

# *PV225: Laboratoř systémové biologie*

David Šafránek

6.12.2012

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



# *Obsah*

*Paradigma systémové biologie*

*Dynamické modely biologických procesů*

*Parametrizace modelu*

*Statická analýza*

*Modelování fotosyntézy*

# Obsah

*Paradigma systémové biologie*

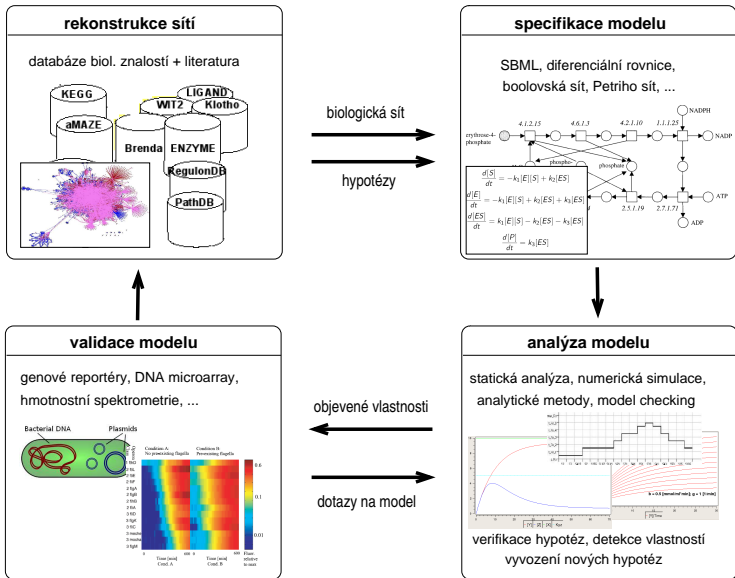
*Dynamické modely biologických procesů*

*Parametrizace modelu*

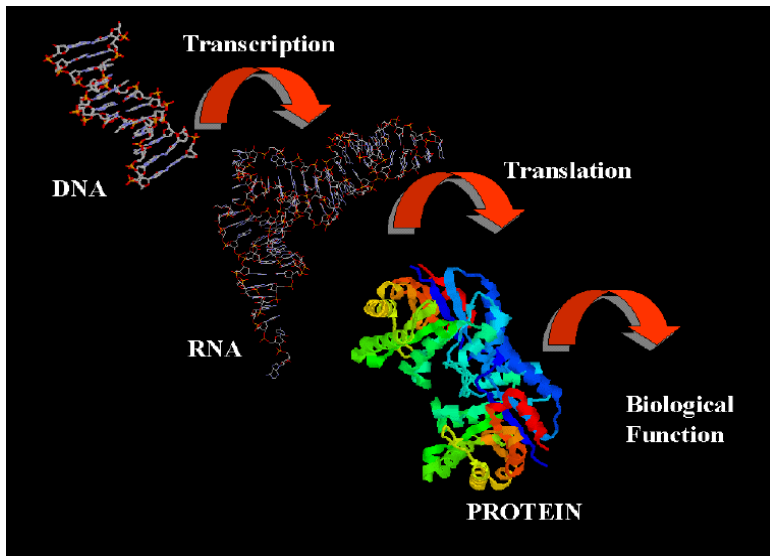
*Statická analýza*

*Modelování fotosyntézy*

# Průběh výzkumu v systémové biologii

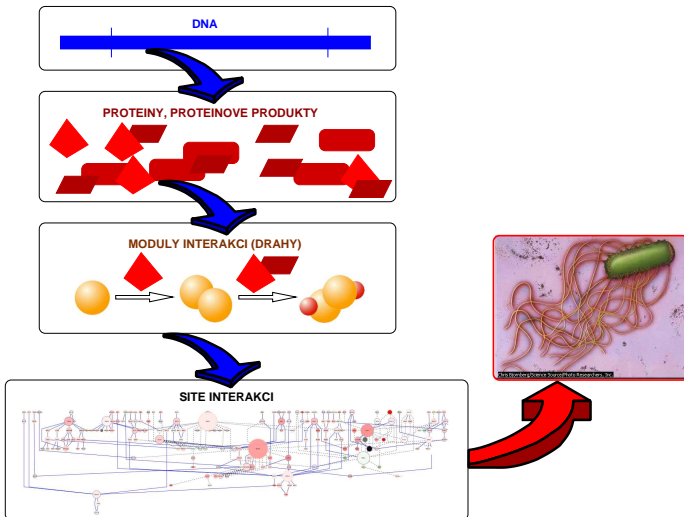


# *Centrální dogma*

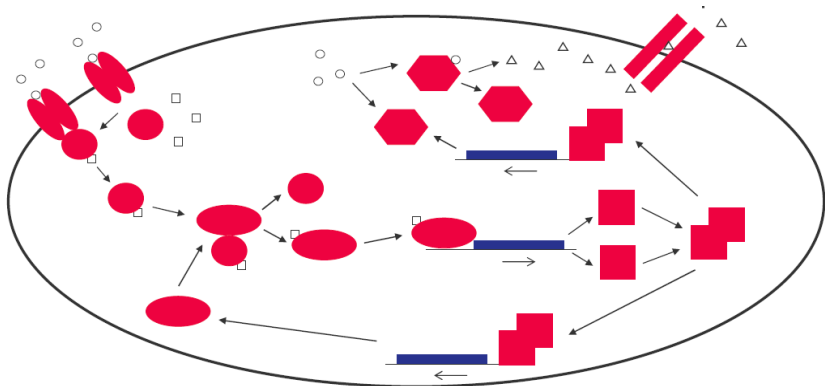


# Genotyp $\rightarrow$ Fenotyp

## Hierarchie interakcí

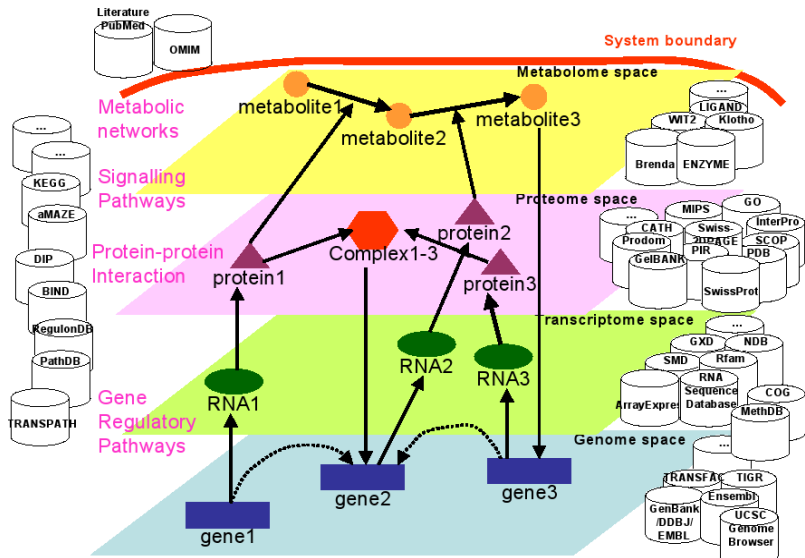


## Biochemické procesy v buňce



- 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

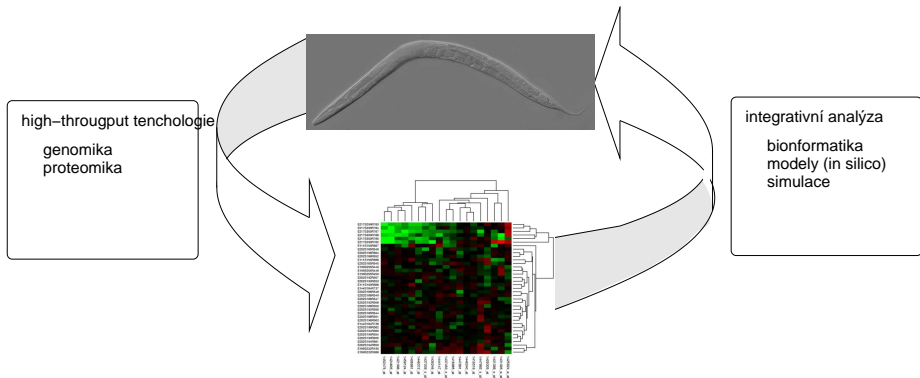
### **interakce proteinů**

- tvorba proteinových komplexů
  - transkripční faktory a enzymy metabolismu

### **transkripční regulace**

- řízení proteosyntézy

# Metody systémového měření



## *Biologické sítě a dráhy*

- biochemická interakce molekul popsána 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>

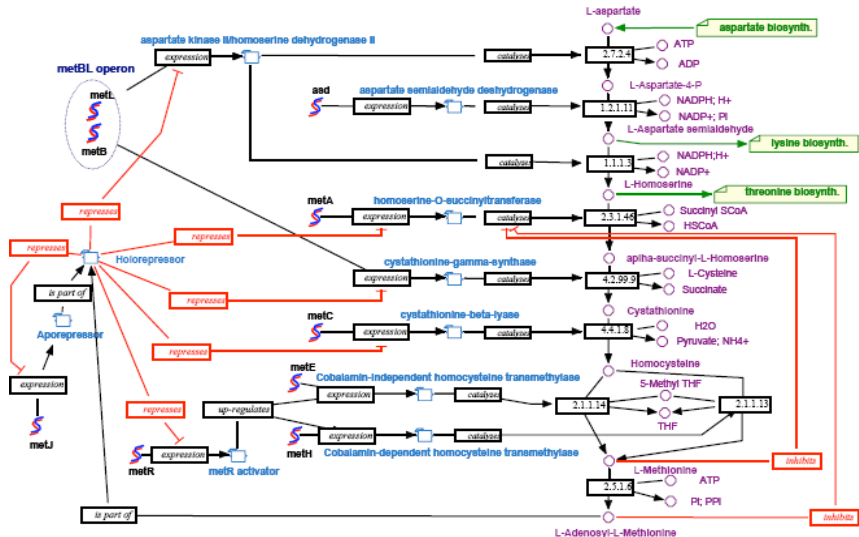
## *Biologická síť jako bipartitní graf*

### *Definition*

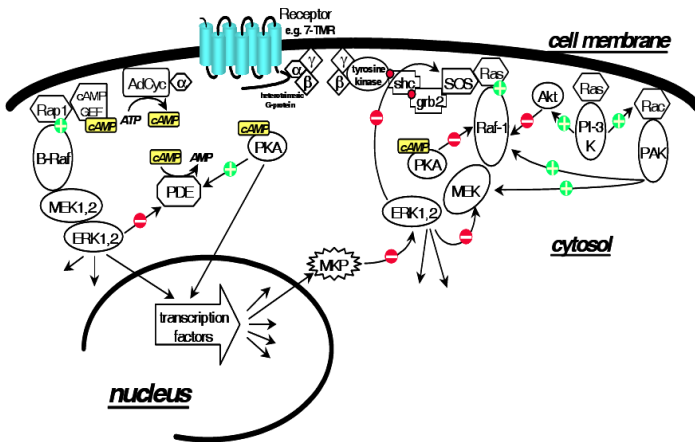
Nechť  $V$  je konečná množina substancí a  $R$  je konečná množina reakcí. Dále nechť  $E_{rct} \subseteq (V \times R) \cup (R \times V)$  a  $E_{reg} \subseteq V \times R$  jsou relace. *Biologickou sítí* nazveme sjednocení **reakčního grafu**  $G_{rct} \equiv (V \cup R, E_{rct})$  a **regulačního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

# Příklad komponenty biologické sítě



# Biologické sítě a dráhy



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

## Reprezentace stechiometrickou 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 stechiometrické 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*

*Dynamické modely biologických procesů*

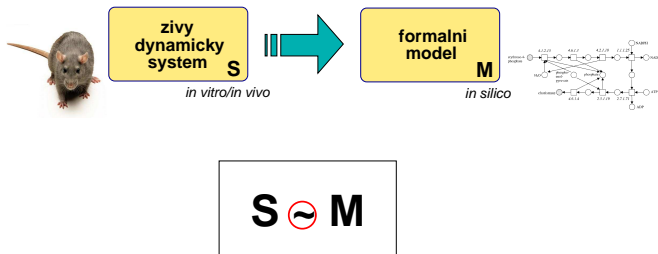
*Parametrizace modelu*

*Statická analýza*

*Modelování fotosyntézy*

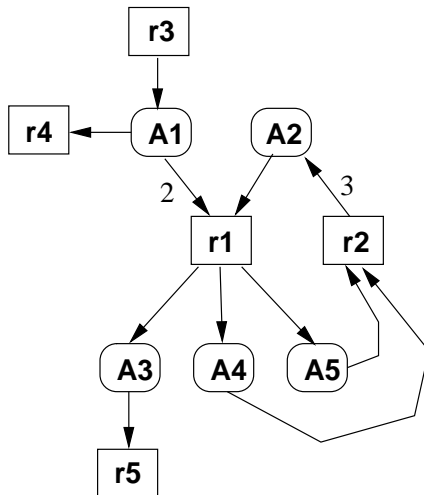


# Model jako abstraktní obraz organismu



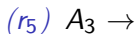
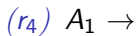
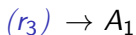
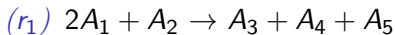
## *Tvorba modelu*

- 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*

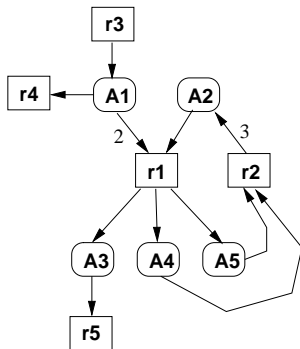
## Specifikace modelu – příklad základní notace

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 stechiometrickou maticí



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

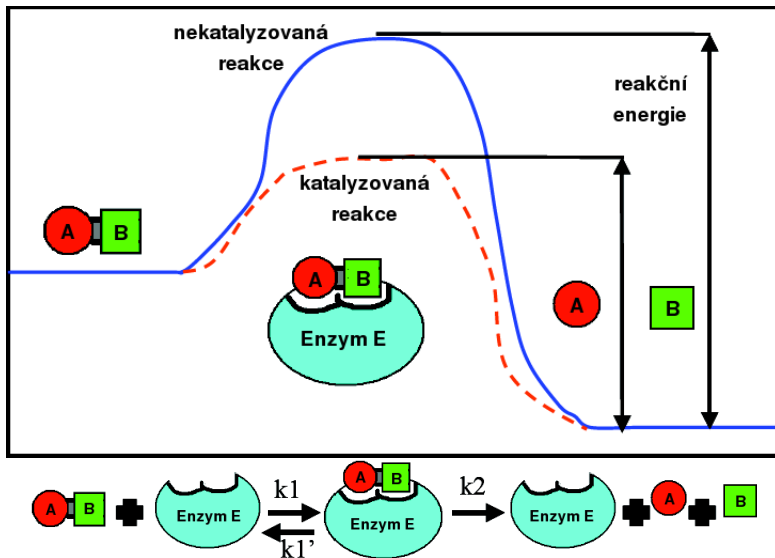
## *Jazyk SBML pro popis modelu*

- 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)

## *Východiska modelování v systémové biologii*

- biologický systém chápán jako rozsáhlý komplex biochemických procesů
- systémové modelování biochemických procesů má hlubokou historii
  - kinetika enzymů vyvíjená v 1.čtvrtině 20.stol.
  - matematické modelování chování populací
  - teorie systémů, řízení a kybernetika
- klíčem je abstrakce na populační úroveň
  - fyzikálně podloženo klasickou mechanikou
  - mnoho zjednodušujících předpokladů, na populační úrovni však výsledky smysluplné a experimentálně ověřené

# *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

## *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 stechiometrické 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:

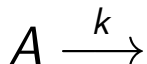
$$X(t) = \langle [S_1](t), \dots, [S_n](t) \rangle$$

- vývoj  $X$  v čase:

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

- průměrné chování lze charakterizovat exponenciální funkcí

## *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*

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

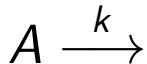
- jaká funkce má stejný tvar jako její derivace?

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

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

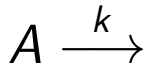
- platí

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

*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  $v_i(t)$

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

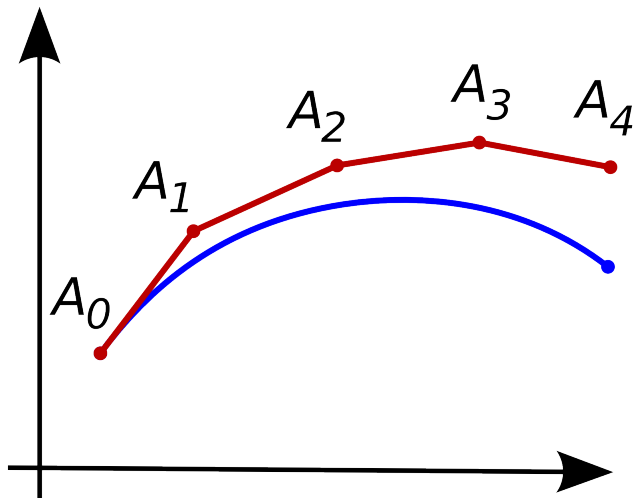


## *Podstata reakční konstanty*

- závislost reakční konstanty  $k$  na absolutní teplotě  $T$  (Arheniův zákon):

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

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

*Eulerova metoda*

## *Eulerova metoda*

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

$$\begin{aligned}y'(t) &= f(t, y(t)) \\ y(0) &= y_0\end{aligned}$$

- přesné řešení  $\varphi(t)$ :

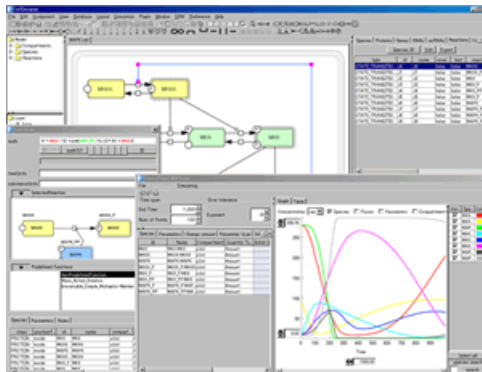
$$\begin{aligned}\varphi'(t) &= f(t, \varphi(t)) \\ \varphi(0) &= y_0\end{aligned}$$

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

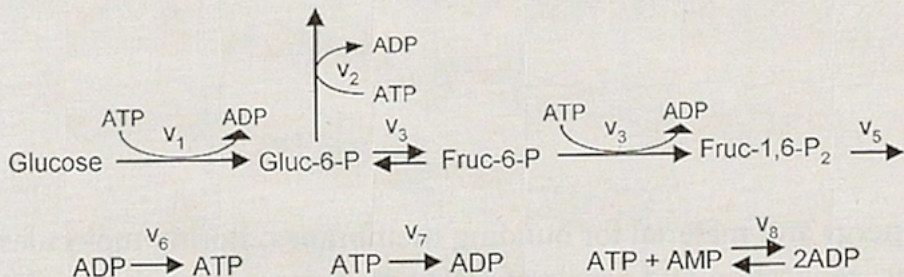
$$y_n \approx \varphi(t_n)$$

# Nástroj CellDesigner

- nástroj pro modelování biologických procesů
- zaměřený na grafickou specifikaci (SBGN)
- napojen na simulátory skrze SB Workbench
- základní metody analýzy zabudovány přímo



## *Cvičení – model metabolismu glukózy*



<http://www.biocyc.org/HUMAN/NEW-IMAGE?type=PATHWAY&object=GLYCOLYSIS&detail-level=2>

[http://www.genome.jp/kegg-bin/show\\_pathway?org\\_name=hsa&mapno=00010](http://www.genome.jp/kegg-bin/show_pathway?org_name=hsa&mapno=00010)

*Cvičení – model metabolismu glukózy*

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

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

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

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

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

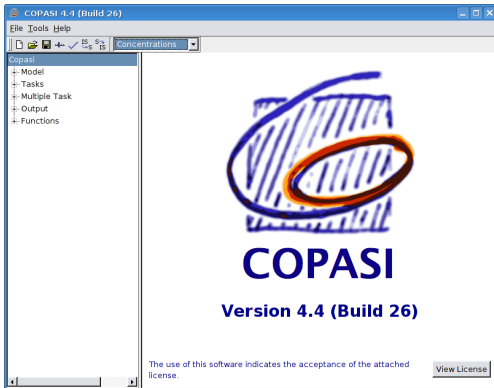
$$\frac{d}{dt} \text{AMP} = -v_8.$$

## *Cvičení – model metabolismu glukózy*

1. Načtěte model [http://www.fi.muni.cz/~xsafran1/PV225/Glucose\\_Metabolism.xml](http://www.fi.muni.cz/~xsafran1/PV225/Glucose_Metabolism.xml) do nástroje CellDesigner.
2. Proveďte simulaci o délce 0.5 časových jednotek s použitím výchozího nastavení parametrů a iniciálních hodnot.

# Nástroj COPASI

- nástroj pro simulaci dynamiky sítí chemických reakcí
- zaměřený především na deterministické solvery
- stechiometrická analýza
- analýza sensitivity parametrů
- estimace parametrů



<http://www.copasi.org/>



## *Cvičení – model metabolismu glukózy*

1. Načtěte model `http://www.fi.muni.cz/~xsafran1/PV225/glycolysis.cps` do nástroje COPASI.
2. Pomocí úlohy Parameter Scan – Time Course proveďte simulaci pro různá nastavení vstupní koncentrace glukózy (simulujte do času 0.5 min).
3. Zobrazte matematický model (diferenciální rovnice).

## *Základní kinetické zákony*

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

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

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

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

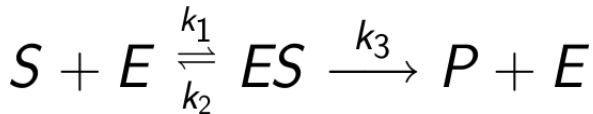
$$v_8 = k_{8f} \cdot \text{ATP} \cdot \text{AMP} - k_{8r} \cdot \text{ADP}^2$$

## *Pokročilé kinetické zákony*

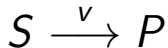
$$v_1 = \frac{V_{max,1} \cdot ATP}{K_{ATP,1} + ATP}$$

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

## Kinetika Michaelis-Menten

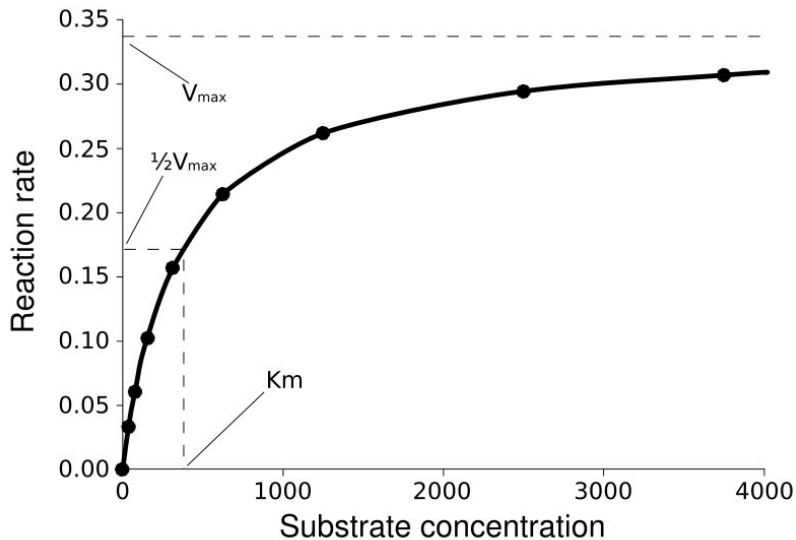


- nutnost znát  $k_1, k_2, k_3$  – obtížně měřitelné
  - $k_1, k_2$  prakticky velmi těžko měřitelné *in vitro*
  - $k_3$  lze získat *in vitro* ale obtížné *in vivo*
- zjednodušení na



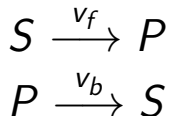
$$v = \frac{V_{max} \cdot S}{K + S}$$

- platí pokud  $S \gg E$

*Kinetika Michaelis-Menten*

## *Kinetika Michaelis-Menten*

- reversibilní reakce:



- model reakčního toku:

$$v_f = \frac{V_{max}^f \cdot S}{K_f + S}, \quad v_b = \frac{V_{max}^b \cdot P}{K_b + P}$$

- pro celkový tok reversibilní reakce platí:

$$v = \frac{\frac{V_{max}^f \cdot S}{K_f} - \frac{V_{max}^b \cdot P}{K_b}}{1 + \frac{S}{K_f} + \frac{P}{K_b}}$$

# *Obsah*

*Paradigma systémové biologie*

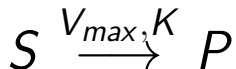
*Dynamické modely biologických procesů*

*Parametrizace modelu*

*Statická analýza*

*Modelování fotosyntézy*

## *Estimace parametrů*



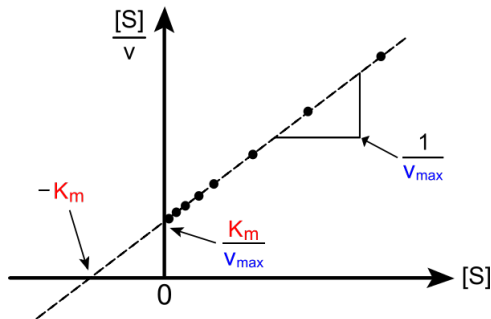
- model je parametrizovaný množinou parametrů  $\{V_{max}, K\}$
- $V_{max}, K$  lze získat nepřímo měřením *in vitro*
  - měří se koncentrace  $S(0)$  a odpovídající výkon reakce  $v$
  - po transformaci systému  $v = \frac{V_{max}S}{K+S}$  se souřadnicemi  $v, S$  na systém v souřadnicích  $\frac{S}{v}, S$  dostáváme:

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

toto lze využít pro lineární regresi (Hanes-Woolf plot)



## *Estimace parametrů*

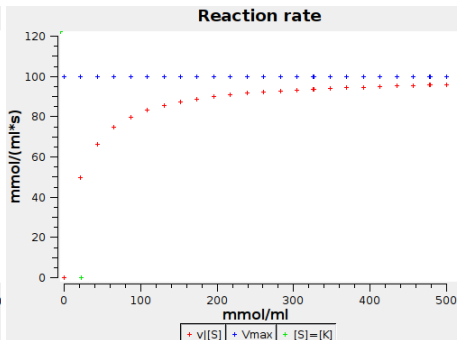
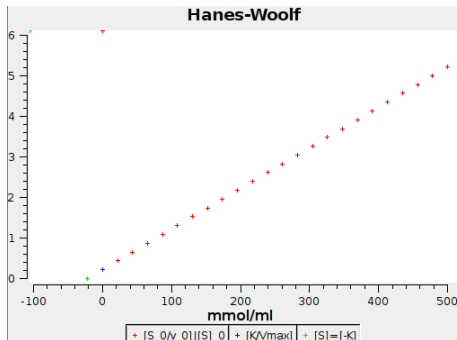


- měří se koncentrace  $S(0)$  a iniciální tok  $v$
- lineární regresí získáme  $K$  a  $V_{max}$
- problémy: chybovost měření, nerealizovatelné *in vivo*

Hanes, CS (1932). "Studies on plant amylases: The effect of starch concentration upon the velocity of hydrolysis by the amylase of germinated barley.". *Biochemical Journal* 26 (5): 1406-1421.

# Estimace parametrů

Demonstrace v COPASI



- závislost iniciálního toku  $v(0)$  na iniciální koncentraci  $S(0)$
- $S(0)$  samplováno rovnoměrně od 0 do 500
- $K$  (resp.  $-K$ ) znázorněno jako bod na ose  $S$

## 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*

- co je cílem?
  - najít takovou valuaci parametrů, která nejlépe odpovídá experimentálně zjištěným time-course datům
  - chceme tedy co nejvíce přiblížit simulaci experimentálním datům (tzv. **fitting**)
- lineární regrese požaduje normální rozložení chyb měření
- transformací se nepřesnosti kumulují
- model je inherentně nelineární
  - nelineární jsou i naměřená data
- problém fittingu chápán jako optimalizační problém
- mnoho heuristických metod pro aproximativní řešení  
⇒ viz COPASI → Parameter Estimation

## *Problém reverzního inženýrství*

- tzv. **inverzní problémy**
  - cílem je získat model z pozorování systému
  - obecně řešeno v teorii systémů (identifikace systémů)
  - pro nelineární systémy obecně neřešitelné
  - viz. IV120
- obecné schema řešení inverzního problému:
  1. identifikace vztahů mezi proměnnými
  2. identifikace funkcí popisujících sémantiku jednotlivých vztahů (např. zákon zachování hmoty, Michaelis-Menten, Hill, ...)
  3. **estimace hodnot parametrů ve funkcích získaných v předch. bodě**

# Problém identifikovatelnosti

## Systém a jeho pozorování

- mějme systém daný vektorovou diferenciální rovnicí:

$$\frac{dx}{dt} = f(x(t), p)$$

kde  $x \in \mathbb{R}^n$  je stavový vektor a  $p \in P$  je **vektor hodnot parametrů** z uvažovaného prostoru parametrů  $P$

- uvažujme  $T = \langle t_1, \dots, t_m \rangle$  rostoucí posloupnost časových bodů (tzv. **časovou řadu**)
- pozorovatelné (měřitelné) chování systému v časové řadě  $T$  je zachyceno funkcí:

$$y^M(t_i, p) = g(x(t_i, p))$$

- předpokládejme posloupnost  $\langle x(t_1, p), \dots, x(t_m, p) \rangle$  je aproximace řešení  $x$  zachyceného v časové řadě  $T$  a parametrizaci  $p$  (**simulace**)

# *Problém identifikovatelnosti*

## *Experimentální měření*

- mějme **experiment** jako posloupnost vektorů naměřených veličin  $\langle y^D(t_1), \dots, y^D(t_m) \rangle$  v časové řadě  $T$ :

$$y^D(t_i) \doteq y^M(t_i, p)$$

- pro jednoduchost uvažujme  $\dim(x) = \dim(y) = 1$  (obecně  $\dim(x) \geq \dim(y)$  libovolné, ale složitější formulace)
- uvažujeme (nereálnou) situaci přesného měření

# *Problém identifikovatelnosti*

## *Experimentální měření s chybou*

- mějme **experiment** jako posloupnost vektorů naměřených veličin  $\langle y^D(t_1), \dots, y^D(t_m) \rangle$  v časové řadě  $T$ :

$$y^D(t_i) = y^M(t_i, p) + \epsilon_i$$

- chyba *itého* časového bodu měření  $\epsilon_i$
- uvažujeme statistickou charakterizaci chyby normálním rozložením:

$$\epsilon_i = N(0, \sigma_i)$$

- $\sigma_i$  lze odhadnout např. opakovaným měřením v daném časovém bodě



## *Problém identifikovatelnosti*

- znalost struktury systému (funkce  $f$ )
- znalost experimentálního protokolu (funkce  $g$ )
- možno určit identifikovatelnost parametrů systému:

Lze najít parametrizaci, která jednoznačně vysvětluje daný experiment nad systémem?

## *Problém identifikovatelnosti*

- **strukturní identifikovatelnost** – teoretická vlastnost systému nezávislá na kvalitě a rozlišení měření
- systém je strukturně identifikovatelný vzhledem k protokolu  $g$  a parametrickému prostoru  $P$  pokud:

$$\forall p_1, p_2 \in P. p_1 \neq p_2 \implies \exists t. g(x(t, p_1)) \neq g(x(t, p_2))$$

## *Problém identifikovatelnosti*

- uvažujeme hodnoty parametrů v okolí určité hodnoty  $p_0$
- **lokální strukturní identifikovatelnost** – teoretická vlastnost systému nezávislá na kvalitě a rozlišení měření, ale uvažuje pouze lokální jednoznačnost
- systém je lokálně strukturně identifikovatelný vzhledem k protokolu  $g$  a  $\epsilon$ -okolí parametru  $p_0$  pokud:

$$\begin{aligned} \forall p_1, p_2 \in \{p \in P \mid \|p - p_0\| < \epsilon\}. p_1 \neq p_2 \\ \implies \exists t. g(x(t, p_1)) \neq g(x(t, p_2)) \end{aligned}$$

## *Problém identifikovatelnosti*

- analytické prokázání strukturní identifikovatelnosti je obtížné pro nelineární systémy
- uvažuje se pouze lokální identifikovatelnost v okolí reálné hodnoty parametru
- prakticky je určení lokální identifikovatelnosti závislé na kvalitě a rozlišení naměřených dat  
⇒ hovoříme o **praktické identifikovatelnosti**
- úspěšnost odhadu parametrů charakterizována statisticky

## *Estimace parametrů optimalizací*

- obecný postup:
  1. srovnej experimentální časovou řadu se simulovanou časovou řadou
  2. pokud rozdíl menší než nastavená tolerance → DONE  
jinak modifikuj parametry modelu
  3. proved time-course simulaci modelu
  4. iteruj (1)

## *Estimace parametrů optimalizací*

- definujeme odchylku experimentu od simulace v časovém bodě  $t_i$  jako tzv. **reziduál**:

$$r(t_i, p) = y^D(t_i) - y^M(t_i, p)$$

- reziduál chápeme jako funkci závislou na nastavení parametrů simulovaného modelu
- srovnání experimentu a simulace je vyjádřeno jako součet čtverců reziduálů přes vš. časové body  $T$ :

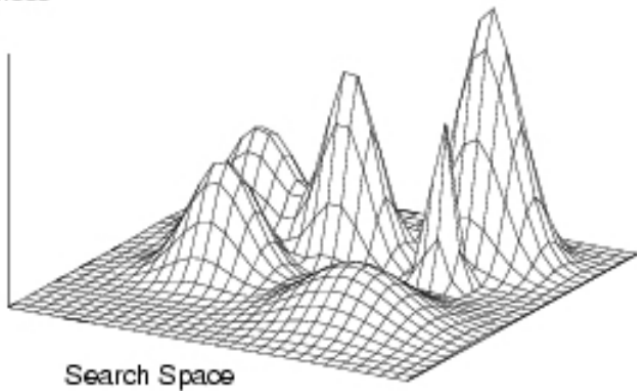
$$S(p) = \sum_i^m (r(t_i, p))^2$$

## *Estimace parametrů optimalizací*

- funkce  $\mathcal{S}(p)$  se nazývá **užitková funkce**
- vystihuje průměrnou odchylku simulace od experimentu přes danou časovou řadu
- **minimální hodnota**  $\mathcal{S}(p)$  určuje optimální vektor hodnot parametrů  $p$ , který globálně minimalizuje rozdíl mezi experimentem a modelem
- jedná se o nelineární funkci
- počet neurčitých parametrů určuje její dimenzi

# *Optimalizační krajina*

Fitness





## *Procházky po optimalizační krajině...*

- CÍL: najít **globální minimum**
- nejpoužívanější jsou stochastické black-box přístupy:
  - náhodné procházení (random search)
  - evoluční strategie (evolution strategy)
  - ...
- black-box znamená absolutní nezávislost na tvaru užitkové funkce
- existují i metody, které využívají znalosti užitkové funkce (např. simulované žíhání, Truncated Newton, ...)

# Procházky po optimalizační krajině...

## Random Search

1. inicializuj náhodně výchozí hodnotu  $p$ 
  - typicky z rovnoměrného rozložení
2. dokud není překročen povolený počet iterací, prováděj:
  - 2.1 sampluj novou pozici  $p'$ 
    - uniformní náhodný výběr z hyperkoule o daném poloměru
  - 2.2 spočítej  $S(p')$
  - 2.3 pokud  $S(p') < S(p)$ , nastav novou pozici  $p := p'$
3.  $p$  nastaveno na nejvýhodnější pozici (z pohledu běhu algoritmu)

*Pozn. Existují varianty s fixním i adaptivním poloměrem.*

# *Procházky po optimalizační krajině...*

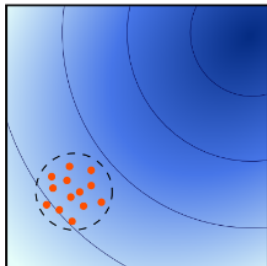
## *Evoluční strategie*

- inherentně adaptivní metoda, redukuje počet iterací
- staví na výběrech z (vícerozměrného) normálního rozložení
- rozměr daný dimenzí vektoru parametrů
- značeno CMA-ES (Covariance Matrix Evolution Strategy)
- postup:
  1. vytvoř generaci
    - sampluj rozložení pozic hodnot  $P$  dle normálního rozložení
  2. pro každé  $p \in P$  spočítej  $S(p)$
  3. adaptace: uprav parametry normálního rozložení pro další iteraci
    - různé varianty, mohou být velice komplexní adaptace citlivé na charakter evoluční krajiny – adaptace kovarianční matice
- metoda se ukazuje výhodná pro biologické modely (vysoká míra neznalosti parametrů)

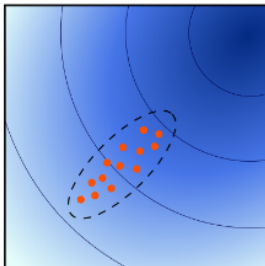
# *Procházky po optimalizační krajině...*

*Evoluční strategie*

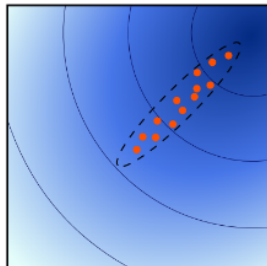
First generation



Second generation



Third generation



## *Estimace parametrů – vážená varianta*

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

$$S(p) = \sum_{i=1}^n w_i r_{i,p}^2$$

- 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>. Použijte metodu "Evolutionary Programming".
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, uvažujte parametrizaci  $V_{max} = 100, K = 22$ .
2. Proveďte estimaci iniciální podmínky  $S(0)$  tak, aby ve stabilním stavu bylo  $P = 400$ .
3. Uvažujte experimentální data <http://www.fi.muni.cz/~xsafran1/PV225/productsteady.csv>.

## 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 vstupní hodnoty koncentrace glukózy a parametru  $V_{max1}$  pro dataset [http://www.fi.muni.cz/~xsafran1/PV225/metaboldata\\_steady2.csv](http://www.fi.muni.cz/~xsafran1/PV225/metaboldata_steady2.csv) naměřená ve stabilním stavu (při zastaveném metabolismu). Parametry jsou uvažovány jako neznámé současně v jednom vektoru.
3. Proveďte estimaci stejné parametrizace pro data set [http://www.fi.muni.cz/~xsafran1/PV225/metaboldata\\_steady3.csv](http://www.fi.muni.cz/~xsafran1/PV225/metaboldata_steady3.csv) (naměřený rovněž ve stabilním stavu) a zjistěte v čem se liší od předchozího datasetu.

# *Obsah*

*Paradigma systémové biologie*

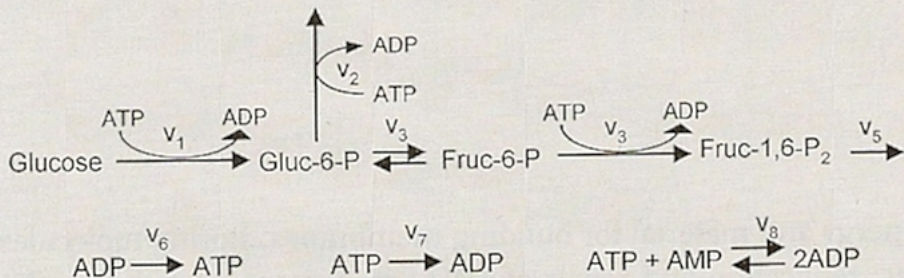
*Dynamické modely biologických procesů*

*Parametrizace modelu*

*Statická analýza*

*Modelování fotosyntézy*

## Příklad



## Příklad

Uvažujeme 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.

$$S = \begin{pmatrix} Gluc6P \\ Fruc6P \\ Fruc_{1,6}P_2 \\ ATP \\ ADP \\ AMP \end{pmatrix}$$

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

## *Hodnost stechiometrické matice*

- dimenze stechiometrického 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 stechiometrické matice

- dimenze stechiometrického 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 stechiometrické 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 stechiometrické 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 stechiometrické 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 stechiometrické matice

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

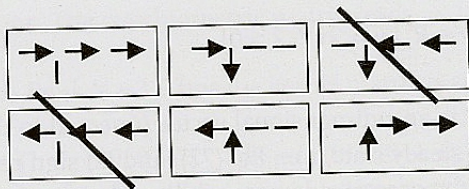
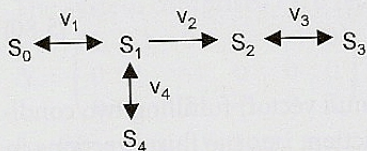
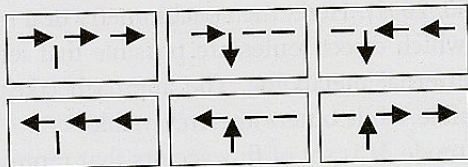
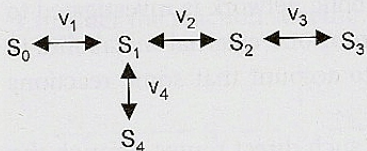
$$J = \sum_{i=1}^{|R|-h(M)} \alpha_i \times K(i)$$

- 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 stechiometrické matice



# *Obsah*

*Paradigma systémové biologie*

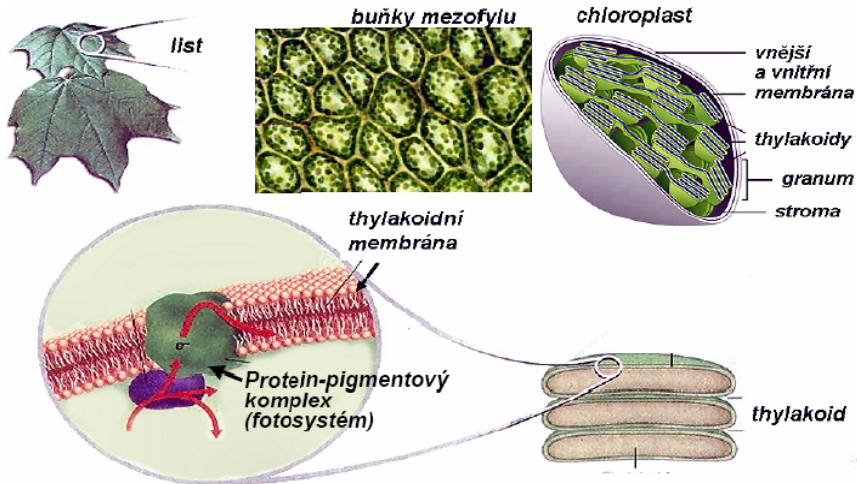
*Dynamické modely biologických procesů*

*Parametrizace modelu*

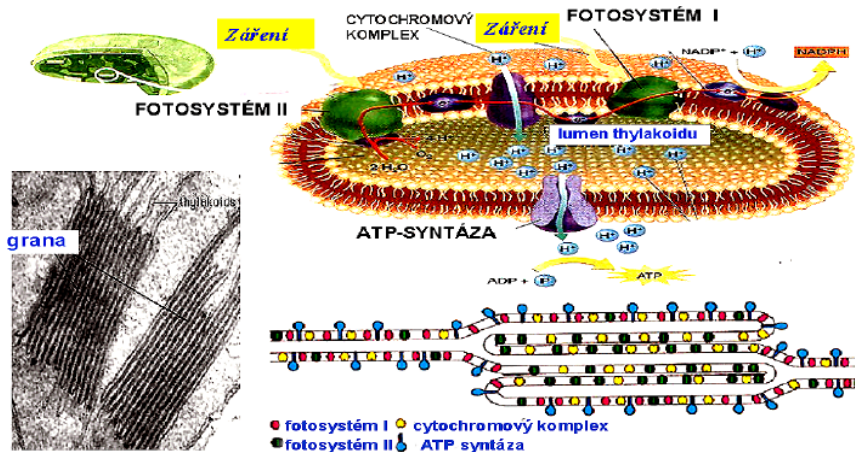
*Statická analýza*

*Modelování fotosyntézy*

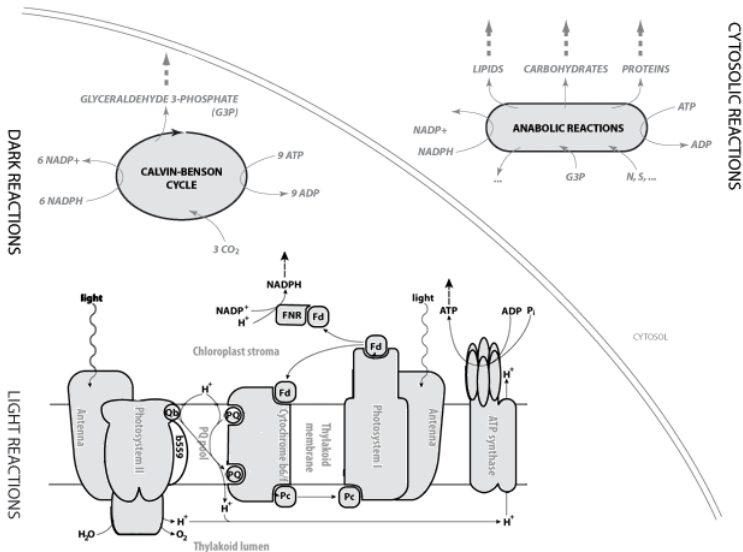
# Photosynthesis Process



# Photosynthesis Process

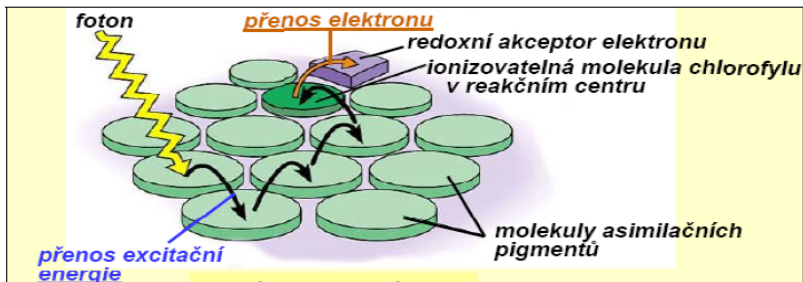
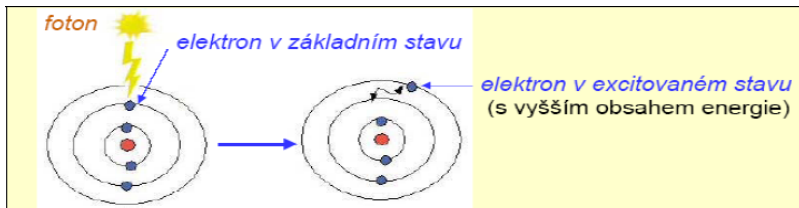


# Photosynthesis Process



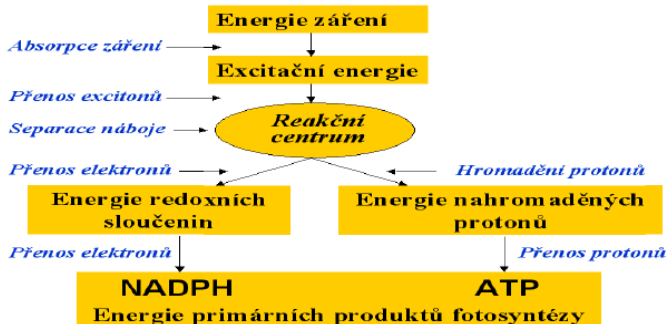


# Photosynthesis Process

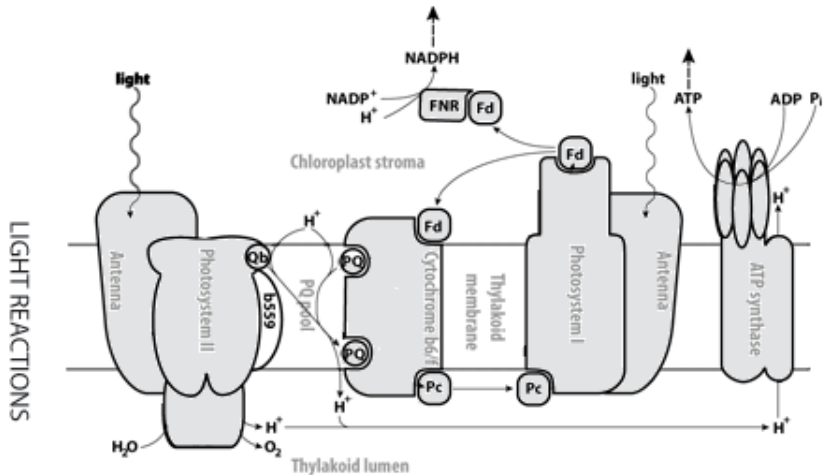


# Photosynthesis Process

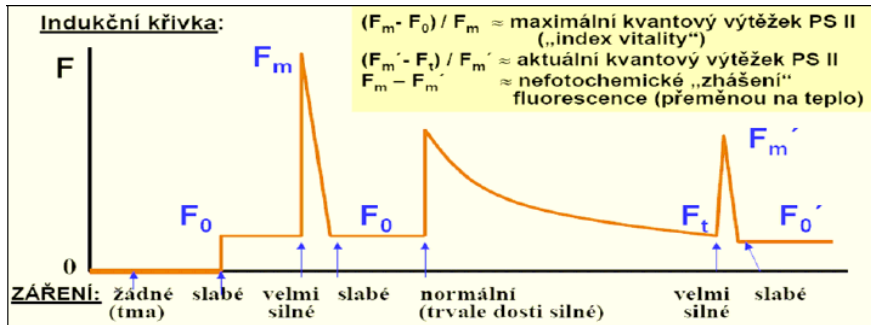
## Přehled typů procesů ve fotochemické části fotosyntézy



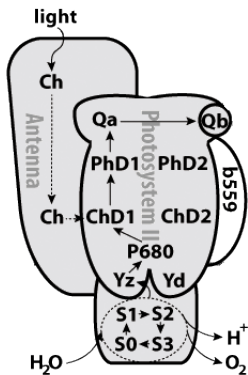
# Photosynthesis Process



# Photosynthesis Process – měření



## Photosystem II



- po excitaci antény dochází k přenosu excitonu:
  - oxidace prim. donoru (chlorofyl d, *ChID*)
  - oxidace sek. donoru (chlorofyl a, *P680*)
  - redukce primárního akceptoru  $Q_a$
  - redukce sekundárního akceptoru  $Q_b$
  - výstup: protonace plastochinonu *PQ*
- oxygen-evolving complex (OEC)
  - donace elektronů z manganového komplexu
  - 5 stavů, poslední nestabilní  
(štěpí  $2H_2O$  na  $4H^+$  a  $O_2$ )

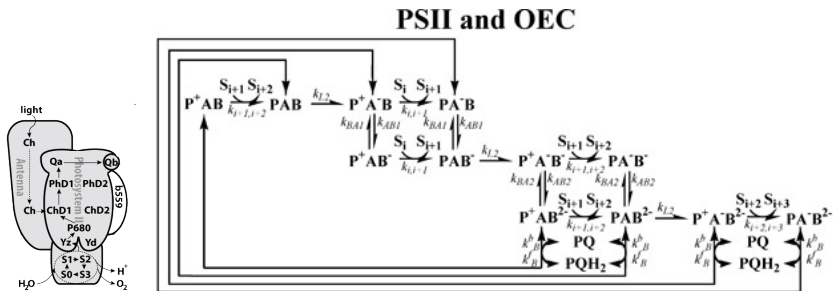


# *Photosystem II – model odezvy světelného impulzu*

## *Cvičení*

1. Na portálu <http://www.e-photosynthesis.org> prolistujte model Holzwarth\_2006.
2. Proveďte simulaci, zobrazte “Default” graphset sledujte vývoj proměnných modelujících excitovanou anténu, neutrální stav a redukovanou komponentu quinonu A.
3. Graphset “Fluorescence” zobrazuje energii excitovanou ve viditelném spektru.

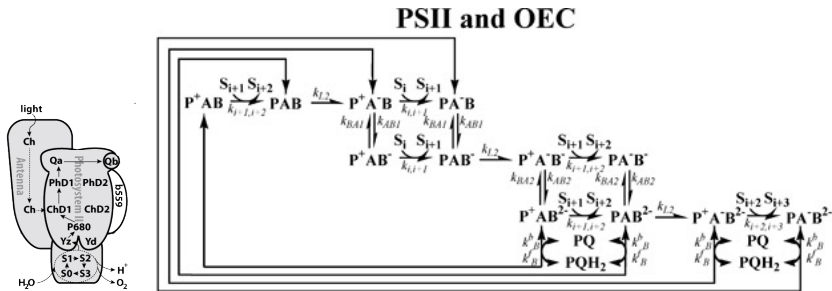
# Photosystem II – model fluorescence



- předpoklad vzájemné inertnosti různých PSII komplexů
- pouze otevřené stavy ( $Q_a$  neredukované) s neexcitovaným reakčním centrem absorbují foton



# Photosystem II – model fluorescence



- náběh fluorescence při dark-light přechodu je odhadován z celkového podílu zavřených stavů

$$F_{unq}(t) = \frac{(1-p) \cdot \sum[Q_a^-](t)}{1-p \cdot \sum[Q_a^-](t)}$$

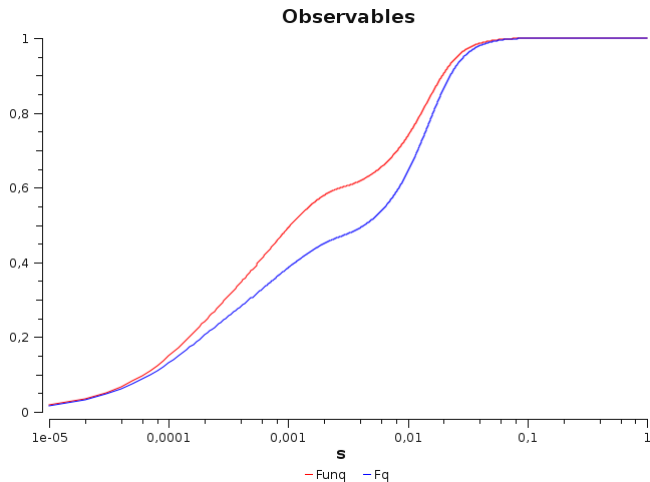
kde  $p$  vyjadřuje míru konektivity komplexů PSII v chloroplastu



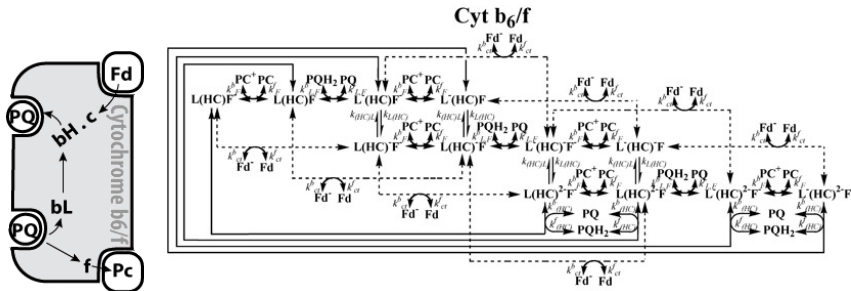
## *Specifika modelů fotosyntézy*

- procesy dynamiky konformací proteinových komplexů v membráně
- nejedná se o kompartmentové modely jejichž kinetika je určena pohybem v určitém objemu
- pro kinetiku těchto komplexů nelze uplatnit kinetiku 2. řádu
- použito procentuálního vyjádření podílu jednotlivých konformací a kinetiky 1. řádu
- přenosy mezi komplexy jsou definovány pohybem v objemu (poolu), proto tyto lze modelovat kinetikou 2. řádu

# Photosystem II – model

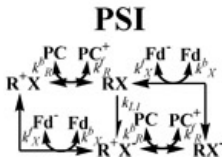
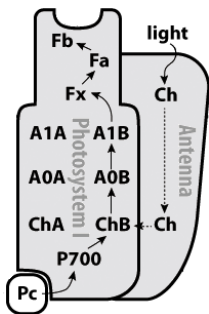


# Cytochrome b6/f



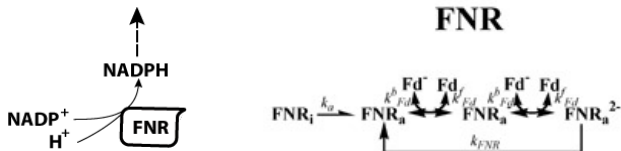
- zachycení plastohydrochinonu  $PQH_2$ 
  1. uvolnění 2 protonů do lumenu
  2. uvolnění 2 elektronů
    - 2.1 přenos na plastocyanin  $PC$
    - 2.2 přenos na hemovou redoxní skupinu a redukce  $PQ$
    - 2.3 donace od feredoxinu  $Fd$
- tzv. Q-cyklus (přečerpávání protonů ze stromatu do lumenu)
- výstup: protonace lumenu, redukce plastocyaninu

# Photosystem I



- absorpce radiace a oxidace reakčního centra  $P700$
- zachycení elektronu z plastocyaninu  $PC$  a donace reakčnímu centru
- dále redoxní řetězec
- výstup: přenos elektronu na feredoxin

# Photosystem I



- přenos  $Fd$  ke konečnému akceptoru  $\text{NADP}^+$
- redukce na  $\text{NADPH}$
- zprostředkováno feredoxin-NADP reduktázou
- alternativně je  $Fd$  použito k redukci plastochinonu v PQ-poolu
- výstup:  $\text{NADPH}$  nebo cyklický přenos (“nabíjení”)

## *Cvičení*

1. Na <http://www.e-photosynthesis.org> prozkoumejte model Lazar\_2009.
2. Proveďte simulaci a pozorujte fluorescenční křivku  $F_q$ .
3. Prostudujte tutoriál k estimaci parametrů pro tento model, viz [http://anna.fi.muni.cz/~xsafran1/PV225/parameter\\_estimation/report.html](http://anna.fi.muni.cz/~xsafran1/PV225/parameter_estimation/report.html).