

# *PV225: Laboratoř systémové biologie*

David Šafránek

15.12.2010

Tento projekt je spolufinancován Evropským sociálním fondem a státním rozpočtem České republiky.



# *Obsah*

*Paradigma systémové biologie*

*Pojem modelu in silico*

*Modelování a simulace dynamiky*

*Analýza modelu*

# *Obsah*

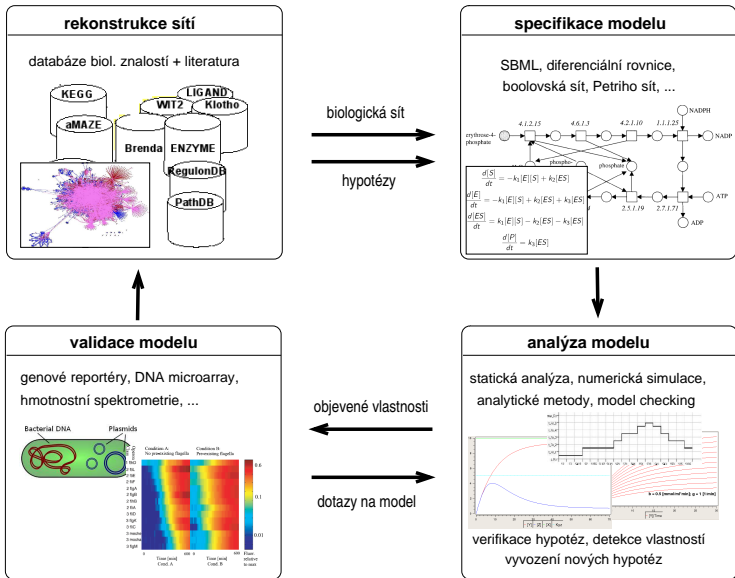
*Paradigma systémové biologie*

*Pojem modelu in silico*

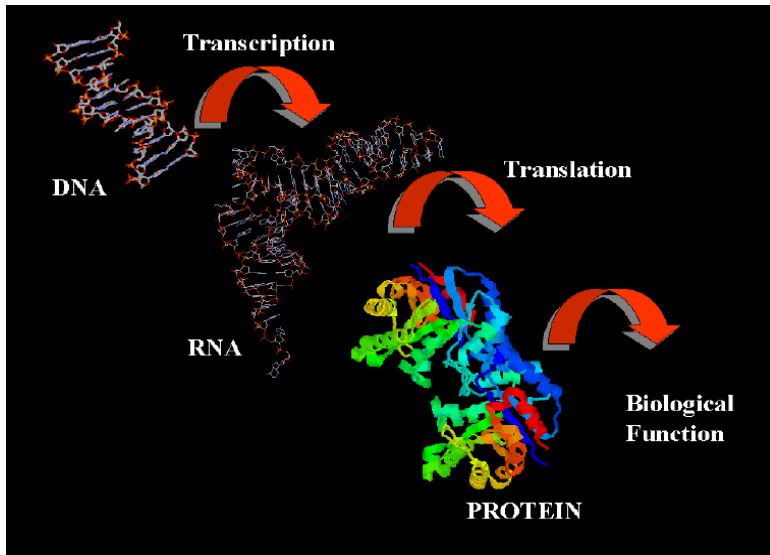
*Modelování a simulace dynamiky*

*Analýza modelu*

# Průběh výzkumu v systémové biologii

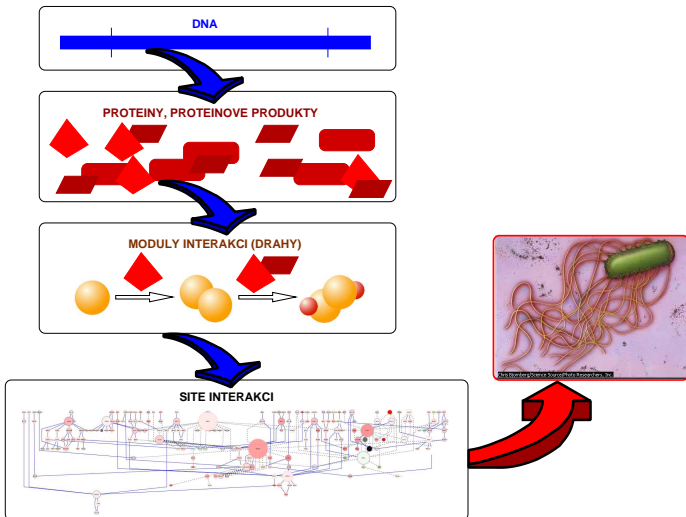


# *Centrální dogma*

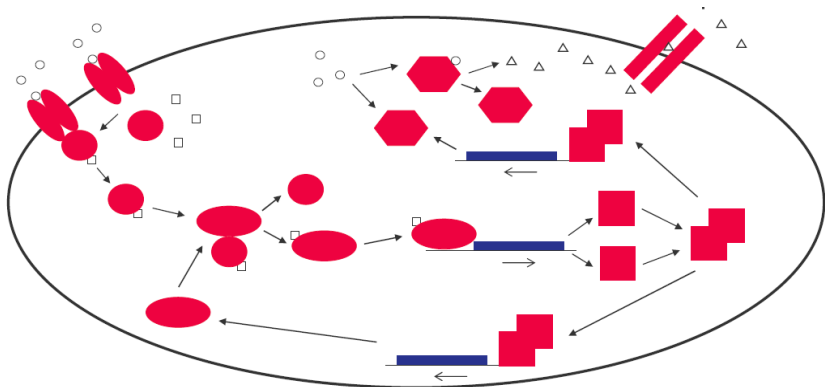


*Genotyp*  $\longrightarrow$  *Fenotyp*

*Hierarchie interakcí*

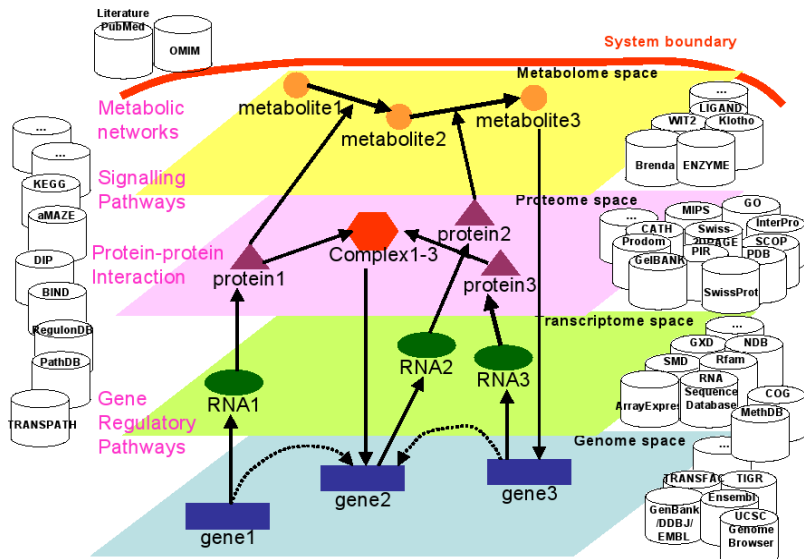


## Biochemické procesy v buňce



- molekulární komponenty – proteiny, DNA, RNA,...
- interakce na různých úrovních (transkripce, metabolismus,...)
- příjem signálů na membráně

# Funkční vsrtyvy buňky





## *Funkční vrstvy buňky*

### **vrstva metabolismu**

- rozsáhlý soubor katalytických (enzymových) reakcí
- příjem a zpracování energie v buňce
- rozklad a syntéza látek

### **transdukce signálů**

- kaskády reakcí zpravovávající externí/interní signál
- receptory externích signálů na membráně

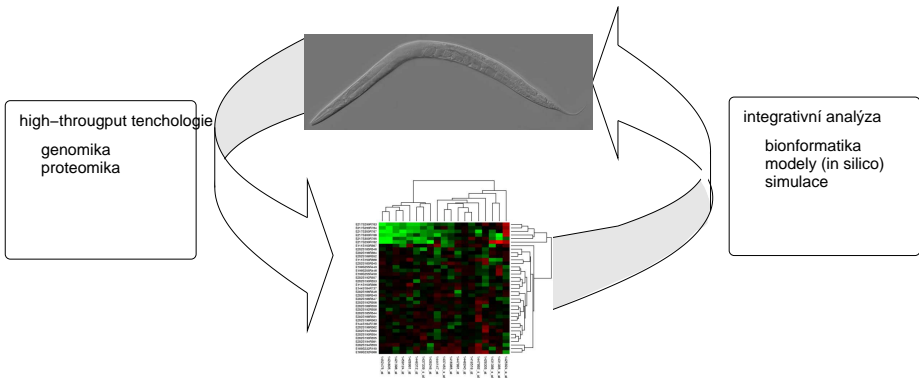
### **interakce proteinů**

- tvorba proteinových komplexů
  - transkripční faktory a enzymy metabolismu

### **transkripční regulace**

- řízení proteosyntézy

# Metody systémového měření



## *Biologické sítě a dráhy*

- biochemická interakce molekul popsaná grafem
- uzly
  - molekuly/komplexy biochemických látek
  - biochemické reakce
- hrany
  - regulace (aktivace, represe, katalýza)
  - příslušnost k reakci (produkt, zdroj)
- dráhy — zaměřené na určitá specifika (látky, reakce)
  - typicky signální dráhy
- sítě — komplexní interakce
- různé úrovně abstrakce, různé notace, např. Kohn's diagrams  
<http://www.nature.com/msb/journal/v2/n1/full/msb4100044.html>

## *Biologická síť jako bipartitní graf*

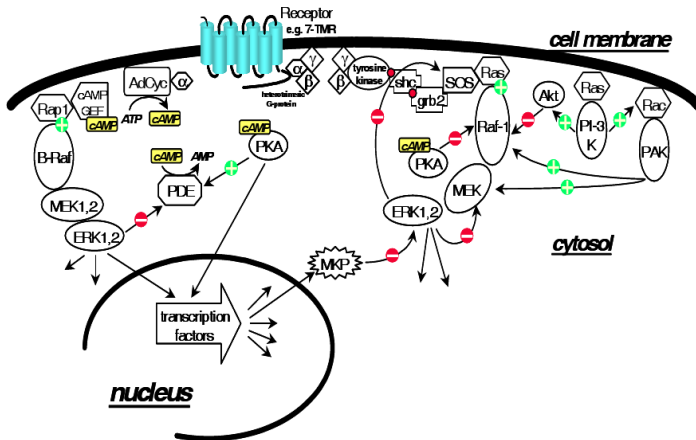
### *Definition*

Nechť  $V$  je konečná množina substancí a  $R$  je konečná množina reakcí. Dále nechť  $E_{rct} \subseteq (V \times R) \cup (R \times V)$  a  $E_{reg} \subseteq V \times R$  jsou relace. *Biologickou sítí* nazveme sjednocení **reakčního grafu**  $G_{rct} \equiv (V \cup R, E_{rct})$  a **regulačního grafu**  $G_{reg} \equiv (V \cup R, E_{reg})$ . Oba dílčí grafy jsou bipartitní.

typ sítě	$V$	$R$	$E$
genové	proteiny	degradace/produkce	regulační interakce
proteinové	proteiny	asociace/disociace	proteinové interakce
metabolické	metabolity	katalytické reakce	tok hmoty
signální	makromolekuly	katalytické reakce	přenos signálu



# Biologické sítě a dráhy



- neformální notace
- vyvíjejí se standardy — SBGN (podporuje např. CellDesigner)

## *Reprezentace stoichiometrickou maticí*

- uvažujme systém  $n$  substancí  $S = \{S_1, \dots, S_n\}$  provázaných  $m$  reakcemi  $R = \{R_1, \dots, R_m\}$
- uvažujeme pouze reakce 1. a 2. řádu
- systém zapisujeme pomocí stoichiometrické matice  $M$  rozměru  $n \times m$ :

$$M_{ij} = \begin{cases} -K, \text{ je-li } K \cdot S_i \text{ reaktantem } R_j \\ K, \text{ je-li } K \cdot S_i \text{ produktem } R_j \end{cases}$$

# Obsah

*Paradigma systémové biologie*

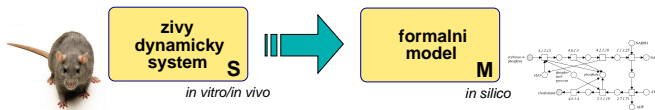
*Pojem modelu in silico*

*Modelování a simulace dynamiky*

*Analýza modelu*



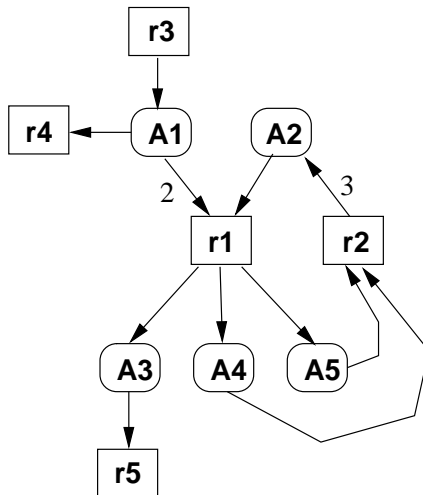
# Model jako abstraktní obraz organismu



## *Tvorba in silico modelu*

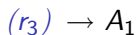
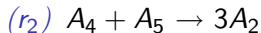
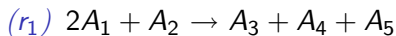
- cílem je modelovat dynamiku organismu
  - nezbytné pro predikci a pochopení fyziologických jevů
- model je definován biochemickými substráty a jejich reakcemi
- model je reprezentován staticky biologickou sítí
- nezávislý na výpočetních (simulačních) nástrojích
- sémantikou modelu je vývoj v čase z daných počátečních podmínek
  - vývoj koncentrací substrátů v čase
- různé přístupy k modelování dynamiky, abstrakce
  - spojité/diskrétní
  - deterministické/stochastické
- chceme vyrobit virtuální laboratoř
  - “náhrada” in vitro/in vivo experimentů analýzou in silico

## Specifikace modelu – příklad



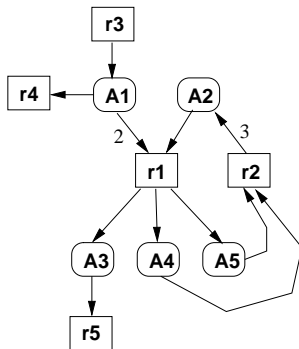
## Specifikace modelu – příklad základní notace

Sada reakcí:



- substráty —  $\{A_1, A_2, A_3, A_4, A_5\}$
- reakční komplexy —  
 $\{A_1, 3A_2, A_3, 2A_1 + A_2, A_3 + A_4 + A_5, A_4 + A_5\}$

# Reprezentace stoichiometrickou maticí



$$M_{SC} = \begin{bmatrix} -2 & 0 & 1 & -1 & 0 \\ -1 & 3 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & -1 \\ 1 & -1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 \end{bmatrix}$$

## *Jazyk SBML pro popis modelu*

- Systems Biology Markup Language (<http://sbml.org/>)
- standard pro biologické modely (XML formát)
- hlavní část SBML popisuje hypergraf (biologickou síť)
- základní elementy:
  - substance (ListOfSpecies) – uzly grafu
  - reakce (ListOfReactions) – hyperhrany
- substance mají význam proměnných (v libovolných jednotkách)
- reakce jsou interakce mezi substancemi
  - reaktanty, produkty, [ modifikátory ]
  - vždy musí být neprázdná alespoň množina reaktantů nebo produktů
  - k reakcím možno definovat sémantiku (kineticLaw)

# Obsah

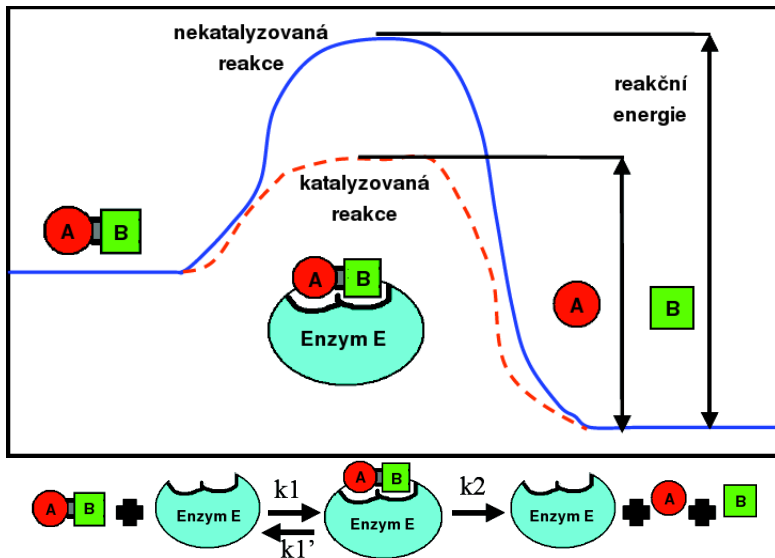
*Paradigma systémové biologie*

*Pojem modelu in silico*

*Modelování a simulace dynamiky*

*Analýza modelu*

# *Energetický proces chemických reakcí*





## *Energetický proces chemických reakcí*

- různé energetické stavy molekuly
  - např. komplex AB méně stabilní než individuální výskyt molekul A, B
  - při přechodu mezi energ. stavy dochází k výměně energie
    - energie požadována pro aktivaci procesu (aktivační energie)
    - energie uvolněna během procesu (volná energie)
- pro biologický systém je zdrojem většiny energie metabolismus
- absolutní teplota ovlivňuje kinetickou energii molekul
- pro reakci (úspěšnou kolizi) musí být splněno:
  - správná prostorová konfigurace (orientace) molekul
  - dostatek kinetické energie

## *Model dynamiky chemických reakcí*

- fixujeme-li konstantní teplotu, objem a tlak, orientace a kinetická energie molekul je stochastickým jevem
  - uvažujeme dobře promíchané médium
  - lze definovat průměrnou pravděpodobnost kolize molekul v časovém okamžiku
  - určeno fyzikálními vlastnostmi reagujících molekul
- máme-li  $N_1$  molekul látky  $S_1$  a  $N_2$  molekul látky  $S_2$ , pak náhodnou proměnnou  $\chi$  charakterizující pravděpodobnost kolize daného počtu molekul  $S_1$  a  $S_2$  v limitním časovém okamžiku lze modelovat:

$$\chi = c \cdot N_1 \cdot N_2$$

- předpokládáme  $S_1$  a  $S_2$  různé látky
- $c$  je stochastická konstanta charakterizující průměrnou frekvenci úspěšných kolizí molekul  $S_1$  a  $S_2$

## *Model dynamiky chemických reakcí*

- závislost frekvence kolizí  $c$  na absolutní teplotě  $T$  (Arheniův zákon):

$$c \propto e^{\frac{-E_A}{RT}}$$

- $E_A$  ... aktivační energie reakce
- $R$  ... plynová konstanta

## *Stochastický model reakční dynamiky*

- počet molekul substance  $S_i$  v čase  $t$  budeme značit  $N_i(t)$
- náhodnou proměnnou rozložení počtů molekul substancí v čase  $t$  charakterizujeme vektorem:

$$X(t) = \langle N_1(t), \dots, N_n(t) \rangle$$

- vývoj tohoto rozložení  $X(t)$  v čase charakterizujeme jako stochastický proces:

$$\{X(t) | t \in \mathbb{R}^+\}$$

## Motivace pro spojitý Markovův řetězec

- stochastický proces  $\{X(t)|t \in \mathbb{R}^+\}$
- spojitý čas pobytu ve stavu
- lze zachytit rozložením  $W$  samplujícím “čekací” dobu mezi změnami stavů
- požadujeme markovskou vlastnost nezávislosti na historii:

$$\Pr\{U > t + \tau | U > \tau\} = \Pr\{U > t\}$$

- tuto vlastnost má exponenciálně distribuovaná proměnná
  - $W \sim \text{Exp}(\lambda)$
  - $\frac{1}{\lambda}$  ... průměrná čekací doba

## Exponenciální rozložení

$$X \sim \text{Exp}(\lambda)$$

pokud:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0, \\ 0, & \text{jinak.} \end{cases}$$

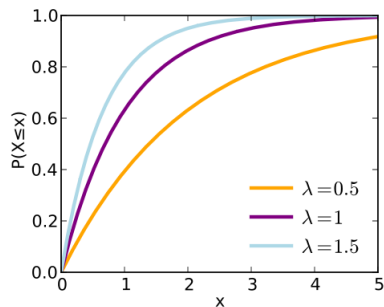
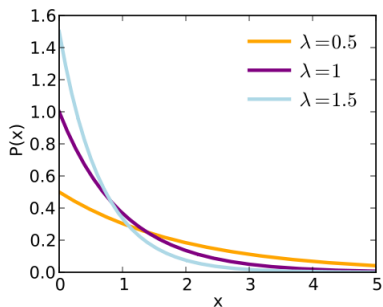
Pro distribuční funkci dostáváme:

$$F_X(x) = \begin{cases} 0, & x < 0, \\ 1 - e^{-\lambda x}, & x \geq 0. \end{cases}$$

Střední hodnota:

$$E(X) = \frac{1}{\lambda}$$

# Exponenciální rozložení



## Stochastický model reakční dynamiky

- interleaving: při přechodu  $X(t) \rightarrow X(t + dt)$  je updatována právě jedna složka  $X$
- provedení právě jedné reakce z  $R$
- provedení reakce uvažováno jako okamžitý jev (trvá nulový čas)
- ve stavu  $X(t) = \langle N_1, \dots, N_n \rangle$  je doba do provedení lib. reakce  $R_i \in R$  charakterizována rozložením  $\text{Exp}(\chi_i(X, c_i))$

$R_i$	$\emptyset \rightarrow *$	$\chi_i(X, c_i) = c_i$
$R_i$	$S_j \rightarrow *$	$\chi_i(X, c_i) = c_i \cdot N_j$
$R_i$	$S_p + S_q \rightarrow *$	$\chi_i(X, c_i) = c_i \cdot N_p \cdot N_q$
$R_i$	$2S_j \rightarrow *$	$\chi_i(X, c_i) = c_i \cdot \frac{N_j \cdot (N_j - 1)}{2}$

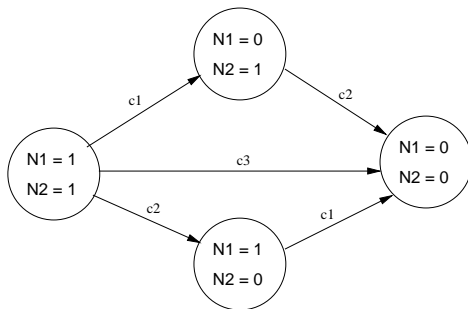
- doba do nejbližší reakce má rozložení  $\text{Exp}(\chi(X, c))$ , kde

$$\chi(X, c) = \sum_{i=1}^m \chi_i(X, c_i), \quad c = \langle c_1, \dots, c_m \rangle$$

- pravděpodobnost provedení reakce  $R_i$ :  $P(R_i) = \frac{\chi_i(X, c_i)}{\chi(X, c)}$



# Stochastický model reakční dynamiky



$R_1$	$S_1 \xrightarrow{c_1}$	$\chi_1 = c_1 \cdot N_1$
$R_2$	$S_2 \xrightarrow{c_2}$	$\chi_2 = c_2 \cdot N_2$
$R_3$	$S_1 + S_2 \xrightarrow{c_3}$	$\chi_3 = c_3 \cdot N_1 \cdot N_2$

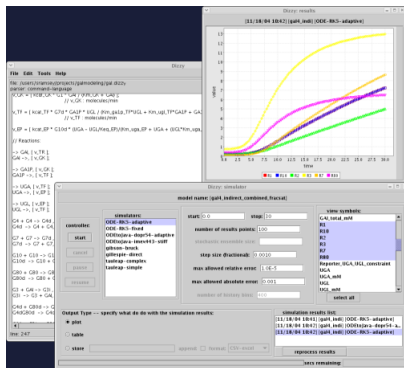
# Monte Carlo simulace

## Gillespiho přímá metoda

1. inicializace  $X(0)$
2. výpočet  $\chi_i(X, c_i) \forall i \in \{1, \dots, m\}$  v aktuálním stavu  $X$
3. výpočet  $\chi(X, c) \equiv \sum_{i=1}^m \chi_i(X, c_i)$
4. simulace doby  $\tau$  do následující události – sampluj  $\tau \in \text{Exp}(\chi(X, c))$
5.  $t := t + \tau$
6. výběr reakce  $R_i$  s pravděpodobností  $\frac{\chi_i(X, c_i)}{\chi(X, c)}$
7.  $X(t) := X^T + M(j)$
8. pokud  $t < T_{max}$ , iteruj (2)

# Nástroj Dizzy

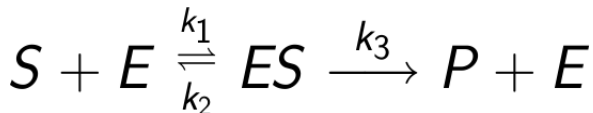
- nástroj pro simulaci dynamiky sítí chemických reakcí
- obsahuje stochastické i deterministické solvery
- mimo přímý Gillespiho algoritmus zahrnuje další varianty stochastické simulace



<http://magnet.systemsbiology.net/software/Dizzy/>

## Cvičení

Uvažujte následující síť chemických reakcí:



1. Vytvořte odpovídající model v nástroji Dizzy.
2. Proveďte několik stochastických simulací o délce 400 časových jednotek pro následující nastavení parametrů:

$$k_1 = 1, k_2 = 0.1, k_3 = 0.01$$

$$S(0) = 20, P(0) = 0, E(0) = 5, ES(0) = 0$$

## *Deterministický model reakční dynamiky*

- uvažujme systém  $n$  substancí  $S = \{S_1, \dots, S_n\}$  provázaných  $m$  reakcemi  $R = \{R_1, \dots, R_m\}$
- uvažujeme pouze reakce 1. a 2. řádu
- systém zapisujeme pomocí stoichiometrické matice  $M$  rozměru  $n \times m$ :

$$M_{ij} = \begin{cases} -K, \text{ je-li } K \cdot S_i \text{ reaktantem } R_j \\ K, \text{ je-li } K \cdot S_i \text{ produktem } R_j \end{cases}$$

## *Deterministický model reakční dynamiky*

- uvažovány vysoké molární koncentrace látek v buňce
- koncentraci substance  $S_i$  v čase  $t$  budeme značit  $[S_i](t)$
- systém v čase  $t$  charakterizujeme vektorem:

$$X(t) = \langle [S_1](t), \dots, [S_n](t) \rangle$$

- vývoj  $X$  v čase:

$$\frac{dX}{dt} = f(X)$$

- průměrné chování lze charakterizovat exponenciální funkcí

## *Převod počtu molekul na molární koncentraci*

- molární koncentrace  $[M]$ :

$$m = \frac{n}{V}$$

kde  $n$  je množství látky  $[mol]$ ,  $V$  je objem roztoku  $[l]$

- vyjadřuje se pomocí Avogadrovy konstanty (počet částic v 1 molu):

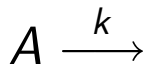
$$m = \frac{N}{N_A \cdot V}$$

kde  $N_A$  Avogadrova konstanta  $[mol^{-1}]$ ,  $V$  objem roztoku  $[l]$  a  $N$  je počet molekul.

- převodní faktor  $\gamma = N_A \cdot V$ :

$$N = m \cdot \gamma$$

## *Deterministický model reakční dynamiky*



- předpokládejme nádobu jednotkového objemu obsahující v čase  $t$  látku  $A$  v molárním množství  $[A]$  [mol]
- kolik množství látky  $A$  “odteče” za jednotku času?
  - hodnota přímo úměrná hodnotě  $[A]$  v daném okamžiku

$$-\frac{d[A](t)}{dt} = k \cdot [A](t)$$

- koeficient úměrnosti je konstanta  $k$  [ $s^{-1}$ ]  
tzv. *reakční konstanta (koeficient)*
  - determinuje rychlost reakce rozpadu (“odtoku”)



## *Deterministický model reakční dynamiky*

$$\frac{[A](t)}{dt} = k \cdot [A](t)$$

- jaká funkce má stejný tvar jako její derivace?

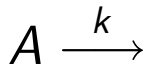
- $f(t) = 1 + t + t^2/2! + t^3/3! + t^4/4! + \dots$

$$f(t) = e^t$$

- platí

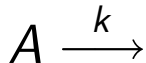
$$\frac{de^t}{dt} = e^t$$

# *Deterministický model reakční dynamiky*



$$-\frac{d[A](t)}{dt} = k \cdot [A](t)$$

## *Deterministický model reakční dynamiky*



$$-\frac{d[A](t)}{dt} = k \cdot [A](t) \Leftrightarrow [A](t) = [A](0) \cdot e^{-kt}$$

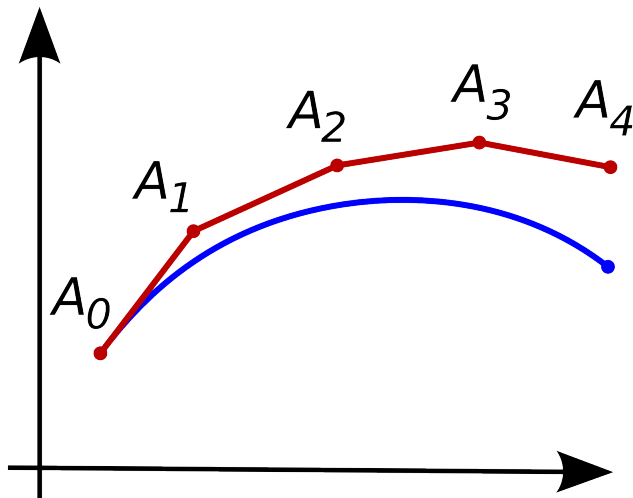
- lineární dif. rce 1. řádu
- jednoznačné řešení
- numericky aproximovatelné

## *Deterministický model reakční dynamiky*

- spojité chování: při přechodu  $X(t) \rightarrow X(t + dt)$  jsou updatovány všechny složky  $X$  (souběžný spojitý tok reakcí)
- časová informace o běhu reakce  $R_i$  promítnuta do okamžitého reakčního toku  $v_i(t)$

$R_i$	$\emptyset \rightarrow *$	$v_i(t) = k_i$
$R_i$	$S_j \rightarrow *$	$v_i(t) = k_i \cdot [S_j](t)$
$R_i$	$S_p + S_q \rightarrow *$	$v_i(t) = k_i \cdot [S_p](t) \cdot [S_q](t)$
$R_i$	$2S_j \rightarrow *$	$v_i(t) = k_i \cdot [S_j]^2(t)$

## *Eulerova metoda*



## *Eulerova metoda*

- aproximativní řešení  $y(t)$  (Euler):

$$y'(t) = f(t, y(t))$$

$$y(0) = y_0$$

- přesné řešení  $\varphi(t)$ :

$$\varphi'(t) = f(t, \varphi(t))$$

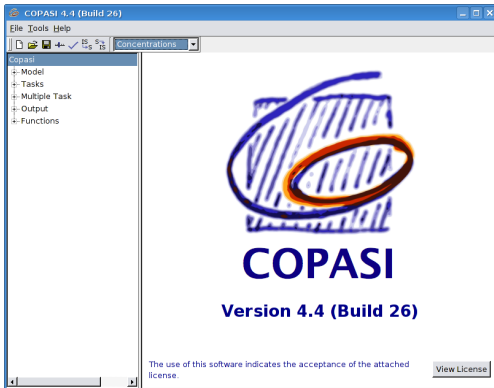
$$\varphi(0) = y_0$$

- pro lib.  $n \geq 0$ ,  $t_n = n\Delta t$ :

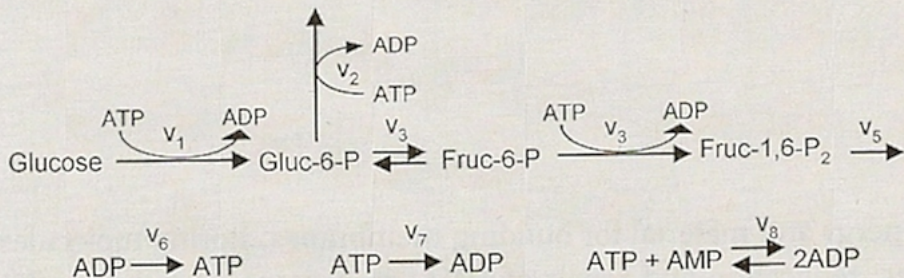
$$y_n \approx \varphi(t_n)$$

# Nástroj COPASI

- nástroj pro simulaci dynamiky sítí chemických reakcí
- zaměřený především na deterministické solvery
- zahrnuje i efektivnější alternativu Gillespiho algoritmu (Gibson-Bruck)
- stoichiometrická analýza
- estimace parametrů



## *Cvičení – model metabolismu glukózy*





## *Cvičení – model metabolismu glukózy*

$$\frac{d}{dt} \text{Gluc6P} = v_1 - v_2 - v_3$$

$$\frac{d}{dt} \text{Fruc6P} = v_3 - v_4$$

$$\frac{d}{dt} \text{Fruc1,6P}_2 = v_4 - v_5$$

$$\frac{d}{dt} \text{ATP} = -v_1 - v_2 - v_4 + v_6 - v_7 - v_8$$

$$\frac{d}{dt} \text{ADP} = v_1 + v_2 + v_4 - v_6 + v_7 + 2v_8$$

$$\frac{d}{dt} \text{AMP} = -v_8.$$

## *Cvičení – model metabolismu glukózy*

1. Načtěte model `http://www.fi.muni.cz/~xsafran1/PV225/glycolysis.cps` do nástroje COPASI.
2. Proveďte simulaci o délce 0.5 min s použitím výchozího nastavení parametrů a iniciálních hodnot.
3. Pomocí úlohy Parameter Scan – Time Course proveďte simulaci pro různá nastavení vstupní koncentrace glukózy.

## *Základní kinetické zákony*

$$v_2 = k_2 \cdot \text{ATP} \cdot \text{Gluc6P}$$

$$v_5 = k_5 \cdot \text{Fruc}_{1,6}\text{P}_2$$

$$v_6 = k_6 \cdot \text{ADP}$$

$$v_7 = k_7 \cdot \text{ATP}$$

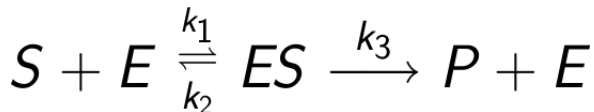
$$v_8 = k_{8f} \cdot \text{ATP} \cdot \text{AMP} - k_{8r} \cdot \text{ADP}^2$$

## *Pokročilé kinetické zákony*

$$v_1 = \frac{V_{max,1} \cdot ATP}{K_{ATP,1} + ATP}$$

$$v_3 = \frac{\frac{V_{max,3}^f}{K_{Gluc6P,3}} Gluc6P - \frac{V_{max,3}^r}{K_{Fruc6P,3}} Fruc6P}{1 + \frac{Gluc6P}{K_{Glu6P,3}} + \frac{Fruc6P}{K_{Fruc6P,3}}}$$

## *Kinetika Michaelis-Menten*

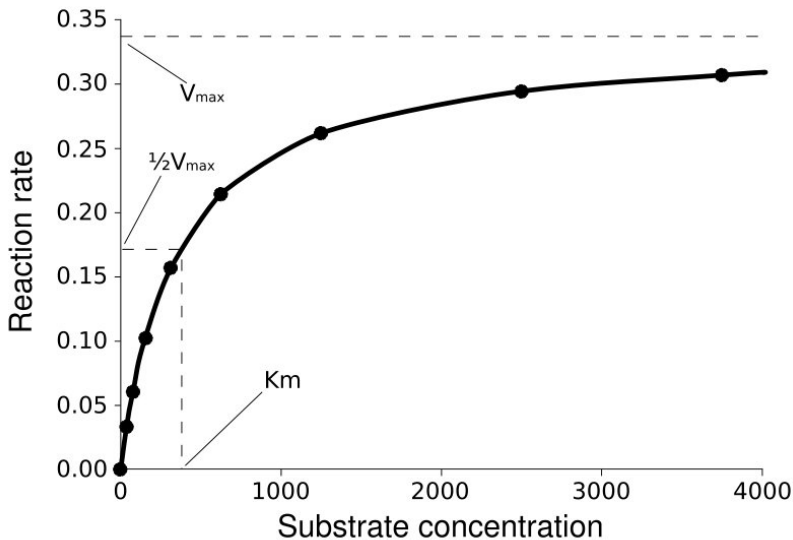


- nutnost znát  $k_1, k_2, k_3$  – obtížně měřitelné
- zjednodušení na

$$S \xrightarrow{v} P$$
$$v = \frac{V_{max} \cdot S}{K + S}$$

- platí pokud  $S \gg E$

## *Kinetika Michaelis-Menten*



## Kinetika Michaelis-Menten

- měření iničiálního  $v$  v *in vitro* podmínkách (izolovaný enzym) pro různá  $S$
- nutno použít nelineární regresi
- pro usnadnění se používají transformace do lineární závislosti mezi proměnnými  $\Rightarrow$  lineární regrese

	<i>Lineweaver-Burk</i>	<i>Eadie-Hofstee</i>	<i>Hanes-Woolf</i>
Transformed equation	$\frac{1}{v} = \frac{K_m}{V_{max}} \frac{1}{S} + \frac{1}{V_{max}}$	$v = V_{max} - K_m \frac{v}{S}$	$\frac{S}{v} = \frac{S}{V_{max}} + \frac{K_m}{V_{max}}$
New variables	$\frac{1}{v}, \frac{1}{S}$	$v, \frac{v}{S}$	$\frac{S}{v}, S$
Graphical representation			

## *Estimace parametrů*

- předpokládejme množinu dvojic  $n$  naměřených hodnot  $\{(x_i, y_i) | 1 \leq i \leq n\}$ 
  - $x$  ... nezávislá proměnná (čas, koncentrace substrátu, ...)
  - $y$  ... závislá proměnná (koncentrace látky, reakční tok, ...)
- dále uvažujme funkci  $f(x, \vec{p})$ , kde  $p$  je vektor parametrů, např.  
 $f(S, \langle V_{max}, K \rangle) = \frac{V_{max}S}{K+S}$
- definujeme reziduál:

$$r_i = y_i - f(x_i, \vec{p})$$

- problém: najít hodnoty parametrů  $\vec{p}$  tak, aby byla  $f$  co nejbližší ("best-fit") naměřeným hodnotám
- např. metoda nejmenších čtverců (OLE):  
optimalizujeme parametry minimalizací  $S = \sum_{i=1}^n r_i^2$



## *Estimace parametrů*

- OLE dává statisticky dobré výsledky pro nezávislá měření se stejnou neurčitostí (chybou)
- pro měření s různou neurčitostí lze použít vážené OLE:

$$S = \sum_{i=1}^n w_i r_i^2$$

- pokud  $\sigma_i$  je standardní odchylka hodnot  $i$ tého měření, lze volit

$$w_i = \frac{1}{\sigma_i^2}$$

## Cvičení

1. Specifikujte pomocí COPASI model reakce  $S \xrightarrow{v} P$  pomocí Michaelis-Menten kinetiky, uvažujte  $V_{max} = 100, K = 22$ .
2. Proveďte simulaci pro  $S(0) = 500, P(0) = 0$ .
3. Proveďte estimaci parametru  $V_{max}$  dle naměřených experimentálních dat <http://www.fi.muni.cz/~xsafran1/PV225/producttimeseries.csv>.
4. Proveďte estimaci parametrů  $V_{max}, K$  dle téhož datasetu.

## Cvičení

1. Specifikujte pomocí COPASI model reakce  $S \xrightarrow{v} P$  pomocí Michaelis-Menten kinetiky.
2. Proveďte estimaci iniciální podmínky  $S(0)$  tak, aby ve stabilním stavu bylo  $P = 400$ .
3. Na základě znalosti  $P(0)$ ,  $S(0)$  a  $P, S$  ve stabilním stavu proveďte estimaci parametrů  $V_{max}$ ,  $K$ . Uvažujte např. následující time-course data:

Time	P
0	0
6	495
8	500

## Cvičení

1. Specifikujte pomocí COPASI model reakce  $S \xrightarrow{v} P$  pomocí Michaelis-Menten kinetiky.
2. Předpokládejte iniciální toky  $v$  naměřené pro různé výchozí koncentrace  $S$ . Data jsou k dispozici v souboru <http://www.fi.muni.cz/~xsafran1/PV225/vmaxes.tar>.
3. Proveďte estimaci parametrů  $V_{max}$ ,  $K$ .

## Cvičení

1. Uvažujte model glykolýzy <http://www.fi.muni.cz/~xsafran1/PV225/glycolysis.cps>.
2. Proveďte estimaci
3. Proveďte estimaci parametrů  $V_{max}$ ,  $K$ .

# *Obsah*

*Paradigma systémové biologie*

*Pojem modelu in silico*

*Modelování a simulace dynamiky*

*Analýza modelu*

## Model jako síť reakčních komplexů

Uvažujme reakční síť  $\mathcal{M} = (S \cup R, E_{rct})$ ,  $V = \{s_1, \dots, s_n\}$  množina substrátů,  $R = \{r_1, \dots, r_m\}$  množina reakcí.

Pro reakci  $r_i \in R$  tvaru  $c_i \rightarrow c_j$  definujeme  $c_i, c_j$  jako vstupní (resp. výstupní) *komplex*  $r_i$ .

Pro reakci  $r_i \in R$  musí existovat  $l_1, l_2 \leq n$  tak, že:

$$c_i = \sum_{k=1}^{l_1} \alpha_k \cdot s_k$$

$$c_j = \sum_{k=1}^{l_2} \alpha_k \cdot s_k$$

Předpokládáme  $l_1 + l_2 > 0$ .

## *Model jako síť reakčních komplexů*

Definujeme *matici komplexů*  $M_C$  následujícím předpisem:

$$M_C(j, i) = \begin{cases} \alpha \in \mathbb{N}, & \text{pokud } \alpha \cdot s_j \text{ je členem komplexu } c_i, \\ 0, & \text{jinak.} \end{cases}$$

Pro libovolné  $i$ ,  $M_C(i)$  odpovídá vektoru všech substrátů obsažených v komplexu  $c_i$ .

Množinu všech komplexů modelu  $\mathcal{M}$  značíme  $\text{complexes}(\mathcal{M})$ .



## *Třídy souvislosti*

Množinu  $L \subset \text{complexes}(\mathcal{M})$  nazýváme *třídou souvislosti* pokud splňuje následující podmínky:

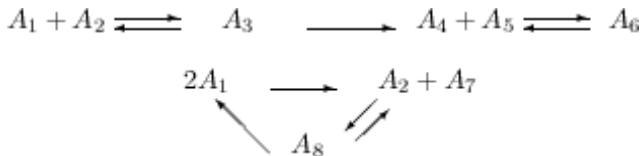
1.  $c \in L \Rightarrow (\forall c' \in \text{complexes}(\mathcal{M}))(\exists r \in R. r \equiv c \rightarrow c') \Rightarrow c' \in L$
2.  $c \in L \Rightarrow (\forall c' \in \text{complexes}(\mathcal{M}))(\exists r \in R. r \equiv c \rightarrow c') \Rightarrow c' \in L$

Počet tříd souvislosti modelu  $\mathcal{M}$  značíme  $\lambda_{\mathcal{M}}$ .

M. Feinberg: Lectures on Chemical Reaction Networks

<http://www.che.eng.ohio-state.edu/~FEINBERG/LecturesOnReactionNetworks/>

## Třídy souvislosti – příklad



$$\text{complexes}(\mathcal{M}) = L_1 \cup L_2 :$$

$$L_1 = \{A_1 + A_2, A_3, A_4 + A_5, A_6\}$$

$$L_2 = \{2A_1, A_2 + A_7, A_8\}$$

$$\Rightarrow \lambda_{\mathcal{M}} = 2$$

## *Komplexnost modelu*

Definujeme *komplexnost modelu*  $\mathcal{M}$  jako nezáporné číslo  $\kappa \in \mathbb{N}_0$  definované vztahem:

$$\kappa = |\text{complexes}(\mathcal{M})| - \lambda_{\mathcal{M}} - h(M)$$

kde  $h(M)$  je hodnost stoichiometrické matice

## *Komplexnost modelu*

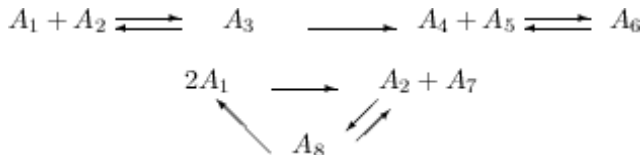
Definujeme *komplexnost modelu*  $\mathcal{M}$  jako nezáporné číslo  $\kappa \in \mathbb{N}_0$  definované vztahem:

$$\kappa = |\text{complexes}(\mathcal{M})| - \lambda_{\mathcal{M}} - h(M)$$

kde  $h(M)$  je hodnota stoichiometrické matice

Interpretace: *Čím větší je  $\kappa$ , tím méně stoichiometrická matice vystihuje dynamiku modelu (= tím složitější je model).*

## *Komplexnost modelu – příklad*



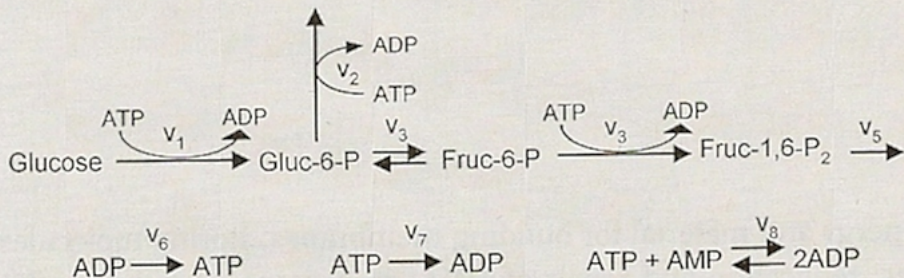
$$|\text{complexes}(\mathcal{M})| = 7$$

$$\lambda_{\mathcal{M}} = 2$$

$$h(\mathcal{M}) = 5$$

$$\Rightarrow \kappa = 7 - 2 - 5 = 0$$

# Cvičení



# Cvičení

$$S = \begin{pmatrix} \text{Gluc6P} \\ \text{Fruc6P} \\ \text{Fruc}_{1,6}\text{P}_2 \\ \text{ATP} \\ \text{ADP} \\ \text{AMP} \end{pmatrix}$$

$$M = \begin{pmatrix} 1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ -1 & -1 & 0 & -1 & 0 & 1 & -1 & -1 \\ 1 & 1 & 0 & 1 & 0 & -1 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{pmatrix}$$

$$h(M) = 5, |\text{complexes}(\mathcal{M})| = 12, \lambda_{\mathcal{M}} = 4 \Rightarrow \kappa = 3$$

## *Hodnota stoichiometrické matice*

- dimenze stoichiometrického prostoru je dána  $h(M)$
- uvažme následující značení:
  - $N_N \in \mathbb{Z}^{h(M) \times |R|}$  matice lineárně nezávislých řádků  $M$
  - $N_D \in \mathbb{Z}^{(|S|-h(M)) \times |R|}$  matice lineárně závislých řádků  $M$

$$M = \begin{bmatrix} N_N \\ N_D \end{bmatrix}$$



## *Hodnost stoichiometrické matice*

- dimenze stoichiometrického prostoru je dána  $h(M)$
- uvažme následující značení:
  - $N_N \in \mathbb{Z}^{h(M) \times |R|}$  matice lineárně nezávislých řádků  $M$
  - $N_D \in \mathbb{Z}^{(|S| - h(M)) \times |R|}$  matice lineárně závislých řádků  $M$

$$M = \begin{bmatrix} N_N \\ N_D \end{bmatrix}$$

Definujeme *spojovací (link) matici*  $L \in \mathbb{Z}^{h(M) \times (|S| - h(M))}$  jako matici splňující vztah:

$$N_D = L \cdot N_N$$

## Význam hodnosti stoichiometrické matice

- konzervace mas/energií — *moiety conservation*
- definováno jako v čase konstantní součet koncentrací substrátů
  - např.  $ATP + ADP$
- zachyceno lineární závislosti řádků v  $M$
- každý substrát v  $N_D$  je konzervován lineární kombinací substrátů v  $N_N$
- spojovací matice zachycuje právě tuto závislost

## *Cvičení*

- Zjistěte konzervační vztahy v modelu glykolýzy.
- Použijte nástroj COPASI (Mass Conservation).

## *Podprostory stoichiometrické matice*

Definujeme *levý nulový podprostor*  $M$ , značíme  $\text{Inp}(M)$ , jako prostor generovaný vektory splňujícími

$$M^T \cdot x = 0$$

- $\dim(\text{Inp}(M)) = |S| - h(M)$
- levý nulový prostor zachycuje konzervační a časové invarianty
- všechny reakce zahrnuté v tomto prostoru manipulují s konzervovanou masou/energií

## *Podprostory stoichiometrické matice*

Definujeme *nulový podprostor*  $M$ , značíme  $np(M)$ , jako prostor generovaný vektory splňujícími

$$M \cdot x = 0$$

- $\dim(np(M)) = |R| - h(M)$
- nulový prostor zachycuje stabilní distribuci reakčního toku (flux)
- báze vektory tohoto prostoru tvoří jádro matice  $M$ :

$$M \cdot K = 0$$

- $K \in \mathbb{N}^{|R| \times (|R| - h(M))}$
- netriviální řešení pro  $h(M) < |R|$ , není obecně určeno jednoznačně

## Podprostory stoichiometrické matice

- ve stabilním stavu lze vyjádřit reakční tok jako lineární kombinaci vektorů v  $K$ :

$$J = \sum_{i=1}^{|R|-h(M)} \alpha_i \times K(i)$$

- báze  $np(M)$  určuje *módy reakčního toku*, které vymezují podsítě modelu se specifickou dynamikou ve stabilním stavu:

$$Em(M) = \{v \in \mathbb{N}^{|R|} \mid v = \gamma \cdot v', \gamma > 0, v' \in np(M)\}$$

- *elementární mód* je reakční mód daný bázevým vektorem  $np(M)$

# Podprostory stoichiometrické matice

