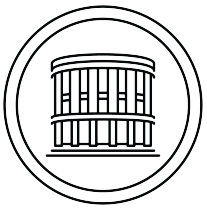
UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY



VARIANCIA V GENETICKÝCH REGULAČNÝCH SIEŤACH

DIPLOMOVÁ PRÁCA

2024 Laura Hajzoková

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

VARIANCIA V GENETICKÝCH REGULAČNÝCH SIEŤACH

DIPLOMOVÁ PRÁCA

Študijný program:	Ekonomicko-finančná matematika a modelovanie
Študijný odbor:	Matematika
Školiace pracovisko:	Katedra aplikovanej matematiky a štatistiky
Vedúci práce:	doc. Mgr. Pavol Bokes, PhD.

Bratislava 2024 **Laura Hajzoková**





Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Bc. Laura Hajzoková
Študijný program: ekonomicko-finančná matematika a modelovanie (Jednoodborové štúdium, magisterský II. st., denná forma)
Študijný odbor: matematika
Typ záverečnej práce: diplomová
Jazyk záverečnej práce: slovenský
Sekundárny jazyk: anglický

Názov: Variancia v genetických regulačných sieťach
Variance in gene regulatory networks

Anotácia: Expresia génov je heterogénna medzi geneticky identickými bunkami a dynamicky v nich fluktuuje. Vzájomné interakcie génov formujú genetické regulačné siete. V tejto súvislosti vyvstáva otázka, nakoľko architektúra genetickej regulačnej siete vplýva na variabilitu génovej expresie, a naopak, nakoľko je možné z variability génovej expresie inferovať architektúru genetickej regulačnej siete. Projekt záverečnej práce sa bude venovať tejto otázke z matematického hľadiska.

Vedúci: doc. Mgr. Pavol Bokes, PhD.
Katedra: FMFI.KAMŠ - Katedra aplikovanej matematiky a štatistiky
Vedúci katedry: prof. RNDr. Marek Fila, DrSc.

Dátum zadania: 10.01.2023

Dátum schválenia: 11.01.2023 prof. RNDr. Daniel Ševčovič, DrSc.
garant študijného programu

.....
študent

.....
vedúci práce

Poďakovanie Týmto ďakujem vedúcemu diplomovej práce, doc. Bokesovi, za jeho konštruktívne pripomienky a trpezlivosť.

Abstrakt

HAJZOKOVÁ, Laura: Variancia v genetických regulačných sieťach [Diplomová práca], Univerzita Komenského v Bratislave, Fakulta matematiky, fyziky a informatiky, Katedra aplikovanej matematiky a štatistiky; školiteľ: doc. Mgr. Pavol Bokes, PhD., Bratislava, 2024, 46 s.

Táto diplomová práca sa zaoberá modelovaním genetickej regulačnej kaskády a jeho stabilitou. Je známe, že v deterministickom prípade vieme určiť stabilitu v závislosti od parametrov chemických reakcií aj príslušné bifurkácie. V stochastickom prípade však vstupuje do systému parameter označujúci celkový počet molekúl a deterministická stabilita sa v závislosti od tohto parametra mení. V tejto práci preto skonštruujeme syntetickú proteínovú kaskádu so siedmimi stupňami, ktorá v deterministickom prípade dosahuje tzv. bistabilitu, pričom ju stochasticky modelujeme pomocou master rovnice. V závislosti od parametra objemu systému potom skúmame jeho stabilitu a prostredníctvom simulácií sa snažíme určiť aj bod zmeny stabilného stavu. Práca detailne popisuje teoretické pozadie deterministického aj stochastického prístupu k modelovaniu genetických regulačných sietí, ktorý je podporený simuláciami a vlastnými experimentálnymi výsledkami.

Kľúčové slová: genetické regulačné siete, bistabilita, master rovnica, stacionárny stav, šum, variancia

Abstract

HAJZOKOVÁ, Laura: Variance in gene regulatory networks, [Master's Thesis], Comenius University in Bratislava, Faculty of Mathematics, Physics and Informatics, Department of Applied Mathematics and Statistics; Supervisor: doc. Mgr. Pavol Bokes, PhD., Bratislava, 2024, 46 p.

This thesis delves into the modeling of genetic regulatory cascades and examines their stability. The stability analysis of the deterministic model allows for the observation of the model's stable state as well as the related bifurcations based on the values of its parameters. However, in stochastic modeling, a parameter representing the total number of molecules enters the system, and deterministic stability changes depending on this parameter. Here, we construct a synthetic seven-stage protein cascade that exhibits bistability under deterministic conditions. Conversely, we employ master equations to model it stochastically. By varying the system volume parameter, we explore its stability and endeavor, through simulations, to pinpoint the transition point of stable states. The thesis outlines the theoretical background of deterministic and stochastic approaches to modeling gene regulatory networks, bolstered by simulations and our own experimental findings.

Keywords: gene regulatory networks, bistability, master equation, stable state, noise, variance

Obsah

Úvod **8**

1 Teoretické pozadie **10**

1.1 Hillova funkcia 11

2 Deterministický model **13**

2.1 Stabilita modelu 13

2.2 Senzitivita 15

3 Stochastický prístup **20**

3.1 Kinetická schéma a master rovnice 21

3.2 Gillespie algoritmus 23

3.3 Tau-Leap metóda 25

4 Experimentálne výsledky **27**

4.1 Dizajn modelu 27

4.2 Výsledky simulácií 30

4.3 Porovnanie s ďalšími modelmi 37

Záver **41**

Zoznam použitej literatúry **43**

Príloha **46**

Úvod

Genetické regulačné siete hrajú dôležitú úlohu vo funkcií živých organizmov a ich vnútrobunkových procesov. Spravidla bývajú súčasťou komplexnejších štruktúr, avšak vzhľadom na ich komplexitu väčšinou nie je možné ich matematicky modelovať úplne presne. Rovnako má však biologický význam daný systém genetickej regulačnej siete čo najlepšie izolovať a modelovať ho separátne [1], [21]. Takéto skúmanie regulačných kaskád a ich vlastností je totiž kľúčové pre pochopenie biologických procesov odohrávajúcich sa v každom živom organizme. Aj vďaka množstvu biologických dát, ktoré sú v súčasnosti k dispozícii, je záujem o modelovanie biologických procesov veľmi vysoký. Skonštruovanie syntetickej regulačnej siete a skúmanie jej správania za rôznych podmienok nám môže poskytnúť cenné informácie o tom, ako tieto procesy prebiehajú v skutočnosti.

Na dosiahnutie reprezentatívnych výsledkov a ich správnu interpretáciu je kľúčové poznať teoretické pozadie modelu. Prvá časť je preto venovaná základným pojmom ako aj detailnejšiemu popisu chemickej reakcie prebiehajúcej v samotnej génovej expresii. Stručne je vysvetlené aj odvodenie Hillovej funkcie a interpretácia jej parametrov [1], [28]. Jej pôvod vychádza z chemickej kinetiky a je jedným zo základných nástrojov či už deterministických, alebo stochastických modelov génovej expresie.

Nasleduje teoretická časť zahŕňajúca deterministický prístup k modelovaniu genetických kaskád pomocou diferenciálnych rovníc a tiež analýza stability všeobecného autonómneho systému. Skúma sa tiež senzitivita stabilného stavu, t.j. možné fluktuácie elementov kaskády (proteínov) v závislosti od parametrov chemickej reakcie za predpokladu, že stav systému sa nachádza v okolí stabilného stavu. Ako totiž poukázal Ferrell v [9], [10], stabilný stav proteínových kaskád vykazuje známky ultrasenzitívneho správania. Hooshangi v [21] tiež spomína, ako senzitivita súvisí s veľkosťou, resp. počtom stupňov takejto kaskády. Nadväzujúc na [33] analyzujeme stabilitu systému, spolu so senzitivitou a jej závislosťou od parametrov a počiatočného stavu.

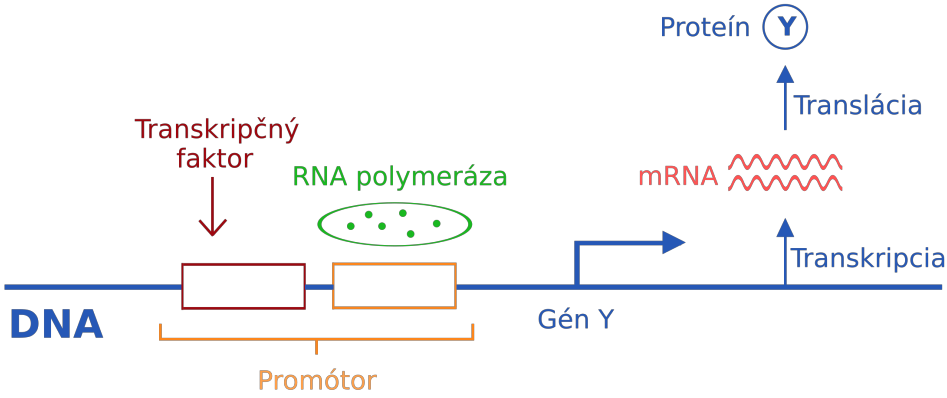
Ďalej predstavíme stochastický koncept modelovania génovej expresie pomocou tzv. master rovnice odvodenej z chemickej kinetiky [15] a zaoberáme sa zmenou stability, ktorá vznikne pridaním šumu do systému, podobne ako v [33], [30]. Kľúčovou otázkou zostáva, ako sa systém zmení pridaním šumu, resp. ako veľkosť systému (inými slovami, počet molekúl v systéme) vplýva na jeho stabilný stav. Očakávame totiž, že pridaním

variability, teda zavedením parametra Ω, ktorý reprezentuje jeho veľkosť, sa môže stabilný stav niektorých deterministicky stabilných systémov zmeniť. Zaujímame sa pritom bližšie o vzťah stability a parametra Ω pre konkrétnu konštrukciu syntetickej regulačnej kaskády a hlbšie analyzujeme systematckú zmenu stability.

V závere práce podporíme hypotézy o stabilite simuláciami stochastického systému genetickej regulačnej siete. Výsledky skúmnania stochastickej stability ilustrujeme na príklade syntetickej regulačnej kaskády a objasníme získané výsledky a ich možné príčiny. Na záver vytýčime možné vedecké ciele v tejto oblasti a spomenieme jej potenciálne rozšírenia.

1 Teoretické pozadie

Samotná génová expresia prebieha za pomoci tzv. transkripčného faktora, ktorý má schopnosť meniť svoj stav z inaktívneho na aktívny (a opačne) v závislosti od rôznych environmentálnych signálov. Rozlišujeme dva typy transkripčných faktorov [1]. Buď tzv. aktivátor, ktorý zvyšuje intenzitu transkripcie génu, alebo represor, ktorý ju, naopak, potláča. Transkripčný faktor sa vo svojom aktívnom stave naviaže na DNA a reguluje prepis génu do mRNA (detailnejšie na Obr. 1). Z mRNA potom vznikajú proteíny, ktoré sú odpoveďou na pôvodný impulz.



Obr. 1: Zjednodušená schéma génovej expresie. Transkripčný faktor sa naviaže na časť DNA (promótor) a reguluje intenzitu, s ktorou RNA polymeráza reguluje transkripciu génu Y. Ten sa prepíše do mRNA a z nej potom vzniká proteín. Tento proces sa nazýva translácia.

Výsledný proteín tiež môže byť vstupným faktorom do ďalšej génovej expresie, teda môže zastávať rolu transkripčného faktora. Takúto postupnosť dvoch alebo viacerých stupňov génovej expresie nazývame genetická regulačná kaskáda. Je to spravidla stochastický proces, avšak existujú deterministické matematické modely [21], [11], [6], ktoré nám pomáhajú pochopiť správanie týchto systémov za rôznych podmienok. Nasledujúca časť teda bude venovaná deterministickému modelu regulačnej kaskády a pozorovaniu vplyvu jednotlivých parametrov na správanie systému.

1.1 Hillova funkcia

V deterministických modeloch génovej expresie sa často objavuje tzv. Hillova funkcia [20]. Jej pôvod siaha do chemickej kinetiky [28], [1], pričom sa uvažuje nasledovná chemická reakcia

$$X_n + D \overset{k_d}{\rightleftharpoons} X_n D. \tag{1}$$

V reakcii (1) sa n molekúl aktivátora (alebo represora) X naviaže na časť promótora D a vytvoria komplex $X_n D$. Je zrejmé, že časť DNA D môže byť buď naviazaná molekulami X alebo voľná, z čoho dostávame rovnosť

$$[D] + [X_n D] = [D_t], \tag{2}$$

kde $[D_t]$ je celková koncentrácia D. Ďalej predpokladáme, že X sa viaže na D s intenzitou k_{on} a naopak, komplex $X_n D$ sa rozpadne s intenzitou k_{off}. To nám dáva nasledovný vzťah

$$\frac{d[X_n D]}{dt} = k_{on}[X]^n[D] - k_{off}[X_n D], \tag{3}$$

kde symbol [.] označuje koncentráciu daného prvku. Za stabilného stavu, t.j. keď sa koncentrácia komplexu $[X_n D]$ už nemení v čase, dostávame položením $d[X_n D]/dt = 0$ nasledovnú rovnicu

$$[X]^n[D] = \frac{k_{off}}{k_{on}}[X_n D], \tag{4}$$

kde $k_d = \frac{k_{off}}{k_{on}}$ je disociačná konštanta z (1). Z rovností (2) a (4) dostaneme pre voľné a naviazané časti D nasledovné vyjadrenia

$$\frac{[X_n D]}{[D_t]} = \frac{[X]^n}{K^n + [X]^n}, \tag{5}$$

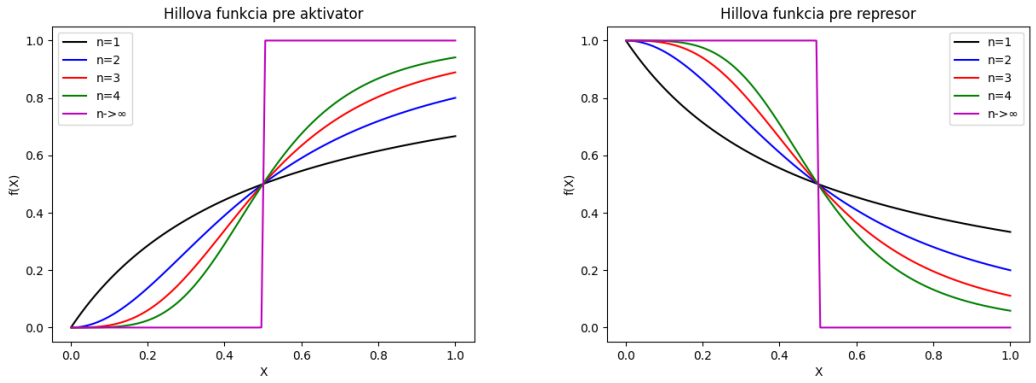
$$\frac{[D]}{[D_t]} = \frac{K^n}{K^n + [X]^n}, \tag{6}$$

kde $K^n = k_d$. Zavedie sa tiež parameter β, ktorý zodpovedá maximálnej intenzite transkripcie, pričom závisí od množstva faktorov vrátane sekvencie DNA a pozície RNA polymerázy viažucej sa na časť promótora. Zahrnutím β do rovníc (5), (6) dostaneme Hillove funkcie

$$H_1([X]) = \frac{\beta}{1 + \left(\frac{K}{[X]}\right)^n}, \tag{7}$$

$$H_2([X]) = \frac{\beta}{1 + \left(\frac{[X]}{K}\right)^n}, \tag{8}$$

kde H_1 zodpovedá prípadu, keď X je aktivátor a H_2 zodpovedá prípadu, keď X je represor. Tvar funkcií H_1 a H_2 je zobrazený na obrázku Obr. 2



(a) Hillova funkcia pre aktivátor X.

(b) Hillova funkcia pre represor X.

Obr. 2: Hillova funkcia pre aktivátor, resp, represor pre rôzne honoty n, pričom $\beta = 1,\ K = 0.5$.

2 Deterministický model

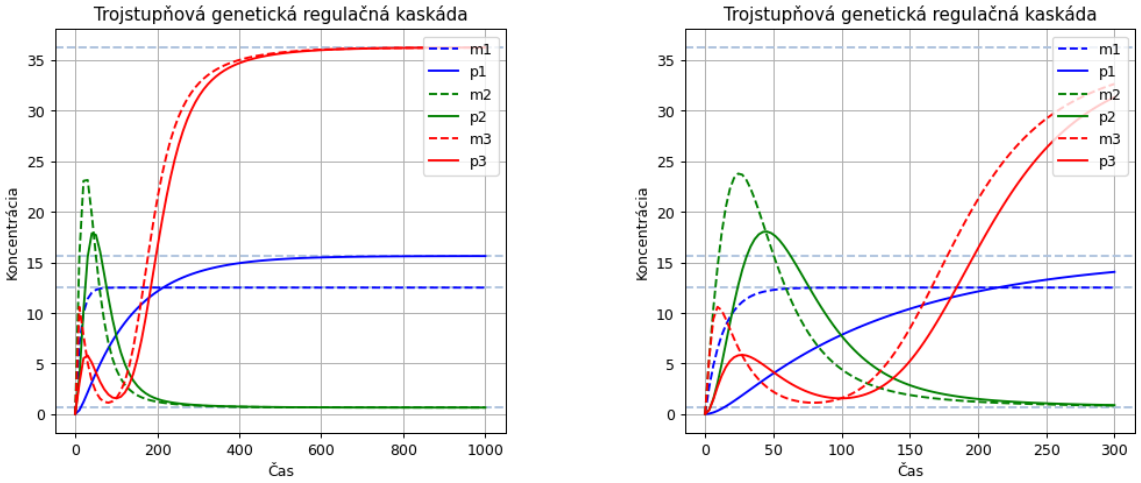
Pre syntetickú regulačnú kaskádu sa dá použiť deterministický model [33], ktorý bol mnohokrát potvrdený experimentálnymi výsledkami, ako spomína aj Hooshangi a kolektív v [21]. V tomto modeli sa uvažuje genetická regulačná kaskáda N génov. Označme koncentráciu mRNA v čase t zodpovedajúcu transkripcií i-teho génu ako $m_i(t)$ a koncentráciu vzniknutého proteínu P_i ako $p_i(t)$, pre $i = 1, ..., N$. Ďalej označme γ_i intenzitu rozpadu mRNA a Γ_i intenzitu rozpadu proteínu a tiež K_i intenzitu translácie proteínu pre i-ty gén. Intenzitu transkripcie mRNA predpokladáme konštantnú pre prvý gén, t.j. k_1 a pre zvyšné gény $i = 2, 3, ..., N$ ju modelujeme pomocou Hillovej funkcie (7) označenej ako h_i. Teda dostávame nasledovný model

$$\begin{aligned}
\tfrac{dm_1}{dt} &= -\gamma_1 m_1 + k_1, \\
\tfrac{dp_1}{dt} &= K_1 m_1 - \Gamma_1 p_1, \\
\tfrac{dm_i}{dt} &= -\gamma_i m_i + \frac{k^i_{max}}{1+\left(\frac{p_{i-1}}{k^i_d}\right)^{n_i}}, \qquad i = 2, 3, ..., N, \\
\tfrac{dp_i}{dt} &= K_i m_i - \Gamma_i p_i, \qquad i = 2, 3, ..., N,
\end{aligned} \tag{9}$$

kde k^i_{max}, k^i_d, n_i sú parametre Hillovej funkcie. k^i_{max} reprezentuje maximálnu intenzitu transkripcie, n_i Hillov koeficient udávajúci sklon Hillovej funkcie (viď. Obr. 2) a k^i_d väzbovú konštantu, resp. mieru koncentrácie proteínu, pre ktorú je intenzita transkripcie na polovici jej maxima. Je dôležité si všimnúť, že expresia dvoch po sebe idúcich génov v regulačnej kaskáde je v modeli previazaná len cez závislosť koncentrácie mRNA m_i prislúchajúcej i-temu génu od koncentrácie proteínu p_{i-1} prislúchajúcej $(i-1)$-ému génu. Jednoduchý príklad trojstupňovej syntetickej regulačnej kaskády modelovanej pomocou (9) je zobrazený na Obr. 3.

2.1 Stabilita modelu

Pevný bod systému (9) dostaneme položením derivácii rovných nule. Je zrejmé, že systém má jediný pevný bod (m^*, p^*), kde $m^* = (m^*_1, m^*_2, ..., m^*_N)$ a $p^* = (p^*_1, p^*_2, ..., p^*_N)$. Ak



Obr. 3: Príklad trojstupňovej regulačnej kaskády pre nasledovné hodnoty parametrov: $k_1 = 1,\ \gamma_1 = 0.08,\ K_1 = 0.01,\ \Gamma_1 = 0.008,\ \gamma_2 = \Gamma_2 = \gamma_3 = \Gamma_3 = K_2 = K_3 = 0.05,\ k^2_{max} = k^2_d = n_2 = k^3_{max} = k^3_d = n_3 = 2$. Je zrejmé, že koncentrácie proteínov a mRNA konvergujú k stabilnému stavu $m^* = (12.50, 0.65, 36.23)^T$, $p^* = (15.63, 0.65, 36.23)^T$.

označíme

$$\begin{aligned} h_i(p_{i-1}) &= \frac{k^i_{max}}{1+\left(\frac{p_{i-1}}{k^i_d}\right)^{n_i}}, & i = 2, 3, ..., N, \\ b_i &= \frac{K_i}{\gamma_i}, & i = 2, 3, ..., N, \end{aligned} \tag{10}$$

dostaneme pevný bod daný rovnicami (11)

$$\begin{aligned} m^*_1 &= \frac{k_1}{\gamma_1}, \\ p^*_1 &= \frac{b_1 k_1}{\Gamma_1}, \\ m^*_i &= \frac{h_i(p^*_{i-1})}{\gamma_i}, & i = 2, 3, ..., N, \\ p^*_i &= \frac{b_i h_i(p^*_{i-1})}{\Gamma_i}, & i = 2, 3, ..., N. \end{aligned} \tag{11}$$

V rovniciach (10) potom $h_i(p_{i-1})$ reprezentuje transkripciu i-teho génu modelovanú Hillovou funkciou a b_i zodpovedá priemernému počtu proteínov vyprodukovaných počas jednej

transkripcie [30], [33]. Jakobián systému (9) potom nadobúda tvar

$$J = \begin{bmatrix} -\gamma_1 & & & & & & & & \\ K_1 & -\Gamma_1 & & & & & & & \\ & \ddots & \ddots & & & & & & \\ & & dh_i(p_{i-1}) & -\gamma_i & & & & & \\ & & & K_i & -\Gamma_i & & & & \\ & & & & \ddots & \ddots & & & \\ & & & & & dh_N(p_{N-1}) & -\gamma_N & & \\ & & & & & & K_N & -\Gamma_N \end{bmatrix} \tag{12}$$

Keďže Jakobián J je bidiagonálna matica a jej vlastné hodnoty $-\gamma_i, -\Gamma_i,\ \forall i = 1, ..., N$ sú záporné, pevný bod (m^*, p^*) systému (9) je asymptoticky stabilný. Inými slovami, koncentrácia proteínu p_i konverguje k svojmu stabilnému stavu p_i^* a teda pre $t \to \infty$ sa koncentrácia ustáli pre $\forall i = 1, ..., N$.

2.2 Senzitivita

Dôležitým faktorom vplývajúcim na stabilitu systému je tiež senzitivita stabilného stavu. Postupnosť pevných bodov $\{p_1^*, ..., p_N^*\}$ sa dá totiž vyjadriť diferenčnou rovnicou [33], ktorá má vlastný pevný bod, pričom ten môže byť citlivý na zmeny počiatočného stavu. V súvislosti s genetickými kaskádami sa množtvo literatúry zaoberalo senzitivitou, napr. Ferrel [9], [10] poukázal na to, že proteínové kaskády sa vyznačujú ultrasenzitivitou, teda malé zmeny počiatočného stavu spôsobujú veľké fluktuácie na výstupe. Tiež, Hooshangi vo svojej publikácii [21] poukázal na skutočnosť, že náchylnosť k ultrasenzitívnemu správaniu narastá s veľkosťou genetickej kaskády, t.j. s počtom stupňov. V tejto časti sa teda pozrieme na to, ako závisí senzitivita stabilného stavu od hodnôt parametrov modelu (9), pričom sa budeme opierať o analýzu [33].

Zo systému (9) sme dostali stabilný stav (m^*, p^*) vyjadrený ako (11), kde $m = (m_1, ..., m_N)$ označuje koncentráciu mRNA a $p = (p_1, ..., p_N)$ označuje koncentráciu proteínov v jednotlivých stupňoch kaskády. Nás v tomto prípade zaujíma hladina proteínov p^*, na ktorej sa systém (9) stabilizuje.

Vieme, že stav systému niekde v okolí pevného bodu p^* môžeme aproximovať funkciou $h_i,\ \forall i = 1, ..., N$ Taylorovým rozvojom do prvého rádu [33]. To pre jednotlivé stupne

genetickej kaskády (9) dáva

$$h_i(p_{i-1}) \approx h_i(p^*_{i-1}) + h'_i(p^*_{i-1})(p_{i-1} - p^*_{i-1}), \tag{13}$$

kde h'_i predstavuje deriváciu funkcie h_i. Pre jednoduchosť označíme konštantné členy (vzhľadom na premennú p_{i-1}) zo vzťahu (13) ako

$$d_i = h_i(p^*_{i-1}) - h'_i(p^*_{i-1})p^*_{i-1}, \qquad i = 2, 3, ..., N. \tag{14}$$

Tak dostaneme zjednodušenú formu (13) v tvare

$$h_i(p_{i-1}) \approx h'_i(p^*_{i-1})p_{i-1} + d_i. \tag{15}$$

Keďže h_i je Hillova funkcia $\forall i = 1, ..., N$, jej deriváciu $h'_i(p_{i-1})$ môžeme vyjadriť v nasledovnom tvare

$$\frac{dh_i(p_{i-1})}{dp_{i-1}} = -\frac{k^i_{max}}{\left[1 + \left(\frac{p_{i-1}}{k^i_d}\right)^{n_i}\right]^2} \frac{n_i}{k^i_d} \left(\frac{p_{i-1}}{k^i_d}\right)^{n_i - 1}, \tag{16}$$

čo je to isté, ako

$$\frac{dh_i(p_{i-1})}{dp_{i-1}} = -\frac{n_i}{k^i_{max} k^i_d} \left(\frac{p_{i-1}}{k^i_d}\right)^{n_i - 1} (h_i(p_{i-1}))^2. \tag{17}$$

Z explicitných vzťahov pre p^*_i (11) a aproximácie funkcie h_i (15) dostaneme nasledovné vzťahy pre pevný bod p^*

$$p^*_1 = \frac{b_1 k_1}{\Gamma_1}, \tag{18}$$

$$p^*_i = p^*_1 \prod_{j=2}^{i} \frac{b_j h'_j}{\Gamma_j} + \sum_{j=2}^{i} \frac{b_j d_j}{\Gamma_j} \Big(\prod_{l=j+1}^{i} \frac{b_l h'_l}{\Gamma_l} \Big), \qquad i = 2, 3, ..., N, \tag{19}$$

kde $h'_j = h'_j(p^*_{j-1})$. Všimnime si, že pevné body p^*_i sú dané obyčajnou diferenčnou rovnicou s nekonštantnými koeficientmi (19).

V tomto prípade teda uvažujeme autonómny systém, čo zodpovedá špeciálnemu prípadu, kedy sú parametre reakcie rovnaké v každom stupni kaskády. Inými slovami, $k^i_{max} = k_{max},\ k^i_d = k_d,\ n_i = n,\ K_i = K,\ \gamma_i = \gamma,\ \Gamma_i = \Gamma$, teda aj $b_i = b,\ h_i = h,\ h'_i = h',\ d_i = d$. Pokiaľ označíme $q_i = p^*_i$, resp. $q_{i-1} = p^*_{i-1}$ dostaneme zo vzťahov (15) a (11) nasledovnú diferenčnú rovnicu pre pevné body p^*_i

$$q_i = \frac{bh'}{\Gamma} q_{i-1} + \frac{bd}{\Gamma}, \qquad i = 2, 3, ..., N, \tag{20}$$

pričom $q_1 = bk_1/\Gamma$. Po dosadení d zo vzťahu (14) do (20) je zrejmé, že systém (20) je možné upraviť na tvar

$$q_i = \frac{bh(q_{i-1})}{\Gamma}, \qquad i = 2, 3, ..., N. \tag{21}$$

Tiež pre pevný bod x^* musí byť splnená podmienka $x^* = f(x^*)$, a teda, ak q^* je pevný bod, potom z (21) dostávame vzťah

$$q^* = \frac{bh(q^*)}{\Gamma}, \qquad i = 2, 3, ..., N. \tag{22}$$

Označme $h' = h'(q^*)$. Je zrejmé, že pevný bod je asymptoticky stabilný, ak platí $|f'(x^*)| < 1$ a nestabilný, ak $|f'(x^*)| > 1$. To v našom prípade znamená, že p^* je

1. asymptoticky stabilný, ak

$$\left|\frac{bh'}{\Gamma}\right| < 1, \tag{23}$$

2. nestabilný, ak

$$\left|\frac{bh'}{\Gamma}\right| > 1. \tag{24}$$

Stabilný prípad (23) vieme rozdeliť na dva prípady:

1. S narastajúcim počtom stupňov kaskády bod q_i konverguje k bodu q^* (21) monotónne, čo zodpovedá prípadu

$$0 < bh' < \Gamma. \tag{25}$$

2. S narastajúcim počtom stupňov kaskády bod q_i konverguje k bodu q^* (21) oscilujúco, čo zodpovedá prípadu

$$-\Gamma > bh' > 0. \tag{26}$$

Nestabilný prípad (23) vieme rozdeliť na dva prípady:

1. S narastajúcim počtom stupňov kaskády bod q_i monotónne diverguje do $\pm\infty$ (v závislosti od q_1), čo zodpovedá prípadu

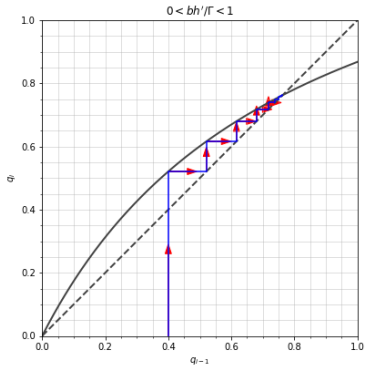
$$0 < bh' < \Gamma. \tag{27}$$

2. S narastajúcim počtom stupňov kaskády bod q_i oscilujúco diverguje do $+\infty$ alebo $-\infty$, čo zodpovedá prípadu

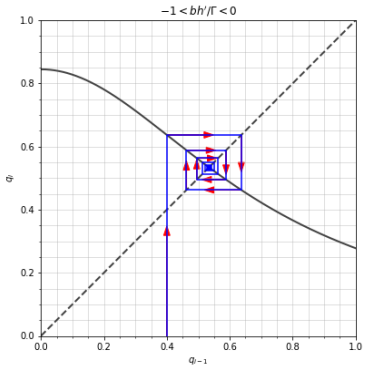
$$-\Gamma > bh' > 0. \tag{28}$$

Stabilita systému a jeho konvergencia sa v deterministickom prípade zvyčajne ilustruje na tzv. Cob-web diagrame, ktorý v našom prípade zobrazuje vzťah q_i od q_{i-1}, ktoré sú previazané funkciou (21) a v pevnom bode q^* platí (22). Na Obr. 5 sú zobrazené Cob-web

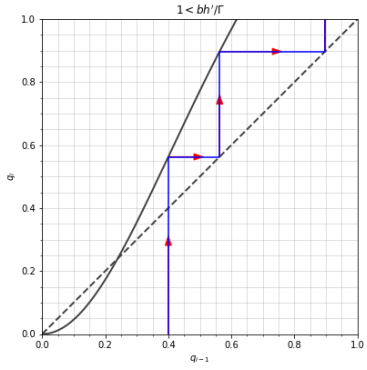
diagramy pre prípady (25)-(28). Vidíme, že stabilita, ako aj typ konvergencie sa zmení v závislosti od parametrov, resp. hodnoty bh'/Γ. V Tabuľke 1 sú zobrazené hodnoty bh'/Γ, ako aj pevný bod q^* pre všetky prípady.



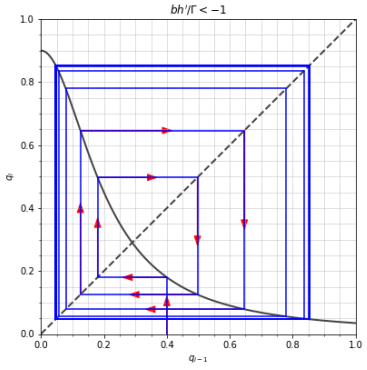
(a) Asymptoticky stabilný pevný bod q^* pre hodnoty parametrov $b = 0.5, \Gamma = 0.8, k_d = 0.8, k_{max} = 2.5, n = 1$



(b) Asymptoticky stabilný pevný bod q^* pre hodnoty parametrov $b = 0.8, \Gamma = 1.8, k_d = 0.7, k_{max} = 1.9, n = 2$.



(c) Nestabilný pevný bod q^* pre hodnoty parametrov $b = 0.8, \Gamma = 0.8, k_d = 0.7, k_{max} = 2, n = 2$.



(d) Nestabilný pevný bod q^* pre hodnoty parametrov $b = 0.5, \Gamma = 1, k_d = 0.2, k_{max} = 1.8, n = 2$.

Obr. 4: Cobweb diagram pre systém (20). Je zrejmé, že pre sytém môžu nastať štyri rôzne scenáre, v závislosti od parametrov $b, \Gamma, k_d, k_{max}, n$.

My sa budeme v práci zaoberať prípadom 4d z obrázka Obr. 5. Všimnime si, že napriek tomu, že systém sa označuje za nestabilný, hodnoty q_i nedivergujú do ∞, ale zostanú stabilne oscilovať, resp. prepínať medzi dvoma stavmi. Tento jav nazývame tzv. bistabilitou, kedy postupnosť pevných bodov $\{p_1^*, ..., p_N^*\}$, resp. $\{q_1, ..., q_N\}$ pre $I = 1, ..., N$

Diagram	Pevný bod	Podmienka stability bh'/Γ
Obr. 4a	0.7625	0.5120
Obr. 4b	0.5339	-0.7355
Obr. 4c	0.2395	1.7905
Obr. 4d	0.2900	-1.3555

Tabuľka 1: Tabuľka hodnôt parametrov a príslušných pevných bodov pre prípady (25)-(28).

vieme rozdeliť na dve podpostupnosti

$$\begin{aligned} &\{q_{2k}\}_{k=1}^{N/2}, \qquad \{q_{2k-1}\}_{k=2}^{N/2} \text{ ak } N \text{ je párne}, \\ &\{q_{2k}\}_{k=1}^{(N-1)/2}, \quad \{q_{2k+1}\}_{k=1}^{(N-1)/2} \text{ ak } N \text{ je nepárne}. \end{aligned} \tag{29}$$

Jedna z nich konverguje k stabilnému stavu vo vyššej koncentrácii (ľavý horný roh Cob-web diagramu 4d) a druhá k stabilnému stavu v nižšej koncentrácii (pravý dolný roh Cob-web diagramu 4d) v závislosti od počiatočného $q_1 = p_1^*$. Inak povedané, stabilné koncentrácie proteínov s párnymi indexmi konvergujú k jednému pevnému bodu a stabilné koncentrácie s nepárnymi indexami konvergujú k druhému pevnému bodu, ak počet stupňov kaskády $N \to \infty$.

V nasledujúcej kapitole sa pozrieme na to, ako sa tieto systémy vo všeobecnosti modelujú stochasticky, a tiež ako vyzerá deterministická bistabilita v prípade, že do systému pridáme šum.

3 Stochastický prístup

Táto kapitola sa zaoberá stochastickým prístupom k modelovaniu genetickej regulačnej siete, pričom tento prístup sa od deterministického líši prítomnosťou šumu v systéme. Keďže génová expresia je prirodzene stochastický proces [23], [22], študovanie šumu je jednou z kľúčových tém pre pochopenie správania vnútrobunkových procesov. Dokonca sa ukazuje, že veľakrát deterministický model nestačí, keďže vnútrobunkové chemické reakcie často prebiehajú v nízkych koncentráciách s pomalou intenzitou, kde deterministický prístup zvykne zlyhať [34], [22].

Za zdroj variability génovej expresie sú považované dva typy šumu - vnútorný a vonkajší [7]. Vnútorný šum je variabilita koncentrácie proteínov spôsobená prirodzenou stochastickou povahou chemickej reakcie a je špecifická pre dané bunky [29]. Inými slovami, vplývajú na ňu faktory ako intenzita transkripcie, intenzita translácie či degradácia proteínov. Na druhej strane, za vonkajší šum sú zodpovedné procesy mimo bunky, rozdiely medzi bunkami alebo vplyv prostredia [7]. V tomto prípade nás bude zaujímať efekt vnútorného šumu na stabilitu genetickej regulačnej kaskády. Ukazuje sa, že pre jednoduchosť je zmysluplným prístupom k takémuto problému pozerať sa na čo možno najlepšie izolovaný systém od vonkajších vplyvov a ostatných regulačných elementov [21], [11], [6]. Zostrojením syntetickej kaskády môžeme získať rozumný náhľad do vnútrobunkových procesov, konkrétne stability genetickej regulačnej kaskády v závislosti od jednotlivých parametrov, počtu stupňov kaskády a tiež od počtu molekúl. resp. objemu systému.

Najčastejšie sa využívajú dva prístupy k stochastickému modelovaniu. Prvý je pomocou stochastických diferenciálnych rovníc, resp. Langevinovej formulácie, ktorá vznikne pridaním bieleho šumu do systému obyčajných diferenciálnych rovníc [18], [31]. Takýto prístup však nemusí dávať dobré výsledky, pokiaľ ide o simulácie na mikroskopickej škále bistabilných systémov [3], ktorými sa budeme zaoberať aj v tejto práci. Druhý spôsob, ktorý aplikujeme v našom prípade, spočíva v skonštruovaní tzv. master rovnice (*angl. master equations*) z kinetickej schémy daných chemických reakcií [31] a následných simulácií pomocou stochastických simulačných algoritmov (SSA) [16].

3.1 Kinetická schéma a master rovnice

V našom prípade budeme modelovať systém pomocou master rovnice. Najskôr však pôvodný systém (9) zjednodušíme nasledovným spôsobom. Ako uvádza Zheng a kolektív v [33], a tiež Thattai a kolektív v [30], molekuly mRNA reagujú veľmi rýchlo oproti proteínom. Inými slovami, pomer intenzít rozpadu $\Gamma_i/\gamma_i,\ \forall i = 1, ..., N$ je veľmi malý, a teda dynamika systému je primárne ovplyvnená zmenou koncentrácie proteínov. V takom prípade môžeme podľa [5] uvažovať koncentráciu mRNA ako konštantnú, čo zodpovedá

$$\frac{dm_i}{dt} = 0, \qquad \forall i = 1, ..., N. \tag{30}$$

Koncentrácie mRNA vieme teda vyjadriť nasledovným spôsobom:

$$\begin{aligned} m_1 &= \tfrac{k_1}{\gamma_1}, \\ m_i &= \tfrac{h_i(p_{i-1})}{\gamma_i}, \qquad i = 2, 3, ..., N. \end{aligned} \tag{31}$$

Systém (9) vieme v takom prípade aproximovať ako

$$\begin{aligned} \tfrac{dp_1}{dt} &= b_1 k_1 - \Gamma_1 p_1, \\ \tfrac{dp_i}{dt} &= b_i h_i(p_{i-1}) - \Gamma_i p_i, \qquad i = 2, 3, ..., N, \end{aligned} \tag{32}$$

čím sa počet rovníc zmenší na polovicu, a teda je značne jednoduchší v zmysle modelovania. A keďže sa predpokladá, že stav systému sa nachádza niekde v okolí stabilného stavu, eliminácia mRNA zo systému predstavuje zmysluplnú aproximáciu systému (9) [33], [5]. Pre počty kópií proteínov potom dostaneme kinetickú schému [33], [31]

$$\begin{aligned} x_1 &\xrightarrow{\Omega b_1 k_1} x_1 + 1, \\ x_1 &\xrightarrow{\Gamma_1 x_1} x_1 - 1, \\ x_i &\xrightarrow{\Omega b_i h_i(x_{i-1}/\Omega)} x_i + 1, \qquad i = 2, 3, ..., N, \\ x_i &\xrightarrow{\Gamma_i x_i} x_i - 1, \qquad i = 2, 3, ..., N. \end{aligned} \tag{33}$$

Vo všeobecnosti sa dá uvažovať množina všetkých reakcií v systéme $\{R_1, R_2, ..., R_M\}$. Avšak systém (32) zodpovedá kaskáde N proteínov, čo predstavuje v našom prípade presne $M = 2N$ chemických reakcií.

Označíme stav systému v čase t ako $X(t) = (X_1(t), ..., X_N(t))^T$, kde $X_i(t),\ \forall i = 1, ..., N$ je náhodná premenná [14] reprezentujúca počet molekúl i-teho proteínu v čase t. Okrem parametrov reakcií sa zavedie parameter Ω, ktorý reprezentuje objem systému

[31]. Vzťah koncentrácie p_i a počtu molekúl x_i jednotlivých proteínov v systéme vieme potom vyjadriť vzťahom

$$x_i = \Omega p_i, \qquad \forall i = 1, ..., N. \tag{34}$$

Predpokladá sa tiež výskyt mnohých nereaktívnych molekulárnych kolízií medzi reaktívnymi. Za takýchto podmienok sa dá ukázať, že každú chemickú reakciu $R_j, \forall j = 1, ..., M$ možno charakterizovať dobre definovanou funkciou w_j (tzv. *propensity function*) a vektorom v_j (tzv. *state-change vector*) [14] definovanými nasledovne:

$$w_j(x)dt = P(R_j|X(t) = x), \text{ t.j. pravdepodobnosť za podmienky} X(t) = x, \tag{35}$$
$$\text{že na intervale } [t, t + dt) \text{ nastane reakcia } R_j,$$

$$v_j \in \mathbb{R}^N \text{ taký, že jeho } i\text{-ty prvok } v_{ij} \text{ reprezentuje zmenu} \tag{36}$$
$$\text{počtu molekúl } i\text{-teho proteínu } \forall i = 1, ..., N.$$

Vzťahy (35) a (36) indikujú, že $X(t)$ je Markovovský proces v $\mathbb{R}^N$, pričom v prípade systému (33) sa jedná o tzv. *Birth-Death* proces [33]. V dôsledku toho vieme skonštruovať chemickú master rovnicu (*angl. chemical master equation*) popisujúcu zmenu pravdepodobnosti $P(x, t|x_0, t_0)$ v čase. Inak povedané, $P(x, t|x_0, t_0)$ označuje pravdepodobnosť nadobudnutia počtu molekúl $x = (x_1, ..., x_N)^T$ v čase t, za podmienky, že je známy počiatočný stav $X(t_0) = x_0$. Odvodenie master rovnice sa dá podrobnejšie nájsť v [15].

Vo všeobecnosti sa dá master rovnica pre M chemických reakcií a N rôznych reaktantov napísať ako

$$\frac{\partial}{\partial t}P(x, t|x_0, t_0) = -\sum_{k=1}^{M} w_k(x)P(x, t|x_0, t_0) + \sum_{k=1}^{M} w_k(x - v_k)P(x - v_k, t|x_0, t_0), \tag{37}$$

pričom v našom prípade, t.j. pre kinetickú schému (33) nadobudne tvar

$$\begin{aligned}
\tfrac{\partial}{\partial t}P(x,t) = \; & -\Gamma_1 x_1 P(x_1, ..., x_N, t) && -\Omega b_1 k_1 P(x_1, ..., x_N, t) + \\
& +\Gamma_1 x_1 P(x_1 + 1, ..., x_N, t) && +\Omega b_1 k_1 P(x_1 - 1, ..., x_N, t) \\
& -\textstyle\sum_{i=2}^{N} \Gamma_i x_i P(x_1, ..., x_N, t) && -\textstyle\sum_{i=2}^{N} \Omega b_i h_i(x_{i-1}/\Omega) P(x_1, ..., x_N, t) \\
& +\textstyle\sum_{i=2}^{N} \Gamma_i x_i P(..., x_i + 1, ..., t) && +\textstyle\sum_{i=2}^{N} \Omega b_i h_i(x_{i-1}/\Omega) P(..., x_i - 1, ..., t).
\end{aligned} \tag{38}$$

Skonštruovanie master rovnice spolu so vzťahmi (35) a (36) vedie k formulácii funkcie hustoty (ozn. $f_{\tau,J|X(t)=x}(\tau, j|x, t)$) pre každú nasledujúcu chemickú reakciu. Podrobné odvodenie sa dá nájsť v publikáciách v [13], [15]. Jedná sa teda o pravdepodobnosť za

podmienky $X(t) = x$, že R_j bude nasledujúcou reakciou, ktorá nastane na intervale $[t + \tau, t + \tau + d\tau)$. Združenú funkciu hustoty pre takéto pravdepodobnostné rozdelenie potom vieme vyjadriť nasledovnou rovnicou:

$$f_{\tau,J|X(t)=x}(\tau, j|x, t) = w_j(x)\exp(-w_0(x)\tau), \qquad \tau \geq 0, \ \forall j = 1, ..., M, \tag{39}$$

pričom

$$w_0(x) = \sum_{j=1}^{M} w_j(x). \tag{40}$$

Master rovnice vo všeobecnosti nemusia mať analytické, či dokonca ani numerické riešenie pre zložitejšie systémy chemických reakcií [14]. Preto vznikla potreba tieto systémy simulovať pomocou rôznych simulačných algoritmov. Vzťah (39) tvorí podstatu jedného z najpoužívanejších stochastických simulačných algoritmov pre chemické reakcie, tzv. Gillespie-ho algoritmus [13].

3.2 Gillespie algoritmus

Existujú rôzne varianty Gillespie-ho algoritmu v závislosti od spôsobu generovania hodnôt náhodných premenných τ a J. V tejto časti niektoré z nich v krátkosti spomenieme, pričom detailnejšie vysvetlenie sa nachádza v článku [14]. Spomínané stochastické simulačné algoritmy vychádzajú pritom zo vzťahu združenej funkcie hustoty (39), ktorú si vieme napísať ako podmienenú hustotu

$$f_{\tau,J|X(t)=x}(\tau, j|x, t) = f_{J|\tau,X(t)=x}(j|\tau, x, t) f_{\tau|X(t)=x}(\tau|x, t). \tag{41}$$

Zo vzťahu (39) je zrejmé, že hustoty $f_{J|\tau,X(t)=x}(j|\tau, x, t)$ a $f_{\tau|X(t)=x}(\tau|x, t)$ sú dané ako

$$f_{\tau|X(t)=x}(\tau|x, t) = w_0(x)\exp(-w_0(x)\tau), \qquad \tau \geq 0, \tag{42}$$

$$f_{J|\tau,X(t)=x}(j|\tau, x, t) = \frac{w_j(x)}{w_0(x)}, \qquad \forall j = 1, ..., M. \tag{43}$$

Rovnica (42) teda hovorí o tom, že τ je náhodnou premennou reprezentujúcou čas do ďalšej reakcie a riadi sa exponenciálnym rozdelením s parametrom $w_0(x)$ [4]. Zo vzťahu (43) zase vieme povedať, že náhodná premenná J pochádza z diskrétneho rozdelenia na množine $\{1, 2, ..., M\}$ s pravdepodobnosťami $\{w_1(x)/w_0(x), w_2(x)/w_0(x), ..., w_M(x)/w_0(x)\}$. Na generovanie náhodných dvojíc τ a j sa využíva nasledujúca veta o inverznej transformácii (Veta 3.1).

Veta 3.1 (Veta o inverznej transformácii 2.1 [19]). *Nech $F : \mathbb{R} \to [0,1]$ je distribučná funkcia a nech $Q = F^{-1}$ je jej kvantilová funkcia, t.j. $Q(u) = sup\{x \in \mathbb{R} : Q(x) \leq u\}$ pre všetky $u \in (0,1)$. Nech náhodná premenná U je z rovnomerného rozdelenia $U(0,1)$. Potom náhodná premenná $Q(U)$ má distribučnú funkciu F.*

Keďže τ je z exponenciálneho rozdelenia s parametrom $w_0(x)$, distribučná funkcia $F(\tau)$ má tvar $F(\tau) = 1 - \exp(-w_0(x)\tau)$. Zodpovedajúca kvantilová funkcia potom nadobudne tvar

$$F^{-1}(\tau) = \frac{1}{w_0(x)} \ln\Big(\frac{1}{1-\tau}\Big). \tag{44}$$

Podľa Vety 3.1 platí, že ak náhodná premenná U je z rovnomerného rozdelenia na intervale $(0,1)$, potom kvantilová funkcia $F^{-1}(U)$ má rovnaké rozdelenie ako τ, t.j. exponenciálne rozdelenie s parametrom $w_0(x)$. Stačí teda vygenerovať náhodnú premennú r z rovnomerného rozdelenia $U(0,1)$ a premennú τ potom môžeme vyjadriť ako

$$\tau = \frac{1}{w_0(x)} \ln\Big(\frac{1}{r}\Big). \tag{45}$$

Využili sme pri tom fakt, že ak $1 - r \sim U(0,1)$ potom aj $r \sim U(0,1)$.

Ďalej vieme, že pre ľubovoľnú distribučnú funkciu F spojitú sprava je jej kvantilová funkcia Q definovaná ako [25]

$$Q(u) = F^{-1}(u) = \inf\{x : F(x) \geq u\}. \tag{46}$$

A keďže pre náhodnú premennú J má distribučná funkcia tvar

$$P(J \leq j) = \frac{1}{w_0(x)} \sum_{k=1}^{j} w_k(x), \tag{47}$$

jej kvantilová funkcia $Q(s)$ sa dá napísať ako

$$Q(s) = \inf\Big\{j : \frac{1}{w_0(x)} \sum_{k=1}^{j} w_k(x) \geq s\Big\}. \tag{48}$$

Podľa Vety 3.1 tiež platí, že ak náhodná premenná S je z rovnomerného rozdelenia $U(0,1)$, tak náhodná premenná $Q(S)$ zo vzťahu (48) má rovnaké rozdelenie ako náhodná premenná J. Náhodným vygenerovaním dvoch hodnôt r, s z rozdelenia $U(0,1)$ potom dostaneme použitím vzťahov (45) a (48) hodnoty (τ, j) z rozdelenia (39). Takýto spôsob generovania (τ, j) sa v literatúre nazýva aj priama metóda [14].

Modifikáciou priamej metódy tiež vznikla tzv. metóda prvej reakcie (*angl. First reaction method*), ktorú tiež vymyslel a publikoval Gillesipe v roku 1976 [13]. Rovnako, ako priama metóda, funguje na princípe generovania z rozdelenia (39). Opäť sa pre každú reakciu $R_j, \forall j = 1, ..., M$ vygenerujú časy $\{\tau_1, ..., \tau_M\}$ podľa (45). Z nich sa potom vyberie najmenšia hodnota $\tau_{min} = \min\{\tau_1, ..., \tau_M\}$, pričom reakcia s rovnakým indexom sa zoberie ako nasledujúca. Idea za tým je, že nasledujúca chemická reakcia je taká, ktorá nastane v najkratšom čase od poslednej reakcie [14]. Náhodne vygenerované dvojice (τ, j) pochádzajú z rozdelenia (39) a teda priama metóda a metóda prvej reakcie sú ekvivalentné, čo je dokázané v [13]. Treba však spomenúť, že v každej iterácii sa vygeneruje M náhodných hodnôt τ, pričom $M-1$ hodnôt sa vôbec nevyužije, čo má za následok značné zvýšenie výpočtovej náročnosti.

Tento problém sa snaží odstrániť metóda nasledujúcej reakcie (*angl. next reaction method*) vymyslená v roku 2000 Gibsonom a Bruckom [12]. Tá funguje na veľmi podobnom princípe ako predošlá metóda, rovnako vygeneruje M hodnôt τ ako v predošlom prípade. Zvyšných $M-1$ hodnôt však nezostane nevyužitých, avšak za špeciálnych podmienok je možné s nimi operovať ďalej a nie je nutné ich generovať v každom kroku. Tiež sa v každom kroku nepočítajú nanovo všetky hodnoty $w_j(x), \forall j = 1, ..., M$, ale skonštruuje sa graf závislosti, vďaka čomu sa v každom kroku spočíta len minimálny počet hodnôt $w_j(x)$. Detailný postup je uvedený v pôvodnom článku [12]. Je dôležité poznamenať, že *next reaction method* prekonáva problémy efektivity výpočtu prítomné v *first reaction method*, z hľadiska trvania výpočtu je dokonca efektívnejšia ako pôvodná priama metóda [14]. Všetky tri vyššie uvedené metódy sú matematicky ekvivalentné a považujú sa za exaktné, keďže produkujú presnú realizáciu procesu $X(t)$ priamym generovaním (τ, j) z rozdelenia (39). Existuje však trieda aproximatívnych metód, ktoré síce presne negenerujú proces $X(t)$, avšak asymptoticky dávajú porovnateľné výsledky a z hľadiska výpočtovej náročnosti prekonávajú exaktné metódy, čo môže byť nevyhnutné pre veľké systémy, t.j. veľký počet reakcií či reaktantov.

3.3 Tau-Leap metóda

Keďže priame metódy generujú každú jednu reakciu v procese $X(t)$, pre systémy väčšieho objemu vyžadujúce veľký počet simulácií sa výpočet stáva z hľadiska trvania veľmi ne-

efektívny. Potreba rýchlych a efektívnych algoritmov podnietila vznik aproximatívnych metód, ktoré už nefungujú na báze generovania jednotlivých reakcií. Jednou z týchto metód je aj Tau-leap metóda, ktorá je vylepšenou verziou Gillespie-ho algoritmu.

Pri τ-leap metóde sa generovanie každej jednej reakcie obchádza. Namiesto toho sa časová os rozdelí na podintervaly, kde nás nebude zaujímať, v ktorom čase presne ktorá reakcia nastala, ale bude nás zaujímať, koľkokrát na danom časovom intervale nastala každá reakcia $R_j, \forall j = 1, ..., M$. V krátkosti sa teda pozrieme na to, ako sa dá tento postup chápať matematicky. Detailne je vysvetlený v pôvodnej publikácii [14].

Pravdepodobnosť za podmienky $X(t) = x$, že na časovom intervale $[t, t+\tau)$ nastane k_j reakcií R_j pre všetky $j = 1, ..., M$ označíme ako $P(k_1, ..., k_M|\tau; x, t)$. Každé $K_j(\tau; x, t)$ je teda diskrétna náhodná premenná označujúca počet reakcií R_j na intervale $[t, t + \tau)$. Pre dobrý odhad $P(k_1, ..., k_M|\tau; x, t)$ pre ľubovolné $\tau \geq 0$ sa zavedie tzv. skoková podmienka (*angl. leap condition*), ktorá predpokladá, že τ je dostatočne malé na to, aby zmena stavu systému bola na tomto intervale taká nepatrná, že hodnota žiadnej funkcie $w_j, \forall j = 1, ..., M$ sa na tomto intervale zásadne nezmení. Za predpokladu, že táto podmienka platí, môžeme uvažovať, že hodnoty $w_j(x), \forall j = 1, ..., M$ budú na každom intervale $[t, t + \tau)$ konštantné. Inými slovami, $w_j(x)dt$ nám dáva pravdepodobnosť, koľkokrát nastane reakcia R_j na danom časovom intervale nezávisle od toho, kedy nastanú ostatné reakcie. Vďaka nezávislosti potom môžeme povedať, že $K_j(\tau; x, t)$ sú z Poissonovho rozdelenia

$$K_j(\tau; x, t) \sim Poiss(\lambda_j), \tag{49}$$

s parametrami $\lambda_j = w_j(x)\tau, \ \forall j = 1, ..., M$. Ďalej, keďže $K_j(\tau; x, t), \forall j = 1, ..., M$ sú diskrétne náhodné premenné, $P(k_1, ..., k_M|\tau; x, t)$ je združenou funkciou hustoty, ktorú vieme vyjadriť ako súčin pravdepodobností

$$P(k_1, ..., k_M|\tau; x, t) = \prod_{i=1}^{M} P_{\mathcal{P}}(k_j; w_j(x), \tau), \tag{50}$$

kde $P_{\mathcal{P}}(k_j; w_j(x), \tau)$ je pravdepodobnosť, že v čase τ nastane k_j reakcií R_j, pričom reakcia R_j nastáva s pravdepodobnosťou $w_j(x)\tau$. Vieme ju vyjadriť ako

$$P_{\mathcal{P}}(k_j; w_j(x), \tau) = \frac{(w_j(x)\tau)^{k_j} \mathrm{e}^{-w_j(x)\tau}}{k_j!}, \qquad \forall j = 1, ..., M. \tag{51}$$

Pre každú reakciu R_j sa teda vygeneruje realizácia náhodnej premennej k_j z Poissonovho rozdelenia (50), ktorá predstavuje počet aktivovaných reakcií R_j na intervale

$[t, t+\tau)$. To je v súčasnosti už veľmi jednoduchá úloha, keďže sú k dispozícií rôzne efektívne algoritmy na generovanie z Poissonovho rozdelenia.

Kameňom úrazu však je voľba "skoku", resp. šírky časového intervalu τ. Totiž, aby platila *leap condition*, musí byť τ dostatočne malé na to, aby zmena stavu na tomto intervale bola len nepatrná. Označíme teda zmenu počtu molekúl i-teho proteínu v dôsledku reakcie R_j na intervale $[t, t+\tau)$ ako ν_{ij}. Potom zmena stavu celého systému na intervale $[t, t+\tau)$ sa dá vyjadriť vzťahom

$$\delta_\tau = \sum_{j=1}^{M} k_j \nu_j, \tag{52}$$

kde $\nu_j = \sum_{i=1}^{N} \nu_{ij}$. Aby bola $\tau - leap$ aproximácia presná, musí byť zmena hodnoty funkcie $w_j(x)$ dostatočne malá pre každú reakciu R_j. Požadujeme teda, aby $|w_j(x+\delta_\tau) - w_j(x)|$ bola nejakým spôsobom ohraničená, resp. "dostatočne malá".

Spôsob, akým voliť τ je popísaný v [14], avšak predstavuje len akýsi návod, nie robustnú podmienku pre τ. Vychádza z neho však sofistikovanejší postup popísaný v [17] a neskôr vylepšený [4] a implementovaný pre Python v knižnici StochPy.

V nasledujúcej kapitole predstavíme experimentálne výsledky pre zvolený model genetickej regulačnej kaskády a prezentujeme vlastné výsledky simulácií. Ukáže sa, ako vplýva veľkosť systému na bistabilitu a čo to znamená v kontexte génovej expresie a chemických reakcií prebiehajúcich v každom živom organizme.

4 Experimentálne výsledky

Ako bolo spomenuté v Kapitole 2, v našich experimentoch sa budeme venovať prípadu, keď sú produkty chemických reakcií (proteíny) v dlhodobom horizonte infinitezimálne blízko ku svojmu stabilnému stavu $p^* = (p_1, ..., p_N^*)^T$ a tieto pevné body konvergujú v zmysle bistability (29). V závislosti od parametra Ω sa bistabilita daného systému mení, čo sa neskôr ukáže aj na výsledkoch simulácií.

4.1 Dizajn modelu

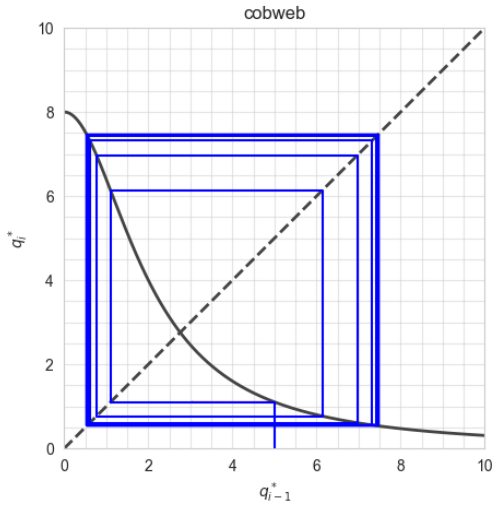
V tejto časti predstavíme model genetickej regulačnej kaskády, ktorý budeme používať v našich experimentoch. Uvažujeme autonómny systém, čo znamená, že parametre chemických reakcií budú rovnaké pre všetky proteíny okrem prvého, u ktorého predpo-

kladáme, že sa produkuje s konštantnou intenzitou. Pracujeme s predpokladom (30), teda pôvodný systém (9) zjednodušíme na (32).

Systém pozostáva zo siedmich stupňov regulačnej kaskády, t.j. $N = 7$ a konštanty vystupujúce v Hillovej funkcii spolu s intenzitami reakcií sme umelo zvolili tak, aby systém vykazoval bistabilné správanie. Hodnoty parametrov sú uvedené v Tabuľke 2. Počiatočná podmienka bola nastavená na $p(0) = \mathbf{0}_{N\times 1}$, teda vychádzame z nulových koncentrácií proteínov v systéme v čase $t = 0$.

Prvý stupeň kaskády (P_1)		Ostatné stupne ($P_2, ..., P_7$)	
Parameter	hodnota	Parameter	hodnota
k_1	1	k_{max}	2
		k_d	2
		n	2
γ_1	0.1	γ	0.05
K_1	0.05	K	0.1
Γ_1	0.1	Γ	0.5
b_1	0.5	b	2

Tabuľka 2: Tabuľka hodnôt parametrov používaných v simuláciách.



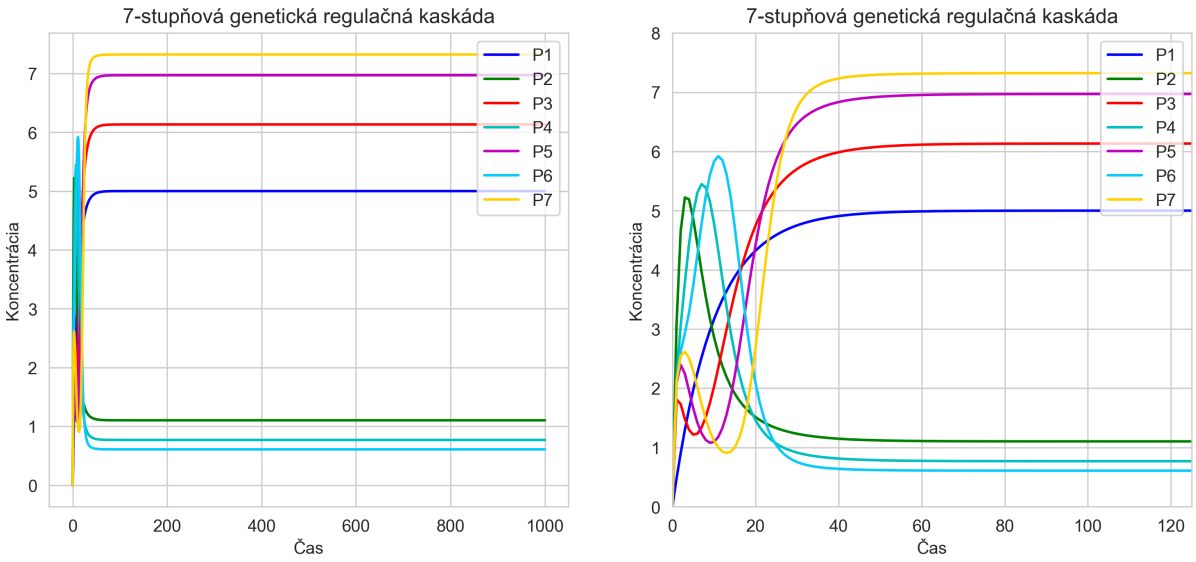
Obr. 5: Cob-Web diagram prislúchajúci systému z Obr. 6.

Na Obr. 5 je zobrazený Cob-web diagram systému (32) pre náš modelovaný systém, t.j. hodnoty parametrov zodpovedajú hodnotám v Tabuľke 2. Ako sa dá všimnúť, diagram vychádza z počiatočného bodu $p_1^* = 5$, ktorý sme dostali dosadením parametrov z Tabuľky 2 do vzťahu (11). Teoretické hodnoty pevných bodov sa dajú jednoducho dopočítať dosadením do vzťahu (21), z čoho dostávame pre náš systém hodnoty zobrazené v Tabuľke 3. Je zrejmé, že hodnoty pevných bodov korešpondujú s Cob-web diagramom z Obr. 5. Pre postupnosť pevných bodov $\{p_1^*, ..., p_N^*\}$, resp. $\{q_1, ..., q_N\}$ pre $N \to \infty$ by sme rozdelením na dve podpostupnosti podľa (29) dostali dva pevné body $q_{(1)}^*, q_{(2)}^*$ zodpovedajúce ľavému hornému a pravému dolnému rohu Cob-Web diagramu. Skutočne teda dostávame bistabilný systém, ako je to zrejmé aj z Obr. 5, kde stabilný stav N-tého proteínu prepína medzi stavmi $q_{(1)}^* \doteq 7.46$ a $q_{(2)}^* \doteq 0.54$ pre $N \to \infty$, v závislosti od toho, či ide o nepárny alebo párny index. Približné hodnoty $q_{(1)}^*$ a $q_{(2)}^*$ sme dostali po sto iteráciách s využitím vzťahu (21).

Pevný bod	Približná hodnota
p_1^*	5
p_2^*	1.10
p_3^*	6.13
p_4^*	0.77
p_5^*	6.97
p_6^*	0.61
p_7^*	7.32

Tabuľka 3: Približné hodnoty pevných bodov pre deterministický systém (32) s hodnotami parametrov z Tabuľky 2.

Koncentrácie proteínov v čase pre vyššie popísaný model je zobrazený na Obr. 6 aj pre dlhšiu, aj kratšiu časovú škálu. Na začiatku možno pozorovať mierne fluktuácie v koncentráciách, kým sa po čase systém ustáli na nejakej stabilnej hladine. Všimnime si, že hodnoty stabilných koncentrácií opäť korešpondujú s dopočítanými približnými hodnotami v Tabuľke 3.



Obr. 6: Príklad sedemstupňovej regulačnej kaskády pre hodnoty parametrov z Tabuľky 2 pre dlhšiu a kratšiu časovú škálu.

4.2 Výsledky simulácií

V tejto kapitole predstavíme vlastné výsledky stochastických simulácií pre model popísaný vyššie v Kapitole 4.1. Zaujímať nás bude závislosť stabilného stavu jednotlivých proteínov od parametra Ω, ktorý označuje veľkosť systému, resp. celkový počet molekúl.

Simulácie stavajú na nasledovnej kinetickej schéme

$$\begin{aligned} x_1 &\xrightarrow{w_1^+} && x_1 + 1, \\ x_1 &\xrightarrow{w_1^-(x_1)} && x_1 - 1, \\ x_i &\xrightarrow{w_i^+(x_{i-1})} && x_i + 1, && i = 2, 3, ..., 7, \\ x_i &\xrightarrow{w_i^-(x_i)} && x_i - 1, && i = 2, 3, ..., 7, \end{aligned} \tag{53}$$

kde prvá reakcia, resp. produkcia proteínu P_1 prebieha s konštantnou intenzitou $w_1^+ = \Omega b_1 k_1$ a jeho rozpad s intenzitou $w_1^-(x_1) = \Gamma x_1$, pričom x_1 je počet molekúl P_1. Translácia každého ďalšieho proteínu $P_i, \forall i = 1, ..., 7$ prebieha s intenzitou $w_i^+(x_{i-1}) = \Omega b h(x_{i-1}/\Omega)$, kde h je Hillova funkcia (10) a ich rozpad s intenzitou $w_i^-(x_i) = \Gamma x_i$. Hodnoty reakčných konštánt vystupujúcich v simuláciách sú uvedené v Tabuľke 2.

Ako aproximáciu stabilného stavu $p_i^*, \forall i = 1, ..., 7$ sa používal odhad strednej hodnoty koncentrácie proteínov $\bar{p}_i = E(p_i)$ získaný zo simulácií, keďže pre dostatočne dlhý časový horizont nemajú počiatočné fluktuácie signifikantný vplyv na strednú hodnotu, čo sa

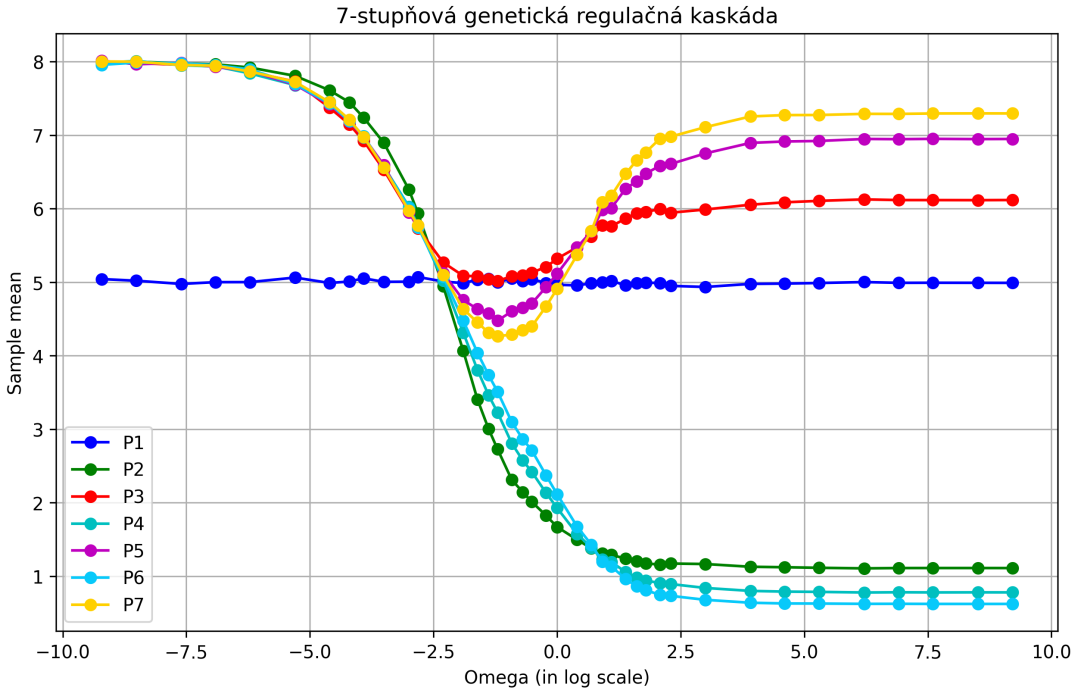
potvrdilo aj v našich experimentoch.

Parameter Ω sme volili najskôr experimentálne ako mocniny 10 na škále od 10^{-4} až po 10^4. Z prvých experimentov sa potom ukázalo rozumné zvoliť jemnejšie delenie v oblastiach, kde zmena stredných hodnôt $\bar{p_i}$, $\forall i = 1, ..., 7$ nastávala rýchlejšie, resp. výraznejšie a vykazovala potenciálne „zaujímavé" správanie. V uvedených experimentoch sme preto pracovali s nasledovnými hodnotami $\Omega \in \{10^{-4}, 2 \times 10^{-4}, 5 \times 10^{-4}, 10^{-3}, 2 \times 10^{-3}, 5 \times 10^{-3}, 10^{-2}, 1.5 \times 10^{-2}, 2 \times 10^{-2}, 3 \times 10^{-2}, 5 \times 10^{-2}, 6 \times 10^{-2}, 10^{-1}, 0.15, 0.2, 0.25, 0.3, 0.4, 0.5, 0.6, 0.8, 1, 1.5, 2, 2.5, 3, 4, 5, 6, 8, 10, 20, 50, 10^2, 2 \times 10^2, 5 \times 10^2, 10^3, 2 \times 10^3, 5 \times 10^3, 10^4\}$.

Na simulácie sme použili implementovaný Gillespie-ho algoritmus v knižnici StochPy v jazyku Python. Pre menšie hodnoty Ω (t.j. $\Omega < 10$), kde reakcie nastávajú s nižšou intenzitou, sme použili priamu metódu, keďže použitie aproximatívnych metód nie je v takýchto prípadoch presné [14]. Pre väčšie hodnoty Ω (t.j. $\Omega \geq 10$), kde je použitie aproximatívnych metód rozumné (aj z hľadiska trvania výpočtu), sme zvolili Tau-leap metódu s dĺžkou kroku $\epsilon = 0.02$ pre $\Omega \in \{10, 20, 50, 100\}$ a $\epsilon = 0.03$ pre $\Omega > 100$. Hranicu pre veľkosť Ω, odkedy sa oplatí použiť Tau-leap sme zvolili experimentálne a tiež podľa výpočtovej náročnosti a vlastných technických možností, aby kvalita výsledkov zostala nezmenená. Na začiatku sme zvolili prednastavenú hodnotu v StochPy podľa [4], t.j. $\epsilon = 0.03$, $\forall \Omega \geq 10$, avšak okolo $\Omega = 10$ sa dal pozorovať mierny výkyv strednej hodnoty, preto sme usúdili, že dĺžka kroku nemusí byť vhodná, keďže $\Omega = 10$ je stále pomerne malá hodnota. Preto sme hodnotu pre $10 \leq \Omega \leq 100$ znížili. Ukázalo sa, že pre takéto hodnoty už nebol prítomný žiadny šum. Časový rozsah sme pritom volili tak, aby sme pre odhady stredných hodnôt dostali približne rovnako reprezentatívnu vzorku pre všetky hodnoty Ω, t.j. t_{max} sme volili spôsobom $t_{max}^{(i)} \doteq 10^5/\Omega^{(i)}$. Pre hodnoty $\Omega > 10$, kde už sa používala Tau-leap metóda, sme však už fixne nastavili $t_{max} = 5000$. Pre každú hodnotu Ω sme tiež vygenerovali 5 trajektórií na celej časovej škále na dosiahnutie lepších odhadov stredných hodnôt $E(p_i)$, resp. $E(x_i)$.

Hlavným motívom našich experimentov bolo zistiť, aký vplyv má objem systému na vzájomnú reguláciu produkcie proteínov v umelo skonštruovanej genetickej regulačnej kaskáde. Pre veľké hodnoty Ω, t.j. veľký počet molekúl v systéme by sa zrejme dalo očakávať, že systém sa bude správať približne deterministicky, keďže v každom časovom

intervale môže nastať veľký počet chemických reakcií. Pre veľký objem systému by sme teda čakali, že bistabilita bude veľmi podobná deterministickému prípadu. Mnoho chemických systémov však prebieha na tzv. nano-škále, teda pri nízkych koncentráciách s nízkym počtom molekúl a v dôsledku toho sú prítomné výrazné fluktuácie v reakčných intenzitách [22]. Pri malých počtoch molekúl totiž intenzita chemických reakcií nie je taká veľká, čo znamená, že už proteín P_1 vzniká s veľmi malou intenzitou (keďže tá závisí od Ω). Regulácia translácie proteínu P_2 proteínom P_1 potom nemusí byť taká výrazná, čo sa premietne aj na ďalších stupňoch kaskády.



Obr. 7: Stabilný stav koncentrácie proteínov $P_i, ..., P_7$ v závislosti od veľkosti systému. Veľkosť systému je daná počtom molekúl Ω (zobrazená na osi x v logaritmickej škále). Stabilný stav koncentrácie jednotlivých proteínov je aproximovaný odhadom strednej hodnoty $p_i^* = \Omega^{-1}E(x_i),\ \forall i = 1, ..., 7$.

Matematicky sa tomu dá rozumieť nasledovne. V deterministickom prípade je stabilný stav koncentrácie proteínu P_1 daný vzťahom (11). Keď do modelu pridáme variabilitu prostredníctvom parametra Ω, zo vzťahu (34) dostaneme stabilný stav počtu molekúl

$$E(x_1) = \frac{b_1 k_1}{\Gamma_1}\Omega. \tag{54}$$

Všimnime si, že pre $\Omega \to 0$ aj $E(x_1) \to 0$. Pre proteín P_2 to potom znamená, že

$$\begin{aligned} E(x_2) &= \lim_{E(x_1)\to 0} \left[\frac{b}{\Gamma} \frac{k_{max}}{1+\left(\frac{E(x_1)}{k_d}\right)^n} \Omega \right] \\ &= \frac{bk_{max}}{\Gamma} \Omega. \end{aligned} \tag{55}$$

Opäť platí, že pre $\Omega \to 0$ aj $E(x_2) \to 0$. Keby sme takto postupovali ďalej, dostali by sme rovnaký vzťah pre každý ďalší stupeň kaskády, t.j.

$$\begin{aligned} E(x_i) &= \lim_{E(x_{i-1})\to 0} \left[\frac{b}{\Gamma} \frac{k_{max}}{1+\left(\frac{E(x_{i-1})}{k_d}\right)^n} \Omega \right] \\ &= \frac{bk_{max}}{\Gamma} \Omega, \end{aligned} \tag{56}$$

pričom $E(x_i) \to 0$ pre $\Omega \to 0$. Pre koncentrácie $p_1^*, ..., p_N$ potom dostaneme z (41) a z hodnôt parametrov v Tabuľke 2 nasledovné pevné body

$$p_1^* = \frac{b_1 k_1}{\Gamma_1} = 5, \tag{57}$$

$$p_i^* = \frac{bk_{max}}{\Gamma} = 8, \qquad \forall i = 1, ..., 7, \tag{58}$$

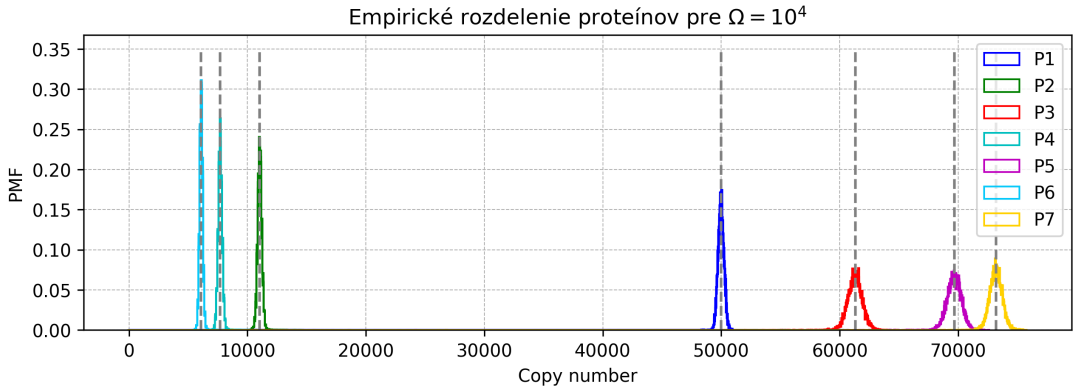
za predpokladu, že $\Omega \to \infty$.

Na Obr. 7 sú zobrazené odhadnuté stredné hodnoty $\bar{p_1}, ..., \bar{p_7}$ v závislosti od Ω v logaritmickej škále, kde

$$\bar{p_i} = \Omega^{-1} \bar{x_i}, \text{ pričom } \bar{x_i} = \frac{1}{T} \sum_{i=1}^{T} x_i \tag{59}$$

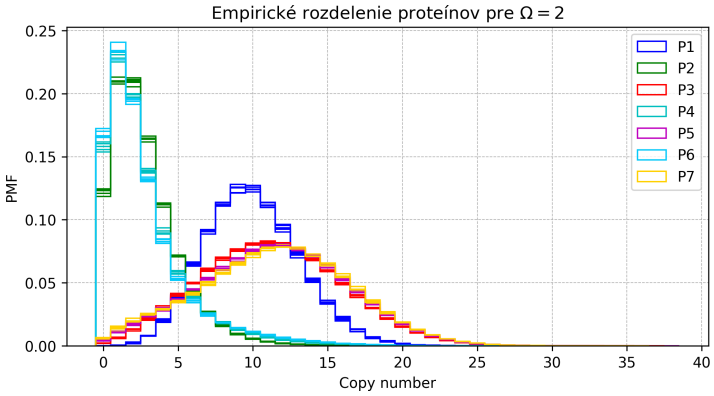
a T je počet časových krokov. Je zrejmé, že pre malé hodnoty Ω sú koncentrácie pre všetky proteíny približne rovné 8, ako sme to aj predpokladali vyššie (vzťahy (57) a (58)). Pre dostatočne malé Ω, t.j. $\Omega < 10^{-3}$ už regulácia prakticky nenastáva. Teda pre náš systém (represor) žiadny predchádzajúci proteín nepotláča transláciu nasledujúceho a teda všetky reakcie prebiehajú s maximálnou intenzitou k_{max}. Naopak, pre veľké Ω systém vykazuje bistabilné správanie, podobne, ako to je v deterministickom prípade (Kapitola 4.1). To sa dá lepšie vidieť aj na Obr. 8, kde sú zobrazené empirické rozdelenia počtov molekúl jednotlivých proteínov, t.j. x_i, $\forall i = 1, ..., 7$. Z obrázka sa dá vidieť, že odhady stredných hodnôt $\bar{p_i}$, $\forall i = 1, ..., 7$ konvergujú k teoretickému stabilnému stavu $p^* = (p_1^*, ..., p_7^*)^T$. Na Obr. 10, kde sú zobrazené počty molekúl v čase, sa skutočne dá vidieť, že systém sa správa skoro deterministicky pre veľký objem, v tomto prípade približne $\Omega \doteq 10^4$.

Pre $\Omega \leq 10$ (približne $\log(\Omega) < 2.5$ na Obr. 7) vidíme, že bistabilita sa postupne vytráca a distribúcie množstva proteínov sa postupne menia. Dokonca sa javí, že postupnosť $E(p_2), E(p_4), ..., E(p_{2N})$, resp. $E(p_3), E(p_5), ..., E(p_{2N+1})$ sa pre $\Omega \leq 10$ správa ako



Obr. 8: Empirické rozdelenie pravdepodobnosti počtu molekúl x_i, $\forall i = 1, ..., 7$ jednotlivých proteínov. Sivou prerušovanou čiarou sú znázornené teoretické pevné body $p_1^*, ..., p_7^*$ preškálované na veľkosť systému $\Omega = 10^4$.

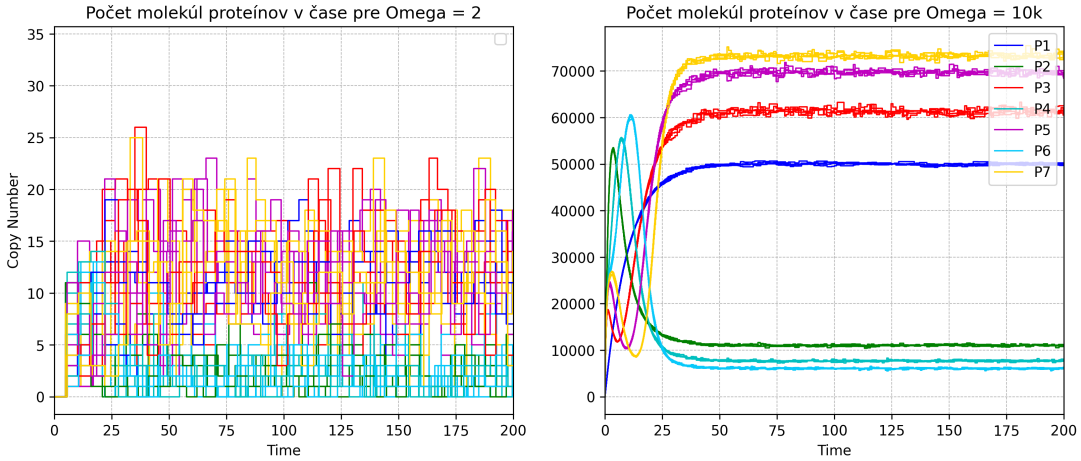
monotónna postupnosť. Avšak, v bode $\tilde{\Omega} \doteq 2$ (t.j. $\log(\Omega) \doteq 0.69$) sa rastúca postupnosť mení na klesajúcu (nepárne indexy), resp. klesajúca sa mení na rastúcu (párne indexy). To znamená, že v tomto bode majú všetky proteíny s párnymi, resp. nepárnymi indexmi rovnaké stredné hodnoty. Pozreli sme sa teda aj na distribúcie, aby sme zistili, či sa aj tie približne rovnajú.



Obr. 9: Empirické rozdelenie pravdepodobnosti počtu molekúl x_i, $\forall i = 1, ..., 7$ jednotlivých proteínov pre veľkosť systému $\Omega = 2$.

Na Obr. 9 sa dá vidieť, že proteíny s párnymi indexmi skutočne majú približne rovnaké empirické rozdelenie, čo platí aj pre proteíny s nepárnymi indexmi. Tiež vidíme, že regulácia pre systém s $\Omega = 2$ vôbec nie je taká výrazná ako pre $\Omega = 10^4$, čo je ešte umocnené faktom, že $\Omega = 2$ je približne bod „zlomu" postupností, kde majú koncentrácie

proteínov s párnymi, resp. nepárnymi indexmi približne rovnaké stredé hodnoty. Kontrast medzi distribúciami, keď $\Omega = 2$ a keď $\Omega = 10^4$ je jasne vidieť aj na Obr. 10, kde sú zobrazené počty molekúl v čase.



Obr. 10: Počet molekúl $x_i(t)$, $\forall i = 1, ..., 7$ jednotlivých proteínov v čase získaný zo simulácií pre $\Omega = 2$ a $\Omega = 10^4$.

Zaujímavé je ešte pozrieť sa na tzv *Fano faktor*, vymyslený taliansko-americkým fyzikom Ugom Fanom pre potreby skúmania fluktuácií častíc v iónových detektoroch [8]. Fano faktor reprezentuje, podobne ako variačný koeficient, akúsi mieru variability v stochastickom procese a často sa využíva práve ako miera variability v genetických regulačných sieťach [30]. Je definovaný ako [8]

$$F(t) = \frac{\sigma_t^2}{\mu_t}, \qquad t > 0, \tag{60}$$

kde, σ_t^2 je variancia a μ_t je stredná hodnota počtu udalostí nejakého stochastického procesu (konkrétne ide o tzv. *counting process* [26]) od času $t > 0$. Ako sa uvádza napríklad v [14], [30], počet molekúl v čase sa štandardne berie ako Poissonoov proces (Def. 2.1.1., [26]). Keďže stredná hodnota aj variancia Poissonovho rozdelenia $Poiss(\lambda)$ je λ, je jednoduché si všimnúť, že Fano faktor pre Poissonov proces bude

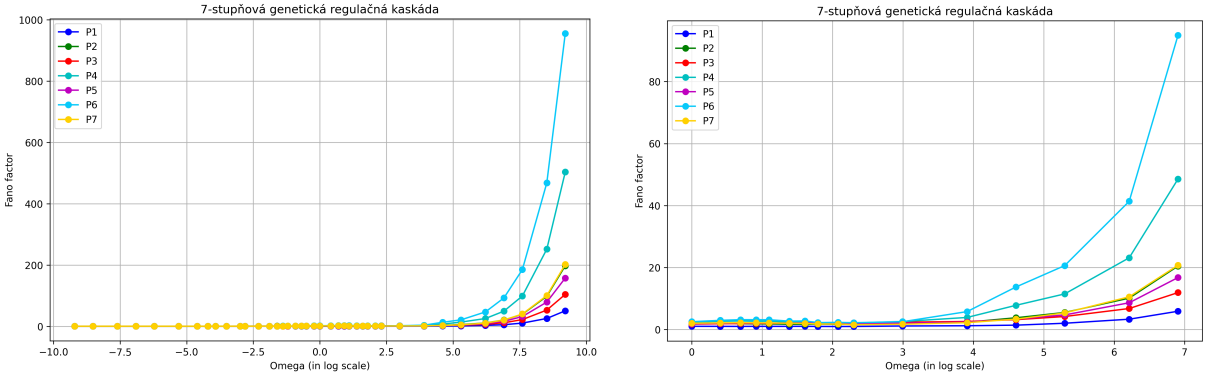
$$F_{\mathcal{P}}(t) = 1. \tag{61}$$

V Tabuľke 4 je uvedený Fano faktor pre každý proteín a pre $\Omega \in \{10^{-4}, 2, 10^3\}$. Prvý prípad, t.j. keď $\Omega = 10^{-4}$, zodpovedá malému počtu molekúl, teda chemické reakcie tam

	Fano faktor pre všetky stupne kaskády						
Ω	P_1	P_2	P_3	P_4	P_5	P_6	P_7
10^{-4}	1.0001	0.9999	0.9999	1.0000	1.0000	0.9999	1.0000
2	0.9834	1.9010	1.9310	2.7320	2.0811	3.1987	2.2031
10^3	2.1149	5.4066	4.3676	11.0615	4.7716	18.7867	5.2980

Tabuľka 4: Fano faktor pre všetky stupne kaskády pre vybrané Ω.

prebiehajú s nízkou intenzitou. Fano faktor sa tam však drží okolo 1, čo približne zodpovedá teoretickej hodnote (61). Pre $\Omega = 2$ už možno pozorovať odchýlku od teoretického Fano faktora, a tiež sa zdá, že proteíny s párnym indexom majú väčšiu variabilitu oproti nepárnym, čo sa prejavuje aj u $\Omega = 10^3$, dokonca oveľa výraznejšie. Keď sa pozrieme na Obr. 11, vidíme, že Fano faktor narastá so zväčšujúcim sa Ω. Tiež sa dá všimnúť, že s narastajúcimi stupňami kaskády narastá variabilita, čo analyzoval vo svojich experimentoch napríklad aj Hooshangi a kolektív [21].



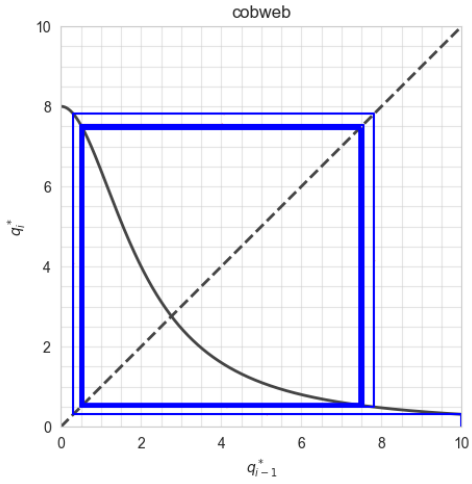
Obr. 11: Fano faktor pre celý časový úsek pre jednotlivé proteíny v závislosti od veľkosti systému Ω (v logaritmickej škále). Na prvom obrázku je Fano faktor pre všetky hodnoty Ω, pričom na druhom obrázku je vykreslený len úsek pre $\Omega \in \{1, 2, 5, 10, 20, 50, 100, 200, 500, 1000\}$.

Z výsledkov uvedených simulácií môžeme usúdiť, že regulačný efekt kaskády sa systematicky mení v závislosti od objemu systému, čo najlepšie vidieť na Obr. 7. Pre veľké Ω pozorujeme skoro deterministické správanie, pričom pre veľmi malé Ω (zrejme keď $\Omega < 10^{-3}$) regulácia úplne zanikne v dôsledku nízkej intenzity reakcií. Ukázalo sa, že v okolí $\Omega = 2$ nastáva zlom, kedy sa monotónne postupnosti koncentrácií proteínov menia z

rastúcich na klesajúce (a opačne). Otázkou je, či takýto jav nastáva vo všeobecnosti alebo je to len určitý špeciálny prípad.

4.3 Porovnanie s ďalšími modelmi

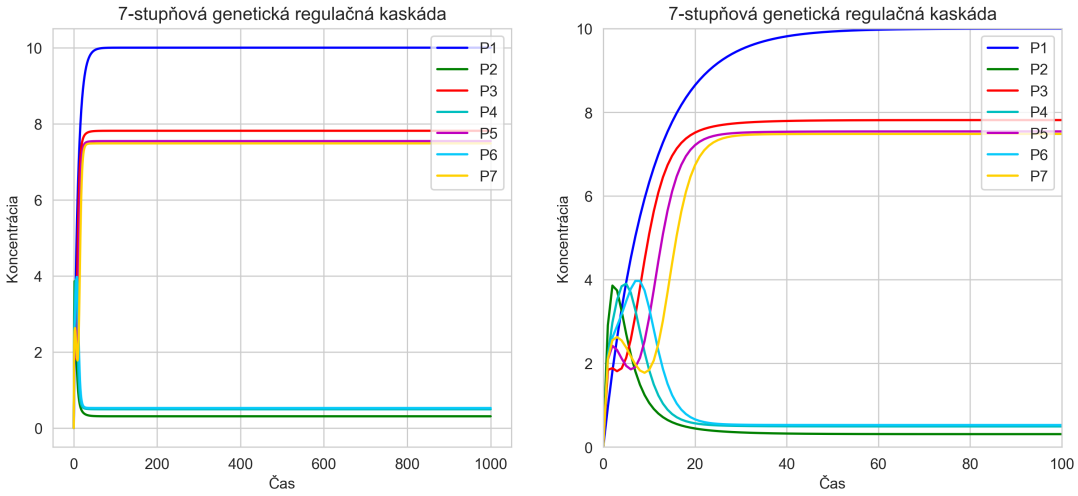
Pre porovnanie sme zobrali alternatívny model, kde sme zmenili konštantu k_1 z Tabuľky 2 na $k_1 = 2$, pričom ostatné hodnoty z tabuľky zostali rovnaké. Za následok to malo iba zmenu stabilného stavu proteínu P_1, čo sa dá vnímať ako zmena počiatočnej podmienky v diferenčnej rovnici (19). Pevné body $p_1^*, ..., p_N^*,\ \forall i = 1, ..., 7$ teda konvergujú v zmysle Cob-web diagramu smerom „dovnútra" (Obr. 12). Tiež to možno vidieť na deterministickom správaní na Obr. 13, kde je zobrazený deterministický priebeh modelu.



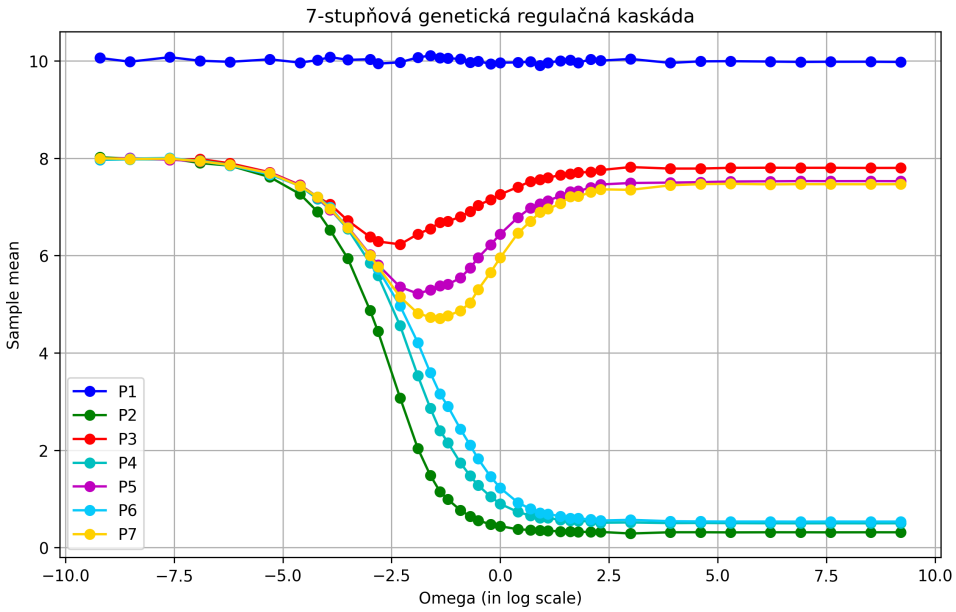
Obr. 12: Cob-web diagram pre alternatívny model.

Na Obr. 14 je zobrazená stredná hodnota v závislosti od Ω pre druhý model (t.j., kde $k_1 = 2$). Vidíme, že v tomto prípade nenastáva situácia, že by existoval nejaký zlomový bod $\tilde{\Omega}$, kde majú proteíny s párnymi, resp. nepárnymi indexmi rovnaké stredné hodnoty, ako to bolo v predošlom modeli pre $\tilde{\Omega} = 2$.

Ako možné rozšírenie sme uviedli model s jedenástimi stupňami kaskády, t.j. $N = 11$. Tam sme však pre zložitosť modelu a vysokú výpočtovú náročnosť značne zredukovali rozsah simulácií. Zvolili sme hrubšie delenie pre Ω na intervale $[10^{-4}, 10^4]$, teda sme uvažovali $\Omega \in \{10^{-4}, 2\times10^{-4}, 5\times10^{-4}, 10^{-3}, 2\times10^{-3}, 5\times10^{-3}, 0.01, 0.02, 0.05, 0.1, 0.2, 0.5, 1, 2, 5, 10, 20, 50, 100, 200, 500, 10^3, 2\times10^3, 5\times10^3, 10^4\}$, pričom pre každú hodnotu Ω sme spustili



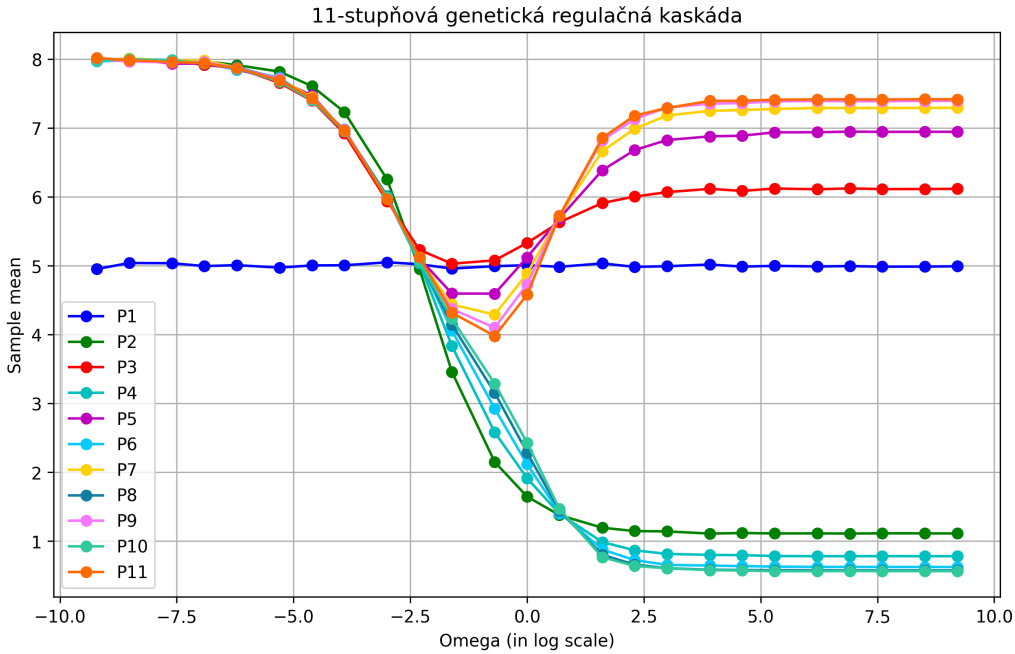
Obr. 13: Deterministické správanie modelu pre $k_1 = 2$.



Obr. 14: Stredná hodnota pre jednotlivé proteíny v závislosti od Ω pre alternatívny model.

len jeden simulačný beh namiesto piatich. Ostatné nastavenia experimentu sme ponechali pôvodné, t.j. rovnako, ako je popísané v Kapitole 4.2 aj s hodnotami v Tabuľke 2.

Z Obr. 15 sa dá ešte jasnejšie vidieť postupnosť $p_2^*, p_4^*, ..., p_{10}^*$, resp. $p_3^*, p_5^*, ..., p_{11}^*$ a rovnako aj „zlomový bod" $\tilde{\Omega} = 2$, kde sa monotónnosť postupnosti mení (z rastúcej na klesajúcu a opačne). Je zrejmé, že v okolí bodu $\tilde{\Omega}$ stredná hodnota klesá, resp. rastie veľmi rýchlo v závislosti od stupňa kaskády, t.j. čím väčší stupeň, tým rýchlejšia zmena

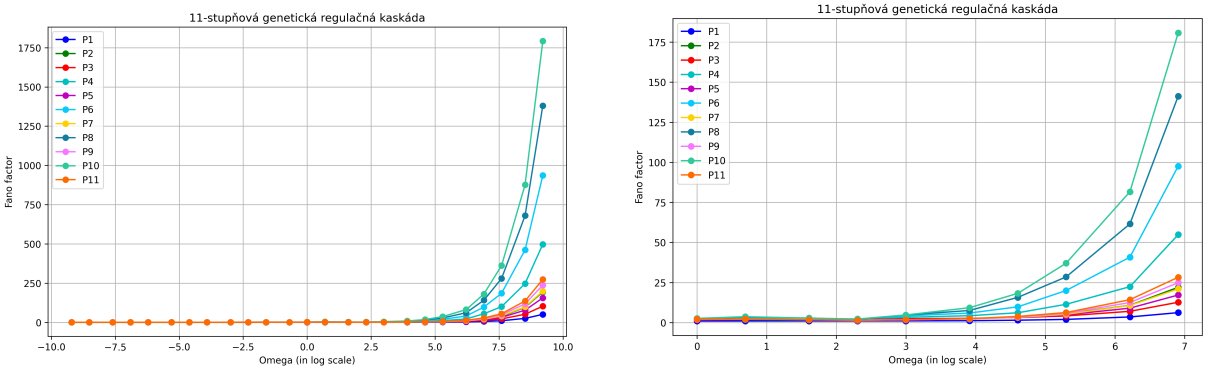


Obr. 15: Stredná hodnota pre jednotlivé proteíny v závislosti od Ω pre rozšírený model o ďalšie štyri stupne kaskády, t.j. pre $N = 11$.

strednej hodnoty v okolí $\tilde{\Omega}$. Keď si porovnáme výsledky s pôvodným modelom z Obr. 7, môžeme povedať, že pridanie ďalších štyroch stupňov kaskády neposkytlo nový pohľad na zmenu strednej hodnoty alebo určenie bodu $\tilde{\Omega}$. Skôr podporilo našu teóriu vyvstávajúcu z Obr. 7, že funkcie stredných hodnôt koncentrácií proteínov v závislosti od Ω pravdepodobne tvoria akúsi postupnosť (samostatne pre proteíny s párnym, resp. nepárnym indexom). Zvyšujúca sa rýchlosť klesania (resp. nárastu) strednej hodnoty pre zvyšujúci sa počet stupňov kaskády môže signalizovať akýsi stochastický ekvivalent k bifurkácii (v deterministickom modeli) v limitnom prípade, t.j. pre $N \to \infty$.

Pozreli sme sa aj na Fano faktor pre rozšírený systém a v podstate je vidieť rovnaký trend ako v pôvodnom modeli (Obr. 11). Fano faktor sa zvyšuje s narastajúcim Ω a tiež so stupňom kaskády. Opäť sa dá všimnúť, že pre proteíny s párnymi indexmi je Fano faktor väčší ako pre druhú skupinu proteínov (teda s nepárnymi indexmi).

Z výsledkov simulácií a porovnaní rôznych systémov je hneď zrejmé, že v závislosti od počiatočnej podmienky, resp. stabilného stavu koncentrácie prvého proteínu môžeme dostať rôzne vzťahy stredných hodnôt a počtu molekúl v systéme. Dokonca aj napriek tomu, že uvažujeme len sytémy, ktoré sú v deterministickom prípade bistabilné. Ako sme



Obr. 16: Fano faktor pre jednotlivé proteíny pre celý pozorovaný časový rozsah v závislosti od veľkosti systému Ω (v logaritmickej škále). Na prvom obrázku je Fano faktor pre všetky hodnoty Ω, pričom na druhom obrázku je vykreslený len úsek pre $\Omega \in \{1, 2, 5, 10, 20, 50, 100, 200, 500, 1000\}$.

si všimli, pre druhý model (Obr. 14) sa postupnosti nijako „nekrížili", vždy išlo buď o rastúcu, alebo klesajúcu postupnosť. Nás však zaujímal scenár z Obr. 7, kde by sme predpokladali, že bod $\tilde{\Omega} = 2$ môže byť indikátorom akejsi „stochastickej bifurkácie", kde sa stredná hodnota v okolí $\tilde{\Omega}$ radikálne mení pre $N \to \infty$ (teda v limitnom prípade). Obr. 15 sa zdanlivo prikláňa k tejto teórii, keďže so zvyšujúcimi sa stupňami sa zvyšuje aj sklon funkcie strednej hodnoty od Ω, až kým tieto postupnosti (t.j. (29)) potenciálne neskonvergujú pre $N \to \infty$. Na potvrdenie takejto teórie by však bolo potrebných niekoľko ďalších experimentov, ako aj zložitejšie simulácie väčších systémov a podrobnejšia analýza rozličných prípadov za rôznych podmienok.

Záver

V tejto práci sme sa sústredili na syntézu teoretických a experimentálnych poznatkov týkajúcich sa modelovania génovej expresie. Zhrnuli sme existujúce poznatky týkajúce sa deterministického modelu regulačných kaskád, pozreli sa bližšie na ich stabilitu a sledovali, ako závisí od jednotlivých parametrov modelu. Položili sme ju do kontextu so stochastickým prístupom k modelovaniu génovej expresie prostredníctvom Markovovských procesov, pričom v simuláciách sme sa zamerali na prípad, kedy systém (teda regulačná kaskáda) dosiahne v deterministickom prípade tzv. bistabilitu [11].

Simulácie bistabilného systému s rovnakými parametrami sme skonštruovali pre rôzny celkový počet molekúl Ω, pričom nás zaujímala zmena stability pre rôzne hodnoty Ω. Stabilitu sme aproximovali strednou hodnotou koncentrácie proteínov, resp. počtom molekúl relatívne k objemu systému z jednotlivých simulačných behov na dostatočne dlhom časovom horizonte. Ukázalo sa, že daný teoreticky bistabilný systém skutočne mení svoje správanie pre rôzne Ω a s klesajúcim počtom molekúl sa deterministická bistabilita postupne vytráca.

Zaujímavým pozorovaním bolo, že stredné hodnoty koncentrácie proteínov sa pre Ω z určitého intervalu správali ako monotónna postupnosť funkcií (v závislosti od objemu systému). Tiež sa dal pozorovať určitý „zlomový bod" $\tilde{\Omega}$, kde sa postupnosť menila z rastúcej na klesajúcu (a *vice versa*, v závislosti od stupňa kaskády). V tomto bode sa stredné hodnoty pre proteíny s párnym, resp. nepárnym indexom približne rovnali, a dokonca sa rovnali aj ich empirické distribúcie. Takýto bod by mohol byť zajímavý pre ďalšie experimenty, keďže by mohol indikovať akúsi bifurkáciu v ďalšom rozmere danom parametrom Ω. Takéto výsledky by mohli byť kľúčovými najmä v rôznych oblastiach biomedicíny, keďže chemické reakcie v živých organizmoch prebiehajú v rôznych koncentráciách a vedomosť o ich stabilite by mohla byť cennou informáciou.

V ďalšom kroku tohto výskumu by bolo vhodné vykonať niekoľko ďalších simulácií za účelom spresnenia odhadu bodu $\tilde{\Omega}$. Iný pohľad v zmysle experimentov by mohol byť naopak taký, že pri pevnom objeme systému skúsiť určiť (simulačne alebo teoreticky) parametrický priestor reakčných konštánt tak, aby sme dostali práve tento „bod zlomu" $\tilde{\Omega}$. Priestor pre ďalšie skúmanie ponúka bod $\tilde{\Omega}$ aj v kontexte s lineárnou aproximáciou šumu (*angl. linear noise approximation*) [5], ozn. LNA. Dalo by sa pozrieť na to, aký

vplyv na varianciu v zmysle LNA má hodnota Ω, špeciálne ak ide o bod $\tilde{\Omega}$.

Práca týmto poskytuje náhľad do vnútrobunkových procesov a príslušných chemických reakcií prostredníctvom simulácií a zaoberá sa ich stabilitou v závislosti od počtu molekúl v systéme. Pomocou simulácií skúma parametrický priestor syntetickej regulačnej kaskády a predstavuje vlastné výsledky v analýze stochastickej stability. Veríme, že tieto výsledky budú cenným prínosom pre širokú škálu biomedicínskych aplikácií a podnietia ďalší rozvoj výskumu v tejto oblasti.

Zoznam použitej literatúry

[1] U. Alon: *An introduction to systems biology: design principles of biological circuits*, CRC press, 2019

[2] U. Alon: *Response Delays and the Structure of Transcription Networks*, J. Mol. Biol., vol. 329 (2003), 645–654

[3] F. Baras, M. M. Mansour, J. E. Pearson: *Microscopic Simulation of Chemical Bistability in Homogeneous Systems*, The Journal of Chemical Physics, 1996, vol. 105, 8257–8261

[4] Y. Cao, D. T. Gillespie, L. R. Petzold: *Efficient step size selection for the tau-leaping simulation method*, The Journal of Chemical Physics, 2006, vol. 124, 044109

[5] J. Elf, M. Ehrenberg: *Fast evaluation of fluctuations in biochemical networks with the linear noise approximation*, Genome Res., 2003, vol. 13, 2475-2484

[6] M. Elowitz, S. Leibler: *A synthetic oscillatory network of transcriptional regulators*, Nature, 2000, vol. 403, 335–338

[7] M. B. Elowitz, A. J. Levine, E. D. Siggia, P. S. Swain: *Stochastic gene expression in a single cell*, Science, 2002, vol. 297, 1183-1186

[8] U. Fano: *Ionization Yield of Radiations. II. The Fluctuations of the Number of Ions*, Physical Review, 1947, vol. 72, 26-29

[9] J. E. Ferrell: *Tripping the switch fantastic: how a protein kinase cascade can convert graded inputs into switch-like outputs*, Trends Biochem Sci., 1996, vol. 21, 460-6

[10] J. E. Ferrell, E. M. Machleder: *The Biochemical Basis of an All-or-None Cell Fate Switch in Xenopus Oocytes*, Science, 1998, vol. 280, 895-898

[11] T. Gardner, C. Cantor, J. Collins: *Construction of a genetic toggle switch in Escherichia coli*, Nature, 2000, vol. 403, 339–342

[12] M. A. Gibson, J. Bruck: *Efficient Exact Stochastic Simulation of Chemical Systems with Many Species and Many Channels*, The Journal of Physical Chemistry, 2000, vol. 104, 1876–1889

[13] D. T. Gillespie: *A general method for numerically simulating the stochastic time evolution of coupled chemical reactions*, Journal of Computational Physics, 1976, vol. 22, 403-434

[14] D. T. Gillespie: *Approximate accelerated stochastic simulation of chemically reacting systems*, The Journal of Physical Chemistry, 2001, vol. 115, 1716–1733

[15] D. T. Gillespie: *A rigorous derivation of the chemical master equation*, Physica A: Statistical Mechanics and its Applications, 1992, vol. 188, 404-425

[16] D. T. Gillespie: *Exact stochastic simulation of coupled chemical reactions*, The Journal of Physical Chemistry, 1977, vol. 81, 2340-2361

[17] D. T. Gillespie, L. R. Petzold: *Improved leap-size selection for accelerated stochastic simulation*, The Journal of Chemical Physics, 2003, vol. 119, 8229–8234

[18] S. Grossmann: *Langevin forces in chemically reacting multicomponent fluids*, The Journal of Chemical Physics, 1976, vol. 65, 2007–2012

[19] R. Harman, S. Rosa: *Stochastické simulačné metódy*, skriptá k predmetu Stochastické simulačné metódy, FMFI UK, Bratislava, 2019

[20] A. V. Hill: *The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves*, The Journal of Physiology, 1910, vol. 40, iv-vii

[21] S. Hooshangi, S. Thiberge, R. Weiss: *Ultrasensitivity and noise propagation in a synthetic transcriptional cascade*, Proceedings of the National Academy of Sciences of the USA, 2005, vol. 102, 3581-3586

[22] H. H. McAdams, A. Arkin: *It's a noisy business! Genetic regulation at the nanomolar scale*, Trends in Genetics, 1999, vol. 15, 65–69

[23] H. H. McAdams, A. Arkin: *Stochastic mechanisms in gene expression*, Proc. Natl Acad. Sci. USA, 1997, vol. 94, 814-819

[24] E. M. Ozbudak, M. Thattai, I. Kurtser, A. D. Grossman, A. van Oudenaarden: *Regulation of noise in the expression of a single gene*, Nature Genetics, 2002, vol. 31, 69–73

[25] E. Parzen: *Nonparametric Statistical Data Modeling*, Journal of the American Statistical Association, 1979, vol. 74, 105–121

[26] S. M. Ross: *Stochastic processes*, John Wiley & Sons, 1995

[27] M. R. Said, A. V. Oppenheim, D. A. Lauffenburger: *Modeling cellular signal processing using interacting Markov chains*, 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003, Proceedings. (ICASSP '03), Hong Kong, China, 2003, pp. VI-41

[28] M. Santillán: *On the Use of the Hill Functions in Mathematical Models of Gene Regulatory Networks*, Math. Model. Nat. Phenom., 2008, vol. 3, 85-97

[29] A. Singh, M. Soltani: *Quantifying Intrinsic and Extrinsic Variability in Stochastic Gene Expression Models*, PLoS ONE, 2013, vol. 8, e84301

[30] M. Thattai, A. van Oudenaarden: *Intrinsic noise in gene regulatory networks*, Proceedings of the National Academy of Sciences of the USA, 2001, vol. 98, 8614-8619

[31] N. G. Van Kampen: *Stochastic Processes in Physics and Chemistry*, Elsevier, 2007

[32] Y. Wang, S. He: *Inference on autoregulation in gene expression with variance-to-mean ratio*, Journal of Mathematical Biology, 2023, vol. 86, 87

[33] X. Zheng, Y. Tao: *Additivity of noise propagation in a protein cascade*, J. Chem. Phys., 2008, vol. 128, 165104

[34] Q. Zheng, J. Ross: *Comparison of deterministic and stochastic kinetics for nonlinear systems*, J. Chem. Phys., 1991, vol. 94, 3644–3648

[35] T. Zhou, J. Zhang: *Analytical Results for a Multistate Gene Model*, SIAM Journal on Applied Mathematics, 2012, vol. 72, 789-818

Príloha

Kódy k simuláciám

Model 7-stupňovej kaskády:

https://github.com/laura-hajzokova/Modelling-of-gene-regulatory-network-using-SSA/tree/d5f2b4e49a0378dfed99d4720aac4b93b69bfe54/7-stage-model-1

Alternatívny model 7-stupňovej kaskády:

https://github.com/laura-hajzokova/Modelling-of-gene-regulatory-network-using-SSA/tree/d5f2b4e49a0378dfed99d4720aac4b93b69bfe54/7-stage-model-2

Model 11-stupňovej kaskády:

https://github.com/laura-hajzokova/Modelling-of-gene-regulatory-network-using-SSA/tree/d5f2b4e49a0378dfed99d4720aac4b93b69bfe54/11-stage-model