

# Stanovení obsahu inosinu, adenosinu a jejich 2'-deoxy- forem v modelové směsi – optimalizace a validace metody

## Úloha do laboratorního cvičení - Speciální metody

*pro magisterské studium Ústav chemie-Analytická chemie, podzimní semestr*

Pro základy kapalinové chromatografie na reverzní fázi (RPLC) prostudujte také návod k úloze do laboratorního cvičení Metody chemického výzkumu v jarním semestru. Znalost těchto informací bude v rámci úlohy ověřena.

### 1 Cíl úlohy

- I. modelová optimalizace HPLC metody s využitím plánu pokusů a umělých neuronových sítí
- II. validace optimalizované HPLC metody

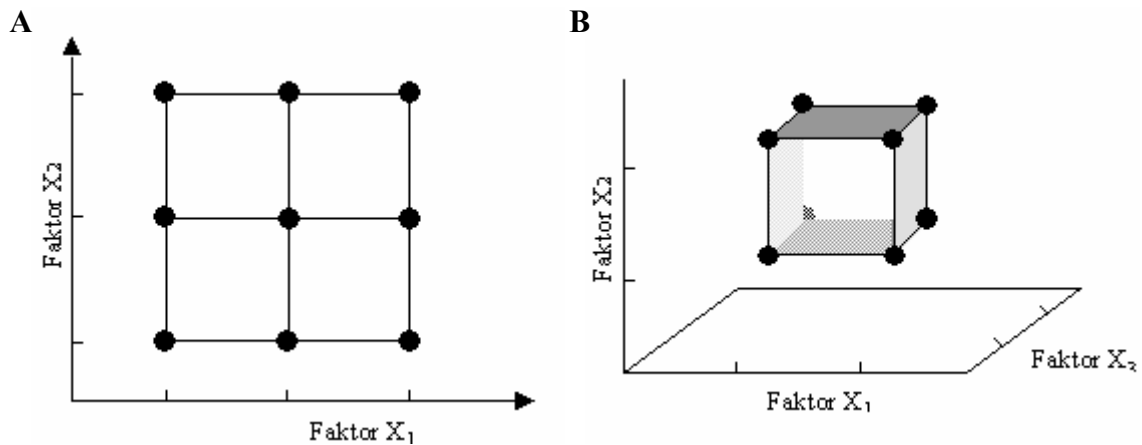
### 2 Teorie

#### 2.1 Optimalizace metody

Při vývoji nové HPLC metody je důležité docílit separace všech sledovaných látek a to za co nejkratší dobu. Separace i celková délka analýzy je ovlivněna mnoha faktory (složení mobilní fáze, pH, teplota, průtoková rychlost, povaha vzorku). Proces hledání nejlepších separačních podmínek pro vybraný analyt/skupinu analytů se nazývá optimalizace. Optimalizace je založena na znalosti použitého systému a povahy analytu. Na základě těchto znalostí je možné navrhnout vhodné parametry k optimalizaci. Nejdůležitější parametry jsou diskutovány v návodu pro podzimní úlohu, kapitola 2.4.

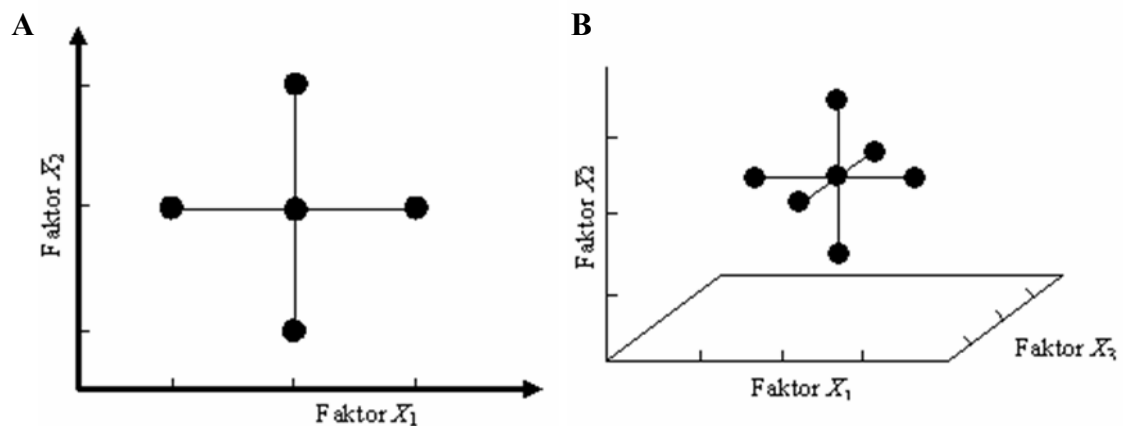
Při optimalizaci lze postupovat „klasickým způsobem“, kdy se při konstantních podmínkách změny úrovně vždy jen jedné optimalizované veličiny (např. pH vodné složky mobilní fáze). Tento postup vyžaduje velké množství experimentů, z čehož plyne jak časová tak materiálová náročnost. Plánování pokusů (experimental design, ED) snižuje počet nutných experimentů cílenou změnou více parametrů zároveň. Experimentátor nejdříve zvolí okrajové hodnoty optimalizovaných faktorů (např. pH by mělo být optimalizováno od 3 do 7, procento organické složky od 0 do 40 %). Plán pokusů pak již určí experimenty, které je nutné provést, abychom pokryly celou optimalizovanou oblast všech faktorů. Mezi častý typ plánu pokusů patří *faktorový plán*, *hvězdicový plán* a jejich kombinace – *centrálně kompozitní plán* (central composite design, CCD).

Předností *faktorového plánu* (factorial design) je možnost odhalení vzájemné interakce jednotlivých proměnných (pokud existují). Počtu vstupních proměnných odpovídá počet rozměrů (optimalizovaných faktorů, např. pH vodného roztoku;  $f$ ). Počtu hodnot každé vstupní proměnné odpovídá počet úrovní (hladin, na kterých je faktor měřen, např. pH 3, 5 a 7 při optimalizaci pH vodné složky;  $L$ ) – viz. Obr. 1. Celkový počet bodů (experimentů) je přitom dán vztahem:  $n = L^f$ .

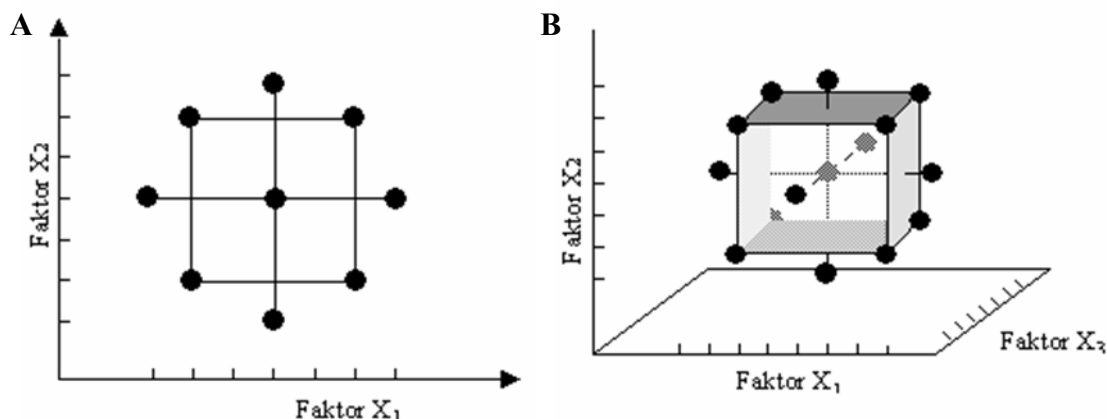


**Obr. 1** Schéma faktorového plánu pro případ optimalizace dvou (např. pH vodné složky a obsah metanolu v mobilní fázi; **A**) a tří faktorů (např. pH vodné složky, koncentrace solí a obsah metanolu v mobilní fázi; **B**). Černé body odpovídají podmínkám jednotlivých experimentů.

*Hvězdicový plán* (star design) je vhodný pro vyjádření hyperplochy a obsahuje  $(2f+1)$  bodů (Obr. 2).



**Obr. 2** Schéma hvězdicového plánu pro dva (**A**) a tři (**B**) optimalizované faktory.

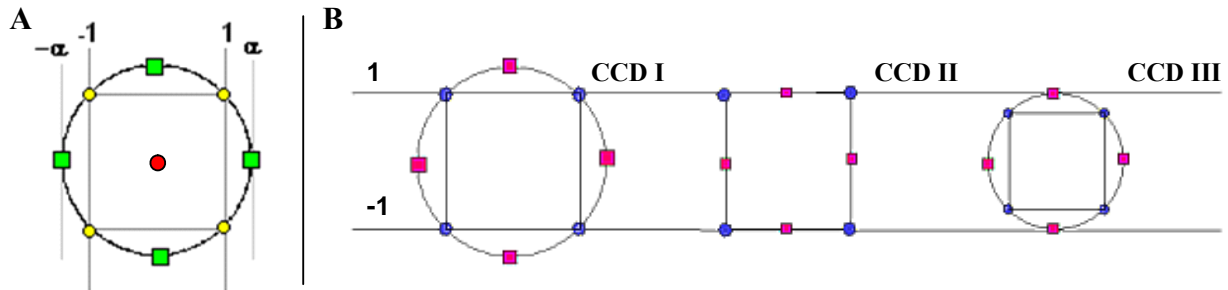


**Obr. 3** Schéma centrálně kompozitního plánu (CCD) pro dva (A) a pro tři (B) rozměry (optimalizované faktory).

Kombinací faktorového a hvězdicového plánu získáme *centrálně kompozitní plán* (central composite design, CCD; Obr. 3). Počet bodů odpovídá  $2^f + 2f + 1$  (Obr. 4A). Body z faktorového plánu (na Obr. 4A zobrazeny žlutě) mají jednotkovou vzdálenost od centrálního bodu (na Obr. 4A vyznačen červeně). Body hvězdicového plánu (na Obr. 4A označené zeleně) mají vzdálenost  $\alpha$ , kdy  $\alpha > 1$  a závisí na počtu faktorů.  $\alpha$  lze vypočítat podle vzorce:  $\alpha = 2^{f/4}$ .

Existují tři typy CCD designu (Obr. 4B). Rozdíl mezi nimi je v interpretaci krajních hodnot optimalizovaných faktorů zadaných operátorem (viz. výše). Výše byl uveden příklad krajních hodnot pro optimalizaci pH vodné složky od 3 do 7. Tyto limity byly zvoleny podle pH stability stacionární fáze a proto není možné jejich překročení. U typu CCD I však design záměrně krajní hodnoty překračuje, aby lépe zmapoval chování systému ve zvolených intervalech. Proto CCD I není možné aplikovat v případech, kdy operátorem zvolené intervaly jednotlivých faktorů jsou limity reálnými (např. obsah organické složky 0 %). V takových případech volíme jeden z alternativních typů CCD. V případě CCD II dělíme body hvězdicového designu (Obr. 4B, růžové body) hodnotou  $\alpha$ . V případě CCD III dělíme všechny body vyjma centrálního hodnotou  $\alpha$ . Tím docílíme „zmenšení“ plánu pokusů tak, že pokrývá námi zadané limity.

Pro vyhodnocení dat získaných z experimentů podle plánu pokusů je výhodné použít umělých neuronových sítí. Výsledkem je aproximace hyperplochy napříč celým intervalem všech optimalizovaných faktorů, ze které lze získat optimální separační podmínky.



**Obr. 4** A – Vyznačení vzdáleností bodů faktorového designu (žluté body) a hvězdicového designu (zelené body) od centrálního bodu (červený bod). B – Tři typy CCD designu. „1“ a „-1“ označují vzdálenost od centrálního bodu designu a operátorem zadané hranice pro jeden z optimalizovaných faktorů.

## 2.2 Umělé neuronové sítě

Umělé neuronové sítě (artificial neural networks, ANNs) byly vytvořeny na základě jednoduchých modelů neuronů – funkčních buněk nervového systému živých organismů. Každá neuronová síť převádí (transformuje) hodnoty vstupních veličin na hodnoty výstupních veličin pomocí transformační funkce. ANN lze využít při modelování a řízení složitých soustav, které nelze s uspokojivou přesností matematicky popsat, nebo matematický model procesu je tak složitý, že jeho případná algoritmizace je buď časově a programově velmi náročná nebo dokonce nemožná.

### 2.2.1 Popis a matematický model umělé neuronové sítě

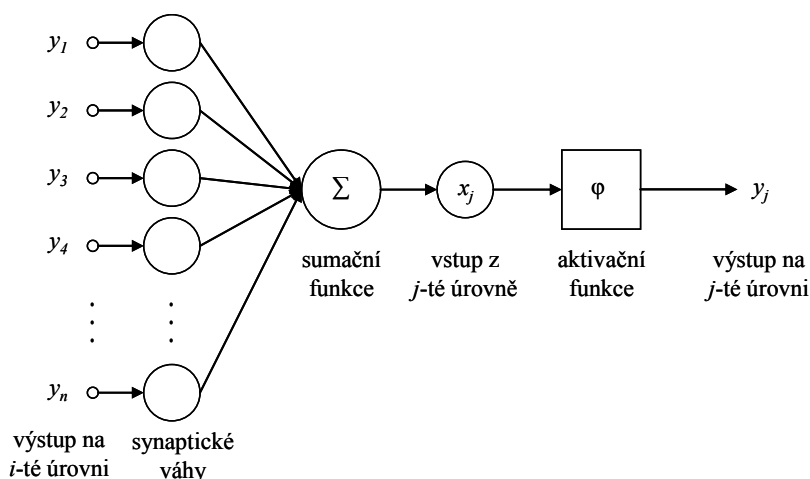
ANN je tvořena neurony nejčastěji uspořádanými do vrstev. Každý neuron přijímá signály z neuronů předchozí vrstvy nebo z vnějšího zdroje (Obr. 5). Význam každého z těchto vstupů je určen vahou vstupu  $w_{ij}$ , kde dochází k tlumení nebo buzení signálu. Vstupní hodnota ( $x_j$ ) neuronu  $j$  je tedy dána součtem výstupů z neuronů  $i$  ( $y_i$ ) vynásobených vahami spojení mezi neurony  $i$  a  $j$  ( $w_{ij}$ ):

$$x_j = \sum (y_i \cdot w_{i,j})$$

Aplikací aktivační funkce je vstupní hodnota  $x_j$  přeměněna na hodnotu výstupní  $y_j$ :

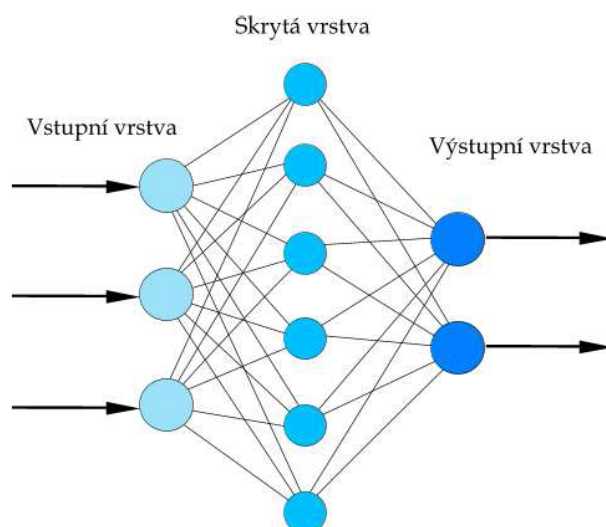
$$y_i = f(x_j)$$

Tímto je korigována amplituda výstupu neuronu do akceptovatelného rozmezí. To je obvykle definováno jako interval  $(-1;1)$ , nebo  $(0;1)$ . Výstup je buď konečným výsledkem, nebo je předán neuronům v dalších vrstvách.



**Obr. 5** Matematický model neuronu. Výstup z neuronů na  $i$ -té úrovni je na základě vah spojení sečten. Suma je dále zpracována aktivační funkcí, která reguluje amplitudu výstupu do akceptovatelné meze (většinou  $(-1;1)$ , nebo  $(0;1)$ )

Vrstvy tvořené neurony lze rozdělit podle jejich umístění na 3 části. První vrstva se nazývá *vstupní* a má za úkol přijímat hodnoty z okolí pro zpracování a přivést je na vstup každého neuronu následující vrstvy. Druhá vrstva je označována jako *skrytá*. Počet skrytých vrstev ve vnitřní části závisí na složitosti funkce, kterou má síť vykonat a na zvoleném typu sítě. Ve většině případů je dostatečná jediná skrytá vrstva. Úkolem skrytých vrstev je zvýšení aproximačních vlastností neuronové sítě. Poslední vrstva, *výstupní*, obsahuje neurony předávající signál do okolí. Tyto výstupní signály jsou pak odezvou neuronové sítě na signály vstupní.



**Obr. 6** Model neuronové sítě s jednou skrytou vrstvou. Jako vstupní vrstva jsou použity např. hodnoty optimalizovaných faktorů naplánované dle plánu pokusů (např. pH, koncentrace solí, obsah organické složky). Dva neurony výstupní vrstvy jsou pak např. retenční časy dvou analytů pozorované za podmínek definovaných pomocí vstupních neuronů.

Podle toku signálu neuronovou sítí pak dělíme sítě na dopředné (feed-forward) a se zpětnou vazbou (rekurentní, feed-back). Při použití struktury s dopředným šířením signálu se signál šíří po orientovaných spojnicích jenom jedním směrem. V rekurentních sítích existují mezi neurony nebo vrstvami zpětné vazby.

### 2.2.2 Trénování umělých neuronových sítí

Základní a velmi podstatnou vlastností ANN je schopnost učit se. Fáze učení předchází fázi vlastní práce a slouží k určení se váhových koeficientů mezi neurony. Fáze učení může probíhat dvěma způsoby, s učitelem a bez učitele. Při metodě učení s učitelem je síť trénovaná pomocí dvojice vstupní vzorek a příslušný, očekávaný výstupní vzorek. Využívají se trénovací data v takovém rozsahu, aby plně popsaly všechny vlastnosti množiny důležité pro danou úlohu. V našem případě se jedná o výsledky z experimentů naplánovaných s pomocí plánu pokusů. Jako vstupní vzorek vezmeme složení mobilní fáze a výstupním vzorkem je některý z parametrů separace, např. rozlišení mezi dvěma sousedními analyty, retenční čas poslední látky atd. Na základě skutečné odezvy trénované sítě a očekávané odezvy se upravují váhové koeficienty. U metody učení bez učitele systém nedostává očekávané výstupy od učitele, ale odvozuje si je od svého výstupu prostřednictvím zpětné vazby.

K trénování sítě existuje několik trénovacích algoritmů. Algoritmus učení dopředné neuronové sítě metodou zpětného šíření chyby (Backpropagation Algorithm) je nejčastěji používaná metoda. Základem je vrstvená síť, kde nejsou žádné zpětné vazby. Chyba se šíří zpětně přes všechny vrstvy k první vrstvě. Musí ale být známa vstupní a výstupní dvojice hodnot.

Během procesů učení je nutno dávat pozor na tzv. přetrénování sítě (popsáno níže). Přetrénovaná síť dává sice přesné výsledky pro trénovací data, ale nerelevantní výsledky pro data, se kterými se doposud nesetkala. Ztrácí tzv. vlastnost zobecnění. Pro zabránění přetrénování sítě se data před trénováním rozdělují na 3 (případně pouze 2) části: data trénovací, data validační (tato data nejsou nutná) a testovací data. Trénovací data se využívají pro nastavení vah sítě a musí pokrývat celý interval optimalizovaných parametrů. Tato data získáme např. z experimentů provedených na základě plánu pokusů, který tyto předpoklady naplňuje. Pro nejvhodnější nastavení parametrů (např. počet skrytých vrstev, počet neuronů ve skryté vrstvě) se využívají validační data, která mohou být součástí dat trénovacích. Testovacími daty se ověří funkce natrénované ANN. Jde o data náhodně vybraná v celém intervalu testovaných parametrů, která nejsou součástí trénovací sady.

### 2.2.3 Kontrola učení umělých neuronových sítí

Míra úspěšnosti natrénované ANN lze hodnotit podle střední kvadratické chyby (RMS, root mean square error), která se vypočte podle vzorce:

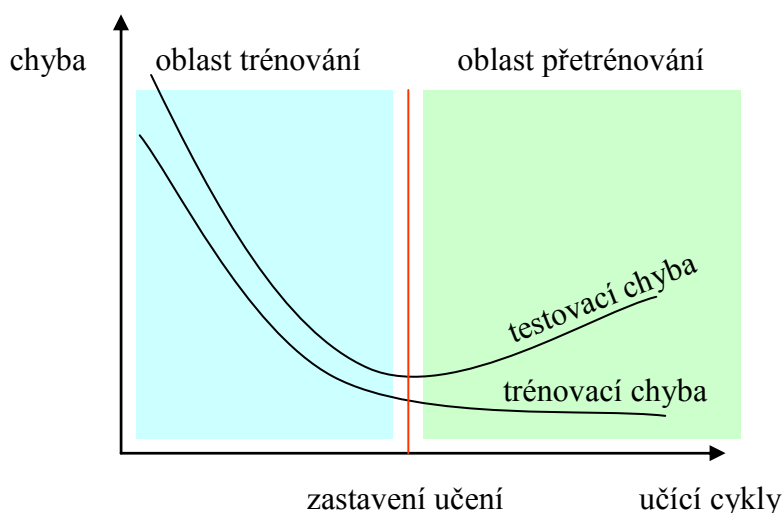
$$RMS = \sqrt{\frac{\sum_{i=1}^m \sum_{j=1}^h (o_{ij} - t_{ij})^2}{mh}}$$

kde  $h$  je počet vzorků (počet řádků),  $m$  je počet výstupních proměnných,  $o$  je vypočítaná hodnota  $j$ -tého neuronu pro  $i$ -tou proměnnou,  $t$  je očekávaná (experimentálně zjištěná) hodnota  $j$ -tého neuronu pro  $i$ -tou proměnnou. Umocněním na druhou RMS zvyšuje váhu větších rozdílů. Pokud máme pouze jednu výstupní proměnnou, vzorec se zjednoduší na:

$$RMS = \sqrt{\frac{\sum_{j=1}^h (o_j - t_j)^2}{h}}$$

RMS je většinou automaticky počítána během trénování sítě pro trénovací i testovací sadu. Díky učení ANN na trénovacích datech se chyba pro tato data s počtem

učících cyklů snižuje. Při trénování sítě jde ve většině případů o její schopnost generalizovat. Kritériem úspěšnosti sítě v tomto ohledu je RMS u testovacích dat. Tato chyba nejdříve klesá, ale po určitém počtu cyklů začne znovu růst (Obr. 7). V takovém případě již hovoříme o přetrénování sítě – síť je schopná přesně předpovědět výstupní neurony pro trénovací data, ale ztrácí již schopnost generalizovat problém na data, se kterými se ještě nesetkala. Učení sítě se proto zastavuje v místě, kde je pozorována nejnižší chyba v předpovědi pro *testovací* data.



**Obr. 7** Průběh chyby předpovědi v průběhu učících cyklů pro trénovací a testovací data. Červená čára označuje počet učících cyklů, kde je pozorována minimální chyba předpovědi pro testovací data. V tomto bodě se zastavuje trénování sítě. Při více učících cyklech se chyba pro testovací sadu zvětšuje a dostáváme se do oblasti přetrénování sítě.

Důležitým kritériem není jen průměrná odchylka, ale také rozptyl chyb. Průměrná absolutní chyba (MAE, mean absolute error) udává průměrnou hodnotu absolutních rozdílů mezi předpovězeným a naměřeným údajem:

$$MAE = \frac{1}{h} \sum_{j=1}^h |o_j - t_j|$$

kde  $h$  je počet vzorků (počet řádků),  $o$  je vypočítaná hodnota  $j$ -tého neuronu,  $t$  je očekávaná (experimentálně zjištěná) hodnota  $j$ -tého neuronu (obdobně platí i pro síť s více výstupními neurony). V případě RMS jsou odchylky před zprůměrováním umocněny a větší odchylky tak mají větší váhu. U MAE jsou si všechny odchylky rovnocenné (průměrují se rovnou bez



umocnění). RMS nabývá stejné nebo větší hodnoty než MAE. Pokud  $RMS = MAE$ , pak všechny odchylky mají stejnou velikost. Čím větší rozdíl mezi RMS a MAE, tím větší je rozdílnost mezi odchylkami.

Linearitu vztahu mezi předpovězeným a pozorovaným retenčním časem lze hodnotit korelačním koeficientem ( $r$ ).

#### **2.2.4 Využití natrénované umělé neuronové sítě**

Natrénovanou ANN lze využít k předpovědi výstupních neuronů pro všechny možné kombinace vstupních neuronů. Jediným limitem je rozsah hodnot jednotlivých vstupních neuronů, který se musí řídit rozsahem použitým pro trénování sítě. Stejně jako u kalibrací, tak i u ANN není možné výsledky extrapolovat. Grafickou závislost hodnoty výstupního neuronu na hodnotě vstupních neuronů nazýváme hyperplochou (Obr. 8). Hyperplochu, případně dvojice vstupních a výstupních dat, je možné využít k nalezení optimálních podmínek separace.

Při hledání optimálních podmínek je vhodné řídit se obecnými doporučeními pro metody kapalinové chromatografie. Těmi základními jsou: 1) rozlišení mezi sousedními analyty alespoň musí být alespoň 1,5 a 2) doba analýzy by měla být co nejkratší. Doporučuje se také, aby retenční čas analytu byl větší než dvojnásobek mrtvého retenčního času. U monolitických kolon je však separace rychlejší v porovnání s částicovými kolonami. Proto tento předpoklad nebývá v některých případech splněn. Předpoklady pro HPLC metodu do určité míry naznačují výhodnost zvolení rozlišení sousedních analytů a retenčního času poslední látky jako výstupní neuronu při trénování ANN. U složitějších směsí je v některých případech zapotřebí počet výstupních neuronů minimalizovat. Platí totiž, že čím více výstupních neuronů, tím méně správné jsou výstupy ANN.

Několik vybraných mobilních fází se nakonec ověřuje experimentálně. Chyby experimentálních výsledků od předpovězených jsou běžně do 30 %. V průměru bývají chyby okolo 10 %. V případě, že chyby jsou větší, je nutné zvážit volbu jiné topologie sítě (více, či méně vnitřních neuronů, více vrstev) včetně zahrnutí dalších vstupních veličin nebo naopak odebrání některého z výstupních neuronů.

### **2.3 Validace metody**

Protože způsob získávání a zpracování experimentálních dat ovlivňuje konečné výsledky z hlediska přesnosti a správnosti, je nutné ověřit funkčnost zvoleného analytického

postupu. Před zavedením analytické metody do praxe je proto nutné zjistit vhodnost metody pro řešení daného analytického problému, statisticky prokázat spolehlivost (tj. přesnost a správnost metody) analytické metody včetně celého obslužného analytického systému a vyšetřit její praktické hranice použitelnosti. Tento proces se nazývá validace metody. Validace se používá nejen po vyvinutí nové analytické metody, ale také při převodu analytické metody z jiné laboratoře.

### **2.3.1 Druhy validací**

Validace metody v rámci jedné laboratoře se nazývá interní (vnitřní) validace a lze ji dále rozdělit:

1. průzkumová validace se stanovuje na omezeném počtu vzorků a slouží ke zjištění vhodnosti analytické metody pro daný úkol
2. plná validace obsahuje vyhodnocení všech validačních parametrů
3. validace při převodu metody slouží pro zavedení již validované metody z jiné laboratoře
4. retrospektivní validaci lze využít pro již dříve naměřená data

Externí (vnější) validace zahrnuje interní validaci společně s validací metody srovnáním výsledků metody z více laboratoří a zahrnuje ověření reprodukovatelnosti metody.

Cílem validace metody je zjistit selektivitu, správnost, přesnost (opakovatelnost, dlouhodobá opakovatelnost), selektivitu, linearitu, dynamický rozsah, limit detekce, limit kvantifikace a robustnost.

### **2.3.2 Test způsobilosti HPLC**

Pokud zvolíme za analytickou metodu HPLC je nutné před zahájením vývoje nové metody, před validací metody i v průběhu studie provádět test způsobilosti HPLC. Tímto testem ověřujeme schopnost HPLC systému dávat relevantní, správné a přesné výsledky. Test zahrnuje:

1. Test opakovaného nástřiku (6 nástřiků, relativní směrodatná odchylka – RSD by měla být menší než 1,5 %)
2. Test linearity (zvolit si min. 5 bodů o různé koncentraci, koeficient determinace by neměl dosahovat menší hodnoty než 0,9900)
3. Test účinnosti chromatografické separace

### 2.3.3 Validační parametry

#### 2.3.3.1 *Selektivita*

Selektivita analytické metody je schopnost přesného a správného určení analytu i v přítomnosti interferujících látek (matrice). Neselektivní metoda dává signál, který neodpovídá skutečnému obsahu analytu, ale je zesílen signálem rušivé složky. Pokud je k dispozici matrice vzorku, analyzujeme matrici, dále vzorek bez matrice i bez analytu (v případě HPLC nastříkneme mobilní fázi), standardní roztoky sledovaných látek, kontrolní vzorky (matrice + standardy stanovované složky), vzorky s přidavkem standardní (stanovované) složky. Zjistíme, zda interference pochází ze srovnávacího vzorku metody, pak se musí odstranit zdroj interference (kontaminace rozpouštědla, laboratorního nádobí, přístrojů atd.) nebo pochází ze srovnávacího vzorku matrice a je nutné optimalizovat extrakci vzorku, předseparaci (čištění vzorku). Odezva interferujících látek musí být menší než 1% odezvy nejnižší koncentrační hladiny analytu ve vzorku.

#### 2.3.3.2 *Přesnost (Precision)*

Přesnost analytické metody je míra shody mezi jednotlivými výsledky získanými z několika měření stejných homogenních vzorků za daných podmínek. Přesnost se vyjadřuje na 3 úrovních. opakovatelnost, dlouhodobá opakovatelnost a reprodukovatelnost.

1. Opakovatelnost (Repeatability) se zkoumá na identickém materiálu (měří se 3-5 vzorků na několika koncentračních hladinách) použitím téže zkušební metody, v téže laboratoři, stejným pracovníkem za použití týchž přístrojů a zařízení, během krátkého časového rozmezí.
2. V dlouhodobé opakovatelnosti (Intermediate Precision) se promítá variabilita v měření. Nejčastěji měřením v delším časovém okamžiku, provedení jiným pracovníkem nebo s jiným laboratorním vybavením.
3. Reprodukovatelnost (Reproducibility) vyjadřuje míru shody mezi jednotlivými laboratořemi a je potřebná při přenosu analytické metody z laboratoře, kde je používána, do jiné laboratoře.

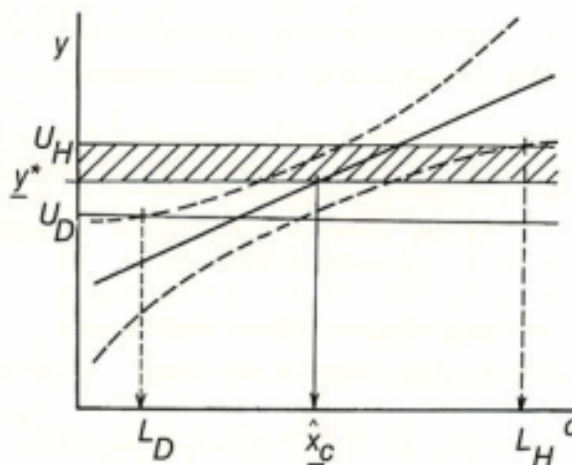
#### 2.3.3.3 *Správnost (Accuracy)*

Správnost je definovaná jako odchylka od správné hodnoty. Vyjadřuje se jako výtěžnost (Recovery) a udává poměr množství analytu zjištěného danou analytickou metodou k přidanému množství. Udává se jako desetinný zlomek nebo v procentech. V HPLC se

vyhodnocuje analýzou vzorku s přidaným standardem na minimálně třech koncentračních hladinách v co nejširším kalibračním rozsahu. Počet paralelních opakování by se měla pohybovat kolem 7 a relativní směrodatná odchylka pro každou koncentrační hladinu nemá být větší než 3%. Výťažnost ( $100 \times \text{zjištěno} / \text{přidáno}$ ) pro koncentrační hladinu se má pohybovat v rozmezí 95% až 105%.

#### 2.3.3.4 Linearita (Linearity)

Linearita vyjadřuje schopnost metody dávat v daném rozsahu koncentrací výsledky přímo úměrné koncentraci stanovované látky ve vzorku. Pro určení linearity se volí minimálně 5 vzorků v celé měřící oblasti. Provede se regresní analýza (zjištění regresních koeficientů  $a$  a  $b$  včetně testování jejich statistické významnosti, zjištění korelačního koeficientu  $-r$  a koeficientu determinace  $-r^2$ ). Ke statistickému potvrzení linearity sledované závislosti je možné



**Obr. 8** Určení intervalu spolehlivosti  $[L_D, L_H]$  koncentrace  $c$  na základě intervalu spolehlivosti signálu  $y$   $[U_D, U_H]$ .

využít např. tabulky kritických hodnot korelačních koeficientů  $r$  pro lineární závislost – viz. Tab. 1. Přesnost kalibrační přímky představuje interval spolehlivosti. Ten je nejužší ve středu kalibrace, a proto je vhodné konstruovat kalibrační přímky s ohledem na tuto skutečnost tak, aby očekávané hodnoty měřené veličiny leželi přibližně uprostřed voleného kalibračního rozsahu. Z obrázku 8 je také patrná nutnost co možná nejpřesnějšího určení  $y$  i  $x$  při kalibračních experimentech. Za podmínek zobrazených na Obr. 8 (velikost intervalu spolehlivosti kalibrační přímky a měřeného signálu  $y$ ) nám totiž vychází interval spolehlivosti zjišťované koncentrace přes téměř celý kalibrovaný interval! Výpočet intervalu spolehlivosti kalibrace není triviálním úkonem, zvládá jej však většina statistických programů, včetně programu Statistica.

**Tab. 1:** Kritické hodnoty korelačního koeficientu  $r$  pro počet stupňů volnosti  $\nu$  a dvě hladiny významnosti  $\alpha$  pro který platí  $\nu = n - 2$ , kde  $n$  představuje počet korelačních bodů

v	$\alpha$		v	$\alpha$		v	$\alpha$	
	0,05	0,01		0,05	0,01		0,05	0,01
1	0,999	999	6	707	834	11	553	684
2	0,950	999	7	666	798	12	532	661
3	0,878	959	8	632	765	13	514	641
4	0,811	917	9	602	735	14	497	623
5	754	875	10	576	708	15	482	606

### 2.3.3.5 Limit detekce (Detection Limit)

Existuje jak horní tak dolní limit detekce. Dolní limit detekce (LOD) je nejnižší množství analytu ve vzorku, který ještě může být detekován. Existuje několik možností jak ho stanovit. U separačních metod lze využít hodnoty signálu slepého pokusu a kalibrační přímky. Z chromatogramu slepého pokusu se určí maximální kolísání základní linie ( $h_{\max}$ ) v oblasti dané 20-ti násobkem pološířky píku stanovovaného analytu. Pro odezvu detekovatelného množství analytu pak platí:  $y_d = 3 \times h_{\max}$  a pro koncentraci na mezi detekce:  $x_d = y_d/b_1$ , kde směrnice kalibrační přímky  $b_1$  musí být z koncentrační závislosti  $y = b_1x$ , kde  $y$  je výška chromatografického píku a ne plocha.

Další možností je metoda postupného zředování, kdy dávkovaný analyt postupně ředíme až do okamžiku, kdy poskytnutý signál má ještě výšku  $3\times$  větší než průměrný šum. Početně lze tento problém řešit z informací dostupných v chromatogramu, kde koncentrace analytu je natolik nízká, že je viditelný i šum základní linie. K výpočtu LOD pak využijeme např. vzorce:

$$LOD = \frac{3 \cdot N \cdot c}{S},$$

kde  $S$  je výška signálu analytu při koncentraci analytu  $c$  a  $N$  je výška šumu základní linie. Jak je z vysvětlení patrné, existují odlišnosti v možných přístupech. Vždy je proto nutné vedle hodnoty LOD či jiných limitů metody, uvést také způsob, jakým bylo výsledků dosaženo.

### 2.3.3.6 Limit kvantifikace (Quantitation Limit)

Limit kvantifikace se určuje stejně jako dolní limit detekce, jen nepředstavuje trojnásobek, ale desetinásobek šumu a určuje množství, které je možné danou metodou ještě nejen spolehlivě detekovat, ale také stanovit.

### 2.3.3.7 Robustnost (Robustness)

Robustnost analytické procedury je míra její kapacity zůstat neovlivněná malými, ale náhodnými změnami v experimentálních parametrech. V případě HPLC by měla robustní metoda dávat přijatelné výsledky (např. retenční čas, plocha píku, rozlišení) i když dochází v experimentálních parametrech (např. složení mobilní fáze, pH, teplota kolony, průtoková rychlost) k mírné odlišnosti od požadovaných v důsledku běžné nepřesnosti (např. nepřesnosti odměrného nádobí, pomocných přístrojů). Robustnost je experimentálně zjišťována pomocí záměrného vnášení malých změn do metody a zkoumání jejich důsledků.

## 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
- píčka 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):
  - 2'-deoxyinosin-5'-monofosfát (dIno)
  - inosin-5'-monofosfát (Ino)
  - 2'-deoxyadenosin-5'-monofosfát (dAde)
  - adenosin-5'-monofosfát (Ade)

- thiomčovina (0,002% roztok ve vodě)
- mobilní fáze pro měření vzorků a standardů (vodný roztok fosfátu sodného o koncentraci 15mM a pH 3,8; metanol pro chromatografii)

### 3.3 Postup – modelová optimalizace metody

V této části postupu je naznačeno, jak lze během optimalizace metody postupovat s využitím ED a ANN. Vzhledem k tomu, že z časových důvodů nelze celý postup zvládnout v laboratorním cvičení, budou zde pouze diskutovány jednotlivé aspekty tohoto přístupu k optimalizaci s využitím reálných výsledků získaných při optimalizaci úlohy. Cílem je získat optimální metodu separace dvou nukleosidů – adenosinu a inosinu – a jejich 2'-deoxy- forem.

#### 3.3.1 Plán pokusů

Navrhnete parametry separace (průtok, teplota, složení mobilní fáze atd.), které je vhodné optimalizovat (očekává se jejich vliv na separaci), a které je možno ponechat konstantní (neočekává se jejich signifikantní vliv na separaci). U vybraných parametrů k optimalizaci navrhnete rozpětí v jakém by měly být testovány a pomocí centrálně kompozitního plánu (typ CCD III; viz. Obr. 4B) navrhnete potřebné experimenty. **~15 minut**

#### 3.3.2 Vytváření umělých neuronových sítí

Na základě plánu pokusů byly provedeny separace směsi (obsahující Ade, dAde, Ino, dIno) za podmínek navržených v plánu pokusů. Určete vstupní neurony ANN. Navrhnete jaké separační charakteristiky z chromatogramů (retenční časy, rozlišení, pološířky píků atd.) lze nejlépe využít jako výstupní neurony umělých neuronových sítí. Důležité je přitom využití předpovězených hodnot výstupních neuronů při posouzení vhodnosti separačních podmínek. **~15 minut**

#### 3.3.3 Trénování umělých neuronových sítí

Do ANN byla vložena vstupní a výstupní data získaná dle plánu pokusů. Architektura sítě obsahovala jednu skrytou vrstvu. K trénování ANN byl použit program Statistica. Natrénováno bylo celkem 100000 sítí. Nastavený rozsah neuronů ve skryté vrstvě byl 3-15. Z natrénovaných sítí bylo programem podle RMS vybráno 100 nejlepších. K těmto

sítím byly vypočítány hodnoty RMS, MAE a jejich rozdíl. Na základě těchto hodnot vyberte nejvhodnější síť pro predikci retenčních charakteristik separované směsi. **~15 minut**

### 3.3.4 Výběr optimální mobilní fáze

Byla provedena separace modelové směsi (obsahující Ade, dAde, Ino, dIno) za třech vybraných podmínek (viz. 3.3.3). Zhodnoťte reálné chování systému v porovnání s předpovězenými hodnotami. Proveďte výběr optimální mobilní fáze. Do protokolu uveďte procentuální odchylky předpovězených hodnot od naměřených a diskutujte rozdíly. **~10 minut**

## 3.4 Postup - Validace metody

V této části budete provádět částečnou validaci optimalizované metody stanovení čtyř analytů v modelovém vzorku. Pro přípravu směsi standardů použijte zásobní roztoky o koncentraci 10mM.

V případě provádění opakování experimentů je vhodné provést také testování na odlehlost krajních výsledků. V případě, že Vám testování odhalí odlehlý výsledek, zvažte jeho vyloučení (vylučovat by se měla pouze ta data, která jsou zatížena hrubou chybou).

Pro intervaly spolehlivosti a testování statistické významnosti rozdílů si zvolte hladinu významnosti 0,05 nebo 0,01. Problém volby hladiny významnosti si vysvětlíme na případu, který budete při vypracovávání protokolu řešit – bod. 3.4.4. Hladina významnosti znamená tzv. *chybu I. druhu*, tzn. pravděpodobnost, že zamítneme *nulovou hypotézu* i přes to, že platí. Nulovou hypotézou může být např. tvrzení „pozorovaná výtěžnost metody je 100%“. Pokud budeme testovat na hladině významnosti 0,05, počítáme s tím, že v pěti procentech případů se stane, že nám vyjde rozdíl statisticky významný (tj. zamítneme rovnost námi naměřenými a teoretickými výtěžnostmi) i přes to, že ve skutečnosti není. V případě, že spočítáme intervaly spolehlivosti pro hladinu významnosti 0,01 vyjdou nám větší – pravděpodobnost zamítnutí pravdivé nulové hypotézy (námi pozorovaná průměrná výtěžnost se neliší od 100 %) tak bude logicky menší. Při výběru hladiny významnosti je proto vhodné mít na paměti definici dané nulové hypotézy, kterou chci testovat a dle toho si pak zvolit patřičnou pravděpodobnost, že ji zamítnu i když bude pravdivá.



### 3.4.1 Linearita

Linearitu ověřte měřeními 7 (případně více) koncentrací standardu. Krajní koncentrace je nutno volit tak, aby bylo možné určit dynamický rozsah metody – minimálně jeden kalibrační bod u hladiny limitu detekce a jeden nad hranicí linearity. V našem případě sestrojte kalibrační křivku v rozsahu 1-0,004mM. Sestrojte kalibrační závislost a určete dynamický rozsah metody (body mimo dynamický rozsah do lineární závislosti nezařazujte, pouze je zobrazte v kalibračním grafu). Pro lineární závislost otestujte významnost obou parametrů lineární rovnice a např. pomocí programu Statistika zobrazte také příslušné intervaly spolehlivosti. Nakonec určete zda je koeficient determinace větší než kritická hodnota (viz. Tab. 1) a byla tak statisticky potvrzena linearita závislosti. **~1 hodina**

### 3.4.2 Opakovatelnost

Připravte si směsi všech čtyř standardních látek v ekvimolárním množství na 3 koncentračních hladinách – 0,5; 0,05 a 0,005 mM. Každou směs změřte 5× za sebou. Pro retenční čas a plochu píků vypočítejte směrodatnou odchylku (SD), relativní směrodatnou odchylku (RSD) a intervaly spolehlivosti. **~1,75 hodiny**

### 3.4.3 Dlouhodobá opakovatelnost

Pro zjištění dlouhodobé opakovatelnosti Vám budou poskytnuta data od kolegů, kteří úlohu vypracovali před Vámi (minimálně 4 skupiny; celkem budete mít k dispozici pětice hodnot). K výpočtu dlouhodobé opakovatelnosti vezměte vždy první měření pro každou koncentrační hladinu. Jako variabilita se do dlouhodobé opakovatelnosti promítne měření v delším časovém úseku a měření jiným pracovníkem. Dlouhodobou opakovatelnost retenčních časů a ploch píků opět vyjádřete jako příslušné směrodatné odchylky (SD), relativní směrodatné odchylky (RSD) a intervaly spolehlivosti na zvolené hladině významnosti. Vyhodnocení této části dodejte

### 3.4.4 Správnost metody

Správnost metody ověřte změřením samotného modelového vzorku a vzorků s přísávkou všech čtyř standardů na dvou koncentračních úrovních (0,1 a 0,01mM). Měření každého vzorku proveďte minimálně třikrát. Přísávek Správnost metody vyjádřete jako výtěžnost v procentech pro kterou určete interval spolehlivosti na zvolené hladině významnosti. Do protokolu uveďte mimo jiné „grafické“, případně početní zhodnocení statistické významnosti rozdílu průměrné výtěžnosti metody a 100% hodnoty (bližší

informace viz. Poznámka ke statistickému zpracování, návod do podzimních laboratoří).

**~1 hodina**

### **3.4.5 Limit metody (LOD a LOQ)**

LOD a LOQ zjistěte dvěma způsoby: 1) z kalibrační přímky a nástřiku slepého vzorku (mobilní fáze, případně destilovaná voda jako rozpouštědlo ve vzorku) a 2) početní metodou z chromatogramu získaného změřením směsi s nízkými obsahy analytů a patrným šumem základní linie jak je popsáno v odstavci 2.3.3.5. Do protokolu uveďte porovnání takto zjištěných limitů a v diskuzi srovnajte jejich správnost.