

Integracija modelov metabolizma in genskega regulatornega omrežja

Uvod

Jernej Puc

IŠRM – UNI-LJ

Marec 2019

Modeliranje bioloških sistemov

Opazovanje bioloških pojavov nas seznaja z različnimi zapletenimi procesi, ki pogosto nimajo očitnega pojasnila in katerih posledice niso enostavno predvidljive.

Naše razumevanje teh procesov se posodablja na podlagi poskusov, ki vodijo do hipotez o posameznih procesih, a za te večinoma ni jasno, ali so smiselne tudi v kontekstu širšega okolja.

Pri tem nam lahko pomagajo matematično modeliranje in računalniške simulacije.

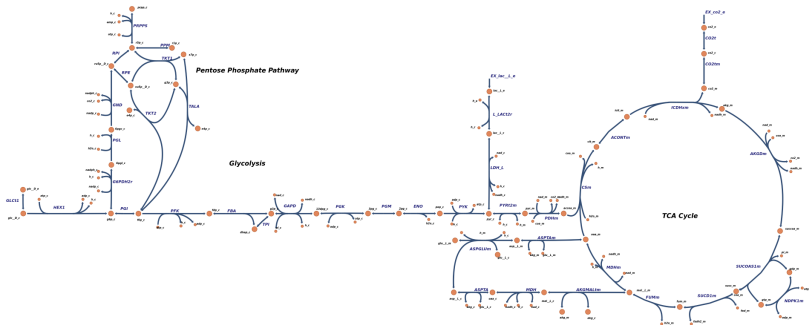
Modeliranje bioloških sistemov

Model je (v splošnem) abstraktna predstavitev predmeta, procesa ali sistema, ki opisane lastnosti povzema v obliki koherentne matematične izrazitve.

Upoštevanje modelov ima številne prednosti pred samimi poskusi:

- ▶ Modeliranje zahteva, da so postavljene hipoteze natančne in dosledne:
 - Tekom formulacije se vidno izpostavijo luknje v razumevanju.
- ▶ Modeliranje omogoča napovedovanje in oblikovanje novih hipotez:
 - Ko je model oblikovan, lahko simulacije (obravnavajo *in silico*) znižajo potrebo po nadaljnjih poskusih ali predlagajo najbolj obetavne smernice nadaljnjih poskusov.
- ▶ Predstavitev v obliki matematičnih izrazov omogoča lažje posploševanje.

Modeliranje (bio)kemijskih reakcijskih omrežij



Slika: V različnih primerih se uporabljajo različni pristopi modeliranja. Splet (bio)kemijskih reakcij je običajno predstavljen z omrežjem, kjer kot vozlišča in povezave nastopajo snovi in reakcije.

Modeliranje (bio)kemijskih reakcijskih omrežij

Najbližje resničnosti bi moral model obravnavati diskretna števila (molekul) ter interakcije med njimi.

Če so dovolj velika, se jih lahko predstavi z nenegativnimi realnimi števili, ki opisujejo njihovo gostoto oz. koncentracijo (število molekul na prostorsko enoto), njihove spremembe (zaradi reakcij) pa z zveznimi funkcijami.

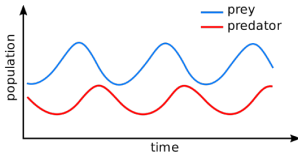
- ▶ Raven abstrakcije načeloma zahteva kompromis – kompleksnost proti natančnosti modela.
- ▶ Pri poenostavitvi se predpostavlja, da je izguba natančnosti zanemarljiva.

Ista omrežja se tako lahko ponazori s sistemi navadnih diferencialnih enačb (angl. *ordinary differential equations* – ODE).

- ▶ Za modeliranje biokemijskih reakcijskih omrežij so ti najbolj pogosto uporabljeni.

Modeliranje (bio)kemijskih reakcijskih omrežij

$$\begin{aligned}\frac{dx}{dt} &= \alpha x - \beta xy \\ \frac{dy}{dt} &= \delta xy - \gamma y\end{aligned}$$



Slika: Primer enostavnega sistema ODE (Lotka-Volterra).

Definicija sistema:

- ▶ Parametri opisujejo lastnosti modela, ki se znotraj ene simulacije ne spreminjajo (npr. hitrost neke sintezne reakcije)
- ▶ Spremenljivke opisujejo lastnosti modela, ki se tekom simulacije spreminjajo (npr. koncentracija nekega proteina)
 - Stanje: množica vrednosti spremenljivk, ki definirajo celoten sistem na nekem časovnem koraku.
- ▶ Pravila oz. zveze, ki določajo spremembe spremenljivk skozi čas.

Modeliranje (bio)kemijskih reakcijskih omrežij

Ista količina je lahko modelirana kot parameter ali spremenljivka.

V določenem primeru se npr. koncentracijo nekega encima privzame za konstantno, v drugem, bolj splošnem, pa je lahko tudi ta spremenljiva, kar se odraža v razširjenem modelu.

Zahteva po kompleksnosti matematičnega modela je odvisna od obravnavanega problema in namena, ki naj bi ga model izpolnjeval.

Modeliranje genskih regulatornih omrežij

Genska regulatorna omrežja so biokemijska omrežja, ki jih sestavljajo osnovni procesi na nivoju genov in imajo pomembno vlogo pri regulaciji oz. nadzoru drugih bioloških procesov (npr. cirkadiana ura).

Mehanizem:

- ▶ Gen: odsek DNK oz. zaporedje parov nukleotidov, ki kodira nek protein (in tako vodi do izražanja lastnosti organizma).
- ▶ Gensko izražanje: aktivatorski ali represorski (inhibitorski) proteini se vežejo na (ti. regulatorni) del gena, ki določa ali se preostalo (kodirajoče) zaporedje (ti. ekspresivni del) prepíše v mRNA (ki se nato prevede v protein) ali ne.

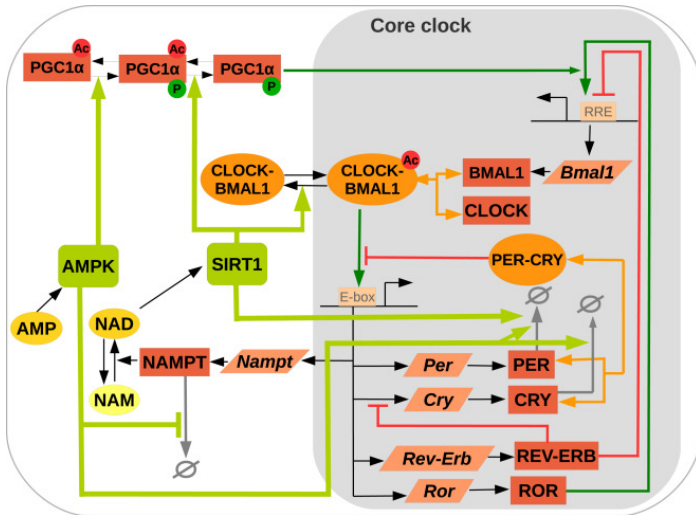
Modeliranje genskih regulatornih omrežij

Osnovni procesi, ki nastopajo pri regulaciji genskega izražanja:

- ▶ prepisovanje (transkripcija) gena v mRNA,
- ▶ prevajanje (translacija) mRNA v ciljni protein,
- ▶ razpad (degradacija) proteina ali mRNA.

Proteini lahko hkrati predstavljajo pogoj in posledico genskega izražanja oz. vplivajo na izražanje drug drugega. Na ta način povezani osnovni procesi lahko tvorijo različno kompleksna omrežja.

Primer: cirkadiana ura – Woller et al., 2016



Slika: Omrežni model sesalske cirkadiane ure.

Primer: cirkadiana ura – Woller et al., 2016

$$\begin{aligned}\frac{d}{dt}y_0(t) &= -\delta_4y_0(t) + V_4 \frac{1 + \alpha_4 act(t, 4, 9)}{1 + act(t, 4, 9)(1 + rep(t, 10, \beta_{12}, 10))}, \\ \frac{d}{dt}y_1(t) &= -\delta_1y_1(t) + \frac{V_1}{rep(t, 7, x_7(t), 4)} \frac{1 + \alpha_1 act(t, 1, 2)}{1 + act(t, 1, 2)(1 + rep(t, 10, \beta_9, 4))}, \\ \frac{d}{dt}y_2(t) &= -\delta_5y_2(t) + V_5 \frac{1 + \alpha_5 act(t, 5, 11)}{1 + act(t, 5, 11)(1 + rep(t, 10, \beta_{13}, 12))}, \\ \frac{d}{dt}y_3(t) &= -\delta_6y_3(t) + V_6 \frac{1 + \alpha_6 act(t, 6, 13)}{1 + act(t, 6, 13)(1 + rep(t, 10, \beta_{14}, 14))}, \\ \frac{d}{dt}y_4(t) &= -\delta_0y_4(t) + V_0 \frac{1 + \alpha_0(1 + x_3(t))rep(t, 8, \beta_0, 1)}{1 + rep(t, 8, \beta_0, 1) + rep(t, 7, x_6(t), 0)}, \\ \frac{d}{dt}y_5(t) &= \alpha_{10}y_0(t) - \delta_{10}y_5(t)(1 + \beta_{18}x_1(t) + \beta_{17}x_2(t)) - \\ &\quad - \kappa_1y_5(t)y_6(t) + \kappa_{-1}y_{10}(t), \\ \frac{d}{dt}y_6(t) &= \alpha_8y_1(t) - \delta_8y_6(t)(1 + \beta_{15}x_2(t)) - \\ &\quad - \kappa_1y_5(t)y_6(t) + \kappa_{-1}y_{10}(t), \\ \frac{d}{dt}y_7(t) &= \alpha_{11}y_2(t) - \delta_{11}y_7(t), \\ \frac{d}{dt}y_8(t) &= \alpha_{12}y_3(t) - \delta_{12}y_8(t), \\ \frac{d}{dt}y_9(t) &= \alpha_7y_4(t) - \delta_7y_9(t) - \kappa_0y_9(t) + \kappa_{-0}y_{11}(t), \\ \frac{d}{dt}y_{10}(t) &= \kappa_1y_6(t)y_5(t) - \kappa_{-1}y_{10}(t) - \delta_{14}y_{10}(t), \\ \frac{d}{dt}y_{11}(t) &= \kappa_0y_9(t) - \kappa_{-0}y_{11}(t) - \delta_{13}y_{11}(t), \\ \frac{d}{dt}y_{12}(t) &= -\delta_3y_{12}(t) + V_3 \frac{1 + \alpha_3 act(t, 3, 7)}{1 + act(t, 3, 7)(1 + rep(t, 10, \beta_{11}, 8))}, \\ \frac{d}{dt}y_{13}(t) &= \alpha_9y_{12} - \frac{\delta_9y_{13}}{1 + \beta_{16}x_2}, \\ \frac{d}{dt}y_{14}(t) &= \frac{V_8y_{13}(c_3 - y_{14})}{\beta_{22} + c_3 - y_{14}} - \frac{\alpha_{13}(y_{14} - \beta_{21})}{\beta_{20} + y_{14} - \beta_{21}}, \\ \frac{d}{dt}y_{15}(t) &= -\delta_2y_{15}(t) + V_2 \frac{1 + \alpha_2 act(t, 2, 5)}{1 + act(t, 2, 5)(1 + rep(t, 10, \beta_{10}, 6))}.\end{aligned}$$

Reševanje sistemov ODE

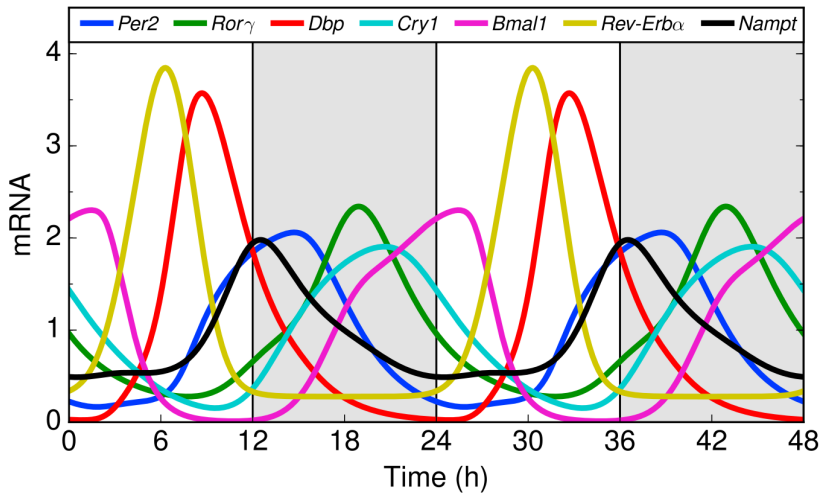
Opisovanje dinamike izrazito prepletenih procesov vodi do kompleksnih nelinearnih sistemov. Te je nepraktično reševati analitično, zato se pri iskanju rešitev poslužuje reševanja preko numerične integracije.

Najbolj enostaven primer numerične integracije je ti. Eulerjeva metoda. Ob $\frac{dy}{dt} = f(y)$ z dovolj majhnim korakom h in danih začetnih pogojih $y(0) = y_0$:

$$y(t + h) = y(t) + h \cdot f(y(t))$$

V Pythonovem okolju sta za numerično reševanje na voljo npr. knjižnici `scipy.integrate` in `JiTCode`.

Primer: circadiana ura – Woller et al., 2016



Modeliranje metabolnih omrežij

Med produkti genskega izražanja so tudi razni encimi – proteinski katalizatorji kemijskih reakcij – in prenašalni proteini.

Ti tvorijo podlago za medsebojne interakcije med snovmi v nekem biološkem okolju (npr. celici).

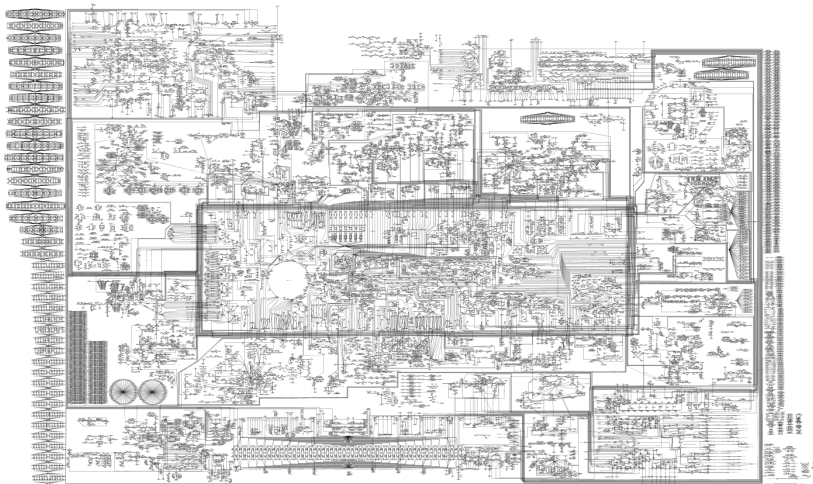
Splet ključnih kemijskih reakcij v organizmu (metabolizem) in snovi, ki pri teh sodelujejo (metaboliti), tvorijo ti. metabolna omrežja.

Modeliranje metabolnih omrežij

Na podlagi kodirajočega zaporedja v genomskem (DNK/RNK) materialu organizma (celice, pripadajoče npr. enoceličnemu organizmu ali nekemu tkivu) se lahko približno izpelje celotno pripadajoče metabolno omrežje, ki povzema vse reakcije in snovi, ki jih lahko celica proizvede.

Podatki, ki predstavljajo metabolno omrežje organizma, so sistematično poenoteni v obliki ti. genomske razsežnih metabolnih modelov (ang. *genome-scale metabolic models* – GEMs).

Modeliranje metabolnih omrežij



Slika: (Ročna) vizualizacija kompleksnega primera metabolnega omrežja.

Modeliranje metabolnih omrežij

GEM-a za vsako encimsko reakcijo v omrežju opisuje:

- ▶ substrate oz. reaktante in produkte, ki sodelujejo v reakciji,
- ▶ razmerja med reaktanti in produkti (stehiometrične koeficiente),
- ▶ usmerjenost reakcije,
- ▶ razdelek celice, v katerem poteka.

Poleg tega vključuje dodatne “reakcije”, ki opisujejo prenos metabolitov med posameznimi razdelki ali izmenjavo z zunanjim okoljem.

Zaradi velikega obsega in proste dostopnosti modelov so GEM-i popularni v modeliranju metabolnih omrežij. Primera javnih podatkovnih baz sta npr.:

- ▶ <http://bigg.ucsd.edu/models/>
- ▶ <https://www.metabolicatlas.org/gems/repository>

Analiza metabolnih omrežij

Za primer reakcije s stehiometričnim opisom



in stehiometričnimi koeficienti $[-1, -1, 2]$, se lahko spremembe koncentracij metabolitov opiše z navadnimi diferencialnimi enačbami:

$$\frac{dS_1}{dt} = -v, \quad \frac{dS_2}{dt} = -v, \quad \frac{dP}{dt} = 2v$$

kjer je v hitrost reakcije oz. ti. metabolni pretok.

Sistem enačb v tem primeru pomeni, da nastaja metabolit P dvakrat hitreje kot razpadata metabolita S_1 in S_2 .

Analiza metabolnih omrežij

Metabolnemu omrežju, sestavljenemu iz m metabolitov in r reakcij, pripada sistem diferencialnih enačb:

$$\frac{dS_i}{dt} = \sum_{j=1}^r n_{i,j} v_j, \quad i = 1, \dots, m$$

kjer so $n_{i,j}$ stehiometrični koeficienti za i -ti metabolit v j -ti reakciji.

Pri tem sta narejeni naslednji predpostavki:

- ▶ Kemijske reakcije so edini vzrok spreminjanja koncentracij.
- ▶ Zunanji metaboliti, ki niso vključeni v model, nimajo znatnega vpliva na opisani sistem.

Analiza metabolnih omrežij

Stehiometrični koeficienti $n_{i,j}$ so lahko poenoteni s ti. stehiometrično matriko \mathbf{N} , kjer vsak stolpec pripada določeni reakciji, vsaka vrstica pa metabolitu:

$$\mathbf{N} = \begin{bmatrix} n_{1,1} & n_{1,2} & \dots & n_{1,r} \\ n_{2,1} & n_{2,2} & \dots & n_{2,r} \\ \vdots & \vdots & \ddots & \vdots \\ n_{m,1} & n_{m,2} & \dots & n_{m,r} \end{bmatrix}$$

Zaradi obsežnosti metabolnih omrežij so lahko tudi stehiometrične matrike precej velike, vendar so redke, saj v posamezni reakciji navadno nastopa le malo število metabolitov.

Predstavitev koncentracij in pretokov z vektorjema \mathbf{S} in \mathbf{v} tako privede do matričnega zapisa sistema ODE:

$$\frac{d\mathbf{S}}{dt} = \mathbf{N}\mathbf{v}$$

Analiza metabolnih omrežij

Tako veliki sistemi ODE zahtevajo določitev velikega števila parametrov, kar je na eksperimentalni osnovi pogosto neizvedljivo.

V ta namen se predpostavi, da so metabolna omrežja organizmov v nespremenljivem okolju večino časa v ravnovesnem stanju oz. da v nekem trenutku opazovanja velja ravnovesje med produkcijo in porabo metabolitov:

$$\frac{d\mathbf{S}}{dt} = \mathbf{N}\mathbf{v} = \mathbf{0}$$

V primeru, da ob spremembi v okolju ravni koncentracij dejansko skonvergirajo proti novemu ravnovesnemu stanju, se predpostavki ugotovi z opazovanjem sistema z dovolj velikim časovnim korakom, da vmesni prehodi izzvenijo.

Tako se obide potrebo po množici eksperimentalno določenih parametrov, s tem pa se pozornost pri obravnavi omrežja preusmeri s koncentracij in procesov, ki jih spreminjajo, na aktivnosti reakcij in procese, ki spreminjanje slednje.

Analiza metabolnih omrežij

Namesto sistema ODE se rešuje matrično enačbo $\mathbf{N}\mathbf{v} = \mathbf{0}$, katere rešitve predstavljajo kombinacije posameznih metabolnih pretokov, ki so dopustne v ravnovesnem stanju.

Ta bo imela netrivialne rešitve, ko bo veljalo $\text{Rank}(\mathbf{N}) < r$. Temu je v splošnem zadoščeno, saj je večinoma r (dosti) večji od m .

Prostor rešitev določa jedrna matrika \mathbf{K} , ki zadošča enačbi $\mathbf{N}\mathbf{K} = \mathbf{0}$, kjer vsak izmed $r - \text{Rank}(\mathbf{N})$ stolpcev predstavlja eno izmed rešitev, ki je linearno neodvisna od ostalih.

Linearne kombinacije stolpcev jedrne matrike prav tako privedejo do dopustne rešitve. Vseh možnih rešitev je torej neskončno.

Sicer neskončni prostor rešitev se lahko omeji oz. privede do enolične rešitve z dodatkom optimizacijskega kriterija.

To je v ozadju ti. analize ravnovesja pretokov (ang. flux balance analysis – FBA), ki je med popularnejšimi metodami v analizi metabolnih omrežij.

Rešitev FBA opisuje porazdelitev pretokov, ki je za celico optimalna po nekem kriteriju, npr. omogoča najvišjo stopnjo celične rasti glede na omejen pritok hranilnih snovi.

FBA – omejitve

FBA vse pretoke obravnava znotraj omejenega intervala dopustnih vrednosti. Pri tem točne vrednosti niso bistvenega pomena, saj poljubno skaliranje privede do enakih razmerij med pretoki – tipično so omejene na $[-1000, 1000]$ oz. $[-1000, 0]$ ali $[0, 1000]$ za reakcije, ki niso obrnljive pod danimi pogoji.

Znotraj teh intervalov se lahko pretoke dodatno omeji glede na dodatne informacije o opazovanem biološkem sistemu, npr.:

- ▶ razpoložljivost oz. koncentracija encimov,
- ▶ razpoložljivost oz. koncentracija reaktantov,
- ▶ obrnljivost reakcij,
- ▶ razporeditev reakcij po razdelkih,
- ▶ regulacija na nivoju genskega izražanja.

Dopustne vrednosti pretokov so tako zgoraj in spodaj omejene:

$$v_i^{\min} \leq v_i \leq v_i^{\max}, \quad i = 1, \dots, r$$

FBA – optimizacijski kriterij

Optimizacijski kriterij je podan z linearno funkcijo, ki opisuje doprinos posamezne reakcije k neki lastnosti opazovanega sistema, ki je relevantna za obravnavani problem – porazdelitev pretokov mora maksimizirati (ali minimizirati) kriterijsko funkcijo:

$$\max \sum_{i=1}^r c_i v_i$$

ki predstavlja npr. sintezo pomembnega metabolita ali tudi bolj abstrakten kriterij, kot je npr. celična rast oz. kopičenje biomase.

FBA – linearni program

Rešitev se poišče s konveksno optimizacijo. V ta namen se vse reakcije obravnava kot neobrnljive (obrnljive se lahko razdeli na ločena neobrnljiva dela). Navedene omejitve, optimizacijski kriterij in nenegativnost baznih vektorjev v_i tako tvori problem v standardni obliki linearnega programa:

$$\max \mathbf{c}^T \mathbf{v}$$

p.p.

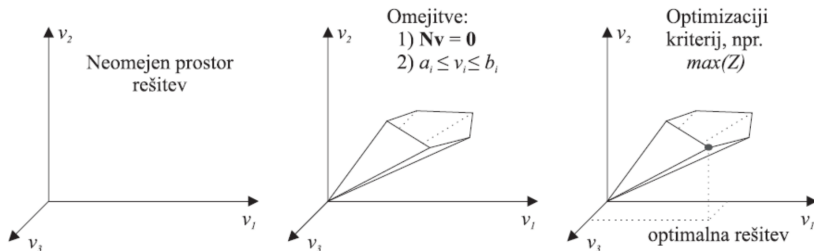
$$\mathbf{N}\mathbf{v} = \mathbf{0}$$

$$v_i^{\min} \leq v_i \leq v_i^{\max}, \quad i = 1, \dots, r$$

$$v_i \geq 0, \quad i = 1, \dots, r$$

Tak linearni program se lahko reši npr. z uporabo simpleksnega algoritma.

FBA – linearni program



Slika: Znotraj večdimenzionalnega prostora, ki ga tvorijo bazni vektorji pretokov v_i , oblikujejo omejitve pretokov konveksni polieder dopustnih rešitev. Lahko se pokaže, da če vsebuje tudi optimalno rešitev po danem linearnem kriteriju, mora ta ležati na njegovi površini – glede na smer gradienta je bodisi enolična in leži v nekem oglišču bodisi je kriterij maksimiziran po celotni stranici.

S FBA se lahko poišče optimalno porazdelitev pretokov v omrežju, vendar ni rečeno, da opazovani biološki sistem deluje enako. Poleg tega je možno, da nekatere rešitve, ki jih model dopušča, nimajo biološkega smisla – to je pričakovana posledica pomanjkljivosti modela.

Pri analizi gre torej bolj za preučitev poti v omrežju, ki prispevajo k danemu kriteriju, in kritičnih elementov, ki ga najbolj omejujejo.

Kar se tiče programskih orodij, je med popularnejšimi COBRA (constraint based reconstruction and analysis), ki je na voljo v (izvorni) matlab in Python implementaciji [COBRApy](#).

- ▶ Edda Klipp, Wolfram Liebermeister, Christoph Wierling, Axel Kowald. *Systems Biology: A Textbook*. John Wiley & Sons, 2016.
- ▶ Živa Rejc, L. Magdevska, T. Tršelič, T. Osolin, R. Vodopivec, J. Mraz, E. Pavliha, N. Zimic, T. Cvitanović, D. Rozman, M. Moškon, M. Mraz. *Computational modelling of genome-scale metabolic networks and its application to cho cell cultures*. Computers in Biology and Medicine, vol. 88, pp. 150–160, 2017.
- ▶ A. Woller, H. Duez, B. Staels, M. Lefranc, *A mathematical model of the liver circadian clock linking feeding and fasting cycles to clock function*. Cell Reports, vol. 17, pp. 1087–1097, 2016.