

Stanovení poměru koncentrací metylovaného a nemetylovaného cytosinu v klinickém vzorku

Úloha do laboratorního cvičení - Metody chemického výzkumu

pro magisterské studium odborné chemie a biochemie, podzimní semestr 2009,

vyučující: Ing. B. Hégrová

1 Cíl úlohy

Ověřit si při řešení praktického úkolu vlastnosti a možnosti kapalinové chromatografie.

1. Příprava kapalinového chromatografu pro měření reálných vzorků
2. Analýza připravených vzorků buněčných linií ovlivněných a neovlivněných 5-aza-2'-deoxycytidinem
3. Identifikace zájmových složek vzorku pomocí retenčních časů standardů a pomocí knihovny UV-Vis spekter.
4. Využitím metody standardního přídatku stanovit poměr mdCMP/dCMP (5-metyl-2'-deoxycytosin-5'-monofosfát a nemetylovaný analog)
5. Na základě získaných poměrů diskutovat vliv 5-aza-2'-deoxycytidinu na stav metylace DNA buněčné linie

2 Teorie

2.1 Kapalinová chromatografie

V kapalinové chromatografii se k separaci složek využívá mnohonásobného opakovaného vytváření rovnovážných stavů mezi stacionární fází a kapalnou mobilní fází na základě různých chemicko-fyzikálních interakcí (velikost částice, polarita, náboj, aj.). V současné době si vysokoúčinná kapalinová chromatografie (HPLC) stále udržuje svůj význam – umožňuje analyzovat prakticky veškeré organické látky s vysokou citlivostí a v rozpětí relativních molekulových hmotností od stovek a.m.u. u iontů až po několik set tisíc u makromolekul až částic. To je umožněno možností volby obou složek separačního systému – mobilní a stacionární fáze. HPLC se oproti jiným separačním technikám vyznačuje robustností (např. opakovatelné retenční časy oproti kapilární elektroforéze) a možností analýzy tepelně nestálých vzorků za laboratorní teploty (např. oproti plynové chromatografii). Nevýhodou HPLC oproti těmto metodám může být nižší účinnost separace (nižší počet teoretických pater) a vyšší spotřeby činidel a vzorků. Objevují se však postupy jako ultravýkonná kapalinová chromatografie (UPLC) a monolitické stacionární fáze, které otevírají kapalinové chromatografii nové možnosti. Nejrozšířenější používanou

kapalinovou chromatografií je chromatografie na obrácených fázích (RPLC). V RPLC se používá nepolární stacionární fáze a mobilní fáze v definovaném stupni polarity.

2.2 Základní schéma kapalinového chromatografu

Požadavky na chromatografické zařízení:

a) umožnit pravidelný a konstantní tok mobilní fáze stacionární fází \Rightarrow *opakovatelnost retenčních časů*,

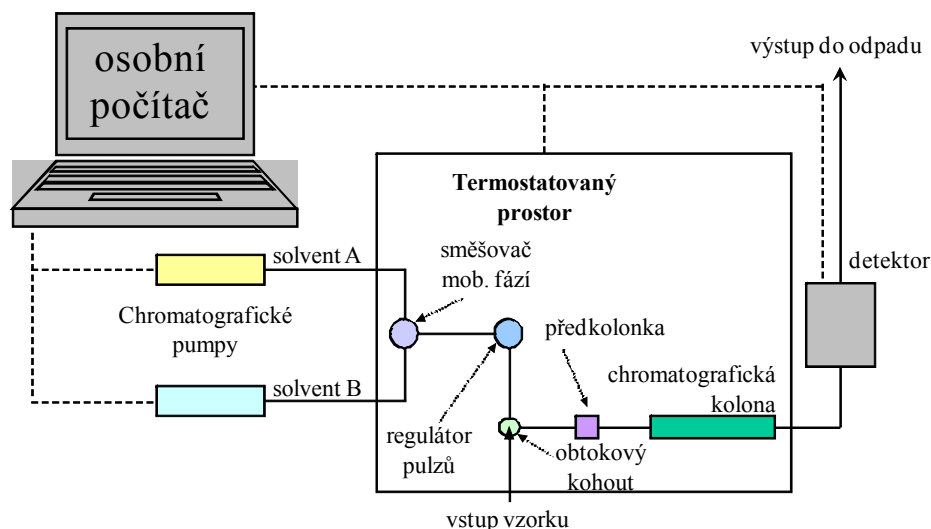
b) zajistit kvantitativní a časově co nejkratší nadávkování analyzované směsi na stacionární fázi \Rightarrow *opakovatelnost kvantitativních i kvalitativních parametrů signálů analytů (výška, plocha, šířka, symetrie, ...)*,

c) vytvořit podmínky k co největšímu počtu opakování sorpčních a desorpčních procesů jako výsledku vzájemných vztahů mezi stacionární fází, molekulami složek mobilní fáze a molekulami solutů v analyzované směsi \Rightarrow *co nejvyšší účinnost kolony a tím rozlišení sledovaných analytů*,

d) uskutečnit průběžné měření a zaznamenávání změn určitých vlastností mobilní fáze vytékající z kolony, které jsou výsledkem přítomnosti separovaných látek \Rightarrow *záznam chromatogramu jako indikace přítomnosti analytických zón separovaných látek*,

e) být schopné selektivně uvedené změny analyzovat a kvantifikovat \Rightarrow *vyhodnocení chromatogramu*.

Pro splnění těchto požadavků jsou konstruovány nebo operátorem voleny jednotlivé části chromatografického zařízení. Příklad složení kapalinového chromatografu je ukázán na Obr. 1.



Obr. 1 Schéma kapalinového chromatografu

2.3 Separace s využitím monolitických kolon

Klasické kolony jsou plněny částicemi sorbentu o definovaném průměru a velikosti pórů. Při zkvalitňování separace zmenšováním rozměru částic v klasických náplňových kolonách pro HPLC se objevují negativní jevy. Zrníčka sorbentu kladou prostupující kapalině značný hydrodynamický odpor. Pro zvýšení účinnosti je proto nutné využít vyšších protitlaků mobilní fáze.

Monolity na druhou stranu vytvářejí jednotný souvislý kus materiálu s vysokou pórovitostí. Monolitické materiály používané k plnění chromatografických kolon lze rozdělit na dvě základní skupiny: monolity na bázi siliky (SiO_2) a z organických materiálů (např. polystyren). Silikové monolitické materiály obsahují dva druhy pórů: makropóry a mezopóry. Makropóry (rozměr ~mikrometry) umožňují vysoký průtok mobilní fáze za nízkého tlaku, zatímco hustá síť mezopórů (rozměr ~nanometry) vytváří velký specifický povrch. Struktura organických polymerů se naproti tomu skládá z málo uspořádaných pospojovaných makroglobulí s makropóry mezi nimi.

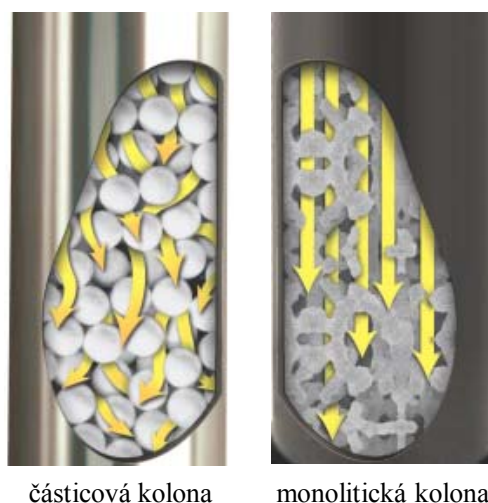
Pórovitost monolitů obecně umožňuje použít vysoké průtoky mobilní fáze za přípustných tlaků. Při vyšší průtokové rychlosti je díky konvekci také rychlejší přenos hmoty. Výsledkem je obdobná separační účinnost monolitických kolon při kratších retenčních časech v porovnání s částicovými kolonami. Nevýhodou monolitických kolon oproti částicovým je z principu přípravy horší reprodukovatelnost retenčních vlastností mezi jednotlivými výrobními sériemi.

2.4 Vliv vybraných vlastností mobilní fáze na průběh separace

Faktory ovlivňující separaci v HPLC je nutno zvážit dle použité stacionární fáze, jinak řečeno dle konkrétního typu chromatografie. V tomto textu bude diskutován pouze případ separace látek na reverzní stacionární fázi (např. ukotvené oktadecylové skupiny C18). Mezi faktory nejvíce ovlivňující separaci patří eluční síla, iontová síla a pH mobilní fáze.

Voda jako nejběžněji používaná složka mobilních fází má nízkou eluční sílu. Eluční sílu zvyšujeme přidávkem organické složky (metanol, aceton), která má eluční sílu vyšší. Např. zvýšením obsahu metanolu o 10 % se sníží kapacitní faktory (k , viz. níže) látek přibližně 2×.

Iontová síla je určena přítomností solí v mobilní fázi a roste s jejich koncentrací. Iontová síla ovlivňuje sílu interakcí separovaných látek se stacionární fází díky



částicová kolona

monolitická kolona

Obr. 2 Zobrazení průtoku mobilní fáze stacionární fází u klasické částicové a monolitické kolony. Převzato a upraveno z brožury ke koloně Onyx Monolithic.

změně solvatačního obalu molekul. S vyšší iontovou silou je solvatační obal molekul tenčí a interakce se stacionární fází se více projevují. Maximální doporučená koncentrace solí je 50 mM.

Koncentrace protonů ovlivňuje ionizovatelné molekuly (např. organické kyseliny a báze). Zde je nutné si uvědomit, že ionizace molekuly vede v případě reverzní stacionární fáze k dramatickému poklesu retence a je proto ve většině případů žádoucí se jí volbou pH vyhnout. Na druhou stranu je možné volbou pH cíleně ionizovat součásti matrice vzorku a zjednodušit si tak výsledný chromatogram. pH mobilní fáze nastavujeme v rozmezí daném většinou pH stabilitou stacionární fáze (např. pH 2-8) použitím vhodného pufru.

2.5 Význam metylace DNA

Metylace DNA (na cytosinu, uhlík C5) patří mezi rozšířené epigenetické procesy u prokaryot i eukaryot. U prokaryot je její funkcí chránit buňku před cizorodou DNA. U eukaryot není funkce metylace DNA zcela známa, ačkoli je již nyní jasné, že zajišťuje přinejmenším dvě základní funkce: kontrolu genové exprese („umlčování“ genů) a ochranu organismu před expresí nežádoucích sekvencí (např. nekódující a parazitické sekvence, repete). Významný je i rozdíl v metylaci normálních a nádorových buněk, kdy u nádorových buněk bývá genom často hypometylován. Důležitou roli v procesu karcinogeneze (tvorba a vývoj nádoru) hraje také metylace CpG oblastí v regulačních oblastech tumor-supresorových genů (geny kódující proteiny se schopností tlumit nádorovou aktivitu). Metylací CpG oblastí těchto genů tak buňka ztrácí jeden z obranných mechanismů proti karcinogenezi (vznik a další vývoj nádorového onemocnění). Metylací cytosinu katalyzují enzymy metyltransferázy (např. EC 2.1.1.37, EC 2.1.1.73). Tyto enzymy používají *S*-adenosylmethionin jako donor methylové skupiny. Jako inhibitor metyltransferáz působí např. 5-aza-2'-deoxycytidin.

3 Praktická část

3.1 Přístroje a zařízení

- HPLC systém 10 AVP fy SHIMADZU:
- odplyňovač GT-154
- systémová řídicí jednotka SCL-10AVP
- pumpa LC-10AVP
- pítka CTO-10ASVP
- PDA detektor SPD-M10AVP
- řídicí software Class-VP 5.02
- 2 monolitické kolony 100x4.6 mm, Onyx C18
- knihovna UV-VIS spekter purinů, pyrimidinů, nukleosidů a nukleotidů
- injekční stříkačka se speciálně upravenou jehlou pro dávkování vzorků (Hamilton)
- mikropipety
- ultrazvuková lázeň
- odměrné baňky a další běžné laboratorní zařízení

3.2 Chemikálie

- standardy bází nukleotidů (koncentrace 10mM):
 - 5-metyl-2'-deoxycytosin-5'-monofosfát (mdCMP)
 - 2'-deoxythymidin-5'-monofosfát (dTMP)
 - 2'-deoxyguanosin-5'-monofosfát (dGMP)
 - 2'-deoxyadenosin-5'-monofosfát (dAMP)
 - 2'-deoxycytosin-5'-monofosfát (dCMP)
- thiomčovina (0,002% roztok ve vodě)
- testovací směs (směs acetonu, benzenu a toluenu v metanolu)
- zpracované vzorky buněčných linií (DNA digesty)
- mobilní fáze pro ekvilibraci a proplach kolony (filtrovaná deionizovaná voda, metanol pro chromatografii)
- mobilní fáze pro měření testovací směsi, vzorků a standardů (vodný roztok fosfátu sodného o koncentraci 50mM a pH 3, metanol pro chromatografii)

3.3 Příprava reálného vzorku (vzorky jsou již připraveny)

Pro jednotlivé vzorky byly použity buňky B-buněčných linií (WSU-NHL, Su-DHL-4, DOHH-2). Tyto buňky byly pěstovány v koncentraci 1×10^7 buněk ve 30 ml média RPMI 1640 (s 10% fetálním bovinním sérem, penicilinem a streptomycinem) při 37°C v atmosféře 5% CO₂. Část těchto buněk byla ošetřena inhibítorem methyltransferáz (1-2 μM 5-aza-2'-deoxycytidin, přidáván každých 24 hod, 3-8 dnů). DNA byla ze vzorku izolována vysolovací technikou (chloroformová extrakce) a dále štěpena endonukleázou *MspI*. Ta štěpí sekvenci CCGG, pokud je vnitřní C v sekvenci methylované i pokud je nemethylované. Část vzorku byla naštěpena enzymem *HpaIII*, který štěpí sekvenci CCGG pouze v případě, že vnitřní C není metylované. Získané fragmenty DNA byly dále purifikovány pro odstranění balastních látek a následně štěpeny Exonukleázou III. Ta enzymaticky štěpí jednotlivé baze ve směru 3'→5'. Na závěr byla naštěpená směs nukleotidů filtrována přes 0,22 μm filtr (odstranění nečistot, které by mohly ucpat póry kolony).

3.4 Obecný postup při práci s kapalinovým chromatografem

3.4.1 Sběr dat

Obsluha řídicího software Class-VP 5.02 a sběr dat se provede dle pokynů vedoucího cvičení. Jako kvantitativní parametr bude odečítána plocha píku. Všechna měření se provádí nejméně třikrát.

3.4.2 Příprava dávkovače a vzorku před nadávkováním vzorku na kolonu

Roztok vzorku se na kolonu dávkuje dávkovacím kohoutem. Před zahájením vlastního měření je nutné propláchnout dávkovací kohout destilovanou vodou (jinak hrozí nebezpečí kontaminace z předchozích vzorků), metanolem nebo mobilní fází. Vzorek dávkovaný na kolonu musí být čirý, částice nebo zákal mohou nevratně znehodnotit kolonu. Vzorky obsahující rozpuštěné plyny (např. metanolicke roztoky) je třeba před nástřikem důkladně odvzdušnit v ultrazvukové lázni.

3.4.3 Naplnění smyčky dávkovače

Při plnění dávkovací smyčky o objemu 20 μl se páčka dávkovače přepne do pohotovostní horní polohy (LOAD). Pro analýzu se pomocí stříkačky bere dostatečné množství vzorku, aby byla naplněna celá dávkovací smyčka (přibližně 25 μl). Do dávkovacího otvoru se jemně zasune jehla injekční stříkačky (nezasouvat násilím na doraz) a tlakem na píst injekční stříkačky se naplní dávkovací smyčka dávkovače (na konci odpadní kapiláry odkápnou přibližně 3-5 kapek). Při

dávkování vzorku do smyčky je nutné kontrolovat přítomnost bublinek. Do systému se nesmí dostat plyn!

3.4.4 Nástřik vzorku na kolonu

V momentě, kdy je software připravený sbírat detektorová data, se páčka dávkovače otočí do pravé spodní polohy. Tím se uvnitř dávkovače přepne směr toku mobilní fáze tak, že na kolonu prochází přes smyčku dávkovacího ventilu, kam jsme před chvílí nadávkovali roztok vzorku. Jehla se poté vytáhne.

3.5 Pracovní postup

3.5.1 Změření testovací směsi

Testovací směs se měří v základní mobilní fázi pro RPLC (70 % MeOH). Měření testovací směsi se optimálně provádí vždy před začátkem a po skončení celého měření. Porovnání píků (retenční čas, rozlišení, počet teoretických pater, asymetrie) testovací směsi se ověřuje, zda nedošlo ke změně vlastností kolony během měření (např. kontaminací kolony těžko eluovatelnými látkami ze vzorku).

3.5.2 Změna pracovní mobilní fáze

Změna mobilní fáze se provede dle instrukcí vedoucího cvičení. Je nutné se řídit obecnými postupy, které zajistí 100% mísitelnost dvou po sobě jdoucích mobilních fází a dostatečnou ekvilibraci kolony mezi dvěma mobilními fázemi. V našem systému se osvědčily následující podmínky pro výměnu mobilní fáze: 1) změnit procento organické složky ze základního stavu (70 % metanolu) na procento použité v cílové mobilní fázi (v našem případě 4 % metanolu); 2) zaměnit vodnou složku z deionizované vody za vodný roztok fosfátu sodného o pH 3. V obou případech ponechat protéct alespoň 10 kolonových objemů pro dostatečnou ekvilibraci kolony – přibližně 15 minut při průtoku 2 ml/min.

3.5.3 Stanovení mrtvého retenčního času (t_M)

Mrtvý retenční čas se stanovuje látkou, která není zadržovaná stacionární fází (v případě RPLC se používá např. močovina) a to v mobilní fázi stejně jako je použita pro měření vzorků. Mrtvý retenční čas udává dobu, za kterou projde látka zcela inertní vůči sorbentu na detektor. Mrtvý retenční čas je nutné znát např. pro výpočet kapacitního faktoru, který má přímý vztah k termodynamickým vlastnostem separace (distribuční konstanty). Je možné jej navíc použít pro porovnání výsledků získaných na jiném chromatografickém systému pomocí totožné metody.

3.5.4 Analýza reálných vzorků

Připravené roztoky digestu DNA izolované z ovlivněných a neovlivněných buněčných linií se 10× naředí mobilní fází. Naředěný roztok vzorku je možné přímo analyzovat.

3.5.5 Identifikace analytů

Identifikace vybraných signálů se provede jak proměřením standardních roztoků tak s pomocí knihovny UV-Vis spekter. Standardní roztoky se připraví naředím (100×) zásobních roztoků (koncentrace 10mM) mobilní fází.

3.5.6 Stanovení poměru dCMP a mdCMP a koncentrace ostatních nukleotidů

Kvantitativní stanovení se provede metodou standardního přídávku. Ke vzorku přidáme standardní přídavek dCMP, dGMP, dTMP a dAMP o koncentraci 0,1 a 0,2 mM a přídavek standardu mdCMP o koncentraci 10 a 20 μ M. Pro přípravu roztoků využijte zásobních roztoků (koncentrace 10 mM), případně již naředěných standardů (koncentrace 0,1 mM)

4 Vyhodnocení naměřených dat

Do protokolu uveďte následující údaje včetně statistického vyhodnocení (průměry nebo mediány, směrodatné odchylky a intervaly spolehlivosti pro hladinu významnosti α 0,05):

- mrtvý čas kolony t_M
- retenční časy separovaných nukleotidů ve směsi
- retenční časy separovaných standardů nukleotidů
- kapacitní poměry separovaných látek

Je nutné si zde uvědomit, že se jedná o vyhodnocení malého počtu opakování a použít příslušné rovnice. Viz. podklady ke statistickému zpracování výsledků níže.

Identifikaci látek proveďte na základě testování na statistickou významnost rozdílů v retenčních časech standardů a signálů v chromatogramu reálného vzorku.

Koncentraci jednotlivých nukleotidů zjistíte konstrukcí závislosti plochy signálu analytu na koncentraci přídávku. Jde o lineární závislost, jejíž průsečík s osou x určuje koncentraci nukleotidu ve vzorku. Poměry metylovaného a nemetylovaného dCMP vyjádřete jako poměr zjištěných koncentrací mdCMP/dCMP. Poměry počítejte pro každé měření vzorku, nikoli jen pro průměry.

5 Otázky k diskuzi (odpovědi a diskuzi uvádějte do protokolu)

- v této úloze není použita korekce výsledných ploch mdCMP a dCMP pomocí vnitřního standardu. Je tento postup správný? Jak by situace vypadala v případě, když bychom chtěli porovnat koncentrace jednotlivých nukleotidů mezi vzorky?
- jaká by byla retence nukleosidů v porovnání s nukleotidovými analogy na nepolární koloně?
- v případě analýzy DNA digestů obsahující v ideálním případě pouze směs nukleotidů je možné provést vnitřní kontrolu přípravy vzorku. V čem tato kontrola spočívá a jak vychází v případě Vámi analyzovaných vzorků?

6 Doplnkové informace k vyhodnocení naměřených dat

Charakteristickou veličinou pro každou chromatografovanou látku je retenční (eluční) čas (t_R). Je to doba, která uplyne od nástřiku vzorku do dosažení maxima eluční křivky (Obr. 3). Lze se také setkat s pojmem eluční objem (V_R), tj. objem mobilní fáze prošlý kolonou, po němž je analyt na konci separační kolony. Je to tedy objem mobilní fáze, který proteče za dobu t_R :

$$V_R = t_R \times F_M$$

kde F_M , je množství mobilní fáze, proteklé kolonou za jednotku času (objemová rychlost toku ml/min).

Mobilní fáze se pohybuje konstantní rychlostí, takže všechny analyty stráví v mobilní fázi stejný čas - mrtvý čas kolony (t_M). Celkový retenční čas analytu zahrnuje redukovaný retenční čas, tj. čas strávený ve stacionární fázi.

$$t_R' = t_R - t_M$$

Kapacitní faktor (poměr, k_i) charakterizuje selektivitu (tedy zadržování analytu na koloně) a slouží též ke srovnávání separace v různých systémech (k_i je v daném systému SF a MF konstantní).

$$k_i = (t_R - t_M)/t_M = t_R'/t_M$$

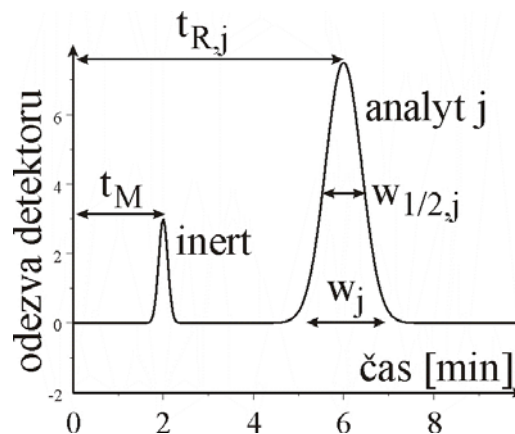
Základním údajem o rozmývání složek vzorku při transportu kolonou (které se projevuje šířkou píku) je počet teoretických pater kolony n . Tento údaj slouží k hodnocení účinnosti kolony podobně, jako odvozená veličina výškový ekvivalent teoretického patra H . Počet teoretických pater se zjistí experimentálně dosazením naměřených parametrů do vztahu

$$n = 16 \cdot \left(\frac{t_{R,j}}{w_j} \right)^2 = 5,545 \cdot \left(\frac{t_{R,j}}{w_{1/2,j}} \right)^2$$

kde w_j je šířka píku při základně a $w_{1/2}$ šířka píku v polovině výšky. Výškový ekvivalent výškového patra H se vypočítá z délky kolony a vypočteného počtu pater n :

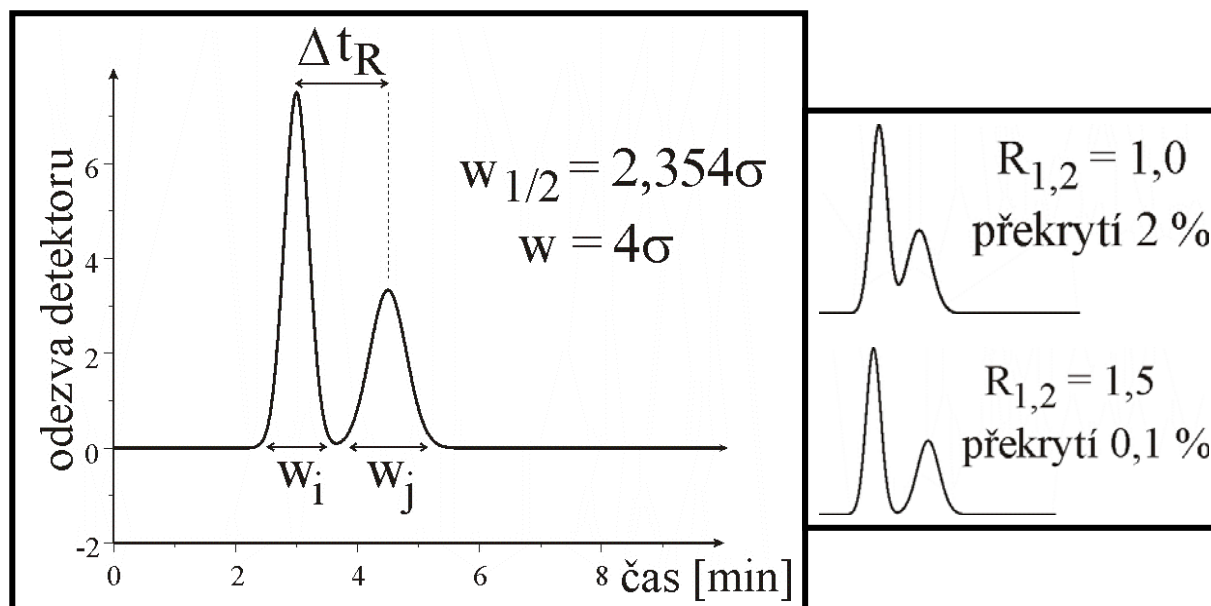
$$H = \frac{L}{n}$$

Pro hodnocení míry vzájemného překrývání dvou sousedních píků se používá veličina rozlišení $R_{i,j}$.



Obr. 3 Ukázkový chromatogram

za



Obr. 4 Ukázková separace dvou analytů s demonstrací hodnoty rozlišení a procenta překrytí píků

$$R_{i,j} = \frac{2 \cdot (t_{R,i} - t_{R,j})}{w_i + w_j} = \frac{2 \cdot \Delta t_R}{w_i + w_j}$$

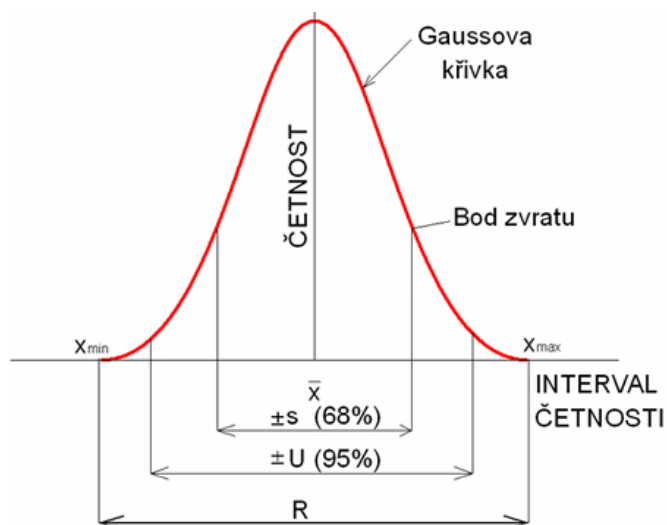
Rozlišení lze vypočítat z retenčních časů analytů a šířek píků. Z hodnoty $R_{i,j}$ lze odhadnout míru překrývání sousedících píků. Při hodnotě $R = 1$ jsou píky překryty z 2 % plochy. Odděleny z 99,9 % a více jsou při hodnotě $R \geq 1,5$.

Pro kvantitativní stanovení se využívá plocha píku, která je úměrná obsahu analytu ve vzorku. V případě, že je tvar píku Gaussovský, je možné brát jako kvantitativní parametr výšku. Tato situace však v kapalinové chromatografii většinou nenastává na rozdíl od plynové chromatografie. Ke zjištění koncentrace analytu ve vzorku lze využít metodu kalibrační křivky, metodu přidavku standardu nebo metodu vnitřního standardu. Metoda přidavku standardu předpokládá linearitu kalibrační závislosti.

7 Poznámka ke statistickému vyhodnocení

Všechny naměřené hodnoty jsou zatíženy chybami. Chyby se obvykle dělí na chyby náhodné, systematické a hrubé. **Hrubé chyby** jsou většinou způsobeny jednorázovým dějem, který vznikne z důvodu chyby přístroje, nebo lidského faktoru. **Systematické (soustavné)** naproti tomu zatěžují výsledek systematicky a určují správnost výsledku. To znamená, na kolik se výsledek blíží skutečné hodnotě. Jsou způsobeny např. nesprávnou kalibrací (pipet, odměrných baněk, analytických vah a podobně). **Chyby náhodné** určují naopak přesnost výsledků. Jde o chyby vzniklé zcela náhodně a to v průběhu v průběhu celého postupu. V dalším textu budeme uvažovat přítomnost pouze náhodných chyb.

Z důvodu výskytu náhodných chyb se dále mluví o tzv. *rozdělení*. Představme si, že budeme měřit danou veličinu nekonečněmnohokrát – získáme tak *soubor dat* (*soubor dat* představuje data, která jsou obrazem celé populace hodnot). Za nepřítomnosti hrubé a systematické chyby dostaneme při vynesení četnosti hodnoty veličiny v závislosti na její velikosti graf, který se označuje jako rozdělení pravděpodobnosti. Příkladem je *rozdělení normální*, které je popsáno Gaussovou křivkou.



Obr. 5 Profil Gaussovy křivky

Plocha pod křivkou vyznačuje všechna měření – 100 % naměřených hodnot. Z tvaru křivky lze následně odvodit vztahy pro určení intervalu v okolí průměrné hodnoty zahrnující např. 95 % hodnot a podobně.

Při statistickém vyhodnocení se ve většině případů u naměřených dat předpokládá právě normální rozdělení (normalitu dat je možné také testovat). Jen nezískáváme celý *soubor* hodnot, ale pouze jejich část – *výběr*. Na základě modelů a testování byly pro *výběry* dat zkonstruovány

tabulky, které umožňují extrapolovat naše výběrová data na data celého souboru. Děje se tak na základě informace o počtu měření a aktuálním výběru. V některých statistikách postačí např. rozptýl všech hodnot, jindy je zapotřebí počítat se všemi získanými hodnotami. Je zapotřebí si uvědomit, že čím menší je *výběr* dat, tím hůře z něj bude definován celý *soubor*. Mluví se o tzv. odhadech vlastností *souboru* (např. střední hodnota) na základě *výběru*.

Pro naměřený blok dat (v našem případě např. naměřené retenční časy, plochy píků a podobně) můžeme vypočítat základní popisné statistiky (směrodatná odchylka, maximum, minimum, průměr, medián, ...). Jak bylo vysvětleno výše, jelikož počítáme z dat, které tvoří *výběr* ze *souboru*, mluvíme o tzv. výběrových odhadech např. skutečné hodnoty (μ ; výběrovým odhadem je průměr nebo medián), rozptylu (σ^2 , odhadem je druhá mocnina směrodatné odchylky) atp.

Jako odhad skutečné hodnoty se nejčastěji uvádí průměr. V případech, kdy je podezření na přítomnost extrémních hodnot, je vhodné použít medián, který je vůči nim více robustní. Užívá se často u dat, kde není potvrzena jejich normalita (normální rozdělení kolem střední hodnoty). Co se týče přesnosti výsledků, ta se nejčastěji uvádí ve formě směrodatné odchylky. Tento přístup však není zcela korektním. Nejlepším způsobem jak vyjadřovat výsledky je pomocí intervalu spolehlivosti pro zvolenou hladinu významnosti α (např. 0,05; 0,01; 0,001). Ten udává meze, ve kterých je skutečný výsledek z definovanou pravděpodobností (tzv. hladina spolehlivosti) $100 \times (1 - \alpha)$ (pro $\alpha = 0,05$ je hladina spolehlivosti 95 %).

Při porovnání dvou bloků dat (např. porovnání výsledků dvou různých metod) je možné dosadit hodnoty do příslušných vzorců na testování významnosti rozdílů mezi dvěma soubory (viz. níže). Stejně tak můžeme postupovat na základě znalosti *intervalů spolehlivosti* „grafickou“ metodou. Ty, jak je uvedeno výše, znamenají interval, ve kterém se s určitou pravděpodobností nachází skutečný výsledek. Pro porovnání dvou bloků dat potom stačí zjistit, zda-li se tyto intervaly protínají, či nikoliv. To nám dá o porovnání dvou *výběrů* kvalitativní informaci – intervaly se protínají \Rightarrow *výběrová data* pochází ze stejného *souboru* = průměry se neliší; případně se intervaly neprotínají \Rightarrow *výběrová data* pochází z různých *souborů* = průměry se liší. Pro toto porovnání platí hladina významnosti α taková s jakou jsou uvedeny intervaly spolehlivosti pro oba průměry/mediány. Výhodou výpočtu pomocí příslušných rovnic je, že jsou schopny nám porovnání průměrů kvantifikovat. Jinak řečeno jsme schopni získat hodnotu α , při které jsou mediány právě odlišné (např. 0,000002 nebo 0,2). Na základě námi předem zvolené α (např. 0,05) pak rozhodujeme o výsledku statistického porovnání – jsou nebo nejsou shodné výběry. (Nutno dodat, že tato výhoda platí pro zpracování většího množství dat, než budete provádět v rámci tohoto cvičení; zde se aplikují statistické testy také pouze s kvalitativním zhodnocením.)

Statistické vyhodnocení záleží do značné míry na počtu opakování (viz výše). Pro statistiku obecně platí čím více dat, tím lépe. Naopak v případě malého počtu opakování může selhat. Z tohoto důvodu se v případech, kdy máme k dispozici data z malého počtu opakování ($n \leq 10$) aplikují odlišné statistické přístupy. Je proto důležité vždy sledovat předpoklady statistických testů a v případě jejich nedodržení testování provádět odlišně, nebo jej neprovádět vůbec! V opačném případě se můžete dopustit chybného závěru. O statistice se potom bude mluvit tak jako v některých vtipech: „Existují tři stupně lži: lež, sprostá lež a *statistika*.“ Statistika je mocným nástrojem, ale pouze v rukou vzdělaného a zodpovědného člověka...

Vylučování výsledků měření se v případě malého počtu opakování provádí pomocí Dean-Dixonova testu na odlehlost krajních hodnot. U větších výběrů se aplikují jiné vztahy a vylučování jako takové je poměrně složitou záležitostí. Spíše než-li mechanické dosazení hodnot do vzorce se uplatňuje porovnávání souborů s a bez testovaného bodu. V principu by se měly totiž vyloučit pouze ta data, která byla naměřena s hrubou chybou. V opačném případě můžeme vyloučením dosáhnout deformace rozdělení našich dat.

K základnímu statistickému vyhodnocení lze použít klasický tabulkový procesor (pozor, Excel pracuje pouze se statistikou pro velká čísla. V některých případech také počítá nepřesně a někdy dokonce uvádí chybné označení!). Větší množství funkcí než např. Excel nabízí Gnumeric (opensource program podobný Excelu). Statistické balíčky, které je možné získat na MU zdarma (viz. <http://www.muni.cz/ics/services/software?lang=cs>) jsou Statistica („klikací“ program, obsahuje množství statistických analýz pro parametrická i neparametrická data) a Matlab (často používaný v publikacích, rozhraní je „programovací“ pro pokročilejší uživatele, ale s pokročilými funkcemi co se týče např. dávkového zpracování dat atd.).

Vybrané rovnice pro statistické vyhodnocení naměřených dat (jsou zde uvedené vzorce používané v případě malého počtu opakování):

- vyloučení odlehlých dat – Q-test

$$Q_n = \frac{m_n - m_{n-1}}{R}; Q_1 = \frac{m_2 - m_1}{R},$$

kde R je rozpětí a m jsou naměřené hodnoty. Index určuje jejich pořadí při jejich seřazení od nejmenší po největší hodnotu: n – poslední (největší) hodnota, $n - 1$ – předposlední hodnota atd. Vypočtené hodnoty Q_n a Q_1 porovnáme s tabelovanými kritickými hodnotami Q_α pro daný počet měření n a zvolenou hladinu významnosti α (Tab. 1). Pokud je Q_n či Q_1 nižší než kritická hodnota Q_α , nejvyšší, respektive nejnižší změřená hodnota není zatížena hrubou chybou a nevylučuje se. Naopak, pokud je vypočtené Q_n nebo Q_1 větší, je nutno výsledek vyloučit. Po vyloučení odlehlého

výsledku je nutné data opět otestovat na přítomnost odlehle hodnoty. Q-test se používá, když je počet měření 3 až 10. V případě tří hodnot platí navíc podmínka, že všechny hodnoty musí být různé. V opačných případech nelze Q-test použít.

Tab. 1 Statistické konstanty dle Dean-Dixona pro hladinu významnosti α 0,05.

počet měření, i	k_i	K_i	Q_i
2	0,886	6,4	
3	0,591	1,3	0,941
4	0,486	0,72	0,765
5	0,430	0,51	0,642
6	0,395	0,40	0,560
7	0,370	0,33	0,507

- výběrový odhad σ - směrodatná odchylka dle Dean-Dixona s_R :

$$s_R = k_n \times R,$$

kde k_n je koeficient závislý na počtu měření n (pro hladinu významnosti 0,05 platí $k_2 = 0,8862$; $k_3 = 0,5908$; $k_4 = 0,4857$), R je rozpětí dat

- výběrový odhad rozdělení konečných výsledků – interval spolehlivosti dle Dean-Dixona ($L_{1,2}$)

$$L_{1,2} = \bar{x} \pm K_n \times R,$$

kde \bar{x} je průměr, K_n je koeficient pro dané n (pro hladinu významnosti 0,05 platí $K_2 = 6,40$; $K_3 = 1,30$; $K_4 = 0,92$; hodnoty uvedeny pro $\alpha = 0,05$) a R je rozpětí

- porovnání dvou výběrů – stejný počet měření – Lordův test

$$u = \frac{|\bar{x}_A - \bar{x}_B|}{R_A + R_B},$$

kde $\bar{x}_{A,B}$ jsou průměry výběrů a $R_{A,B}$ rozpětí. Vypočtené u se porovnává s kritickou hodnotou Lordova testu u_{α} (Tab. 2). Je-li $u < u_{\alpha}$, je rozdíl $|\bar{x}_A - \bar{x}_B|$ statisticky nevýznamný na zvolené hladině významnosti α , a že je možné jej vysvětlit přítomností náhodných chyb obou výsledků.

- porovnání dvou výběrů – různý počet měření – Moorův test

$$U = \frac{|\bar{x}_A - \bar{x}_B|}{R_A + R_B},$$

kde $\bar{x}_{A,B}$ jsou průměry výběrů a $R_{A,B}$ rozpětí. Vypočtené U se porovnává s kritickou hodnotou Moorova testu u_α . Kritické hodnoty jsou uvedeny v tabulce 2. Moorův i Lordův test využívá pro výpočet testové statistiky totožný vzorec. Moorův test tak v případě stejného počtu měření přechází na test Lordův. Zhodnocení výsledku je stejné jako u Lordova testu

Tab. 2 Tabelované hodnoty pro Lordův a Moorův test pro hladinu významnosti α 0,05

n_A	n_B	$u_{i,j}, U_{i,j}$
	2	1,714
2	3	0,915
	4	0,732
	5	0,619
3	3	0,635
	4	0,511
	5	0,429
4	4	0,407
	5	0,353
5	5	0,307