, где значения сомножителей формул (1) и (2) изображены в логарифмической шкале. Из рис. 1а видно, что чем меньше т, тем выше K(t) в области t=0. При t=0,04 значение K(t) превышает допустимый диапазон чисел в формате двойной точности для t<1. Подобные проблемы возникают для уравнения (2) при больших t и t (рис.1б).

Для того чтобы избежать вычислительных недостатков обеих формул, мы использовали для вычислений ЭМГ формулу (1), когда z отрицательна, и формулу (2), когда z положительна. Для очень больших z (>6,71×10<sup>7</sup>) ЭМГ вычисляется по формуле:

$$F(t) \approx \frac{he^{\frac{-(\mu-t)^2}{2\sigma^2}}}{\left(1 + \frac{(\mu-t)\tau}{\sigma^2}\right)}.$$
(3)

При этом в зависимости от значения t различные части ЭМГ с фиксированными значениями о и т вычисляются по разным формулам. Мы показали [14], что временная координата вершины пика ЭМГ t<sub>0</sub> может быть вычислена по формуле:

$$t_0 = \mu + y \, \sigma \sqrt{2} - \frac{\sigma^2}{\tau},$$

где у – корень уравнения  $\operatorname{erfcx}(y) = \frac{\tau}{\sigma} \sqrt{\frac{2}{\pi}}$ .

www.j-analytics.ru



Рис.1. Логарифмическое представление сомножителей в уравнениях (1) и (2); μ=0, σ=1. Большая высота h на рис.16 используется для сдвига вверх кривой G(t)

При малых значениях т  $\left(\frac{\sigma}{\tau} > 10000\right)$  сдвиг положения вершины относительно гауссианы можно оценить величиной  $\mu - t_0 \approx \tau$ , и эта оценка полезна в некоторых практических случаях, вплоть до

$$\frac{\sigma}{\tau} > 10$$
, так как µ-t<sub>0</sub> отличается от т для  $\frac{\sigma}{\tau} = 10$ 

всего лишь на 1%.

Высота вершины пика ЭМГ будет определяться как

$$F(t_0) = h e^{\frac{-(\mu - t_0)^2}{2\sigma^2}}$$

**Оналитика** 4/2012(5)

= 63 <sub>Eu</sub>

 $\bigcirc$ 



Рис.2. Окно программы с оригинальным и восстановленным пиками. Пик восстановлен с использованием точек, не превышающих 10% его высоты. Исходные данные изображены пунктирной линией, восстановленные – сплошной. Для восстановления пика используются только точки, находящиеся между базовой линией и горизонтальной линией, проходящей над базой

и, как и ожидалось [13], вершина будет всегда располагаться на исходной немодифицированной гауссиане.

#### Условия эксперимента

Хроматограф (Bischoff Chromatography): детектор – модель Lambda 1010, насос – модель 2250; 0,01-4,99 мл/мин, термостат колонки Variotherm, 35°С.

Элюент: метанол-вода-уксусная кислота (45:55:0,1 по объему), изократический режим. Скорость потока 1 мл/мин. Колонка 1: Reprosil pur C18aq, 5 мкм, 4×150 мм. Колонка 2: Kromasil 100C18, 6 мкм, 4×150 мм. Длина волны детектирования 250 нм.

Программное обеспечение: МультиХром.

Проба: нипагин (метилпарабен, метиловый эфир параок-

сибензойной кислоты), разведение 1 мг/мл в метаноле.

# РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Для оценки предложенной процедуры восстановления формы пика мы выбрали 44 отдельно стоящих пика с приемлемым уровнем сигнала (высота пика была не менее 10% линейного диапазона детектора), взятых из различных ЖХ-хроматограмм. Формы пиков близки к ЭМГ: остаточное относительное стандартное отклонение не превысило 3%, среднее относительное стандартное отклонение оказалось равным 1,4%. Каждый пик аппроксимирован ЭМГ с использованием точек со значением отклика, не превы-

Gd<sup>64</sup>

۲

шающим 10, 25, 50 и 100% от истинной высоты пика. Пример окна программы с наложенными исходным и восстановленным пиками показан на рис.2. Для каждой аппроксимации измерялась относительная ошибка предсказания высоты и площади:

$$\delta V = 100 \quad \frac{abs(V_r - V)}{V}, \qquad (4)$$

где V соответствует измеренному параметру (площади или высоте) и V<sub>r</sub> – тому же самому параметру восстановленного пика. Средние значения ошибок предсказания высоты и площади пика как функция от использованного процента высоты представлены на рис.3. Видно, что в оптимальных случаях, к которым относятся наши примеры, можно надеяться на достаточно хорошее



Рис.3. Ошибка предсказания высоты и площади для 44 отдельно стоящих пиков на различных хроматограммах (все ЖХ)

предсказание параметров пика. При этом ординаты использованных для реконструкции точек ниже 0,1 высоты пика. Средняя ошибка предсказания высоты пика составляет около 25%.

Для проверки схемы восстановления в реальном эксперименте проведены две серии анализов нипагина при помощи двух различных колонок (см. врезку). Серия 1 включала четыре градуировочные хроматограммы и три хроматограммы с зашкаленными пиками. Наложенные хроматограммы этой серии представлены на рис.4. С целью восстановления пиков в качестве модельного выбран пик, соответствующий градуировочной точке 4 (наибольший по высоте пик в наборе градуировочных точек). Другими словами, этот пик принят нами за пример "истинной" формы пика. Высота этого пика

www.j-analytics.ru

 $\bigcirc$ 

**О**налитика 4/2012(5)

۲

۲

# МЕТОДОЛОГИЯ

۲



#### Рис.4. Серия 1 хроматограмм нипагина;

относительное количество вещества 1:2:3:4:10:20:30. Временная шкала некоторых хроматограмм сдвинута не более чем на 3 с для более наглядного сравнения



**Рис.5.** Восстановление пика для анализов нипагина серии 1

составляла 53% от линейного диапазона детектора, относительное стандартное отклонение 1,5% и т/σ=0,54. Каждый зашкаленный пик восстановлен несколькими различными способами:

- ЭМГ-аппроксимация без ограничений;
- ЭМГ-аппроксимация с фиксированными т и о, характеризующими модельный пик (фиксированная форма ЭМГ);
- аппроксимация гауссианой без ограничений. Проведено сравнение результатов восстановлеима защизающието и молодиието пиков с исполь-

ния зашкаленного и модельного пиков с использованием различных уровней сигнала (рис.5.) Для фиксированной формы ЭМГ параметры о и т равны величинам, полученным для модельного пика; свободная форма ЭМГ позволяет оптимизировать о и т. Модельный пик – это наибольший по высоте "нормальный" пик (пик с относительным количеством нипагина 4 в серии 1). "Истинные" высота и площадь определены экстраполяцией градуировочной зависимости. Результаты восстановления для этой серии оказались недалеки от ожидаемых (рис.5, линия 4) как для аппроксимации без ограничений (рис.5, линия 1), так и для аппроксимации с использованием фиксированной формы ЭМГ (рис.5, линия 2). Аппроксимация с использованием гауссианы (рис.5, линия 3) оказалась значительно хуже.

В серии 2 предпринята попытка перейти на более значительные уровни перегрузки. Эта серия состояла из пяти анализов, первый из которых (модельный) соответствовал 91% линейного диапазона детектора, относительное стандартное отклонение составляло 0,93% и τ/σ=0,84. В других анализах инжектированный объем был увеличен в 2, 3, 4 и 5 раз и пики получались зашкаленными. Результирующие пики обрабатывались аналогично пикам серии 1. Результаты восстановления представлены на рис.6. "Истинная"высота пика рассчитана масштабированием высоты модельного пика (относительное количество вещества равно 1). Видно, что результаты предсказания для больших пиков хуже, чем ожидалось, хотя вплоть до перегрузки в четыре раза (25%) они приемлемы.

Существует множество причин искажения формы пиков при их зашкаливании: нелинейность детектора; уширение пика вследствие увеличения инжектированного объема; уширение или заострение пика вследствие изменения состава раствора, в котором приготовлена проба; перегрузка колонки. Большинство из этих причин изменяют форму пика таким образом, что она не может быть описана ЭМГ и, следовательно, восстановление пика на основе неправильных предположений о его форме даст увеличенную ошибку. Мы не претендуем на перечисление всех возможных причин изменений формы пика, и выясне-



Рис.6. Восстановление пика для анализов нипагина серии 2. Относительное количество вещества 1:2:3:4:5

www.j-analytics.ru

• 65 <sub>ть</sub>

۲

**О́налитика** 4/2012(5)

ние главной причины ошибки в нашем случае не является предметом обсуждения в данной статье.

При поиске оптимального способа восстановления пиков отмечено, что восстановление с использованием фиксированной формы пика дало гораздо лучшие результаты, чем оптимизация без ограничений: оно никогда не приводит к ошибочному локальному минимуму, в то время как неограниченная оптимизация очень легко может скатиться в локальный минимум и, соответственно, требует осторожности при выборе начальных условий. Глобальный минимум оптимизационного процесса всегда близок к минимуму оптимизации с использованием фиксированной формы, так что окончательная процедура состоит из "грубой" оптимизации с использованием фиксированной формы (если она известна) и "тонкой" оптимизации с использованием полного набора переменных, включая о и т.

Моделирование с заранее известной формой пика дает хорошие результаты, но мы должны быть уверены, что форма пика действительно постоянна. Проверка постоянства формы пика может быть выполнена при помощи градуировочных анализов в случае доступности таковых. Хороший индикатор формы пика ЭМГ – отношение т/σ: если оно остается постоянным при градуировочных анализах, можно ожидать, что форма пика не изменится в рабочих анализах.

Мы провели небольшое исследование на тему постоянства формы пика при градуировочных



Рис.7. Зависимость формы пика от количества анализируемого вещества. Градуировки 1 и 2 отражают крайние случаи, найденные в результате исследования более сотни градуировок различных типов

**О**налитика 4/2012(5)

анализах, которые были в нашем распоряжении. В большинстве случаев отношение т/σ оставалось практически постоянным, показывая характер, подобный нипагину (рис.7, линия 3), иногда с меньшей асимметрией для более низких концентраций. Отмечены также некоторые случаи, где форма пика далека от постоянной (рис.7, линии 1 и 2). Обратите внимание, что в ряде случаев при градуировке (рис.7, линия 2) пик, который был "хвостатым" при низких количествах, становился "брюхатым" при высоких количествах, т.е. меняется даже знак т. При аппроксимации с использованием заранее известной формы пика о таких, хотя и редких, случаях следует помнить.

#### выводы

۲

Улучшенный алгоритм вычисления ЭМГ во всем диапазоне параметров позволил нам создать методику разложения перекрывающихся хроматографических пиков, основанную на методе линейной оптимизации. Процедура разложения пика по форме работает даже в тех случаях, когда некоторые пики выходят за пределы линейного диапазона детектора (зашкаливают), поскольку в ней не используются отклонения в точках, где детектор перегружен. Процедура применима для восстановления формы отдельно стоящего зашкаленного пика.

Программа позволяет выполнять оптимизацию с фиксацией выбранных параметров, и этот способ оптимизации помогает избежать локальных оптимизационных минимумов, если σ и т заранее известны. Во многих случаях оценка высоты и площади восстановленного пика с заранее заданными σ и τ может использоваться сразу без дальнейшей оптимизации по этим параметрам, но тогда надо проверить гипотезу о независимости формы пика от концентрации. Необходимо также понимать и другие ограничения методологии. Например, в отличие от перегрузки детектора перегрузка колонки в большинстве случаев не может быть описана ЭМГ, и восстановление такого пика в предположении о том, что его форма описывается ЭМГ, даст ошибочные результаты.

Технология восстановления формы пика может использоваться в повседневной работе при оценке требуемого разведения на ранних стадиях разработки метода или в случае большой ошибки в концентрации. Это разведение должно устанавливать амплитуду пика в таком диапазоне, где весь сигнал будет доступен для определения.

www.j-analytics.ru

Dv 66 =

( )

Численные результаты предсказания площадей и высот по восстановленному пику при малой (до двукратной) перегрузке имеют достаточную точность для внутрилабораторного использования.

# ЛИТЕРАТУРА

۲

- J.C. Sternberg. Advances in Chromatography. Vol. 2. New York, Marcel Dekker, 1966, p.205–270.
- McWilliam, I., Bolton H. Instrumental Peak Distortion. I. Relaxation Time Effects. - Analytical Chemistry, 1969, v.41, p.1755-1762.
- Grushka E. Characterization of Exponentially Modified Gaussian Peaks in Chromatography. – Analytical Chemistry, 1972, v.44, p.1733–1738.
- Yau W. Characterizing Skewed Chromatographic Band Broadening. - Analytical Chemistry, 1977, v.49, p.395-398.
- Pauls R., Rogers L. Band Broadening Studies Using Parameters for an Exponentially Modified Gaussian. – Analytical Chemistry, 1977, v.49, p.625–628.
- Foley J., Dorsey J. A Review of the Exponentially Modified Gaussian (EMG) Function: Evaluation and Subsequent Calculation of Universal Data. – Journal of Chromatographic Science, 1984, v.22, p.40–46.
- 7. **Delley R.** Series for the Exponentially Modified Gaussian Peak Shape. Analytical Chemistry, 1985, v.57, p.388.

ЗАО «Амперсенд» представляет

МультиХром

версия 3

- Hanggi D., Carr P. Errors in Exponentially Modified Gaussian Equations in the Literature. -Analytical Chemistry, 1985, v.57, p.2394-2395.
- NaishP., HartwellS. Exponentially Modified Gaussian Functions - a Good Model for Chromatographic Peaks in Isocratic HPLC? - Chromatographia, 1988, v.26, p.285-296.
- Delley R. Modifying the Gaussian Peak with More Than One Time Constant. - Analytical Chemistry, 1986, v.55, p.2344-2346.
- 11. Felinger A., Pap. T., Inczédy J. Curve Fitting to Asymmetrical Chromatograms by the Extended Kalman Filter in Frequency Domain. – Talanta, 1994, v.41, p.1119–1126.
- Felinger A. Deconvolution of Overlapping Skewed Peaks.
   Analytical Chemistry, 1994, v.66, p.3066–3072.
- Felinger A. Data analysis and signal processing in chromatography. – Data Handling in Science and Technology, ELSEVIER: Amsterdam, the Netherlands, 1998, v.21, p.43–78.
- Kalambet, Y., Kozmin, Y., Mikhailova, K., Nagaev, I., Tikhonov, P. Reconstruction of chromatographic peaks using the exponentially modified Gaussian function. – Journal of Chemometrics, 2011, v.25, p.352–356.
- 15. IEEE Standard for Binary Floating-Point Arithmetic (IEEE 754-2008), IEEE, New York, 2008, 58 p.

 Управление более чем 200 хроматографами и отдельными устройствами различных производителей: Agilent, Altex, Gilson, Knauer, Metrohm, Люмэкс, Аквилон, Хроматек, ПортЛаб, Химавтоматика и др.

- Оригинальные методы предельного подавления шумов, адаптивной интерполяции пиков и разделения перекрывающихся пиков по форме.
- Спектральный анализ и многоканальная хроматография: идентификация и расчет количества компонента по спектру пика, разделение перекрывающихся пиков с помощью факторного анализа.

# support@ampersand.ru

- Новый редактор ручной разметки.
- Поддержка капиллярного электрофореза с дополнительными методами обработки данных.
- Специализированный модуль для гель-проникающей хроматографии.
- Новое в отчетах: широкие возможности форматирования — создание документов заданного образца; выдача отчетов в форматах PDF, RTF, HTML; статистические отчеты для группы хроматограмм.
- Поддержка требований «Надлежащей лабораторной практики» (GLP, GMP) и электронного документооборота.

#### www.multichrom.ru

# +7(499)196-52-90

www.j-analytics.ru

= 67 <sub>Ho</sub>

۲

