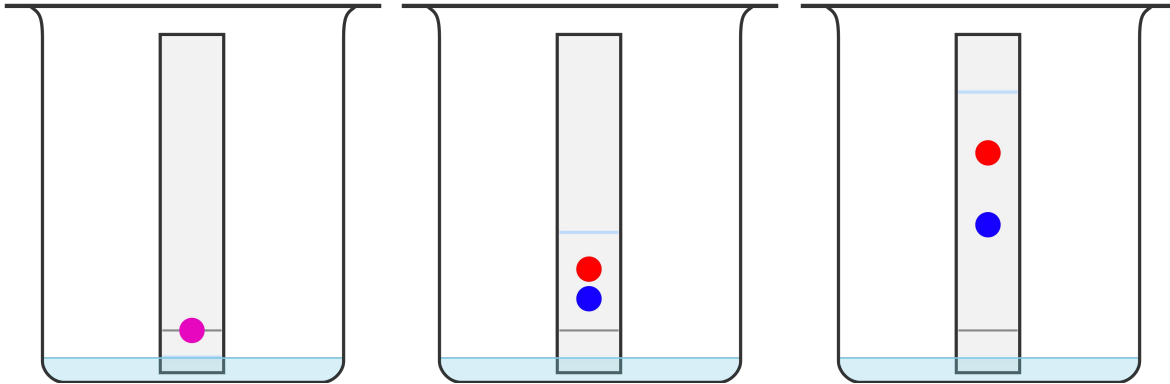


# Kémiai vegyületek kromatográfiás elválasztásának sztochasztikus elmélete és implementálása **R** alatt

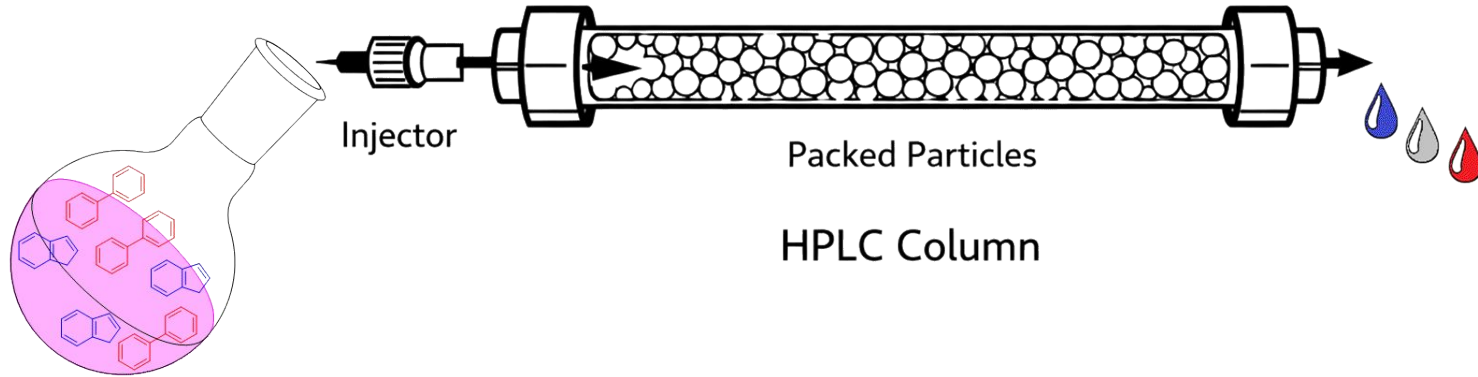
Mirzahosseini Arash - Semmelweis Egyetem GYKI

# Kémiai elválasztás – kromatográfia

- 1950-es években indult
- Papír(vékonyréteg)kromatográfia
- Növényi pigmentek elválasztására

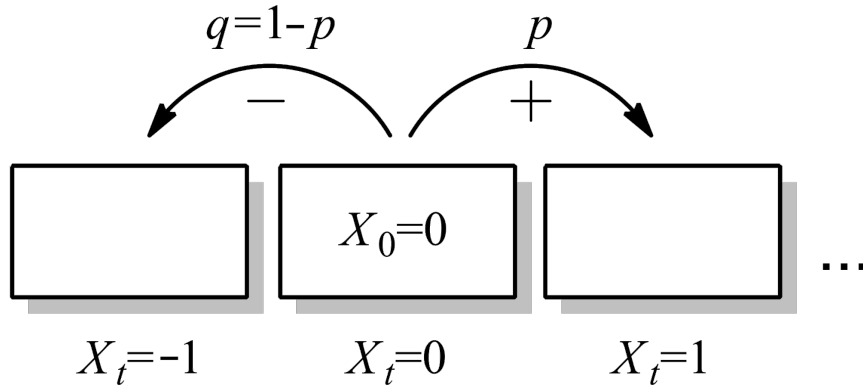


# Folyadék(oszlop)kromatográfia – HPLC



- Mozgófázis (folyadék) pumpával áramoltatva
- Állófázis oszlopban (általában szilika szemcsék)

# Kromatográfia mint sztochasztikus folyamat



Binomiális eloszlású folyamatjelzők

$$\mathbb{P}(X_r = l) = \binom{r}{n_+} p^{n_+} q^{n_-} = \binom{r}{\frac{r+l}{2}} p^{\frac{r+l}{2}} q^{\frac{r-l}{2}}$$

$$\mathbb{P}(T_L = r) = \frac{L}{r} \binom{r}{\frac{r+L}{2}} p^{\frac{r+L}{2}} q^{\frac{r-L}{2}}$$

*First-passage/hitting-time eloszlás*

*Azonban ez csak üres oszlopra érvényes*

- **1D véletlen bolyongás** (random walk on  $\mathbb{Z}$  with absorbing boundary at  $L$ )
- Ha  $p = q = 0.5 \Rightarrow$  diffúzió
- Ha  $p > q \neq 0 \Rightarrow$  diffúzió + drift  
(diszperzió) (advekción)

# 1D véletlen bolyongás – levezetés

## (1) Kombinatorikai levezetés

A jobbra és balra lépések után helyzetünk  $l$ -nél van, ehhez  $r + l$  páros számúnak kell lennie

$$n_+ + n_- = r$$

$$n_+ - n_- = l$$

—————→  
Egyenletrendszer megoldása

Ebből következik a binomiális eloszlás

$$n_+ = \frac{r+l}{2}$$

$$n_- = \frac{r-l}{2}$$

## (2) Levezetés karakterisztikus függvénnyel

Az 1- és  $r$ -lépéses karakterisztikus függvény:

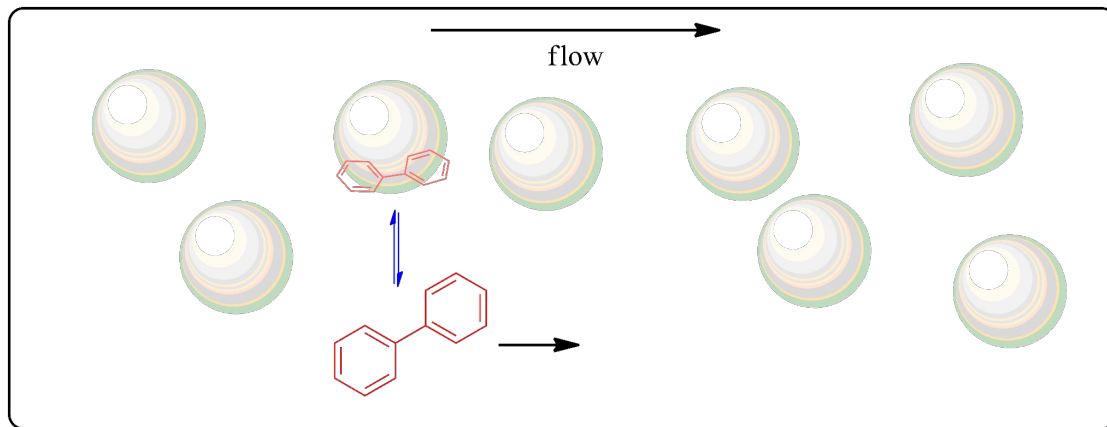
$$\phi_1(k) = \mathbb{E}[e^{i k \Delta X}] = p e^{i k} + q^{-i k} \quad \phi_r(k) = (\phi_1(k))^r = (p e^{i k} + q^{-i k})^r$$

Az inverz Fourier tétel és az alábbi azonosságból levezethető a keresett binomiális alap

$$\mathbb{P}(X_r = l) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-i k l} (p e^{i k} + q^{-i k})^r dk$$

$$(p e^{i k} + q^{-i k})^r = \sum_{m=0}^r (p e^{i k})^m (q e^{-i k})^{r-m}$$

# Összetett Poisson-folyamat



- **Adszorpció:**  $\text{Pois}(\lambda_M)$
- **Deszorpció:**  $\text{Pois}(\lambda_S)$
- Szabad állapotban  $v$  sebességgel drift
- Diffúziótól eltekintünk most

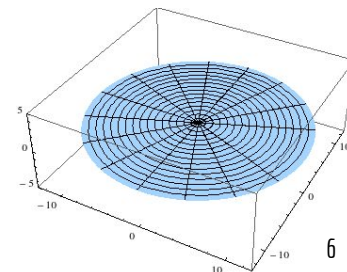
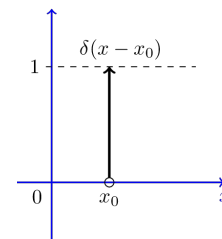
elsőrendű reakciókinetikát és független eseményeket feltételezünk (állandó hazard)

First-passage eloszlás

Giddings & Eyring (1955)

$$f_L(t) = e^{-\lambda_M t_0} \delta(t - t_0) + e^{-\lambda_M t_0 - \lambda_S (t-t_0)} \frac{\lambda_S \lambda_M t_0}{\sqrt{\lambda_S \lambda_M t_0 (t-t_0)}} I_1(2 \sqrt{\lambda_S \lambda_M t_0 (t-t_0)})$$

$$t_0 = \frac{L}{v} \text{ oszlop hossza holtidő}$$



# Összetett Poisson-folyamat – levezetés

1. Mivel  $t_0$  időt tölt egy részecske a mozgófázisban, az adszorpciók számának várható értéke:  $\lambda_M t_0$

2.  $t_0$  idő alatt  $r$  adszorpció előfordulásának valószínűsége:

$$\frac{(\lambda_M t_0)^r e^{-\lambda_M t_0}}{r!}$$

3.  $r$  adszorpciót  $r$  deszorpció kell, hogy kövessen (mivel a részecskék a mozgófázisban indulnak).

$r - 1$  deszorpció valószínűsége  $t$  idő alatt:

$$\frac{(\lambda_S t)^{r-1} e^{-\lambda_S t}}{(r-1)!}$$

4. Pontosán 1 deszorpció előfordulásának valószínűsége  $t$  és  $t + dt$  időintervallumban:  $\lambda_S dt$

5. A 2.-4. tagok szorzata megadja annak valószínűségét, hogy a részecske  $t$  és  $t + dt$  időintervallumban ér az oszlop végéhez:

$$f_{L|r}(t)$$

6. A first-passage eloszlás:

$$f_L(t) = \sum_{r=0}^{\infty} f_{L|r}(t)$$

7. Felismerve az elsőfajú  $n$ -ed rendű módosított Bessel függvény sorfejtését eljutunk a keresett kifejezéshez:

$$\frac{I_n(2\sqrt{z})}{\sqrt{z}} = \sum_{m=0}^{\infty} \frac{z^{m+\frac{n-1}{2}}}{m!(m+n)!}$$

8. A fenti eloszlás csak az adszorbeált állapotban történő várakozási időket veszi figyelembe, ezért kell a végleges formában alkalmazni egy  $t_0$  időeltolás.

# Felírás karakterisztikus függvénnyel

Legyen az adszorpció/deszorpció események száma  $N$ , köztük  $Y_j$  idejű ugrásokkal:

$$N \sim \text{Pois}(\lambda_M t_0) \quad X_t = \sum_{j=1}^N Y_j$$

Az események függetlensége feltétele miatt az  $L$  eléréséig eltelt idő karakterisztikus függvény:

$$\phi_L(k) = e^{i k t_0} \sum_{n=0}^{\infty} \mathbb{P}(N = n) [\phi_Y(k)]^n$$

Behelyettesítve a poisson eloszlás súlyfüggvényét:

$$\phi_L(k) = e^{i k t_0} e^{-\lambda_M t_0} \sum_{n=0}^{\infty} \frac{(\lambda_M t_0)^n}{n!} [\phi_Y(k)]^n$$

Átírjuk a kifejezést az exponenciális forma sorfejtésének segítségével:

$$\sum_{n=0}^{\infty} \frac{a^n}{n!} = e^a$$
$$\phi_L(k) = e^{i k t_0 + \lambda_M t_0 (\phi_Y(k) - 1)}$$

McQuarrie (1963-1967)

Dondi, Cavazzini, Felinger (1990-)

Végük az exponenciális eloszlású ugrásokat behelyettesítve kapjuk az alábbi alakot:

$$Y \sim \text{Exp}(\lambda_S) \quad \phi_Y(k) = \frac{1}{1 - i k / \lambda_S}$$

$$\phi_L(k) = e^{i k t_0 + \lambda_M t_0 \left( \frac{1}{1 - i k / \lambda_S} - 1 \right)}$$

# Felírás karakterisztikus függvénnnyel

Betesszük végül a diffúzió hatását is a karakterisztikus függvénybe. Azonban nem egyszerű konvolúcióval, mivel a diffúzió és advekció nem független folyamatok: diffúzió csak deszorpció állapotban valósul meg. Tehát be kell ágyazni a két folyamatot egymásba.

Ún. subordination módszerrel ha egy folyamat  $R = M(U(t))$  alakban írható fel, akkor azt mondjuk, hogy az  $M(t)$  folyamat (itt a teljes diffúzió + drift) akkor valósul meg amikor az  $U(t)$  egy kitüntetett állapotában vagyunk (deszorpció). A karakterisztikus függvényekkel az alábbi alakban írhatók fel:

$$\phi_{M(U)}(\mathbf{k}) = \phi_U(-i \log(\phi_M(\mathbf{k})))$$

Az alábbi 2 kifejezést használva kapjuk végül a teljes karakterisztikus függvényt:

$$\phi_U(\mathbf{k}) = e^{\frac{vL}{2D} \left(1 - \sqrt{1 - \frac{4Dik}{v^2}}\right)}$$

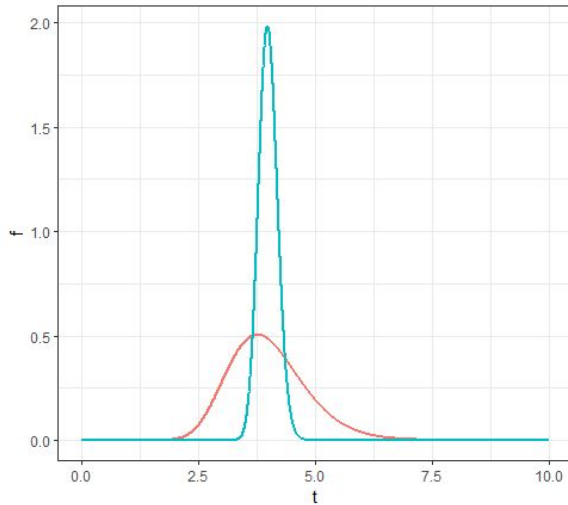
Einstein képlete a diffúzió + drift folyamatra  
inverz Gauss eloszlással felírva, ahol  $D$  a diffúziós állandó

$$\log(\phi_M(\mathbf{k})) = i\mathbf{k} + \lambda_M \left( \frac{1}{1 - i\mathbf{k}/\lambda_S} - 1 \right)$$

Az előző diában levezetett karakterisztikus függvény  
 $t_0$ -ra normalizálva

$$\phi_L(\mathbf{k}) = e^{\frac{vL}{2D} \left(1 - \sqrt{1 - \frac{4Dik t_0}{vL} \left(1 + \frac{\lambda_M/\lambda_S}{1 - i\mathbf{k}/\lambda_S}\right)}\right)}$$

# Implementálás R alatt



```
cf_1site_t0_disp <- function(omega, pars) {  
  
  with(pars, {  
  
    t0 <- L / v  
  
    exp(  
      v * L / (2 * D) * (1 - sqrt(  
        1 - 4 * D * 1i * omega * t0 / (v * L) *  
        (1 + (lambda_M / lambda_S) /  
        (1 - 1i * omega / lambda_S))  
      ))  
    )  
  
  })  
}
```

# ...megegyezik a makroszkopikus modellel

Statisztikai mechanikában a **Fokker-Planck** egyenlet írja le ezt a folyamatot, amelynek több változata is ismert: *Langevin, Einstein-Szmolucsovskij, Klein-Kramers, Navier-Stokes, Black-Scholes*

$$\frac{\partial}{\partial t} f_X(\mathbf{x}, t) = -\frac{\partial}{\partial x} [\mu(\mathbf{x}, t) f_X(\mathbf{x}, t)] + \frac{\partial^2}{\partial x^2} [D(\mathbf{x}, t) f_X(\mathbf{x}, t)]$$

Kromatográfiában az alábbi egyenleteket kapjuk (ún. Lumped kinetic model):

$$\frac{\partial}{\partial t} c(x, t) + F \frac{\partial}{\partial t} q(x, t) + v \frac{\partial}{\partial x} c(x, t) = D \frac{\partial^2}{\partial x^2} c(x, t)$$

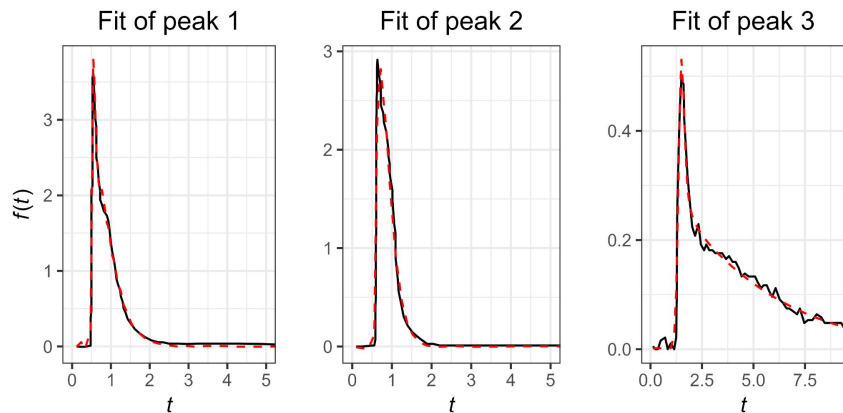
$$\frac{\partial}{\partial t} q(x, t) = \lambda_M c(x, t) - \lambda_S q(x, t)$$

...mint Lévy-folyamat

$$X_t = b t + \sigma B_t + \sum_{n=1}^{N_t} J_n$$

$$\phi_{X_t}(u) = e^{t \left( i b u - \frac{1}{2} \sigma^2 u^2 + \lambda_M (\phi_J(u) - 1) \right)}$$

# Lévy-Hincsin reprezentáció R-ben



```
# Gamma jump distribution
phiJ_gamma <- function(u, shape, rate) {
  exp(shape * (log(rate) - log(rate - 1i * u)))
}

# Compound Poisson + Brownian + drift using Lévy-Khintchine formulation
compPois <- function(t, b, sigma2,
  lambda,
  jump_shape, jump_rate,
  Umax = 150, N = 2^13) {

  phi_levy <- function(u, t, b, sigma2, lambda, phiJ) {
    # phiJ(u): characteristic function of jump size J
    psi_drift <- 1i * b * u
    psi_brownian <- -0.5 * sigma2 * u^2
    psi_compPois <- lambda * (phiJ(u) - 1)
    psi <- psi_drift + psi_brownian + psi_compPois
    exp(t * psi)
  }

  phi_levy_compPois <- function(u, t) phi_levy(
    u, t, b = b, sigma2 = sigma2, lambda = lambda,
    phiJ = function(uu) phiJ_gamma(uu, shape = jump_shape, rate = jump_rate)
  )

  return(pdf_from_cf_fft(phi = phi_levy_compPois, t, Umax, N))
}
```

# ...mint Markov-lánc folyamat

$$M \xrightarrow{\lambda_M} S \text{ (mobility = 0)}$$

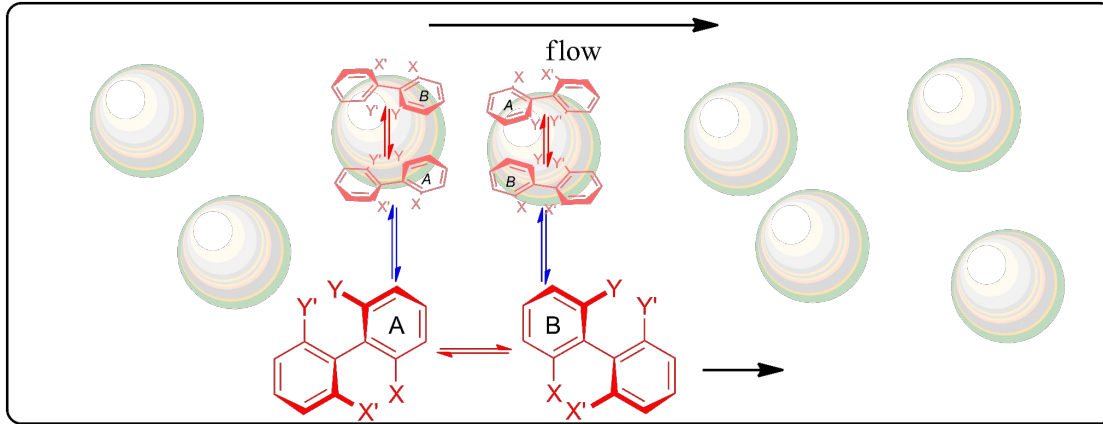
$$S \xrightarrow{\lambda_S} M \text{ (mobility = } \mathcal{N}(vt, 2Dt))$$

[animáció](#)

$$\frac{d}{dt} \hat{\mathbf{p}}(k, t) = A(k) \hat{\mathbf{p}}(k, t)$$

$$A(k) = \begin{pmatrix} -\lambda_M - i k v - D k^2 & \lambda_S \\ \lambda_M & -\lambda_S \end{pmatrix}$$

# Kromatográfia + interkonverzió = **Batman**



Keller & Giddings (1960)

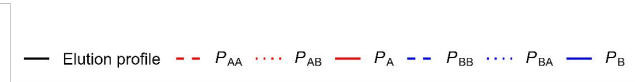
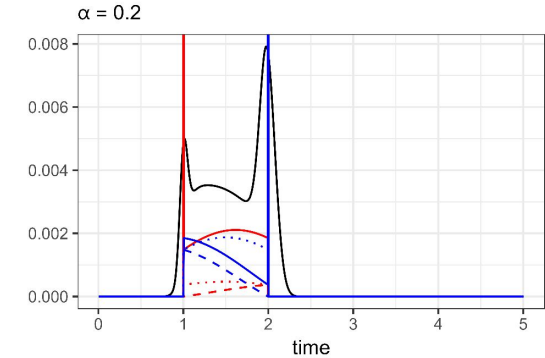
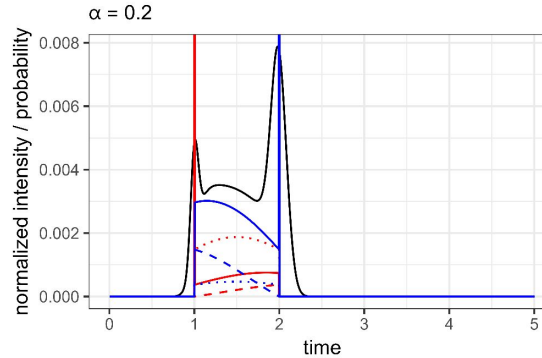
- **Adszorpció:**  $\text{Pois}(\lambda_{M,A})$  és  $\text{Pois}(\lambda_{M,B})$
- **Deszorpció:**  $\text{Pois}(\lambda_{S,A})$  és  $\text{Pois}(\lambda_{S,B})$
- Szabad állapotban  $v$  sebességgel drift
- **Interkonverzió:**  $\text{Pois}(\lambda_f^{\text{stat}})$  és  $\text{Pois}(\lambda_r^{\text{stat}})$  az állófázisban, valamint egységes  $\text{Pois}(\lambda^{\text{mob}})$  a mozgófázisban

# Giddings eloszlás

$$p_1(x) = a e^{-a(1-x)-bx} I_0(2\sqrt{abx(1-x)})$$

$$p_2(x) = \sqrt{\frac{ab(1-x)}{x}} e^{-a(1-x)-bx} I_1(2\sqrt{abx(1-x)})$$

$$p_3(x=0) = e^{-a}$$



Hasonló megfontolásokból, amelyekkel az összetett Poisson-folyamatot levezettük, itt is hasonlóan járunk el.

Vesszük az alábbi 3 valószínűséget, ahol egy részecske útja során az oszlopban jelölje  $x$  azt az időhányadot amikor a részecske a 'B' állapotban van:

1. Egy részecske 'A' állapotban indult és 'B' állapotban ért célba.
2. Egy részecske 'A' állapotban indult és 'A' állapotban ért célba, miközben legalább 2 interkonverzió átesett.
3. Egy részecske 'A' állapotban indult és 'A' állapotban ért célba, miközben nem esett át interkonverzió.

A fenti 3 valószínűség a feltüntetett képletekkel írható fel, ahol  $a = \lambda_f^{app} T_L$  és  $b = \lambda_r^{app} T_L$  a várható oda-vissza interkonverziók száma. A kiindulási 'B' állapotra azonos módon felírható a 3 valószínűség, csak  $a$  és  $b$ , valamint  $x$  és  $1-x$  szerepét kell megcserélni.

Végül a kiindulási 'A' és 'B' formák valószínűségeit összegezni kell (súlyozva a kiindulási arányaikkal) és az interkonverzió nélküli kromatogramokkal vett konvolúció adja meg a Batman csúcsokat.

# Automatizált kromatogram elemzés R-ben

## Adatok beolvasása

- Különböző műszerekből származnak adatok.
- Egységesítés: CSV.
- Automatikus encoding felismerés és **beolvasás**.

## Preprocessing

- Kromatogramok zajszűrése (loess), alapvonalai korrekciója (cubic poly).

## Kezdő paraméterek

- Csúcsok automatikus felismerése (`pracma::findpeaks`) és első **2 momentum** becslése.

## 2 lépéses optimalálás

- Sztochasztikus modell szimuláció.
- *Differential evolution* (`DEoptim`), lassú de stabilan konvergál.
- *Levenberg-Marquadt* warm-start-ból még közelebb húzza az illesztést.

## Export

- Paraméterek és görbék exportja **CSV** és **PNG** formátumba.
- 400 kromatogram (több ezer adatpont / mérés) ~ 12 óra.

# Látszólagos vs fázis-specifikus interkonverzió

Az egyes kromatogramokból látszólagos interkonverziós rátát kapunk, amely a valódi rátákból tevődik össze:

$$\lambda_f^{\text{app}} = \frac{1}{1+k_A^{\text{fact}}} \lambda^{\text{mob}} + \frac{k_A^{\text{fact}}}{1+k_A^{\text{fact}}} \lambda_f^{\text{stat}}$$

$$\lambda_r^{\text{app}} = \frac{1}{1+k_B^{\text{fact}}} \lambda^{\text{mob}} + \frac{k_B^{\text{fact}}}{1+k_B^{\text{fact}}} \lambda_r^{\text{stat}}$$

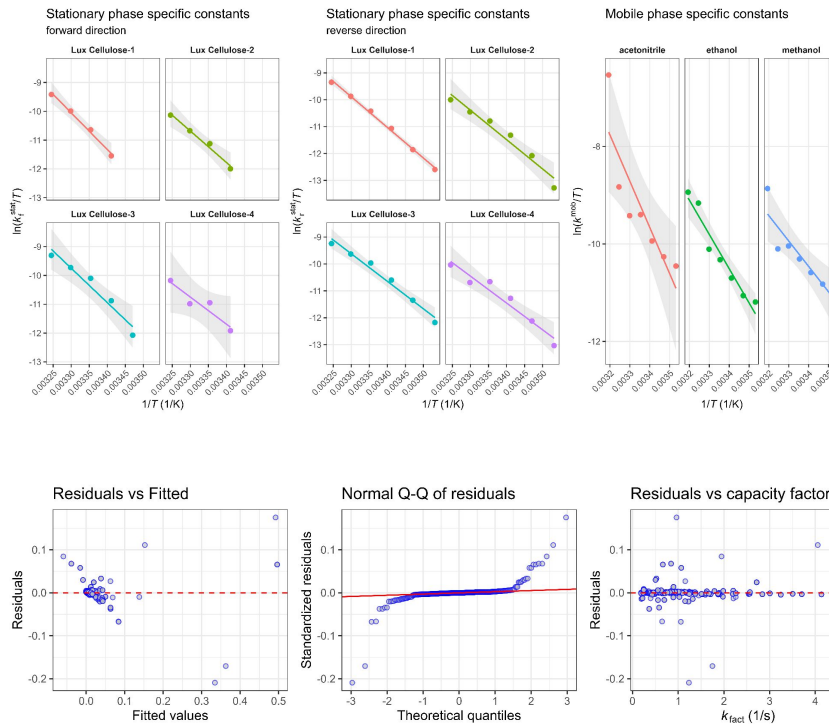
$$k^{\text{fact}} = \frac{T_L - t_0}{t_0}$$

Szükséges kísérleti elrendezés:

- Több **mozgófázis**: acetonitril, etanol, metanol
- Több **állófázis**: 4 fajta cellulóz
  - Feltételezzük, hogy a mozgófázis- és állófázis-specifikus ráták között nincs interakció
- Több **áramlási sebesség**
  - Segít az interkonverziós ráták pontosabb becslésében
- Több **hőmérséklet**
  - Az interkonverziós ráták hőmérsékletfüggése jobban látszik

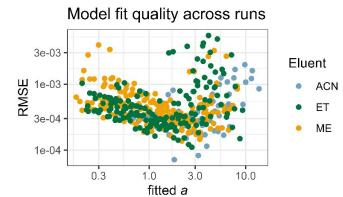
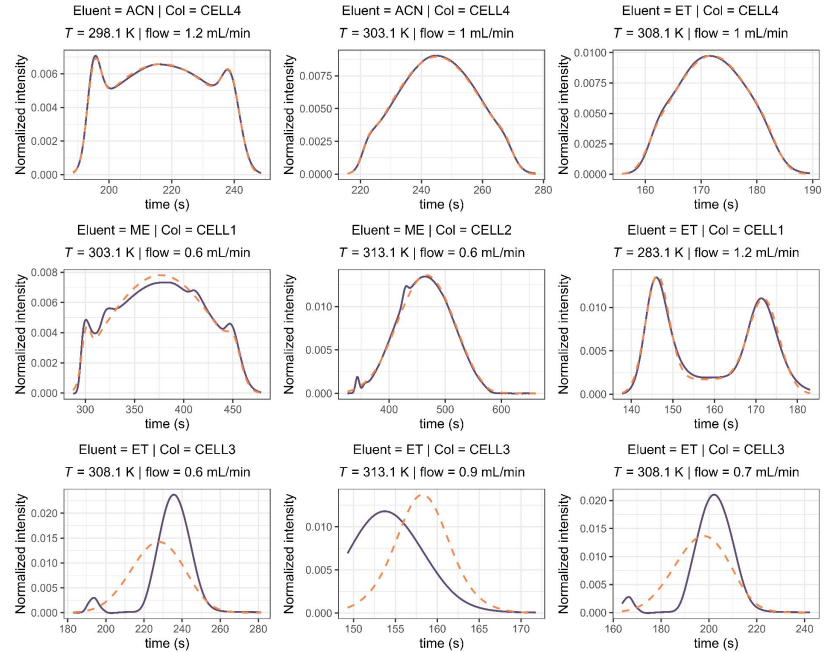
# Kevert modell

```
fit_lmer <- lmer(  
  response_tr ~ k_fact + k_fact:direction +  
    (1 | eluent:Temp) +  
    (0 + k_fact:direction | col_name:Temp),  
  data = dat,  
  weights = weight,  
  control = lmerControl(  
    check.conv.singular = "ignore",  
    optimizer = "bobyqa"  
  )  
)
```

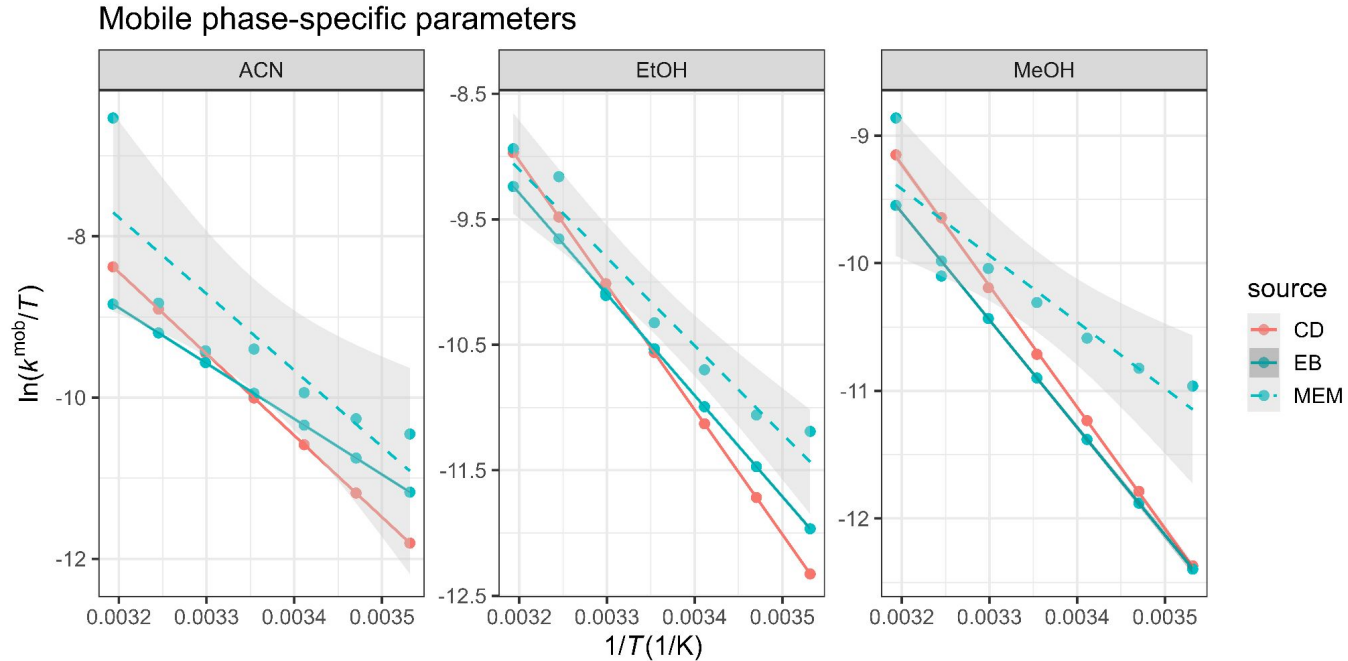


# Bayesi modell

```
neg_log_post <- function(par) {  
  pars <- unpack_par(par, C, E)  
  sse_vec <- future_sapply(1:N, function(i) {  
    yhat <- predict_run(i, pars, alpha = 0.5)  
    if (any(!is.finite(yhat))) {  
      message("Non-finite yhat at run i = ", i)  
      return(1e100)  
    }  
    r <- y_mat[i, ] - yhat  
    if (any(!is.finite(r))) {  
      message("Non-finite residual at run i = ", i)  
      return(1e100)  
    }  
    sum(r * r)  
  })  
  data_term <- inv_2sig2 * sum(sse_vec)  
  prior_term <- prior_penalty(pars)  
  val <- data_term + prior_term  
  if (!is.finite(val)) 1e100 else val  
}
```



# Validáció független kísérletekkel



# Köszönetnyilvánítás

ÁTE- Biostatisztika Tanszék

Lang Zsolt

Semmelweis Egyetem - GYKI

Ali Mhammad

Dombi Gergely

Pálla Tamás

Tóth Gergő

Semmelweis Egyetem -SZVI

Balogh Balázs

Pécsi Tudományegyetem - Bioanalitikai Intézet

Sepsey Annamária

Prof. Felinger Attila

Egis Gyógyszergyár

Horváth Simon

Tóth Eliza

Németh Gábor

LMU München - Department of Chemistry

Prof. Oliver Trapp

## Köszönöm szépen a figyelmet!

# Irodalomjegyzék

Giddings & Eyring (1955):

<https://doi.org/10.1021/j150527a009>

Keller & Giddings (1960):

[https://doi.org/10.1016/S0021-9673\(01\)96983-3](https://doi.org/10.1016/S0021-9673(01)96983-3)

Characteristic function treatment:

<https://doi.org/10.1063/1.1733677>

Stochastic-dispersive theory:

<https://doi.org/10.1021/ac990412u>

Multisite stochastic theory:

<https://doi.org/10.1021/ac990282p>

Review of stochastic theory of chromatography:

<https://doi.org/10.1016/j.chroma.2007.12.066>

**R** workflow:

<https://doi.org/10.1016/j.chroma.2025.466558>