

UNIVERZITA PALACKÉHO V OLMOUCI

Přírodovědecká fakulta

Katedra biochemie



**Změny metabolomu buněk v kultuře účinkem
léčiv**

BAKALÁŘSKÁ PRÁCE

Autor:	Jaroslava Mokrá
Studijní program:	B1406 Biochemie
Studijní obor:	Biochemie
Forma studia:	Prezenční
Vedoucí práce:	doc. RNDr. Tomáš Adam, Ph.D.
Konzultant:	RNDr. Karel Hron, Ph.D.
Termín odevzdání práce:	Srpen 2011

Prohlašuji, že jsem předloženou bakalářskou práci vypracovala samostatně za použití citované literatury.

V Olomouci dne

Jaroslava Mokrá

Chtěla bych poděkovat svému vedoucímu práce doc. RNDr. Tomáši Adamovi, PhD., RNDr. Karlu Hronovi, PhD. za odborné konzultace ve věcech statistiky, Mgr. Petru Wojtowiczovi za poskytnutí dat a všem ostatním, kteří mě ve psaní podporovali.

Tato práce vznikla za podpory projektu, jehož infrastrukturální část (Ústav molekulární a translační medicíny) byla pospořena Operačním programem Výzkum a vývoj pro inovace (projekt CZ.1.05/2.1.00/01.0030).

Bibliografická identifikace:

Jméno a příjmení autora	Jaroslava Mokrý
Název práce	Změny metabolomu buněk v kultuře účinkem léčiv
Typ práce	Bakalářská
Pracoviště	Katedra biochemie
Vedoucí práce	doc. RNDr. Tomáš Adam, Ph.D.
Rok obhajoby práce	2011
Abstrakt	Látka 5-fluorouracil se používá jako cytostatikum v léčbě onkologicky nemocných pacientů. Nepůsobí pouze na buňky postižené nádorovým bujením, ale i na buňky zdravé – má vliv na organismus jako celek, nejen na postiženou tkáň. Fibroblasty jsou primárně nesespecializované buňky vazivové tkáně, které jsou ideální pro laboratorní účely, protože dobře rostou a jedná se o lidský model. V této práci je popsán vliv 5-fluorouracilu na zdravé fibroblasty. Pomocí pokročilých statistických metod (shluková analýza, analýza hlavních komponent) jsou vyhodnoceny změny v metabolismu (převážně aminokyselin a organických kyselin) těchto fibroblastů po jejich kultivaci s 5-fluorouracilem.
Klíčová slova	5-fluorouracil, Metabolomika, Shluková analýza, Analýza hlavních komponent, Software R
Počet stran	42
Počet příloh	0
Jazyk	Český

Bibliographical identification:

Autor's first name and surname	Jaroslava Mokrá
Title	Metabolome changes in cultured cells effected by drugs.
Type of thesis	Bachelor
Department	Department of Biochemistry
Supervisor	doc.RNDr. Tomáš Adam, Ph.D.
The year of presentation	2011
Abstract	A compound 5-fluorouracil is used as a cytostatic agent in the treatment of oncology patients. It does not affect only the cells with cancer proliferation but it also affects the healthy cells – 5-fluorouracil has an effect on the organism as a whole, not just on the affected tissue. Fibroblasts are primarily non-specialized cells of connective tissue that are ideal for laboratory use. Fibroblasts grow well and represent a human model. This work describes the effect of 5-fluorouracil on healthy fibroblasts. Using advanced statistical methods (cluster analysis, principal components analysis) changes in metabolism (of amino acids and organic acids mainly) of fibroblasts after cultured with 5-fluorouracil are evaluated.
Keywords	5-fluorouracil, Metabolomics, Cluster Analysis, Principal Component Analysis, Software R
Number of pages	42
Number of appendices	0
Language	Czech

Cíle práce

- Zpracování literární rešerše na téma:
 - vliv 5-FU na metabolom fibroblastů,
 - hodnocení rozdílu v metabolomu pomocí pokročilých statistických metod.
- Teoretické zvládnutí statistického zpracování dat: PCA a shluková analýza.
- Statistické zpracování experimentu: vliv 5-FU na intracelulární obsah fibroblastů analyzovaný GCxGC-TOF.

Obsah

Cíle práce	6
Obsah.....	7
Teoretická část	8
1. 5-Fluorouracil (5-FU)	9
2. Fibroblasty	11
3. Metabolom, metabolomika	12
3.1. Úvod.....	12
3.2. Měření metabolomických dat.....	13
3.3. Zpracování metabolomických dat	15
4. Použité statistické nástroje.....	17
4.3. Shluková analýza	17
4.4. Analýza hlavních komponent (Principal Component Analysis, PCA)	21
4.5. Zobrazování pomocí biplotů	23
4.6. Software „R“	25
Praktická část	27
5. Postup.....	28
5.1. Instrumentace.....	28
GC x GC-TOF:.....	28
5.2. Buňky	28
Kultivace fibroblastů	28
Quenching a extrakce	29
Derivatizace	29
5.3. Statistické vyhodnocení	29
6. Diskuse	37
7. Závěr.....	39
Seznam použité literatury	40

Teoretická část

1. 5-Fluorouracil (5-FU)

5-FU je derivát pyrimidinu a je řazen do indikační skupiny cytostatik. Po metabolické transformaci inhibuje syntézu DNA, protože blokuje metylaci kyseliny deoxyuridilové na kyselinu thymidylovou. Inhibuje i syntézu RNA, protože způsobuje inkorporaci 5-fluoro-deoxyuridylmonofosfátu jako falešného prekurzoru do RNA. Působí specificky na buněčný cyklus v S-fázi. Maximální účinnost je pozorována ve tkáních s rychlou proliferací (např. kostní dřeň). Používá se pro léčbu různých karcinomů, například tlustého střeva, kůže, hlavy a krku.

Na trhu je k dispozici přes 40 let a byl vyvinut na základě pozorování chování maligní a nemaligní tkáně. Maligní tkáň zužitkovává uracil a tudíž i jeho deriváty ve větší míře než tkáň nemaligní. Jako analog uracilu slouží 5-FU jako substrát pro stejné transportní procesy zahrnující stejné enzymy metabolických drah. K projevení cytotoxicity vyžaduje, aby buňka, do níž vstupuje především usnadněným nukleosidovým transportním systémem, byla metabolicky aktivní.

5-FU je součástí anabolických i katabolických procesů. Cytotoxická aktivita záleží na anabolismu 5-FU do nukleotidů, kde inhibuje aktivitu thymidylát synthasy, nebo se inkorporuje do RNA nebo DNA. Enzymem dihydropyrimidin dehydrogenasou je inaktivován na dihydrofluorouracil. Degradace probíhá pomocí katabolických procesů.

5-FU je reversibilně za katalýzy thymidylát fosforylasou přeměněn na 5-fluoro-2'-deoxyuridin. Ten je poté fosforylován na 5'uhlíku deoxyribosového kruhu za vzniku 5-fluoro-2'-deoxyuridin-5-monofosfátu (FdUMP), který inhibuje enzym thymidylát synthasu (důležitá při DNA replikaci) a brání vzniku thymidinmonofosfátu. Mezi enzymem thymidylát synthasou, FdUMP a redukováným folátovým kofaktorem vzniká ternární komplex, který je reverzibilní. Stabilita tohoto komplexu je dána vnitrobuněčným obsahem 5,10-methyltetrahydrofolátu. Zlepšení cytotoxicity 5-FU je reprezentováno expanzí redukováného folátu a 5-formyltetrahydrofolátu. Důsledkem inhibice thymidylát synthasy je inhibice DNA syntézy jako následek vyčerpání thymidin-5'-monofosfátu a thymidin-5'-trifosfátu nebo po inkorporaci 2'-deoxyuridin-5'-trifosfátu do vznikající DNA. Genotoxický stres, který je vyvolán inhibicí thymidylát synthasy, může vyvolat programovanou buněčnou smrt.

5-FU je také inkorporován jak do nukleární RNA tak do cytoplasmatické. Cytotoxicita je závislá na rozsahu inkorporace 5-FU do RNA. Může dojít k inhibici konverze vysokomolekulární jaderné RNA na nízkomolekulární mRNA, k inhibici polyadenylace RNA se sníženou mRNA stabilitou a k změně sekundární struktury

RNA. Inhibice RNA syntézy nebo změna kvality RNA je závislá na koncentraci 5-FU, ale také na době expozice. Změna v RNA vede ke kvalitativním i kvantitavním změnám v syntézách proteinů. To opět může vést k apoptóze.

Příspěvek cytotoxického vlivu na RNA nebo na DNA je ovlivněn intracelulárním metabolismem buňky, který je jiný u nádorové buňky, než u zdravé buňky. Některé látky mohou účinnost 5-FU ovlivňovat. Jedná se například o methotrexát, cisplatinu nebo radiační záření, které cytotoxicitu zvyšují (Grem, 2000).

Clearance je řízena aktivitou cytosolového enzymu dihydropyrimidin dehydrogenasy, který katalyzuje reakci přeměny 5-fluorouracilu a dihydrofluorouracil. 5-FU je látka, která je velmi rychle metabolicky eliminována a v plazmě je největší množství detekováno mezi 5–15 minutami. Poté koncentrace v plazmě klesá. Jde o tzv. nelineární eliminační kinetiku.

Toxicita 5-FU je detekována v primárně rychle proliferujících tkáních. Většinou se týká gastrointestinálního traktu, kde jsou nežádoucí účinky popisovány záněty trávicí trubice doprovázené průjmy, zvracením a nauzeou. V krevním obraze se toxicita projevuje granulocytopenií, trombocytopenií a megaloblastickou anémií. Dále se může jednat o nežádoucí účinky projevující se na kůži (např. vyrážka, fotosenzitivita), záněty spojivek, poruchami nervového systému, srdečními arytmiemi a bolestmi na hrudi. Největšími komplikacemi při léčbě trpí ženy a starší osoby v důsledku jiné aktivity enzymu dihydropyrimidin dehydrogenasy (Grem, 2000; Kun, 2001).

2. Fibroblasty

Fibroblasty jsou základní buňky vazivové tkáně, jsou primárně nesespecializované, odvozené od mezenchymu, syntetizují mnoho látek (extracelulární matrix, glykoproteidy, atd.), sekretují cytokiny a organizují molekuly fibrózních proteinů – kolagenu a elastinu. Mají protáhlý, vřetenovitý až hvězdicovitý tvar s oválným jádrem, které obsahuje jedno až dvě jádérka.

Jejich význam spočívá v procesech tkáňové reparace. Jsou ideální pro laboratorní účely, protože mají dobře definovaný metabolom, jsou jednoduché, můžeme předpovídat jejich chování, dobře rostou a hlavně se jedná o lidský model (Prachařová, 2009).

3. Metabolom, metabolomika

3.1. Úvod

Metabolomika se zabývá identifikací a kvantifikací všech metabolitů v buňce za daných podmínek. V roce 2002 formuloval Fiehn první definici, podle které je metabolomika všeobecným a přesným vědním oborem, využívajícím analytické nástroje ke studiu metabolitů určitého biologického systému a statistické nástroje k vyhodnocení získaných výsledků.

Metabolom je definován jako kvalitativní a kvantitativní soubor všech metabolitů, které se nacházejí v buňce a odráží fyziologický stav této buňky, protože se účastní metabolických reakcí. Je nezbytný pro udržování aktivity, růst a normální fungování buňky. Spolu s dalšími vědními disciplínami (genomika, transkriptomika a proteomika) napomáhá k porozumění buněčné funkce. Metabolom hraje důležitou roli při kontrole buněčných funkcí a při regulaci metabolismu. Je reprezentován mnoha skupinami látek, které se liší chemickými a fyzikálními vlastnostmi. Rozlišujeme endometabolom (úplný soubor intracelulárních metabolitů) a exometabolom (soubor metabolitů vylučovaných do růstového média nebo extracelulární tekutiny). Charakteristika těchto dvou pojmů je důležitá, protože exometabolom má často velmi odlišnou fyziologickou úlohu než endometabolom. Velikost metabolomu je u různých biologických systémů velmi rozmanitá a závisí na konkrétních vlastnostech organismu.

Metabolomický výzkum používá několik hlavních přístupů – přímou analýzu, metabolické profilování, metabolický fingerprinting a footprinting a metabonomiku. Studium přímého vlivu hlavních efektů na genetické změny se zabývá přímá analýza (Prachařová, 2009).

Přímá analýza studuje přímý vliv hlavních efektů na několik málo metabolitů, které souvisejí se specifickou metabolickou reakcí.

Metabolické profilování je zaměřeno na analýzu daného a předem definovaného souboru metabolitů, například aminokyselin, nebo na cílenou specifikaci vybraných látek např. biomarkerů onemocnění, toxických agens nebo substrátů a produktů enzymových reakcí. Ve farmacii se používá k pozorování katabolismu léků a drog.

Metatabolický fingerprinting srovnává vzory – „fingerprinty“ látek, které se mění při výskytu nemocí, či jako odezva na expozici toxinů nebo vlivem dalších změn.

Metabolický footprinting se zabývá analýzou pouze extracelulárních metabolitů prostřednictvím stanovení určitých metabolitů nebo spekter, které neposkytují informaci o určitém metabolitu.

Metabonomika hodnotí změny hladiny endogenních metabolitů analýzou tkání a tekutin a uplatňuje se při studiu vlivu léčiv na organismus.

Oproti jiným -omickým vědám má metabolomika mnoho výhod. Jelikož je počet metabolitů v buňce menší než počet genů (popř. proteinů), je snazší získat celý metabolom v rozsáhlých analýzách. Koncentrace metabolitů se může měnit, i když se koncentrace enzymů a metabolický tok v biochemických reakcích nemění. Metabolom je konečným produktem genové exprese, ale zároveň odráží i fenotyp. Je levnější.

Metabolomické informace neslouží jen pro hlubší porozumění metabolickým drahám a jejich odpovědí na změny (např. genetické), ale mimo jiné poskytuje náhled na princip fenotypu v závislosti na vývoji, fyziologii, tkáňové identitě, atd. (Pilátová, 2008; Wojtowicz, 2009).

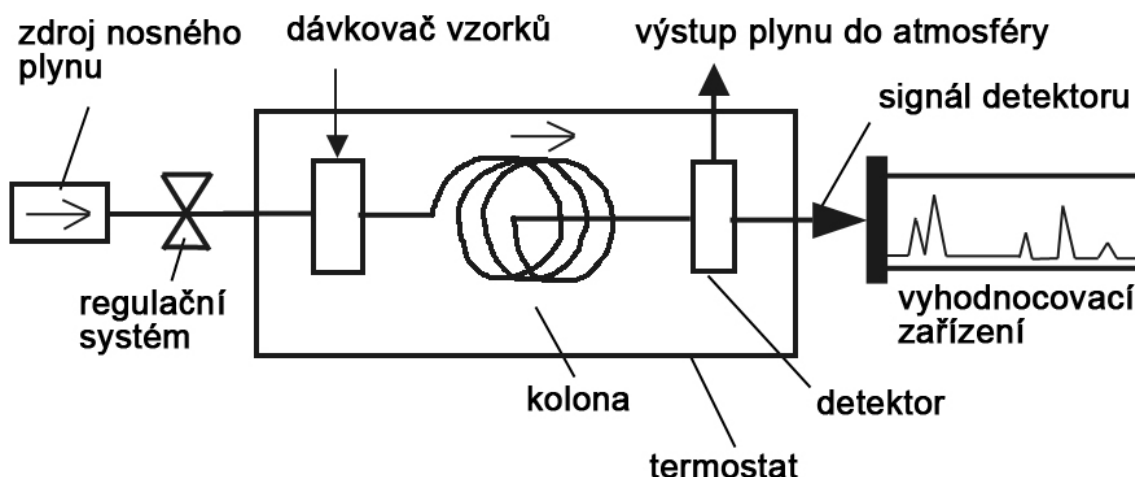
3.2. Měření metabolomických dat

K vlastnímu měření metabolomu je používána řada separačních a detekčních technik, například plynová chromatografie (GC), kapalinová chromatografie (HPLC) a kapilární elektroforéza (CE), které se dají spojit s hmotnostní spektrometrickou detekcí (MS). Vzorky pro tuto bakalářskou práci byly měřeny metodou GC × GC–TOF MS (dvojdimenzionální plynová chromatografie ve spojení s hmotnostní spektrometrickou detekcí s průletovým analyzátozem).

Chromatografie je separační metoda, při níž se oddělují složky obsažené ve vzorku, tzv. analyty. Vzorek je umístěn na počátek separační kolony, nepohyblivé stacionární fáze, a je unášen mobilní fází. Různé analyty jsou ke stacionární fázi poutány různými silami. Složky, které nejsou ke stacionární fázi poutány vůbec nebo minimálními silami jsou mobilní fází unášeny rychleji, eluují se z kolony dříve a jsou dříve detekovány, mají kratší retenční čas (čas mezi nástřikem vzorku na kolonu a její detekcí) (Klouta, 2003).

V případě plynové chromatografie je vzorek zplyněn v dávkovacím prostoru (inlet) a poté je dávkován do proudu nosného plynu. Jako nosné plyny se nejčastěji používají argon, dusík, vodík nebo helium. Analyty se rozdělují mezi stacionární a mobilní fázi na základě rozdílů v těkavosti a struktuře látek. Jako stacionární fáze se používají různé alifatické uhlovodíky a polymery na bázi dimethylsiloxanu (nepolární stacionární fáze) nebo ethylenglykolu (polární stacionární fáze). Další možností jsou směsné fáze, které však slouží spíše pro speciální účely (Wojtowicz, 2009).

Plynový chromatograf (viz obr. 1) se skládá ze zdroje nosného plynu, dávkovače vzorku, kolony a detektoru, na který navazuje vyhodnocovací zařízení. Dávkovač vzorku, kolony a detektor jsou termostatovány.



Obr. 1 Schéma plynového chromatografu, převzato a upraveno z (Klouda, 2003)

Jako zdroj nosného plynu slouží tlaková láhev s některým z permanentních plynů. Na zdroj nosného plynu bývá připojeno čistící zařízení, které zbavuje nosný plyn vlhkosti a nečistot. Následuje regulační systém, který zajišťuje potřebný tlak nosného plynu.

S pomocí dávkovače je vzorek zaváděn do proudu nosného plynu. Nástřik vzorku se provádí mikrostříkačkou několika metodami: nástřik na kolonu – on column (objem vzorku je v rozmezí 1–10 μl), nástřik pomocí děliče toku pro malé objemy vzorku (split, přibližně 0,1–2 μl) a nástřik bez děliče toku pro větší objemy vzorku (splitless, přibližně 0,5–5 μl).

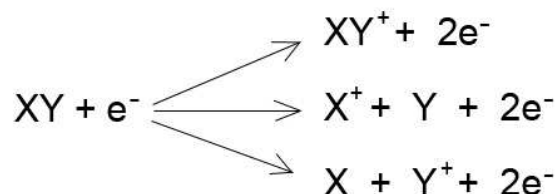
Kolony se používají dvojího typu:

- Kolony náplňové jsou naplněné adsorbenty (silikagel, grafitizovaný uhlík, oxid hlinitý) nebo molekulovými síty (hlinitokřemičitany). Nosiče (např. oxid křemičitý) jsou pokryty kapalnou fází. Jsou to trubice vyrobené z oceli nebo skla o průměru 2 až 3 mm a délce 1 až 3 metry.
- Kapilární kolony jsou vyráběny z taveného křemene a pokryty polyimidovou vrstvou, která má ochrannou funkci. Stacionární fáze je navázána přímo na vnitřní stěnu kapiláry, která má průměr 0,1 až 0,6 mm a délku 15–60 m. Podle uložení stacionární fáze rozlišujeme kolony WCOT (kapalná stacionární fáze tvoří tenký film na vnitřní stěně kapiláry), SCOT (na vnitřní stěně kapiláry je vrstva nosiče s navázanou kapalnou fází) a PLOT (na vnitřní stěně kapiláry se nachází jako adsorbenty pórovité materiály).

Plyn vycházející z kolony je transportován k detektoru, kde jsou analyty detekovány na základě vysílaného signálu v závislosti na čase. Detektory musí mít dostatečnou citlivost a selektivitu pro stanovované analyty. Detektory dělíme na univerzální a selektivní. Mezi univerzální patří tepelně vodivostní detektor a plamenově

ionizační detektor. Mezi selektivní patří detektor elektronového záchytu (Klouda, 2003; Opekar. et al., 2003; Wojtowicz, 2009).

Hmotnostní spektrometrie je založena na principu ionizace a rozštěpení molekul analyzovaných látek na soubor fragmentů, které se liší poměrem hmotnost/náboj (m/z). Při ionizaci látky XY (viz obr. 2) vzniká molekulový ion (XY^+), který může dle množství dodané energie, dále fragmentovat (X^+ , Y^+).



Obr. 2 Ionizace XY , převzato a upraveno z (Opekar. et al., 2003)

Hmotnostní spektrometr se skládá z iontového zdroje, hmotnostního analyzátoru a detektoru.

V případě spojení GC/MS (plynová chromatografie jako separační technika a MS jako detektor) vzorek v podobě par vstupuje do iontového zdroje, kde je bombardován elektrony. Součástí iontového zdroje je i akcelerační komora, kde jsou vzniklé ionty urychlovány do analyzátoru. Jako hmotnostní analyzátor jsou používány kvadrupólový analyzátor, kvadrupólová iontová past, průletový analyzátor. Iontový zdroj i analyzátor se nachází v evakuovaném prostoru, aby nedocházelo k nežádoucím srážkám iontů. Hmotnostní spektrum se zobrazuje jako závislost relativního zastoupení fragmentů na poměru m/z . V případě tvrdých ionizačních technik (je dodán nadbytek energie) dochází k bohaté fragmentaci a vzniklá MS spektra jsou pro danou látku charakteristická.

Spojením GC a MS vzniká velmi účinný přístroj pro identifikaci i kvantitativní stanovení nejrůznějších analytů (Wojtowicz, 2009).

3.3. Zpracování metabolomických dat

Po sběru dat následuje jejich úprava, vyhodnocení a interpretace. Celkovým zpracováním výsledků se zabývá bioinformatika, která využívá různých softwarových programů, informačních databází a nejrůznějších statistických metod (Wojtowicz, 2009).

V první fázi, tzv. pre-processing, dochází k dekonvoluci dat (rozdělení překrývajících se píků), určení jednotlivých sloučenin vzorku (kvalitativní analýza) a jejich relativního zastoupení (kvantitativní analýza). Cílem této fáze je získat tabulku dat $n \times p$ (n je počet objektů = počet vzorků, p je počet proměnných, např. metabolitů).

Vyhodnocení výsledků provádíme především pomocí statistických metod. Ty jsou určeny jak pro analýzu jednorozměrných dat, tak pro analýzu mnohorozměrných dat (Goodacre et al., 2007).

S analýzou jednorozměrných dat se při vyhodnocování a zpracování metabolomických dat téměř nesetkáváme, protože metabolomické experimenty poskytují mnohorozměrná data. Ale i tak je možné jednorozměrnou analýzu použít k hodnocení jednotlivých metabolitů, jejichž hodnoty se mezi různými skupinami významně mění. K tomu se používají například testy shodnosti.

Jelikož metabolomická data jsou zpravidla mnohorozměrná, je k jejich vyhodnocení používána mnohorozměrná analýza. Ta vychází z tabulky dat $n \times p$. Z důvodu velkého množství vstupních proměnných je nezbytné extrahovat základní informaci v datech obsaženou. Ke snížení rozměrnosti dat slouží např. metoda hlavních komponent, shluková analýza nebo korelační analýza. V metabolomice se navíc setkáváme s problémem, kdy množství naměřených proměnných je větší než počet vlastních měření, například z důvodů nedostatečného množství vzorku a hlavně se většina metabolitů v závislosti na čase přeměňuje (Johnson & Wichern, 2007).

4. Použité statistické nástroje

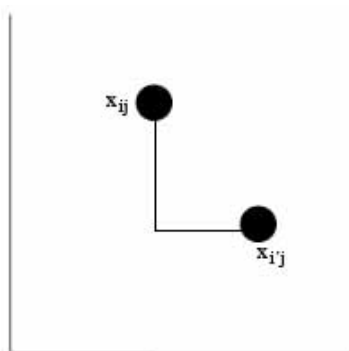
4.3. Shluková analýza

Shluková analýza je metoda zabývající se podobností vícerozměrných objektů, které klasifikuje do shluků, právě na základě jejich podobnosti. Využíváme ji zejména tam, kde se množina objektů sama o sobě rozpadá do tříd, tj. objekty se přirozeně seskupují do shluků. Cílem je, aby si objekty uvnitř shluku byly co nejvíce podobné a objekty z různých shluků co nejvíce rozdílné. Shluková analýza byla primárně vytvořena a vyvinuta pro účely biologie k porovnávání rozdílů mezi organismy a posléze se její využití rozšířilo i na chemii, molekulární biologii a biochemii.

Základem je datová matice $n \times p$ (n je počet objektů, p je počet proměnných). Pro hodnocení vzdálenosti či podobnosti objektů se používá výpočet příslušných měř pro všechny páry objektů, kdy vzniká symetrická čtvercová matice typu $n \times n$. V matici měř vzdáleností jsou na diagonále nuly, v matici měř podobnosti jsou na diagonále jedničky. K určení vzdálenosti či podobnosti objektů (proměnných) využíváme:

- *Hemmingova vzdálenost (Manhattan, viz obr. 3)* se počítá jako součet absolutních hodnot rozdílů složek vektorů mezi i -tým a i' -tým objektem:

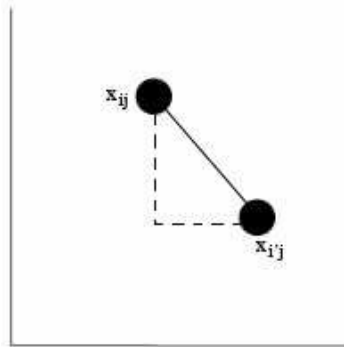
$$D_H(x_i, x_{i'}) = \sum_{j=1}^p |x_{ij} - x_{i'j}|, \quad x_i = (x_{i1}, \dots, x_{ip}), \quad x_{i'} = (x_{i'1}, \dots, x_{i'p}).$$



Obr. 3 Hemmingova vzdálenost, převzato a upraveno z (Hebák et al., 2007)

- *Euklidovská vzdálenost (viz obr. 4)*. Obdobně jako u Hemmingovy vzdálenosti i zde platí, že čím menší je její hodnota, tím více jsou si proměnné (objekty) podobné. Počítá se jako druhá odmocnina ze součtu čtverců rozdílů složek

vektorů mezi i -tým a i' -tým objektem:
$$D_E(x_i, x_{i'}) = \sqrt{\sum_{j=1}^p (x_{ij} - x_{i'j})^2}.$$



Obr. 4 Euklidovská vzdálenost, převzato a upraveno z (Hebák et al., 2007)

- *Mahalanobisova vzdálenost* je podobná euklidovské, ale respektuje rozdílnost ve variabilitě a korelaci proměnných jednotlivých objektů, jinak řečeno, přihlíží ke kovarianční struktuře datového souboru. Tato vzdálenost by neměla být použita v těch případech, kdy je počet proměnných větší než počet objektů.

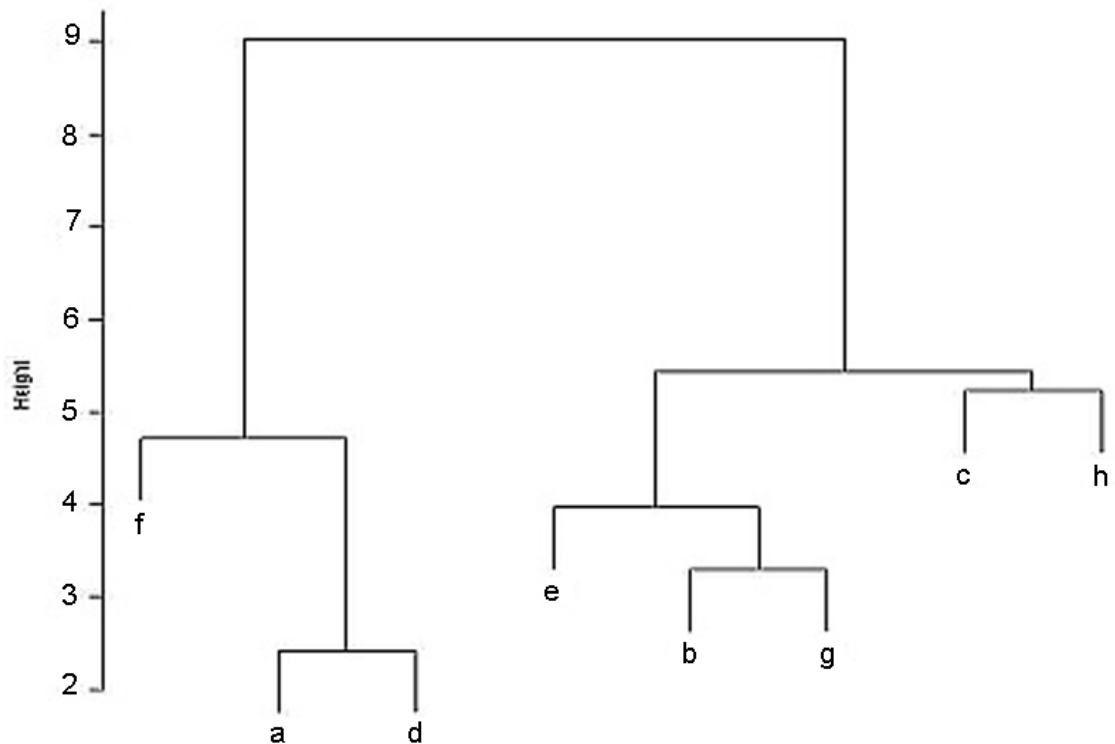
$$D_M(x_i, x_{i'}) = \sqrt{(x_{ij} - x_{i'j}) \cdot S^{-1} \cdot (x_{ij} - x_{i'})} \dots$$
 matice S představuje varianční matici datového souboru (Hebák et al., 2007).

Všechny míry jsou závislé na použitých měřicích jednotkách. Pro úpravu naměřených proměnných se často využívá vhodné transformace, např. mocninné transformace, clr (centred log-ratio) transformace, normalizační transformace, symetrizující transformace, atd.

Postupy způsobu shlukování můžeme dělit na *hierarchické* a *nehierarchické*. Hierarchické postupy dále dělíme na *aglomerativní* a *divizní*.

Nehierarchické metody rozkládají výchozí množinu objektů do několika podmnožin, které se dále mění. Cílem je dosáhnout optimální hodnoty kritéria kvality rozkladu specifického pro danou metodu (matice vnitroskupinové, meziskupinové a celkové variability).

U hierarchických metod dochází k postupnému spojování objektů do větších shluků. Vždy se nejprve vypočítá základní matice vzdáleností mezi objekty. Výsledky zaznamenáváme graficky v podobě stromového diagramu, který nazýváme dendrogram. Na vertikální ose je stupnice pro hladiny spojování a dole začíná strom n větvemi, z nichž dvě v každém kroku spojíme v bodě, který odpovídá příslušné hladině spojení (viz obr. 5). V tomto dendrogramu jsou patrné dva shluky, z nichž každý se dělí na podshluky. První shluk je tvořen podshlukem objektů „a“ a „d“ a dále objektem „f“ s tím, že objekty „a“ a „d“ jsou si podobné, ale v rámci shluku jsou nepodobné s objektem „f“. Ale při hodnocení celého dendrogramu, jsou si tyto tři objekty podobné. Druhý shluk je tvořen dvěma podshluky. Objekty „e“, „b“ a „g“ tvoří první podshluk a jsou si podobné, ale zároveň jsou v rámci shluku nepodobné s objekty „c“ a „h“, které tvoří druhý podshluk.



Obr. 5 Dendrogram

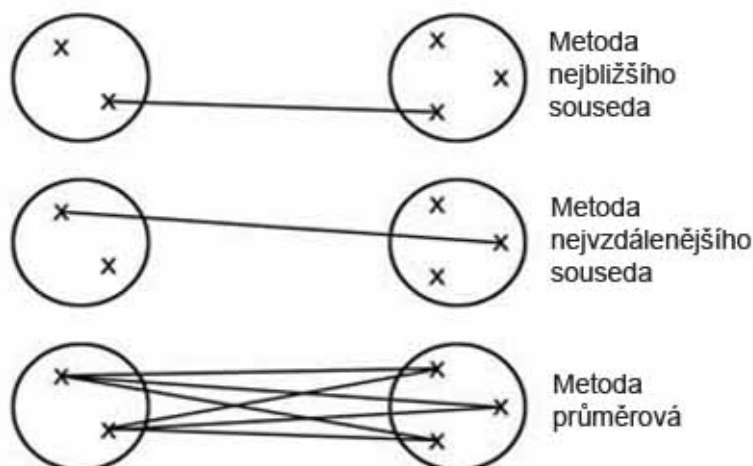
U aglomerativních metod je před zahájením shlukování počet shluků roven počtu objektů (proměnných). Proces shlukování začíná nalezením a spojením dvou objektů, u nichž je vzdálenost minimální (shlukování na základě podobnosti), do prvního shluku. Poté se matice vzdáleností mezi objekty přepočítá s tím, že objekty z prvního shluku jsou vynechány a shluk je zařazen jako celek. Tím dojde ke snížení řádu matice vzdáleností o jedničku. Postup opakujeme do té doby, než zůstane předem daný počet shluků nebo dokud nevznikne shluk jeden. U divizního shlukování vycházíme z jednoho shluku a v každém kroku je jeden ze shluků rozštěpen na dva na základě co největších vzdáleností (nepodobností) mezi shluky. Postupným dělením skončíme až u jednotlivých objektů.

U aglomerativních postupů používáme několik metod shlukování:

- *Metoda single linkage* (metoda nejbližšího souseda, viz obr. 6) shlukuje jednotlivé subjekty podle vzdálenosti mezi nejbližšími členy. Pojem nejbližší člen poukazuje na nejmenší vzdálenost členů nebo největší podobnost mezi nimi. Neumí rozeznat slabě separované shluky, ale umí detekovat i shluky, které mají neelipsovité tvar. V případě této metody může dojít k řetězení objektů, a tudíž se v jednom shluku mohou objevit i objekty značně vzdálené, pokud mezi nimi další objekty vytvoří jakýsi most.
- *Metodou complete linkage* (metoda nejvzdálenějšího souseda, viz obr. 6) jsou objekty shlukovány na základě jejich maximální vzdálenosti. V každém kroku je

vzdálenost (nepodobnost) mezi shluky je určována vzdáleností (nepodobností) mezi dvěma nejbližšími objekty, přičemž je každý z jiného shluku. Je jednou z nejčastěji užívaných metod shlukování.

- Při použití *metody average linkage* (metoda průměrová, viz obr. 6) počítáme vzdálenost mezi dvěma shluky jako průměr vzdáleností mezi všemi členy z daných dvou shluků. Shluky s nejmenší průměrnou vzdáleností všech svých objektů jsou si nejbližší.



Obr. 6 Metody shlukování, převzato a upraveno z (Johnson & Wichern, 2007)

- *Wardova metoda* používá jako kritérium pro spojování shluků přírůstek celkového vnitroskupinového součtu čtverců odchylek pozorování od odhadu střední hodnoty datového souboru. Má tendenci odstraňovat malé shluky a vytvářet shluky přibližně stejné velikosti.
- *Centroidní metoda* vyjadřuje vzdálenost shluků jako euklidovskou vzdálenost jejich těžišť.
- *Metoda mediánová* je „vylepšená“ centroidní metoda, protože odstraňuje rozdílné „váhy“, které centroidní metoda přikládá různě velkým shlukům (Hebák et al., 2007; Meloun & Militký, 2002)

Pomocí hierarchického shlukování můžeme zkoumat podobnost objektů (pomocí dendrogramu objektů) nebo také proměnných (dendrogram proměnných). Dendrogram proměnných odhaluje skupiny proměnných, které jsou si podobné a silně spolu souvisí (korelují). Ty jsou ve společném shluku, jsou tudíž navzájem nahraditelné. Míra věrohodnosti dendrogramu je dána kofenetickým korelačním koeficientem. Jedná se o Pearsonův korelační koeficient mezi skutečnou a předpokládanou vzdáleností. Dalším věrohodnostním kritériem je kritérium delta.

Mezi nejznámější nehierarchické metody patří shlukování *metodou nejbližších středů* (*k-means*). Tuto metodu používáme při vytváření malého počtu shluků z velkého

počtu objektů. Metoda rozděluje n objektů o p proměnných do k shluků, jejichž počet je předem dané a součet čtverců mezi shluky musí být minimalizován. Počet možných uspořádání je obrovský, optimum je většinou nalezeno na lokální úrovni. Rozklad je zlepšován přesouváním objektů mezi shluky, dokud se nesníží hodnota stopy matice vnitroskupinové variability.

Na konci této kapitoly zmiřme též *fuzzy shlukování*, kdy může být jeden objekt součástí více shluků. Jednotlivým objektům u každého shluku je přiřazena příslušnost v intervalu $\langle 0;1 \rangle$, součet dosažených hodnot se u každého objektu musí rovnat jedné. Nula přitom znamená, že objekt nenáleží danému shluku a jednička znamená jednoznačné přiřazení k danému shluku (Hebák et al., 2007; Meloun & Militský, 2002; Johnson & Wichern, 2007; Metody shlukové analýzy)

4.4. Analýza hlavních komponent (Principal Component Analysis, PCA)

Některé analytické metody vedou k datové tabulce s velkým množstvím proměnných, které jsou pro interpretaci nepřehledné, proto je zapotřebí je nahradit menším počtem jiných proměnných aniž by došlo k významné ztrátě původní informace, která je reprezentována rozptýly (variabilitou) původních proměnných. K redukci vícerozměrných dat byla vytvořena metoda hlavních komponent, která se snaží o nalezení skrytých veličin, tzv. hlavních komponent, které jsou přímo neměřitelná a jsou lineárními kombinacemi původních proměnných. Nové proměnné musí co nejlépe reprezentovat variabilitu proměnných původních., jde přitom o to, aby již několik málo hlavních komponent vysvětlilo většinu variability původního datového souboru. Metoda hlavních komponent může být používána jako samostatná statistická metoda nebo jako pomocník při třídění dat pro jinou statistickou metodu (např. regresní analýza, shluková analýza).

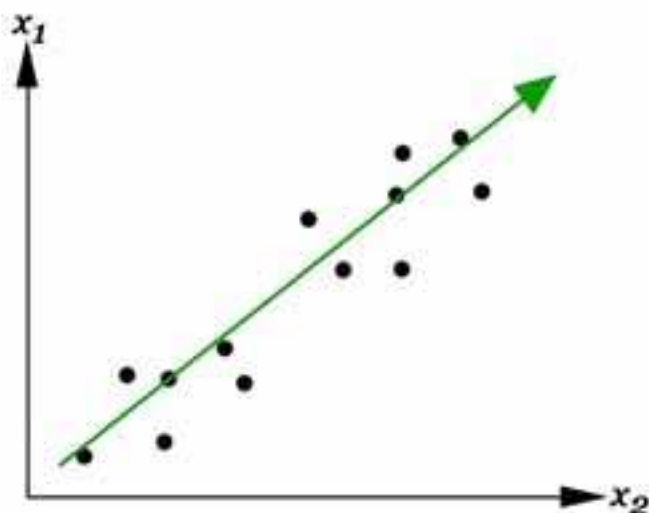
Data z tabulky s p proměnnými mohou být zobrazena v p -dimenzionálním prostoru a, jak již bylo zmíněno, analýza hlavních komponent tuto dimenzionalitu redukuje. Ovšem musíme předpokládat, že touto redukcí se ztratí i část vstupní informace, proto musí být provedena správně a logickým způsobem, aby ztracená informace byla co nejmenší.

V následujícím textu bude ukázán postup konstrukce hlavních komponent pro data v rovině, tedy reprezentovaná souřadnicemi x_1 a x_2 (viz obr. 7). Výsledkem budou dvě hlavní komponenty, PC1 a PC2 (viz obr. 8 a 9), které vytvoří nový souřadnicový systém. První hlavní komponenta, která je hlavní nositelkou původní informace, je vytvořena ortogonální projekcí bodů z roviny do přímky. Ta vede směrem největší změny (variability) v datech, protože PCA předpokládá, že největší informace je skryta

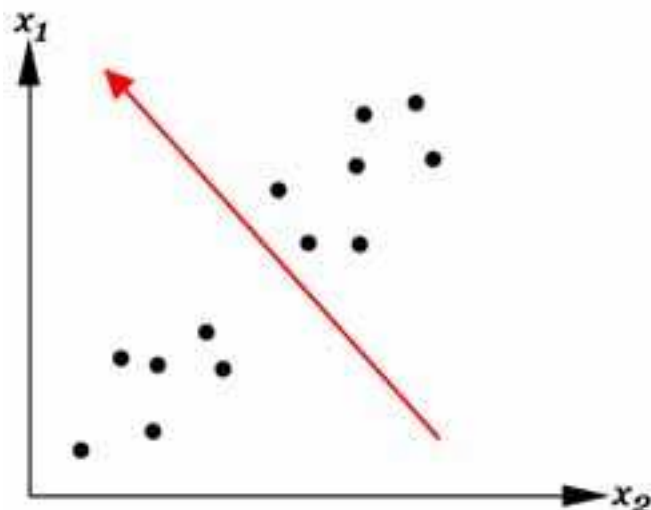
právě v co největší změně v datech. Druhá hlavní komponenta reprezentuje tu část informace, která není obsažena v PC1, tudíž je na PC1 závislá a kolmá. Po převedení dat do souřadnic PC1 (horizontální osa) a PC2 (vertikální osa) vznikne „nový“ graf (viz obr. 10). Je přitom zřejmé, že ke zjištění struktury zobrazeného datového souboru (dva shluky) nám postačí již první hlavní komponenta (Massart & Heyden, 2004; Massart & Heyden, 2005)



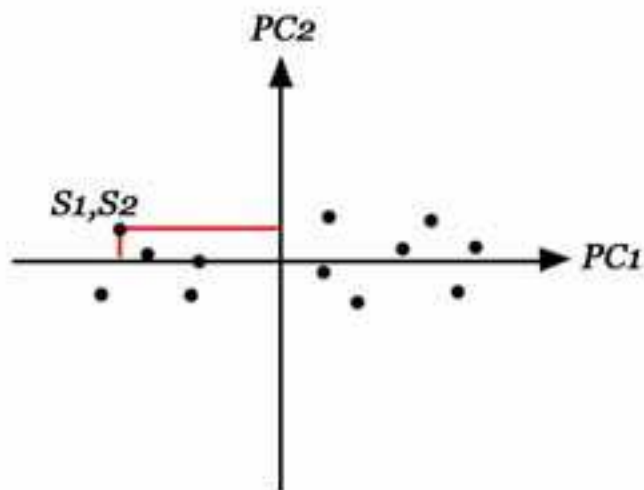
Obr. 7 Zdrojová data ve dvourozměrném prostoru, převzato a upraveno z (Massart & Heyden, 2004)



Obr. 8 PC1, převzato a upraveno z (Massart & Heyden, 2004)



Obr. 9 PC2, převzato a upraveno z (Massart & Heyden, 2004)



Obr. 10 Zobrazení v souřadnicích PC1 a PC2, převzato a upraveno z (Massart & Heyden, 2004)

4.5. Zobrazení pomocí biplotů

Metoda hlavních komponent je teoretickým základem pro konstrukci biplotu. Jedná se o dvojdimenzionální grafické znázornění objektů a proměnných ze zdrojové tabulky v jednom grafu. Objekty jsou přitom jako obvykle seřazeny v n řádcích a proměnné jsou umístěny ve p sloupcích s tím, že datová matice musí obsahovat nejméně dva sloupce.

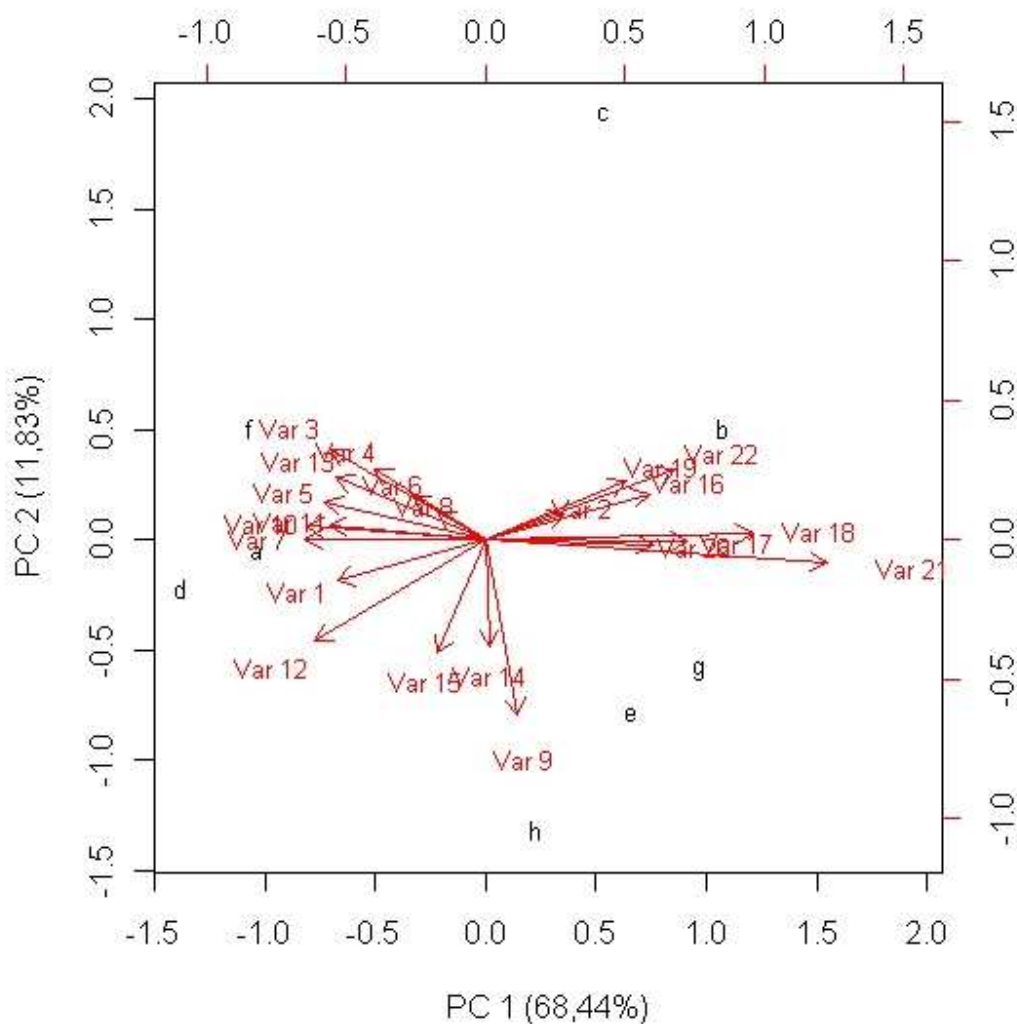
Pokud pracujeme jen se dvěma proměnnými (zdrojová matice obsahuje právě dva sloupce), rovinný graf reprezentuje informaci nesenou v obou proměnných. S narůstajícím množstvím proměnných dochází v biplotu k redukci dat pomocí metody hlavních komponent a k tvorbě plotu využíváme informaci z prvních dvou hlavních komponent.

Objekty jsou zobrazeny jako body a proměnné jsou reprezentovány šipkou (viz obr. 11). Čím je šipka delší, tím je vliv dané proměnné na uspořádání dat větší a čím je

úhel mezi šipkami menší, tím je lineární vztah odpovídajících si proměnných těsnější (konkrétně, délka šipky je přibližně rovna rozptylu příslušné proměnné a kosinus úhlu mezi šipkami reprezentuje hodnotu korelačního koeficientu mezi dvěma proměnnými). Procenta v biplotu značí procento původní informace, které je v biplotu zobrazeno v souřadnicích PC1 a PC2. V případě clr transformovaných dat je naopak těsnost vztahu proměnných reprezentována prostřednictvím délky vrcholů šipek (Kalivodová, 2010).

V ilustračním biplotu je procento původní informace PC1 = 68,44 % a PC2 = 11,83 % a je možné pozorovat:

- pro objekt „h“ jsou determinujícími proměnnými proměnné „14“ a „9“,
- pro objekt „a“ je významnou proměnnou proměnná „7“,
- objekt „b“ je určen proměnnými „2“, „16“, „19“ a „22“, které jsou v těsném vztahu a proměnná „22“ má největší vliv na jejich uspořádání,
- objekt „d“ je určen proměnnou „1“ a
- objekt „f“ je určen proměnnými „8“, „6“, „4“, „15“ a „3“, které jsou opět v těsném vztahu a proměnná „3“ má největší vliv na jejich upořádání.



Obr. 11 Biplot

4.6. Software „R“

R je software určený hlavně pro statistickou analýzu dat a jejich grafické znázornění, používající vlastní programovací jazyk, který se pro začátečníky může jevit poněkud obtížný. Ale tím, že existuje mnoho návodů, které se dají volně získat ze speciálních webových stránek, určených přímo programu R, není složité se základy práce s R naučit. Jazyk, ve kterém se v R programuje, je podobný dříve vytvořenému jazyku S. R je volně distribuovaný pod GNU licencí (General Public Licence) a instalační soubor je opět možné stáhnout z webových stránek www.r-project.org (Venables & Smith).

R bylo vytvořeno Robertem Gentlemanem a Rossem Ihakaem ze statistického oddělení univerzity v Aucklandu. Ovšem na jeho dnešním rychlém vývoji a zdokonalování funkcí se podílí speciální tým vědců po celém světě. Všechny funkce jsou dokumentovány formou pomocných webových stránek a většina užívaných funkcí

je popsána ve volně stažitelných člancích a manuálech. Za všechny jmenujme alespoň "An Introduction to R", který poskytuje základní uživatelské rady.

Součástí R je mnoho funkcí pro statistické analýzy a grafické zobrazování. Funkce vyžadují použití závorek, dokonce i když zůstávají prázdné. S daty a proměnnými se pracuje také pomocí operátorů (logické, aritmetické, porovnávací, atd.). Všechny akce se dějí na datech uložených v aktivní paměti počítače, nejsou vytvářeny dočasné soubory. Četby a zápisy souborů jsou používány pro vstup a výstup dat a výsledků. Uživatel vyvolá funkce pomocí příkazů. Výsledky jsou zobrazeny přímo na obrazovce, uloženy v objektu a nebo zapsány na disk (hlavně ty grafické).

R pracuje s objekty, které jsou charakterizovány svými jmény a obsahy, ale také atributy, které specifikují druh dat, které jsou reprezentovány objekty. Atributy v případě R poskytují nezbytnou informaci pro správné použití funkcí na objekty závisí na attributech. Všechny objekty mají dva vnitřní (vlastní) atributy: mode (režim) a length (délka). Mode je základní typ prvků objektu. Existuje režim numerický, logický, znakový, atd. Délka označuje počet prvků objektu.

Načítání dat se děje pomocí příkazů:

```
setwd("...")
x=matrix(scan("... .txt",dec=","),ncol=X,byrow=T)
x=read.table("... .txt")
x
```

Do uvozovek v prvním řádku píšeme přesnou cestu k uloženému souboru, který chceme načíst. Do uvozovek ve druhém a třetím řádku píšeme název konkrétního souboru. Data pro zpracování jsou většinou uložena jako textový soubor s příponou .txt. Chceme-li pracovat s daty obsahujícími desetinná místa, je nutné desetinné čárky nahradit tečkami.

R také nabízí pozoruhodnou grafiku. Výsledky grafických funkcí nejsou přiřazeny k objektu, ale jsou zobrazeny pomocí speciálního grafického okna nebo formou souboru (jako dokument .pdf, .jpeg, atd.). Existují dva druhy grafických funkcí. Funkce vykreslování na vysoké úrovni, která vytvoří nový graf, a funkce vykreslování na nízké úrovni, která přidává další prvky do stávajícího grafu. Grafy jsou vytvářeny s ohledem na grafické funkce, které jsou definovány výchozím nastavením a mohou být modifikovány.

Grafické výstupy mohou být „vylepšeny“ pomocí grafických parametrů. Mohou být využity jako možnosti grafických funkcí nebo lze trvale změnit grafické parametry podle přání uživatele. Mezi grafické parametry se řadí změna barvy (pozadí, písma, šipek,...) velikost a font písma, vzhled značek a šipek, atd. v biplotu (Paradis).

Praktická část

5. Postup

5.1. Instrumentace

GC × GC-TOF:

Pro analýzy byl použit systém Pegasus 4D sestávající se z plynového chromatografu Agilent 7890 (Agilent Technologies, Palo Alto, USA) a Pegasus HT hmotnostního spektrometru (LECO Co., St. Joseph, USA). GC × GC systém byl umožněn dvoustupňovým čtyřtryskovým modulátorem a sekundární pecí, oboje zabudováno do pece primární. Jako modulační plyn byl použit stlačený vzduch, pro horký pulz je vzduch ohříván teplotou modulátoru, pro chladný prochází nejprve filtrem pro odstranění stop vlhkosti a ochlazován je imersně (−80 °C).

Jako primární kolona byla použita BPX-5 (30 m × 0,25 mm × 0,25 μm) pro druhou dimenzi BPX-50 (2,0 m × 0,1 mm × 0,1 μm, obě SGE, Ringwood, Austrálie). Kolony byly spojeny pomocí spojky SilTite Mini Union (SGE, Ringwood, Austrálie) a to mezi modulátorem a sekundární pecí – modulace tedy probíhala na primární koloně.

Program primární pece byl následující: 40 °C (2 min), 8 °C/min 155 °C (0,2 min), 10 °C/min 255 °C (0,2 min), 100 °C/min 300 °C (2 min). Program sekundární pece kopíruje primární s offsetem +15 °C, program modulátoru s offsetem +50 °C. Modulační perioda 4 s. Jako nosný plyn bylo použito helium (99,999 %, Messer, ČR) v režimu korigovaného konstantního průtoku 1 ml/min. Byl použit splitless nástřik (1 μl, 250 °C, 1 min).

Time-of-flight hmotnostní spektrometr pracoval v režimu elektronové ionizace (−70 eV), s teplotou iontového zdroje a transferové linie 250 °C. Snímáno bylo hmotové rozmezí m/z 35–440 při rychlosti 100 Hz. Napětí na detektoru bylo −1 500 V, solvent delay 650 s.

Software ChromaTOF 4.33 (LECO Co., St. Joseph, USA) je použit jak pro ovládání přístroje a sběr dat, tak pro jejich základní zpracování (výpočet baseline, TruePEAK dekonvoluce signálu, automatická detekce píků a jejich kombinace, identifikace na základě knihovny spekter, výpočet plochy píků, export dat) (Wojtowicz et al., 2010).

5.2. Buňky

Kultivace fibroblastů

Linie fibroblastů (získáno od dobrovolného dárce z řad laboratorního personálu) byla kultivována v inkubátoru (37 °C, 5 % CO₂) standardním postupem v DME (Dulbecco's Modified Eagle's) médiu, které bylo doplněno na 10 % fetálního hovězího

séra a antibiotiky – streptomycinem (100 µg/ml) a penicilinem (100 U/ml). Buňky byly kultivovány v kultivačních lahvičkách (25 cm²) do 90% konfluence a poté pasážovány. Po rozkultivování na dostatečný počet lahviček byly buněčné linie sklizeny quenchingem.

Působení 5FU bylo umožněno přidáním látky (o koncentraci 20 µmol/l) do kultivačního média po dobu 1 (1_FU), 2 (2_FU), respektive 3 dnů (3_FU) před quenchingem. Experiment byl proveden v triplicátu.

Quenching a extrakce

Quenching probíhal pomocí 20ml plastové stříkačky se zahnutou jehlou přímo v kultivační nádobce. Do nádoby bylo vstříknuto 20 ml quenčovacího roztoku (60% methanol, -50 °C). Quenčované buňky byly poté dvakrát extrahovány 1 ml 80% methanolu. Spojené extrakty byly centrifugovány (5 min, 10 000 g), supernatant byl lyofilizován a uložen v mrazicím boxu (-20 °C) do analýzy.

Derivatizace

Před derivatizací bylo k suchým vzorkům (odfoukáno dusíkem) přidáno 50 µl methanolu, dále byly 5 min sonikovány a centrifugovány (1 min, 10 000 g). Supernatant byl přenesen do insertu a odfoukán dusíkem dosucha.

Vlastní derivatizace vzorků byla prováděna dle (Koek et al., 2006). K odfoukané směsi standardů insertu bylo přidáno 10 µl ethoxylamin-hydrochloridu (roztok v pyridinu, 56 mg/ml) a 40 µl pyridinu. Ethoxymace probíhala při teplotě 40 °C po dobu 90 min, následně bylo přidáno 30 µl silylačního činidla (MSTFA), silylace probíhala taktéž při teplotě 40 °C po dobu 50 min.

5.3. Statistické vyhodnocení

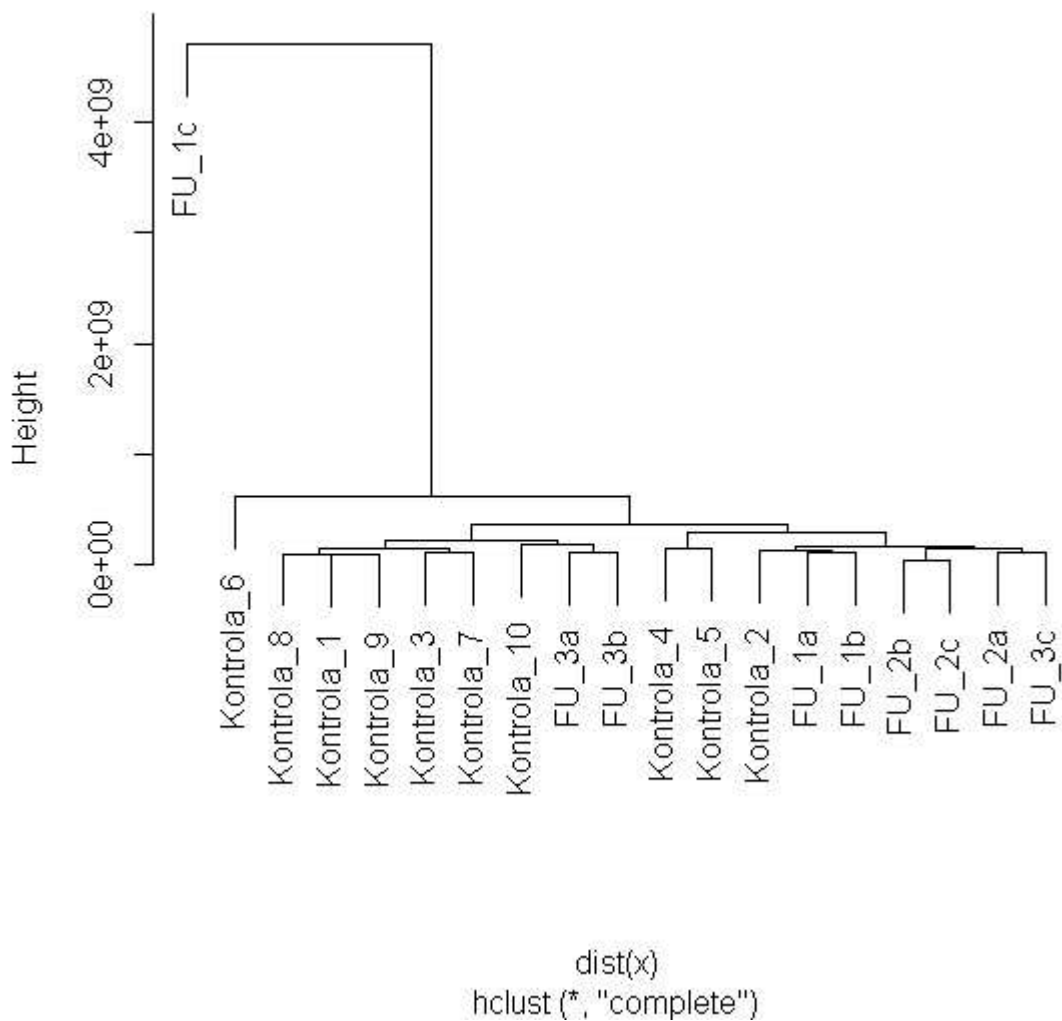
Pro statistické vyhodnocení dat byl použit program R (<http://www.r-project.org/>).

Výsledky byly vyhodnocovány podle plochy pod křivkou píku. K vyhodnocení byla použita shluková analýza (dendrogram) – metoda complete linkage (metoda nejbližšího souseda) a metoda hlavních komponent prostřednictvím biplotu.

Nejdříve byl sestaven dendrogram původních dat (viz obr. 12). Bylo zjištěno, že původní data nejsou pro účely zpracování vhodná, jsou nečitelná a dendrogram tak neposkytuje kvalitní informaci. Příčinou je odlišná geometrie výchozích pozorování, která souvisí s faktem, že tato nesou pouze relativní informaci (tzv. kompoziční data (Aitchison, 1986; Korhoňová et al., 2009)). Proto byla původní data transformována pomocí clr transformace (centred log-ratio). Tato transformace je tvořena logaritmy podílů původních proměnných s jejich geometrickými průměry. Dendrogram (viz obr.

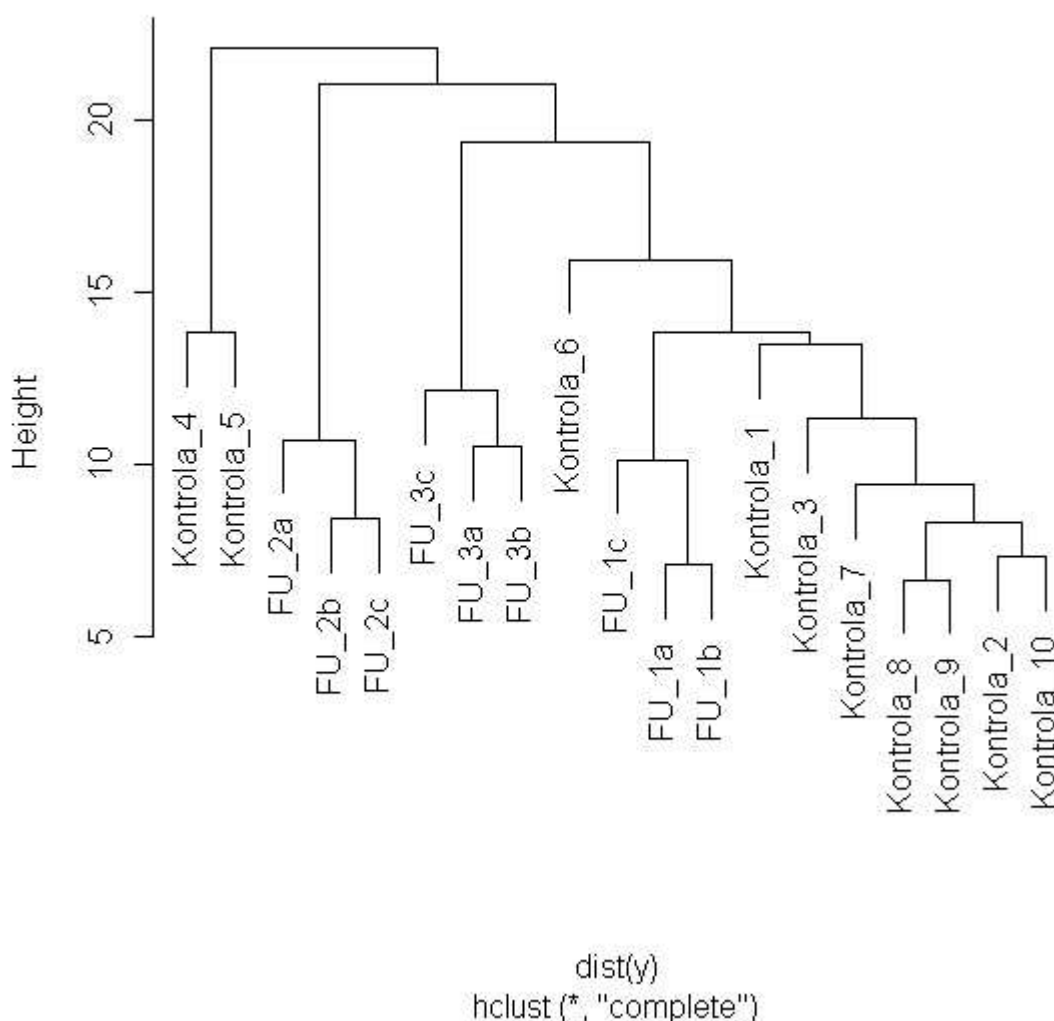
13) je přehledný a lze z něj vyčíst, že fibroblasty kultivované s fluorouracilem se shlukují do samostatných shluků tvořených FU_1, FU_2 a FU_3. Pouze kontroly 4 a 5 významně vybočují a tvoří vlastní shluk, a dále lehce vybočuje kontrola 6. FU_1 tvoří podshluk v rámci shluku kontrol, z čehož vyplývá, že nemá moc odlišné chování od kontrol (metabolom fibroblastů nebyl dostatečně ovlivněn fluorouracilem).

Cluster Dendrogram



Obr. 12 Dendrogram původních dat (vzdálenost původních dat – dist(x), shluková analýza – hclust, metoda nejvzdálenějšího souseda – „complete“)

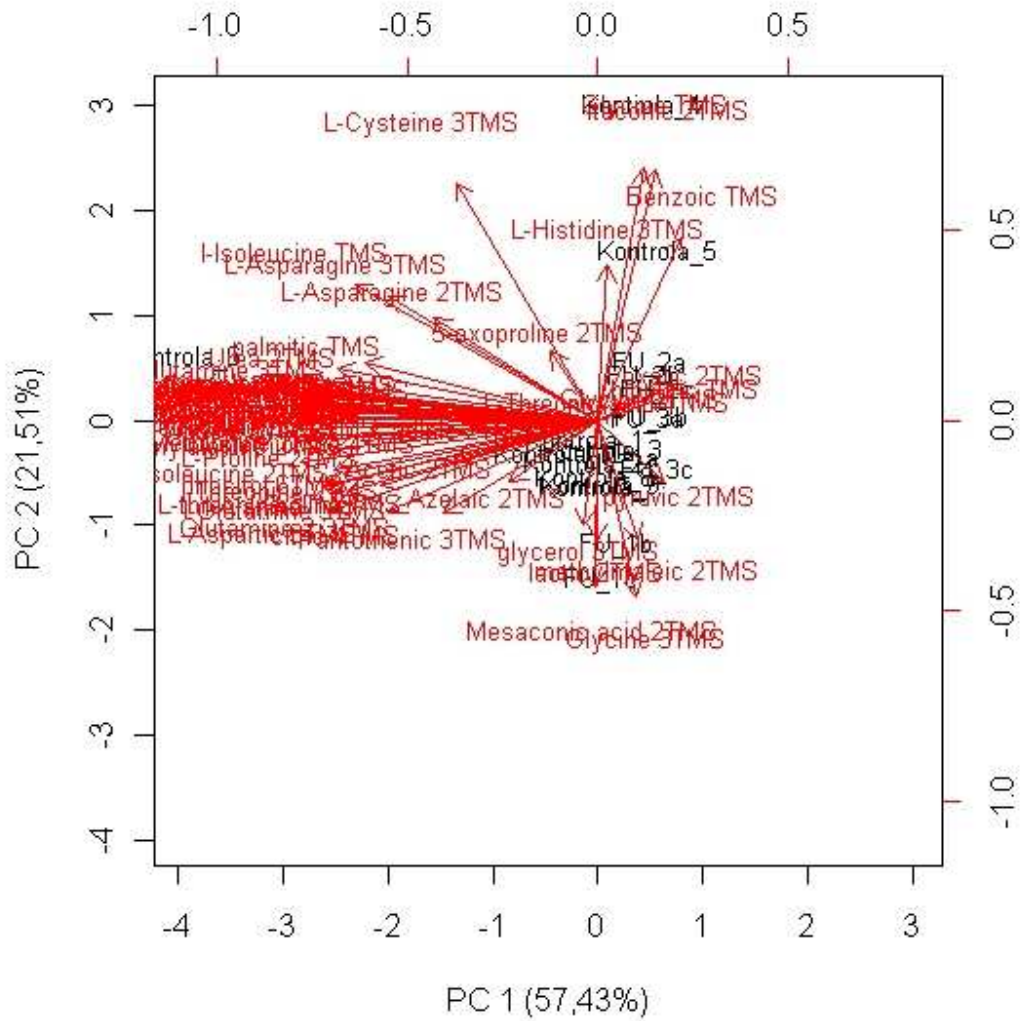
Cluster Dendrogram



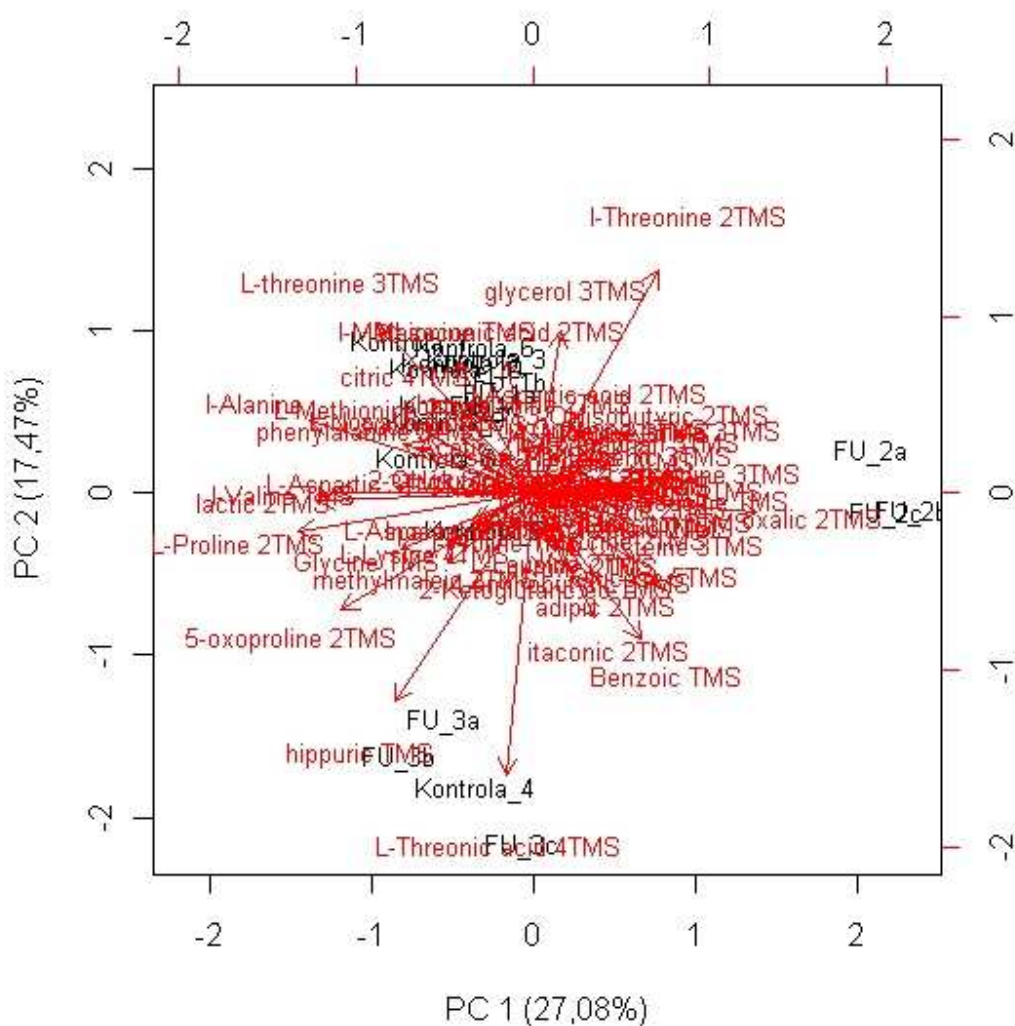
Obr. 13 Dendrogram dat po clr transformaci (vzdálenost transformovaných dat – $\text{dist}(y)$, shluková analýza – hclust , metoda nejvzdálenějšího souseda – „complete“)

Po analýze shluků v datech pomocí dendrogramů, byl sestrojen také biplot (Aitchison & Greenacre, 2002). Opět nejdříve s původními daty a poté po clr transformaci. Biplot (viz. obr 14) vytvořený bez použití transformace je nositelem vysokého procenta původní informace (PC1 = 57,43 % a PC2 = 21,51 %), ale je velmi nepřehledný. Biplot sestrojený s pomocí transformace (viz obr. 15) sice nenese tak vysoké procento původní informace (PC1 = 27,08 %, PC2 = 17,47 %), ale je v něm rozlišeno, že FU_2 a FU_3 se shlukují samostatně a kontroly spolu s FU_1 vytváří vlastní shluk. Ovšem i tak je stále dost nepřehledný a nelze z něj vyčíst, které metabolity nebo která skupina metabolitů by se mohly vlivem 5-fluorouracilu měnit nejvíce (na které metabolity má 5-fluorouracil vliv). Proto byly sestrojeny biploty pro

aminokyseliny (viz obr. 16), organické kyseliny (viz obr. 17) a pro zbytek metabolitů (viz obr. 18).



Obr. 14 Biplot bez clr transformace

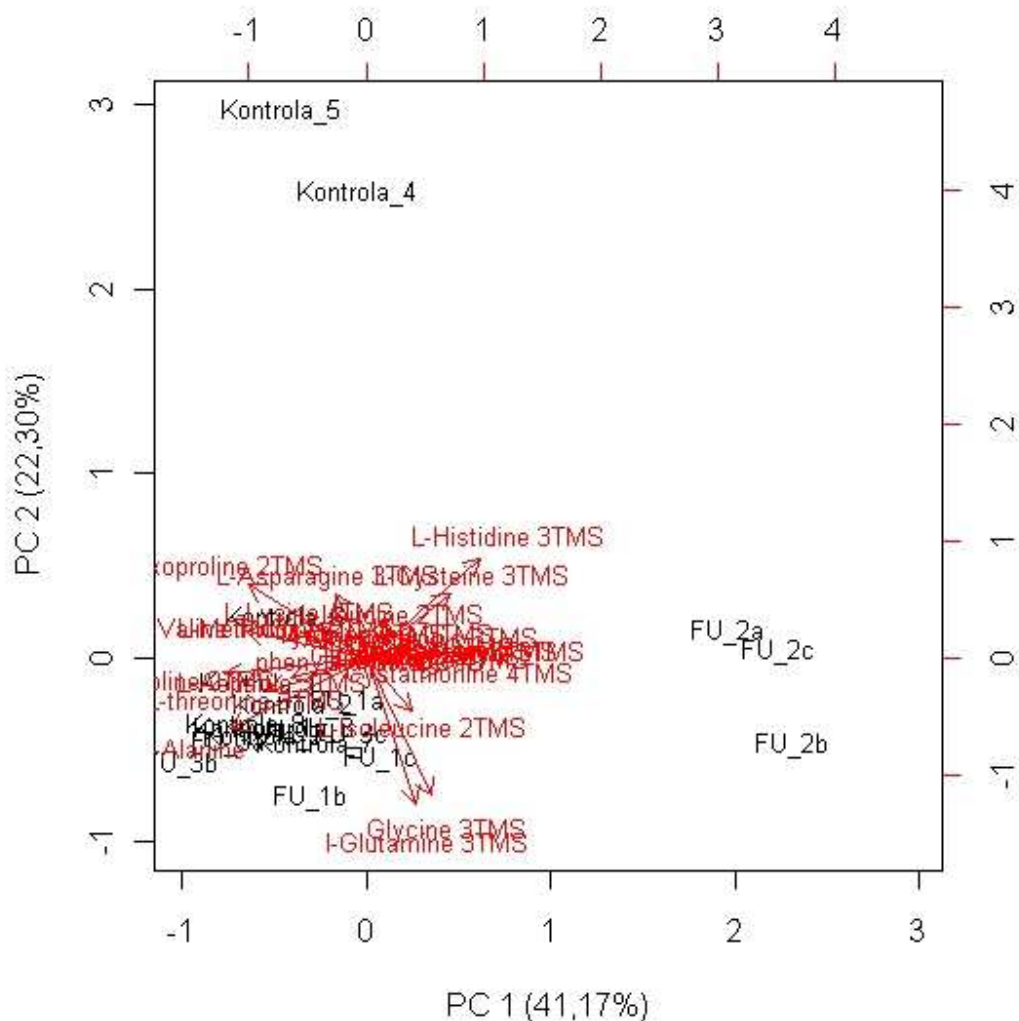


Obn. 15 Biplot s clr transformací

Mezi aminokyseliny byly zařazeny glycin, L-alanin, L-asparagin, kyselina L-asparagová, L-cystein, L-glutamin, L-histidin, L-isoleucin, L-leucin, L-lysin, L-methionin, L-prolin, L-threonin, L-tryptofan, L-tyrosin, L-valin, L-ornithin, L-fenylalanin, L-tyrosin, L-4-hydroxyprolin, L-5-oxoprolin, kyselina aminomalonová, beta-alanin, L-cystathionin a L-glutamin.

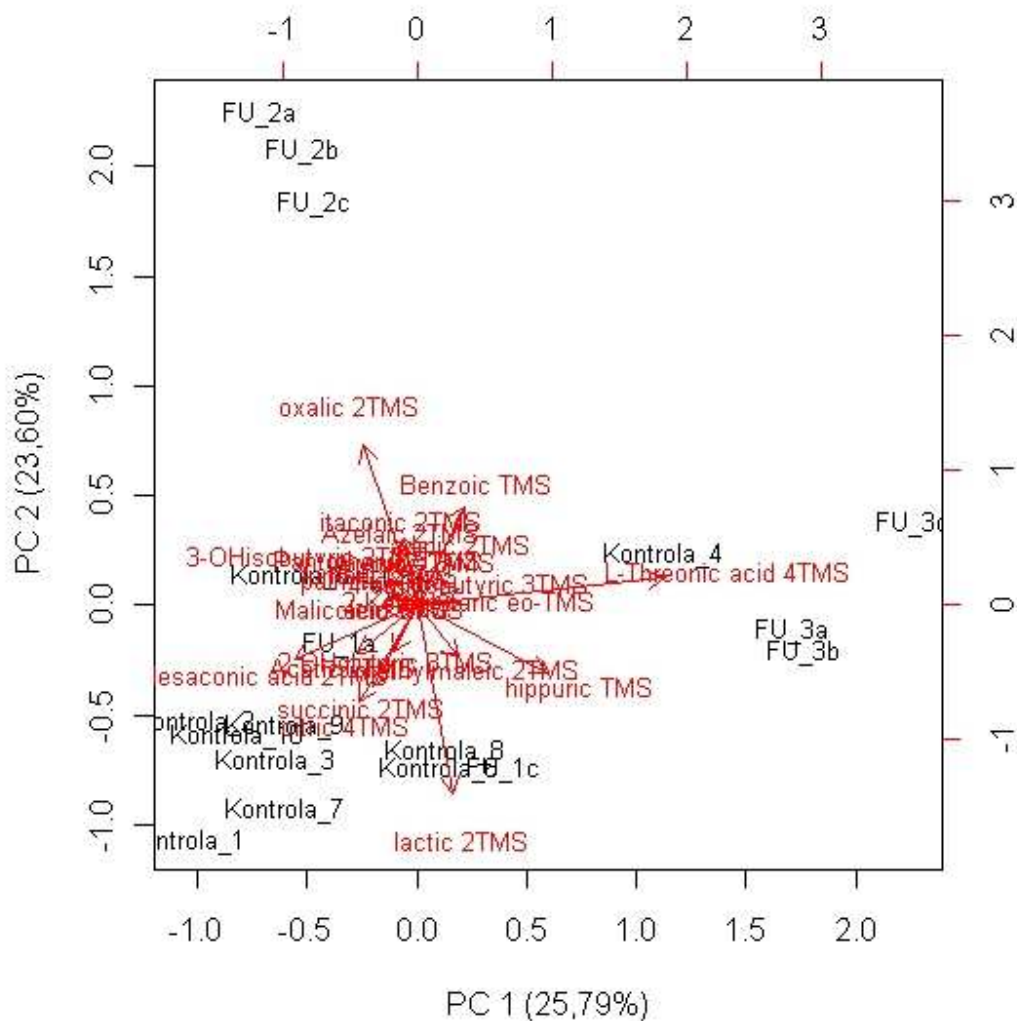
Mezi organické kyseliny byly zařazeny kyselina adipová, kyselina 2-ketoglutarová, kyselina 2-hydroxyglutarová, kyselina 3-hydroxyisobutanová, kyselina 4-aminobutanová, acetát, kyselina azelaová, kyselina benzoová, kyselina citronová, kyselina fumarová, kyselina hippurová, kyselina itakonová, laktát, kyselina L-threonová, malát, kyselina mesakonová, kyselina methylmaleinová, kyselina olejová, kyselina šřavelová, kyselina palmitová, kyselina pantothenová, pyruvát a sukcinát.

Do zbytku látek je řazena glukosa, glycerol, inositol, myo-inositol, fosfát a močovina.



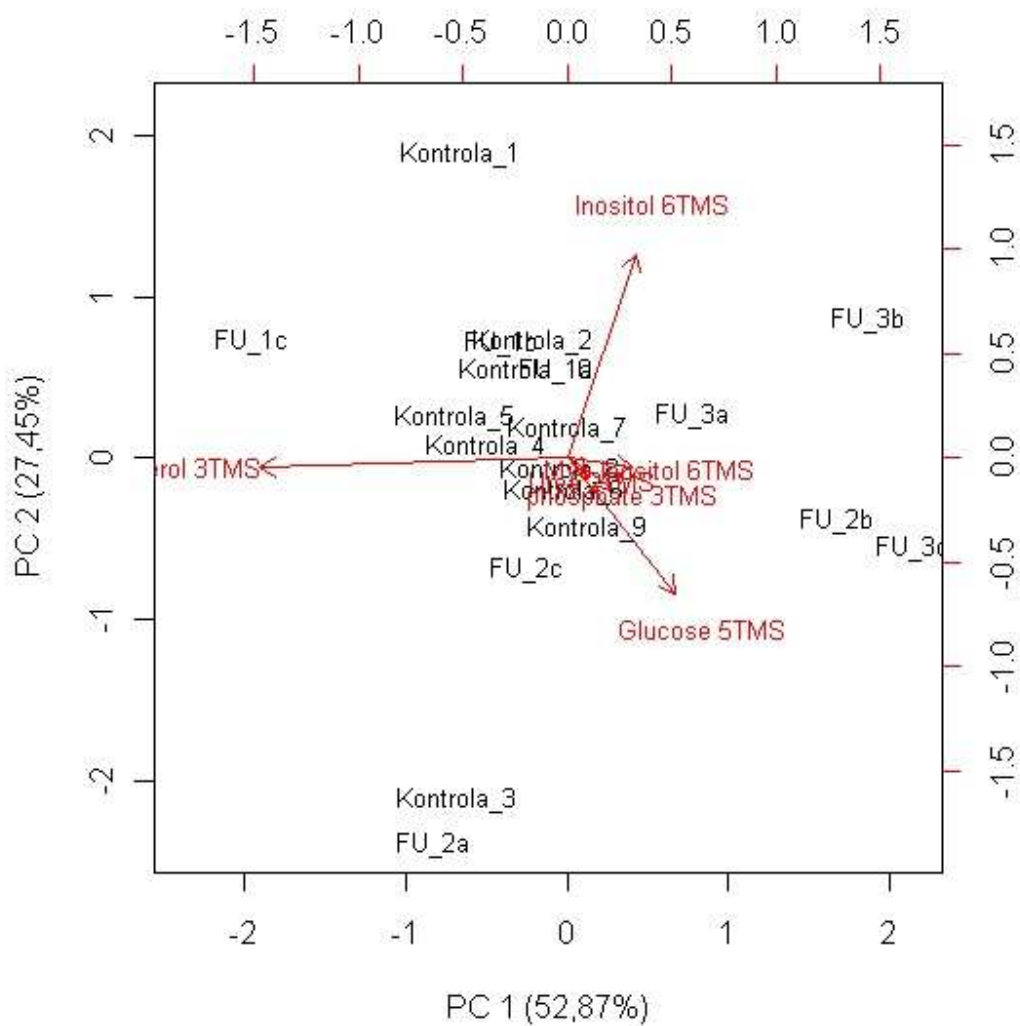
Obr. 16 Biplot s clr – aminokyseliny

U biplotu pro aminokyseliny (viz obr. 16) je patrný shluk FU_2. Pro FU_1 jsou významnými látkami glycin a L-glutamin.



Obr. 17 Biplot s clr – organické kyseliny

V biplotu organických kyselin (viz obr. 17) je zřetelně odlišen shluk FU_2 a FU_3. Pro fibroblasty inkubované s fluorouracilem dva dny je významným metabolitem oxalát (kyselina šťavelová) a pro fibroblasty inkubované s fluorouracilem tři dny je významným metabolitem kyselina L-threonová (kyselina 2,3,4-trihydroxybutanová).



Obr. 18 Biplot s clr – zbytek metabolitů

U biplotu zbytku látek (viz obr. 18) je patrný pouze shluk FU_3, do kterého by bylo možné přiřadit i FU_2b. Další významný shluk tvoří kontroly. Metabolicky nejvýznamnějšími látkami jsou inositol, glycerol a glukosa.

6. Diskuse

Pomocí shlukové analýzy, jejímž výstupem byl dendrogram, bylo zjištěno, že se data transformaci shlukují do 4 shluků. Shluk kontrol 4 a 5, shluk FU_3, shluk FU_2 a poté shluk kontrol, který je tvořen dvěma podshluky: FU_1 a zbytkem kontrol. Podshluk FU_1 je ve shluku kontrol z důvodu, že fibroblasty byly kultivovány s 5-fluorouracilem nejkratší dobu (jeden den) a došlo jen ke slabému ovlivnění metabolomu fibroblastů.

Biplot sestrojený z kompletních dat, na základě analýzy hlavních komponent, byl velmi nepřehledný, proto byly sestrojeny biploty aminokyselin, organických kyselin a zbytku metabolitů. V biplotu aminokyselin je významně oddělen shluk FU_2 a je znatelný i shluk FU_1, jehož významnými metabolity jsou glycin a glutamin. V biplotu organických kyselin jsou významné shluky FU_2 s metabolitem oxalátem a FU_3 s metabolitem kyselinou threonovou. U zbytku metabolitů jsou nejvýrazněji se měnícími metabolity glukosa, inositol a glycerol. Shluk vytváří pouze FU_3.

Glycin (Gly, G) je opticky neaktivní aminokyselina s nepolárním postranním řetězcem. V živočišném organismu slouží jako prekursor pro syntézu hemu, je součástí mysoglobinu a cytochromu c. V nervové soustavě slouží jako inhibiční přenašeč. Je neesenciální, tudíž tělo si jej dokáže vytvořit samo.

Glutamin (Gln, Q) je neesenciální opticky aktivní aminokyselina.

Glycerol (propan-1,2,3-triol) je ve formě svých esterů spolu s mastnými kyselinami součástí lipidů. Je produktem štěpení glukosy a také prekuzorem glukoneogeneze.

Glukosa je šestiuhlíkatý redukující cukr díky své aldehydické skupině. Je nejrychlejším a nezákladnějším zdrojem energie pro živočišné organismy. Pro mozek a erytrocyty (červené krvinky) je zcela nezbytná. Lidské tělo je schopno si ji vytvořit z necukerných prekurzorů pomocí glukoneogeneze. Ve svalech a játrech je uložena ve formě zásobního polysacharidu, glykogenu. Glukosa je odbourávána glykolýzou na pyruvát, který poté může vstupovat do citrátového cyklu (aerobní).

Inositol (cyklohexan-1,2,3,4,5,6-hexol) je ve formě 1,4,5-trifosfátu (IP3) u eukaryotických buněk intracelulárním sekundárním membránovým poslem, který uvolňuje vápenaté kationty z endoplasmatického retikula (po zachycení vnějšího podnětu povrchovým membránovým receptorem). Inositol a jeho deriváty jsou také přítomny v mozkomíšním moku a mimo jiné se podílejí na degradaci tuků a snižování hladiny cholesterolu.

Oxalát (kyselina šťavelová) může ve vysokých koncentracích vyvazovat vápník a tudíž dochází v organismu k hypokalcémii. Dalo by se říci, že vysoké hodnoty oxalátu

mají neurotoxické a kardioinhibiční účinky, přičemž se krystalky šťavelanu vápenatého usazují v ledvinách a podílí se na tvorbě ledvinových kamenů (Murray et al., 2003; Šípál et al., 1992; Voet & Voetová, 1995).

7. Závěr

V této práci byla statisticky zpracována data původní a data po clr transformaci. Bylo zjištěno, že data původní, ač nesou vysoké procento původní informace, nejsou vhodné pro správné statistické vyhodnocení. Proto byla původní data transformována. Procento původní informace, které je neseno daty transformovanými, není tak vysoké, ale dendrogram i biplot jsou přehledné a čitelné a proto jsou transformovaná data vhodná pro statistické vyhodnocení.

Cílem práce bylo pomocí statistických metod zjistit, které z metabolitů se vlivem 5-fluorouracilu významně mění a tudíž mohou způsobovat nežádoucí vedlejší účinky při léčbě onkologicky nemocných pacientů. Potvrdilo se, že 5-fluorouracil nemá vliv jen na nukleové kyseliny, ale ovlivňuje i metabolismus aminokyselin, organických kyselin a dalších látek, které nelze zařadit ani do jedné z jmenovaných skupin.

Seznam použité literatury

- (1) Aitchison J. (1986): The statistical analysis of compositional data. Chapman and Hall: London.
- (2) Aitchison J, Greenacre M. (2002): Biplots of compositional data. *Applied Statistics* **51**: 375–392.
- (3) Gabriel K. R. (1971): The biplot graphic display of matrices with application to principal component analysis. *Biometrika* **58**: 453–467.
- (4) Goodacre R., Broadhurst D., Smilde A. K. (2007): Proposed Minimum Reporting Standards for Data Analysis in Metabolomics. *Metabolomics* **3**, 321-241.
- (5) Grem J. L. (2000): 5-Fluorouracil: Forty-Plus and Still Ticking. A Review of its Preclinical and Clinical Development. *Investigational New Drugs* **18**, 299-313.
- (6) Hebák P. a kolektiv (2007): Vícerozměrné statistické metody 3, pp. 53-80, 120-144, Informatorium, ČR.
- (7) Johnson R, Wichern D. (2007): Applied Multivariate Statistical Analysis (6th edn), pp. 1-7, 671-703, 726-733, Prentice-Hall: London.
- (8) Kalivodová A. (2010): Biplot a jeho aplikace. Bakalářská práce, Přírodovědecká fakulta Univerzity Palackého v Olomouci.
- (9) Klouda P. (2003): Moderní analytické metody (2. přepracované vydání), pp. 10-24, 50-53, Pavel Klouda, ČR.
- (10) Koek M. M. et al.: *Analytical Chemistry* **78** (2006) 1272–1281.
- (11) Korhoňová M., Hron K., Klimčíková D., Müller L., Bednář P., Barták P. (2009): Coffee aroma – statistical analysis of compositional data. *Talanta* **80**, 710-715.
- (12) Kun J. G. (2001): Fluorouracil and the New Oral Fluorinated Pyrimidines. *The Annals of Pharmacotherapy* **35**, 217-227.
- (13) Massart D. L., Heyden Y. V. (2004): From Tables to Visuals: Principal Component Analysis, Part 1. *LC.GC Europe* **17**, 586-591.

- (14) Massart D. L., Heyden Y. V. (2005): From Tables to Visuals: Principal Component Analysis, Part 2. *LC.GC Europe* **18**, 84-89.
- (15) Meloun M., Militský J. (2002): Kompendium statistického zpracování dat, pp. 227-239, 268-290, Academia, ČR.
- (16) Murray R. K., Granner D. K., Mayes P. A., Rodwell V. W. (2003): Harper's Illustrated Biochemistry (26th edition) pp. 14-20, 102-106, 114, 464-465, 477, 606, 607, McGraw-Hill Companies, USA.
- (17) Opekar F., Jelínek I., Rychlovský P., Plzák Z. (2003): Základní analytická chemie, pp. 37-39, 149-164, Karolinum, ČR.
- (18) Pilátová K. (2008): Metabolomika. Bakalářská práce, Přírodovědecká fakulta Masarykovy univerzity v Brně.
- (19) Prachařová J. (2009): Intracelulární nukleotidy v analýze metabolomu. Diplomová práce, Přírodovědecká fakulta Univerzity Palackého v Olomouci.
- (20) Šípal Z., Anzenbacher P., Peč P., Pospíšil J., Růžička I. (1992): Biochemie, pp. 18, 34-36, 99-105, 122, 262, 446, 449, Státní pedagogické nakladatelství, ČR.
- (21) Wojtowicz P. (2009): Optimalizace teplotně programovatelného nástřiku velkého objemu vzorku v plynové chromatografii metodou plánovaného experimentu. Diplomová práce, Přírodovědecká fakulta Univerzity Palackého v Olomouci.
- (22) Wojtowicz P., Zrostlíková J., Kovalczuk T., Schůrek J., Adam T. (2010): Evaluation of comprehensive two-dimensional gas chromatography coupled to time-of-flight mass spectrometry for the diagnosis of inherited metabolic disorders using an automated data processing strategy. *Journal of Chromatography A* **1217**, 8054-8061
- (23) Metody shlukové analýzy. Stránky katedry aplikované matematiky a informatiky, Metody shlukové analýzy, Jihočeská univerzita. Dostupné z http://www2.zf.jcu.cz/public/departments/kmi/MSMT_05/metody%20shlukove%20analyzy.pdf (31. 8. 2011).
- (24) Paradis E.: R for Beginners. Stránky projektu R, dostupné z <http://www.r-project.org/> (31.7. 2011).

- (25) Venables W. N., Smith D. M.: An Introduction to R. Stránky projektu R, dostupné z <http://www.r-project.org/> (31.7. 2011).
- (26) Voet D., Voetová J. G. (1995): Biochemie, pp. 59-78, 267-340, Victoria Publishing, USA, Praha (překlad Arnošt Kotyk), ČR.