

# Chemometrie in der Analytischen Chemie<sup>1</sup>

## 3. Peakform-Analyse zur Verbesserung der chromatografischen Trennleistung

Florian Klingler, Eckhard Reh\*

\* Korrespondenz: Prof. Dr. E. Reh, Technische Hochschule Bingen, Email: e.reh@th-bingen.de

### Einleitung

Eine vollständige chromatografische Trennung komplexer Proben stellt aufgrund limitierender physikochemischer Parameter eine große Herausforderung dar. Durch Analyte mit ähnlichen physikalisch, chemischen Eigenschaften resultieren insbesondere bei der HPLC dabei überlappende Peaks. Insbesondere bei der Gelpermeationschromatografie mit dem begrenzten Elutionsbereich, der Ionenaustauschchromatografie von Proteinen mit den ungünstigen Diffusionskoeffizienten oder der Reversedphase-Trennung von vielen hundert Peptiden ist die HPLC oft nicht in der Lage, die Analyte zu trennen. Ist mittels Optimierung der Chromatografie keine Verbesserung zu erreichen, kann eine Dekonvolution der überlappenden Messdaten mittels Peakform-Analyse genutzt werden.

### Modelle für chromatografische Peaks

Zur mathematischen Modellierung von chromatografischen Peaks sind diverse Modelle beschrieben. Geeignete 4-Parametermodelle sind beispielsweise die exponentiell modifizierte Gaußfunktion [1], die pPoisson-Funktion [2], und die Fraser-Suzuki-Funktion [3] (siehe Anhang), die auf den Parametern für Position, Höhe und Breite und einem Asymmetrie-Parameter beruhen. Weiterhin werden auch Modelle mit mehr als vier Parametern eingesetzt, wie die generalisierte exponentiell modifizierte Gauß-Funktion [4]. Wichtig ist, dass aufgrund der variierenden Peak-Asymmetrien mehrere Modellfunktionen zur Anwendung kommen sollten, die dies berücksichtigen.

### Algorithmen zur Modellierung chromatografischer Peaks

#### a) Nichtlineare Regression (NLR) nach Levenberg-Marquardt

Der Levenberg-Marquardt-Regression [5], auf der auch der nichtlineare Regressionsansatz des in Microsoft Excel implementierten Solvers basiert bzw. die kommerzielle Implementierung in PeakFit (Systat Software), liegt zum Einen ein faktorieller Versuchsplan mit der Differenz zwischen angepassten und gemessenen Daten als Zielgröße und zum Anderen eine Linearisierung durch Näherung des Modells mit Hilfe einer Taylor-Reihe zugrunde. Die nichtlineare Regression benötigt dabei zwingend zu Beginn der Anpassung Startwerte für alle Parameter des Modells.

Daraus resultieren diverse Probleme:

- ungünstige Startwerte können zu falschen Ergebnissen führen,
- die Startwerte werden bei der Anpassung verworfen werden, so dass keine Vorgaben / Grenzen angegeben werden können,
- es besteht die Gefahr auf Nebenoptima zu stoßen und nicht das globale Optimum zu erreichen.

#### b) Iterative lineare Approximation (ILA)

Bei der iterativen linearen Approximation [2] wird zunächst die Spitze der Messreihe durch ein inverses Polynom 3. Grades angepasst. Das Maximum der Funktion liefert die Parameter für Position und Höhe eines Peaks. Asymmetrie und Breite werden aus dem Abfall der Polynomfunktion ermittelt. Der modellierte Peak wird von den Messwerten subtrahiert, danach können in den verbliebenen Rest-Werten sukzessive weitere Peaks angepasst werden. Großer Vorteil dieses Algorithmus ist, dass keine Startwerte benötigt werden, allerdings können aus Voruntersuchungen, Parameter als Grenzen vorgegeben werden, um die Anpassung zu verbessern. So wird die Gefahr auf ein Nebenoptimum aufzulaufen, minimiert. Auf diesem Algorithmus basiert das Programm PeakCalc [6], das für die folgenden Ausführungen eingesetzt wurde.

Zur Bewertung der Peakanpassung dient die relative Residuensumme,  $R\text{Sum}_{\text{rel}}$  (siehe Anhang), sie repräsentiert die prozentuale Abweichung (Flächendifferenz) der angepassten Peaks (graue Kurve) von den Original-Daten (schwarze Kurve). Eine detaillierte Einführung in die theoretischen Grundlagen soll an dieser Stelle nicht erfolgen, es wird auf entsprechende Fachliteratur verwiesen [7].

## Beispiel 1: Trennung von Proteinen mittels GPC

Prinzipiell lässt sich die Analyse der Messdaten mithilfe von herkömmlichen Datenverarbeitungstools wie etwa Tabellenkalkulation Excel oder Statistikpaket R realisieren, jedoch zeigte sich, dass die Implementierung der Modellfunktionen und Algorithmen in kommerzieller Software zu weitaus besseren Ergebnissen bei der Anpassung führt. Grundsätzlich kann die Peakform-Analyse mittels PeakCalc ohne Angaben erfolgen, in komplexen Fällen ist es sinnvoll, Peakparameter in der Iteration vorzugeben, wenn die Daten der einzelnen Analyte vorliegen. Je mehr valide Vorgaben in die Iteration eingebracht werden, desto besser sind die Resultate. Die Vorgaben (für Position, Breite, Asymmetrie bei verschiedenen Retentionszeiten, vgl. Abbildung 1) werden während der Iteration als Grenzen ( $\pm$  Abweichung) eingehalten.

### Fit-Vorgaben

Peakzahl:  Peaktyp:

	Pos. 1	Pos. 2	Pos. 3	Pos. 4	Pos. 5	Pos. 6	Pos. 7	Pos. 8	Pos. 9	Pos. 10
x-Position	<input type="text"/>									
x-Wert	<input type="text"/>									
Breite	<input type="text"/>									
x-Wert	15,10	15,50	16,20	<input type="text"/>						
Asymmetrie	67,00	78,00	102,00	<input type="text"/>						

Abweichung

Position:  Breite:  Asymmetrie:

### Fit-Parameter

Einheit x-Wert:  Einheit y-Wert:   
[s, min, h etc.] [V, mA etc.]

Abb. 1: Fit-Vorgaben PeakCalc für die Dekonvolution der GPC-Trennung

An dem Beispiel der Trennung von Insulin, Aprotinin und Lysozym mit Hilfe einer Gelpermeations-Chromatografie (GPC), kann verdeutlicht werden, wie die Dekonvolution der überlappenden Peaks funktioniert, wenn die Parameter der Peaks zunächst durch eine vorherige Analyse der reinen Einzelkomponenten vorliegen. Das errechnete Modell für die Dekonvolution, weicht dabei nur geringfügig von den tatsächlichen Messdaten ab und trägt so zu einer Verbesserung der Anpassung bei (siehe Abbildung 2). Im vorliegenden Fall wird deutlich, dass die Abweichung auf Grund der großen Asymmetrien gravierender sind ( $R_{Sum_{rel}} > 5\%$ ).

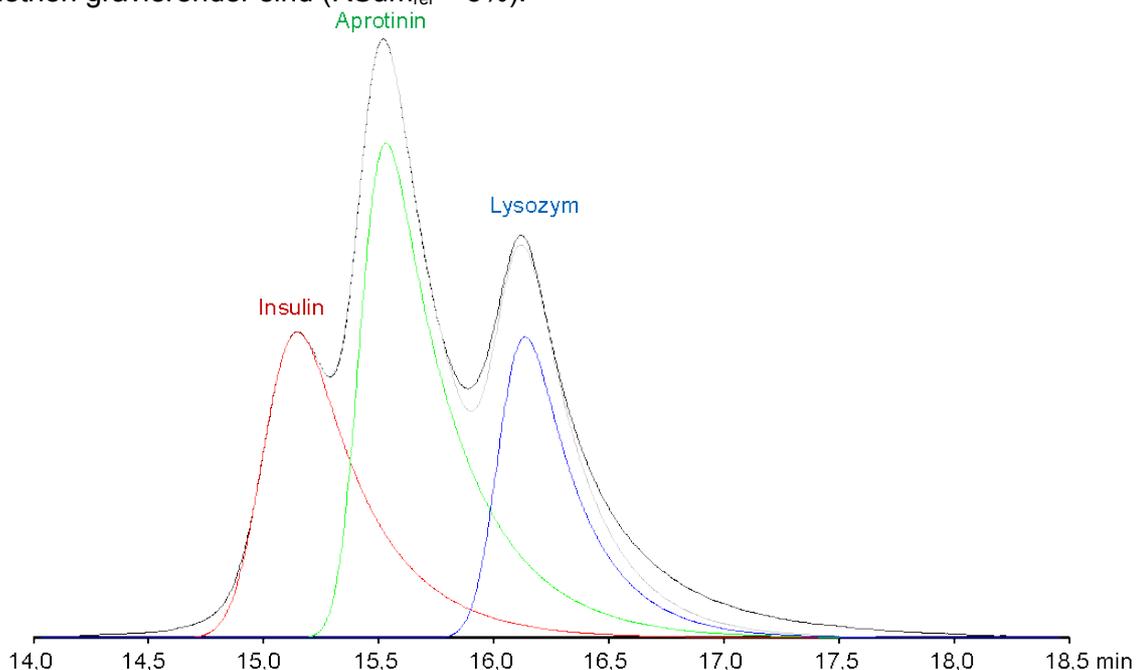


Abb. 2: Dekonvolution einer GPC-Peakgruppe mit Vorgaben aus Einzelkomponentenanalysen (3 x exp. mod. Gauss-Modell, Asymmetrie-Werte),  $R_{Sum_{rel}} = 6,9\%$  (schwarz: Originaldaten, grau: Summe iterierte Peaks).

## Beispiel 2: Trennung von Cytochrom C-Varianten mittels IEC

Auch anhand der Mess- und Modelldaten bei einer Trennung verschiedener Cytochrom C-Varianten mittels Ionenaustausch-Chromatografie (IEC), lässt sich zeigen, wie gut sich die überlappenden Peaks modellieren lassen (siehe Abbildung 4). Allerdings lässt sich dabei auch ein Problem feststellen, welches häufig zu beobachten ist. Je nach verwendetem mathematisches Modell, lassen sich Peak-Asymmetrien nur schwer modellieren. Eine exakte Abbildung von Peak-Leading oder wie in diesem Beispiel Peak-Tailing ist oft nicht gänzlich möglich. Diesen Nachteil gilt es, gerade bei großen Peak-Asymmetrien, bei der Beurteilung der Dekonvolutions-Ergebnisse zu beachten. Durch eine Variation der Anpassung mit Hilfe verschiedener Modelle kann die Anpassung aber verbessert werden (siehe Abbildung 3).

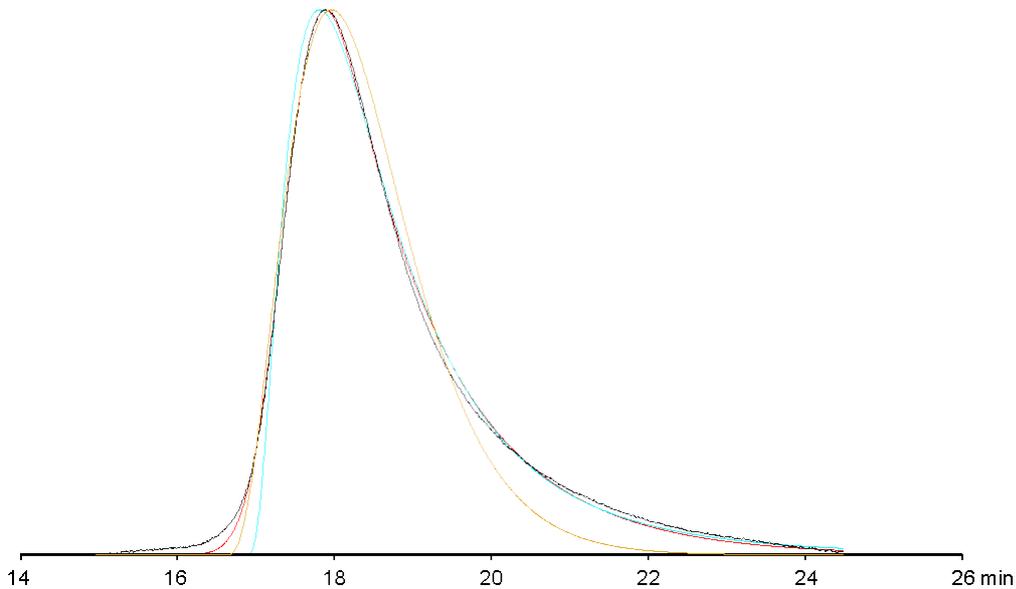


Abb. 3: Vergleich der Anpassung des Einzelpeaks von Cytochrom C (Pferd) mit verschiedenen Modellen (rot: exp. mod. Gauss-Modell, grün: Fraser-Suzuki-Modell, orange: pPoisson-Modell)

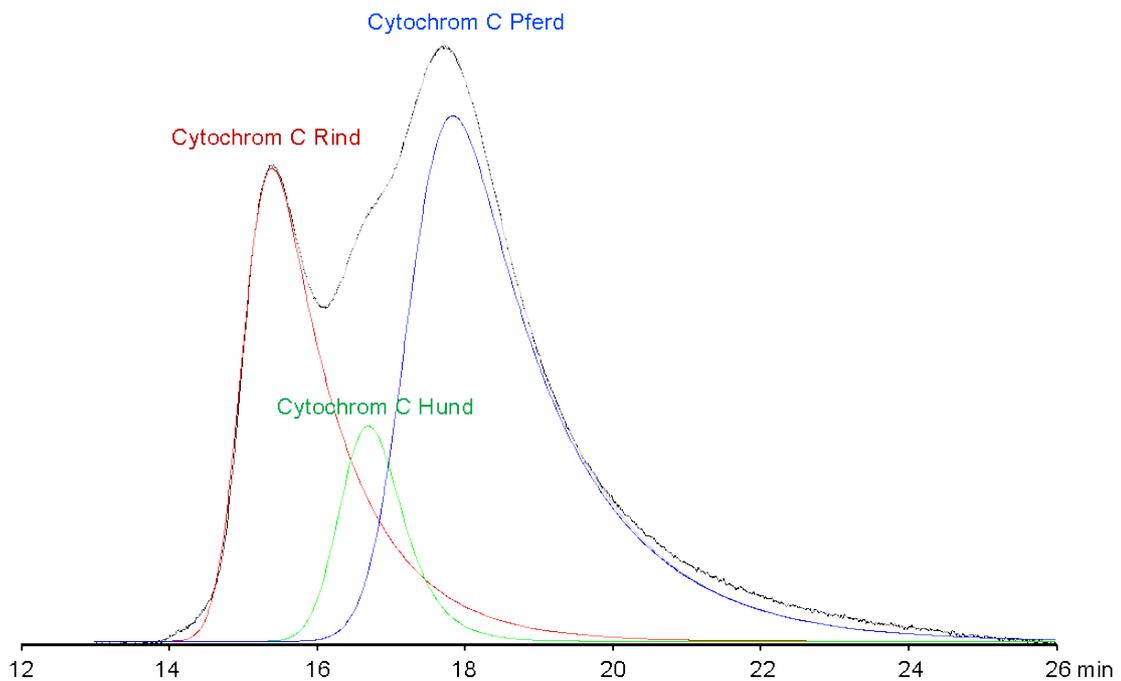


Abb. 4: Dekonvolution einer IEC-Peakgruppe mit Vorgaben aus Einzelkomponentenanalysen (3 x exp. mod. Gauss-Modell, Asymmetrie-Werte),  $R_{\text{Sum,rel}} = 3,0 \%$  (schwarz: Originaldaten, grau: Summe iterierte Peaks).

### Beispiel 3: Trennung von Peptiden mittels RP-HPLC

Ein Problem bei der Dekonvolution tritt dann auf, wenn die einzelnen Analyte in der Mischung nicht bekannt sind. Am Beispiel der Peptidtrennung nach einem tryptischen Verdau von Hämoglobin mittels Reversed-Phase-HPLC (RP-HPLC) lässt sich diese Problematik verdeutlichen. In diesem Fall können aus singulären Peaks vor bzw. nach der Peakgruppe akzeptable Vorgaben abgeleitet werden, eine validierte Aussage mittels Dekonvolution ist hier schwierig.

Bei Vorgabe von 2 Peaks war der Fehler deutlich ( $R_{\text{Sum,rel}} = 7,9\%$ ) und die Differenz an der rechten Flanke (Abbildung 5a) deutete an, dass ein weiterer Peak hier vorhanden ist. Dies wurde durch die Anpassung mit 3 Peaks mit kleinerer rel. Residuensumme ( $R_{\text{Sum,rel}} = 2,5\%$ ) bestätigt (Abbildung 5b).

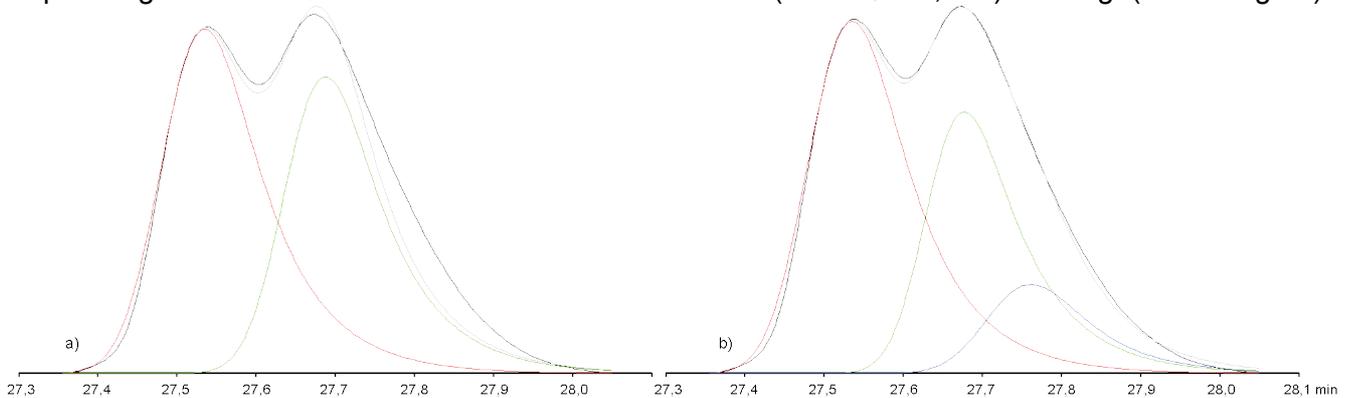


Abb.5: Dekonvolution einer RPC-Peakgruppe mit Vorgaben aus Peaks vor, nach Peakgruppe  
a) 2 Peaks,  $R_{\text{Sum,rel}} = 7,9\%$ , b) 3 Peaks,  $R_{\text{Sum,rel}} = 2,5\%$  (exp. mod. Gauss-Modell, Asymmetrie-Werte; schwarz: Originaldaten, grau: Summe iterierte Peaks).

### Zusammenfassung

Wie an den Beispielen der GPC-, IEC- und RPC-Trennung gezeigt werden konnte, ist die Dekonvolution mit Hilfe einer Peakform-Analyse ein gutes Werkzeug, um die chromatografische Auflösung zu verbessern, wenn dies durch technische Optimierung des Trennverfahrens nicht zu erreichen ist. Zu beachten ist allerdings, dass starke Peak-Asymmetrien oft nicht perfekt modellierbar sind und dies bei der Beurteilung des Dekonvolutions-Resultats berücksichtigt werden muss. Kritisch ist, wenn bei komplexen Peakgruppen keine Peakparameter-Vorgaben möglich sind, hier ist eine Validierung der Dekonvolution nicht möglich. Es konnte jedoch gezeigt werden, dass die konventionelle Auswertung (Lot-Fällung, Tal/Tal-Integration) nahezu immer schlechtere Resultate erbringt [8]. Abschließend bleibt festzuhalten, dass durch die Analyse der Peakform und durch die Dekonvolution bei bekannten überlappenden Analyten die Auflösung eines chromatografischen Trennverfahrens entscheidend verbessert werden kann. Trotzdem ist die Peakform-Analyse keine generelle Problemlösung in der HPLC und bedarf immer einer besonderen Validierung.

### Literatur

- [1] Kalambet, Y., et al. Reconstruction of chromatographic peaks using the exponentially modified Gaussian function. *Journal of Chemometrics*, 25/7, 352-356, 2011
- [2] Reh, E. An algorithm for peak-shape analysis for differentiating unresolved peaks in chromatography. *TrAC Trends in Analytical Chemistry*, 12/5, 192-194, 1993
- [3] Fraser, R. DB; SUZUKI, E. Resolution of overlapping bands. Functions for simulating band shapes. *Analytical chemistry*, 41/1, 37-39, 1969
- [4] Nikitas, P.; Pappa-Lousi, A.; Papageorgiou, A. On the equations describing chromatographic peaks and the problem of the deconvolution of overlapped peaks. *Journal of Chromatography A*, 912/1, 13-29, 2001
- [5] Marquadt, D. W. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the society for Industrial and Applied Mathematics*, 11/2, 431-441, 1963
- [6] [www.chemometrie.info](http://www.chemometrie.info)
- [7] Reh, E. *Chemometrie: Grundlagen der Statistik, Numerischen Mathematik und Software Anwendungen in der Chemie*. Walter de Gruyter GmbH & Co KG, 2017
- [8] Reh, E.; Vetter, A. Peakform-Analyse in der Chromatographie, ein Weg zu verbesserter Trennleistung. *GIT SPEZIAL CHROMATOGRAPHIE*, 72-77, 1997

## Anhang

Es gilt  $t_0$ : Peak-Position;  $w$ : Peakbreite;  $H$ : Höhe;  $A$ : Fläche;  $\tau$ ,  $\sigma$ ,  $\beta$ ,  $\mu$ ,  $b$ : Asymmetrie-Parameter und  $z = [(t-t_0)/w] - [w/\tau]$

### Exponentiell modifizierte Gaussfunktion

$$E(t) = \frac{A}{\tau} e^{0,5 \left(\frac{w}{\tau}\right)^2 - \frac{t-t_0}{\tau}} \int_{-\infty}^{z/\sqrt{2}} e^{-z^2} dz$$

### Poisson-Funktion

$$P(t) = H \frac{e^\mu}{\mu^\mu} \left( \frac{t-t_0}{w} + \mu \right)^\mu e^{-\left(\frac{t-t_0}{w} + \mu\right)}$$

### Fraser-Suzuki-Funktion

$$F(t) = H e^{-\frac{\ln 2}{\beta^2} \left[ \ln \left( 1 + 2\beta \frac{t-t_0}{w} \right) \right]^2}$$

### Generalisierte exponentiell modifizierte Gaußfunktion

$$H(t) = \frac{A}{1+b} \left( e^{q_1 \frac{I_1}{\tau_1}} + e^{q_2 \frac{I_2}{\tau_2}} \right)$$

$$\text{mit } q_i = \frac{\sigma^2}{2\tau_i^2} - \frac{t-t_0}{\tau_i} \quad I_i = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_i} e^{-x^2/2} dx \quad z_i = \frac{t-t_0}{\sigma} - \frac{\sigma}{\tau_i}$$

### Relative Residuensumme $RSum_{rel}$

$$RSum_{rel} = 100 \frac{\left| \sum_{i=1}^n y_i - \sum_{j=1}^k \sum_{i=1}^n \hat{y}_{ij} \right|}{\sum_{i=1}^n y_i} \quad [\%]$$

mit  $n$ : Zahl Messwerte,  $k$ : Zahl Komponenten,  
 $y_i$ : Messpunkt  $i$ ,  $\hat{y}_{ij}$ : modellierter  $y$ -Wert Messpunkt  $i$  bei Komponente  $j$