

Pokroky matematiky, fyziky a astronomie

Tomáš Vejchodský

Stochastické modely v biochemii. Je život jen náhoda?

Pokroky matematiky, fyziky a astronomie, Vol. 59 (2014), No. 4, 277--284

Persistent URL: <http://dml.cz/dmlcz/144079>

Terms of use:

© Jednota českých matematiků a fyziků, 2014

Institute of Mathematics of the Academy of Sciences of the Czech Republic provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these *Terms of use*.



This paper has been digitized, optimized for electronic delivery and stamped with digital signature within the project *DML-CZ: The Czech Digital Mathematics Library* <http://project.dml.cz>

Stochastické modely v biochemii.

Je život jen náhoda?

Tomáš Vejchodský, Praha

V současné době jsme svědky nebývale prudkého rozvoje biochemie a molekulární biologie. Téměř denně čteme o pokroku v sekvencování DNA, o výzkumu kmenových buněk, o geneticky modifikovaných organizmech, o klonování a o řadě dalších dříve nemyslitelných výsledků. Nicméně naše porozumění toho, jak fungují živé organizmy na molekulární úrovni, je stále velmi omezené. Důvodem je množství, složitost a vzájemná provázanost biochemických pochodů, které se v živých buňkách odehrávají. Proto dodnes není jasný mechanismus působení mnohých léků, jako například paracetamolu, který patří k celosvětově nejrozšířenějším lékům proti horečce a bolesti a všichni ho užíváme.

Vývoj nových léků je oblast, která velmi trpí naší omezenou znalostí dynamických biochemických pochodů v živých organizmech. V dnešní době jen jedna z 5 000 až 10 000 molekul navržených jako potenciální léčivo projde úspěšně všemi stadii testů a začne se jako lék skutečně používat. Všechny předklinické i klinické testy úspěšného léčiva mohou stát až miliardu dolarů a trvají běžně i patnáct let. Vývoj nových léků je tedy velmi neefektivní a hlavním důvodem je skutečnost, že biochemickým pochodům v živých organizmech stále rozumíme jen velmi málo.

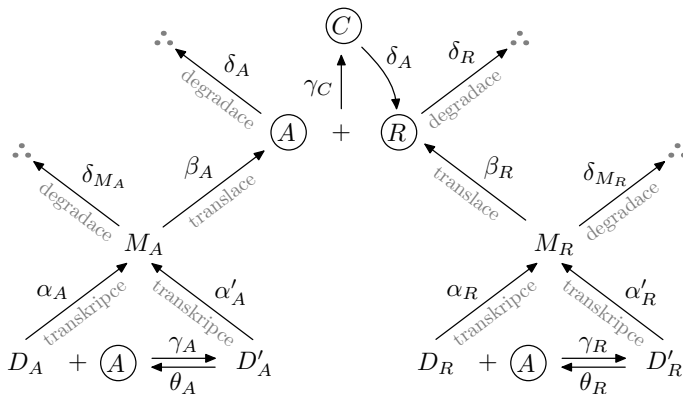
V roce 2001 byl přečten lidský genom [7], [8]. Odhaduje se, že obsahuje 20–25 tisíc genů. Většina genů kóduje nějaký protein (bílkovinu) a proteiny zajišťují téměř všechny funkce v těle. Znalost všech genů a příslušných proteinů dává vědcům katalog potenciálních léčiv. Nicméně bude trvat ještě dlouhou dobu, než se podaří tuto znalost zužitkovat. Přestože máme seznam všech genů, nevíme, jak spolu interagují a jaké z nich jsou důležité při kterých nemocech.

Jednotlivé geny spolu nepřímo interagují. Jeden gen produkuje protein, jenž má schopnost změnit rychlost transkripce jiného genu. Mluvíme o tzv. genových regulačních sítích, které mohou být velmi rozsáhlé a mít komplexní chování. Více informací o genových regulačních sítích je možné najít například v přehledu [1].

Obr. 1 ukazuje jednoduchý příklad takové regulační sítě. Jde o schéma teoretického modelu cirkadiánních (tj. zhruba denních) rytmů [9], který funguje díky zpětnovazebné regulaci dvou genů. Nicméně v reálných systémech uvnitř buněk navzájem interagují desítky až stovky různých genů. To naznačuje, jak obtížné je pochopit dynamické chování takto komplexních systémů. Je zřejmé, že bez pomoci matematického popisu, matematické analýzy a moderních výpočetních metod nemáme šanci tento úkol zvládnout.

Kromě rozsáhlosti a komplexity se u těchto systémů potýkáme i s významnými stochastickými efekty, způsobenými náhodným pohybem molekul. Některé biochemické

Doc. RNDr. TOMÁŠ VEJCHODSKÝ, Ph.D., Matematický ústav AV ČR, v. v. i., Žitná 25, 115 67 Praha 1, e-mail: vejchod@math.cas.cz



Obr. 1. Schéma biochemického modelu cirkadiálních rytmů z [9]

látky se totiž v buňkách vyskytují jen ve velmi malých koncentracích. Někdy jde jen o jednotky molekul a potom vůbec nemá smysl mluvit o koncentraci (např. DNA). Pokud je reagujících molekul tak málo, dochází k jejich srážkám zřídka a začnou se významně projevovat náhodné jevy. Stochastické efekty se v biochemických systémech skutečně pozorují a ukazuje se, že mohou mít významný vliv na dynamiku biochemických procesů a potažmo na chování celé buňky.

V tomto článku ukážeme, jaké matematické nástroje můžeme použít pro popis dynamického chování biochemických procesů, jak modelovat pozorované náhodné efekty a jaké důsledky tyto náhodné jevy mohou mít. Vše vysvětlíme na příkladu výše zmíněného modelu cirkadiálních rytmů.

1. Model cirkadiálních rytmů

Řada živých organismů mění své chování v rytmu s periodou zhruba 24 hodin. Typicky se střídá doba bdění a spánku. Tyto rytmy přetrvávají i když se okolní prostředí nemění. Např. lidé žijící v neustále osvětlené místnosti budou i nadále střídát bdění a spánek. Jde tedy o vnitřní cyklický mechanismus, který se označuje jako cirkadiální rytmus.

Obr. 1 schematicky zachycuje jednoduchý biochemický model takového mechanismu. Tento model je realisticky založen na základních biochemických pochodech. Šipky s rychlostními konstantami α naznačují transkripci genetické informace z DNA do RNA. Konkrétně uvažujeme dva geny (tj. úseky molekuly DNA) a označíme je D_A a D_R . Transkripcí genu D_A vzniká molekula mRNA (mediátorová nebo informační RNA) označená M_A . Z této mRNA vzniká pomocí translace protein, který označujeme jako A . O tomto proteinu předpokládáme, že jde o tzv. transkripční faktor. Ten má schopnost navázat se na tzv. promotor, což je úsek DNA před vlastním genem (reakce s rychlostní konstantou γ_A). Pokud se tak stane, rychlost transkripce se změní. V našem případě budeme uvažovat, že se zvýší, a proto budeme mluvit o proteinu A jako o aktivátoru. Gen D_A s navázaným aktivátorem A označíme jako D'_A a rychlost transkripce pak bude α'_A . Aktivátor se ovšem může od promotoru opět uvolnit a D'_A se rozpadne na molekuly D_A a A , což je reakce s rychlostní konstantou θ_A .

Stejné procesy probíhají i v případě genu D_R . Výsledný protein je označen jako R a budeme mu říkat represor. Všimněme si, že rychlost transkripce genu D_R může být zvýšená, pokud se na jeho promotor naváže aktivátor A . Navíc se aktivátor A a represor R mohou sloučit za vzniku komplexu C , což je reakce s rychlostní konstantou γ_C . Poslední důležitou vlastností je nestabilita všech zúčastněných molekul s výjimkou DNA. Tyto molekuly se samovolně rozpadají na součásti, o kterých předpokládáme, že už dále náš systém nijak neovlivňují. Výjimkou je rozpad komplexu C , při kterém se rozpadá jen složka A a zůstává molekula R . Rychlostní konstanty samovolných rozpadů jsou označeny jako δ s příslušným indexem.

Všechny tyto biochemické procesy modelujeme jako chemické reakce, které probíhají jistou rychlostí (intenzitou). Tyto rychlosti popisujeme rychlostními konstantami, které mají následující hodnoty, viz [9],

$$\begin{aligned}\alpha_A &= 50 \text{ h}^{-1}, & \alpha'_A &= 500 \text{ h}^{-1}, & \beta_A &= 50 \text{ h}^{-1}, & \delta_{MA} &= 10 \text{ h}^{-1}, & \gamma_A &= 1 \text{ Mol}^{-1} \text{ h}^{-1}, \\ \alpha_R &= 0.01 \text{ h}^{-1}, & \alpha'_R &= 50 \text{ h}^{-1}, & \beta_R &= 5 \text{ h}^{-1}, & \delta_{MR} &= 0.5 \text{ h}^{-1}, & \gamma_R &= 1 \text{ Mol}^{-1} \text{ h}^{-1}, \\ \theta_A &= 50 \text{ h}^{-1}, & \theta_R &= 100 \text{ h}^{-1}, & \delta_A &= 1 \text{ h}^{-1}, & \delta_R &= 0.2 \text{ h}^{-1}, & \gamma_C &= 2 \text{ Mol}^{-1} \text{ h}^{-1}.\end{aligned}$$

Symbol Mol označuje počet molekul a h je zkratka pro hodinu.

Dynamickým chováním soustav chemických reakcí se zabývá obor fyzikální chemie zvaný chemická kinetika. Na základě tzv. zákona akce hmoty můžeme snadno odvodit následující soustavu obyčejných diferenciálních rovnic pro množství zúčastněných chemikálií:

$$d\bar{D}_A/dt = \theta_A \bar{D}'_A - \gamma_A \bar{D}_A \bar{A}, \quad (1)$$

$$d\bar{D}'_A/dt = -\theta_A \bar{D}'_A + \gamma_A \bar{D}_A \bar{A}, \quad (2)$$

$$d\bar{D}_R/dt = \theta_R \bar{D}'_R - \gamma_R \bar{D}_R \bar{A}, \quad (3)$$

$$d\bar{D}'_R/dt = -\theta_R \bar{D}'_R + \gamma_R \bar{D}_R \bar{A}, \quad (4)$$

$$d\bar{M}_A/dt = \alpha'_A \bar{D}'_A + \alpha_A \bar{D}_A - \delta_{MA} \bar{M}_A, \quad (5)$$

$$d\bar{M}_R/dt = \alpha'_R \bar{D}'_R + \alpha_R \bar{D}_R - \delta_{MR} \bar{M}_R, \quad (6)$$

$$d\bar{A}/dt = \beta_A \bar{M}_A + \theta_A \bar{D}'_A + \theta_R \bar{D}'_R - \bar{A}(\gamma_A \bar{D}_A + \gamma_R \bar{D}_R + \gamma_C \bar{R} + \delta_A), \quad (7)$$

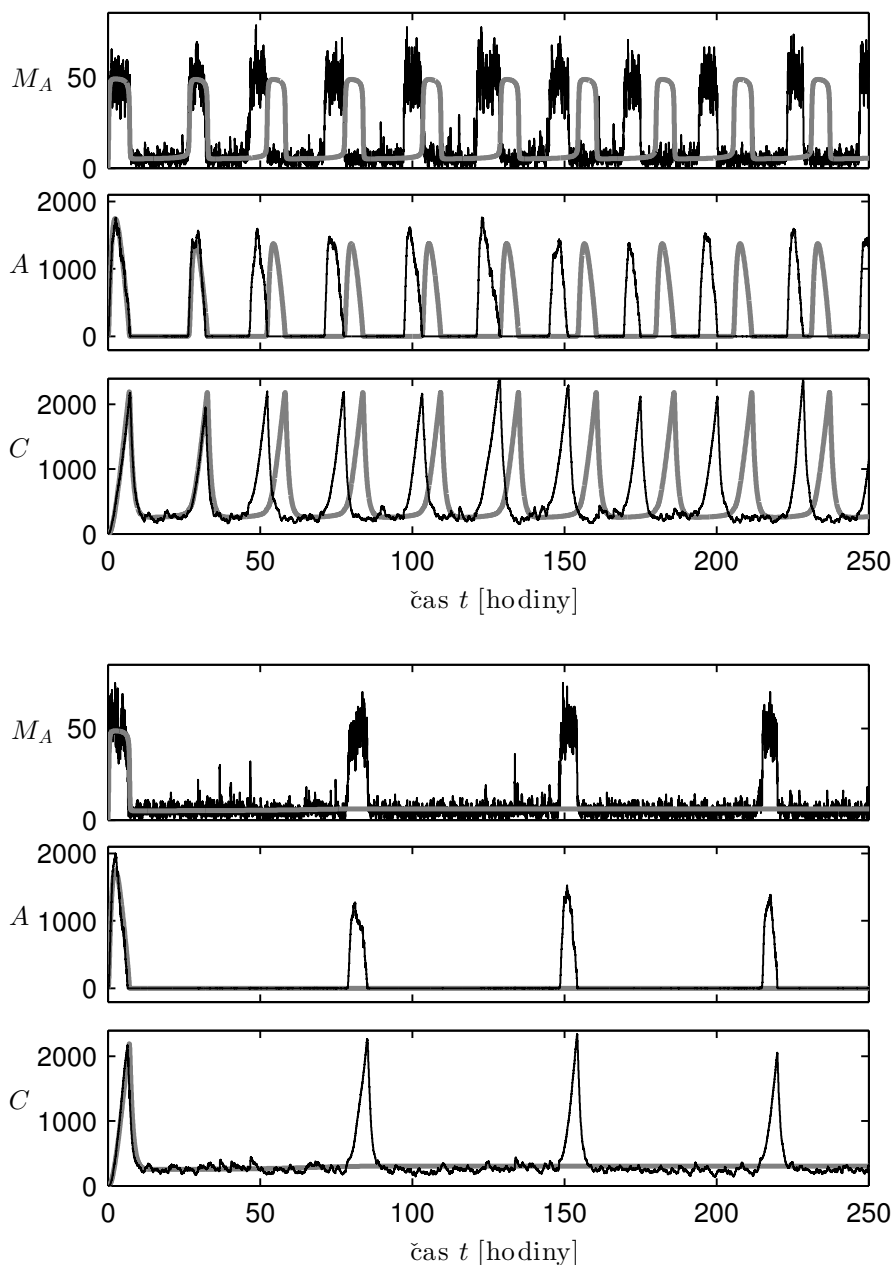
$$d\bar{R}/dt = \beta_R \bar{M}_R - \gamma_C \bar{A} \bar{R} + \delta_A \bar{C} - \delta_R \bar{R}, \quad (8)$$

$$d\bar{C}/dt = \gamma_C \bar{A} \bar{R} - \delta_A \bar{C}. \quad (9)$$

Velká písmena s pruhem označují počty molekul příslušných chemických látek jako funkce času t . Jde o idealizaci, kdy si představujeme, že hmota je nekonečně dělitelné kontinuum a počet molekul tedy uvažujeme i neceločíselný. Pro tuto soustavu předepíšeme nulové počáteční podmínky s výjimkou \bar{D}_A a \bar{D}_R , které budou mít počáteční hodnotu 1 Mol. Soustavu (1) až (9) vyřešíme pomocí standardních numerických metod pro obyčejné diferenciální rovnice. Výsledné řešení je zobrazeno na obr. 2 světlou čarou.

2. Stochastický model

Soustava (1) až (9) je deterministickým modelem biochemického systému z obr. 1. Jak už jsme naznačili v úvodu, některé látky se mohou v biochemických systémech



Obr. 2. Vývoj množství mediátorové RNA aktivátoru M_A , aktivátoru A a komplexu C v čase. Světlá čára odpovídá deterministickému modelu (1) až (9) a tmavá stochastické simulaci. Nahoře vidíme výsledky pro $\delta_R = 0.2$ a dole pro $\delta_R = 0.05$.

vyskytovat v tak malých množstvích, že začínají být významné jednotlivé molekuly a projevují se stochastické jevy. V takovém případě je přirozené sledovat všechny jednotlivé molekuly v systému, jejich brownovské (náhodné) trajektorie a řešit jejich srážky. Tyto tzv. mikroskopické modely molekulární dynamiky jsou velmi přesné, ale náročné na výpočetní kapacitu. Na běžných počítačích bychom těžko mohli modelovat systémy s tisíci molekulami po dostatečně dlouhou dobu. Proto teď vysvětlíme základní principy mezoskopických stochastických modelů, které sice neposkytují tak detailní informace, ale jejich výpočetní náročnost je ve srovnání s mikroskopickými modely podstatně menší.

Pro jednoduchost uvažujme jen slučování A a R za vzniku C , tedy chemickou rovnici



Mezoskopický stochastický model je diskrétní a symboly $A(t)$, $R(t)$ a $C(t)$ budou značit teď už celočíselné počty molekul látek A , R a C . Rychlostní konstanta γ_C v (10) je přímo úměrná pravděpodobnosti, že dvě pevně zvolené molekuly A a R zreagují podle (10). Nás ovšem bude zajímat pravděpodobnost, že libovolné dvě molekuly A a R zreagují podle (10). Tato pravděpodobnost je úměrná výrazu $\gamma_C A(t)R(t)$, kde součin $A(t)R(t)$ odpovídá počtu možných dvojic molekul A a R v systému. Pro názornost, uvažujme (nekonečně) malý časový krok dt . Ten musí být tak malý, aby pravděpodobnost, že v časovém intervalu $[t, t + dt)$ dojde v systému ke dvěma reakcím, byla zanedbatelná vůči pravděpodobnosti, že tam dojde jen k jedné reakci. Výraz $\gamma_C A(t)R(t)dt$ je pak roven pravděpodobnosti, že v časovém intervalu $[t, t + dt)$ dojde v systému k právě jedné reakci (10).

Pokud bychom chtěli být rigorózní, museli bychom uvažovat funkci $\varrho(A(t), R(t)) = \gamma_C A(t)R(t)$ jako hustotu této pravděpodobnosti. Definice s nekonečně malým časovým krokem sice není rigorózní, ale je názorná, a přímo nás vede k algoritmu pro simulování tohoto náhodného (v tomto případě Markovova) procesu.

1. Zvolme počáteční čas t , počáteční stav $A(t)$ a $R(t)$ a dostatečně malý časový krok Δt .
2. Vygenerujme náhodné číslo ξ stejnoměrně rozdělené v intervalu $[0, 1)$.
3. Pokud je $\xi < \gamma_C A(t)R(t)\Delta t$, v systému nastala reakce (10) a jeho stav se změnil na $A(t + \Delta t) = A(t) - 1$, $R(t + \Delta t) = R(t) - 1$ a $C(t + \Delta t) = C(t) + 1$. Jinak stav systému zůstane nezměněný, tj. $A(t + \Delta t) = A(t)$, $R(t + \Delta t) = R(t)$ a $C(t + \Delta t) = C(t)$.
4. Posuňme čas, tj. $t := t + \Delta t$, a pokračujme podle kroku 2.

Podobným způsobem lze modelovat všech 16 biochemických reakcí naznačených na obr. 1. Výsledné stochastické trajektorie jsou zobrazeny na obr. 2 tmavou čarou. Samozřejmě různé realizace tohoto náhodného procesu dávají různé trajektorie, a proto stochastický model na rozdíl od deterministického nemůže předpovědět, v jakém stavu se bude systém nacházet v daném čase. K čemu nám je tedy takový model dobrý? Ano, náhodná trajektorie může skončit prakticky kdekoli a každé spuštění stochastického modelu dá jiný výsledek. Nicméně, pokud vypočteme velké množství náhodných trajektorií, můžeme tato data statisticky zpracovat a získat údaje o *průměrném* chování systému, a to už je reprodukovatelné. Například, v uvedeném modelu cirkadiálních

rytmů je zajímavou veličinou průměrná perioda pozorovaných náhodných oscilací. Bohužel, pro výpočet průměrného chování s dobrou přesností potřebujeme provést velmi mnoho náhodných realizací, a to je v případě rozsáhlejších systémů výpočetně náročné i když se použijí mezoskopické stochastické modely.

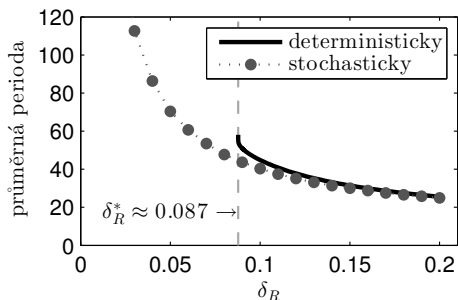
Naopak výhodou je, že tyto výpočty lze snadno paralelizovat. Navíc, existuje tzv. Gillespieho algoritmus [4], který je ekvivalentní výše uvedenému algoritmu, ale je podstatně efektivnější. Kromě toho můžeme pravděpodobnost, že systém najdeme v daném stavu, vypočítat i bez stochastických simulací. Tato pravděpodobnost je totiž dána tzv. základní chemickou rovnicí, což je speciální případ Kolmogorovovy rovnice a obecně to je nekonečná soustava obyčejných diferenciálních rovnic. Avšak ve většině případů nejsme schopni tuto soustavu vyřešit. Analytická řešení jsou známa jen ve výjimečných případech a numerické metody zvládnou jen úlohy s relativně malým stavovým prostorem. Proto se studují přibližné metody, které umožní výpočet výrazně urychlit za tu cenu, že výsledek odpovídá výše definovanému náhodnému procesu jen přibližně. Zmíňme τ -leaping [6] jako metodu pro rychlejší stochastické simulace a chemickou Fokkerovu–Planckovu rovnici [5] pro přibližný výpočet pravděpodobnosti, že se systém nachází v daném stavu. Pro bližší seznámení s těmito zajímavými postupy lze doporučit např. text [3].

3. Oscilace indukované náhodným šumem

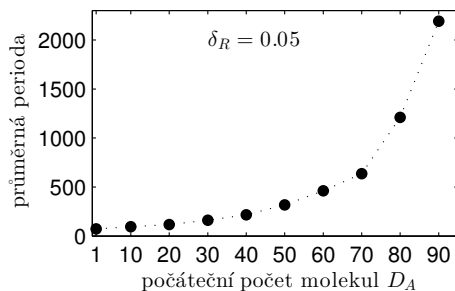
Vraťme se ovšem k příkladu cirkadiánních rytmů. Vysvětlili jsme, jak získat trajektorie zobrazené na obr. 2. Horní část tohoto obrázku porovnává chování deterministického a stochastického modelu pro výše uvedené hodnoty parametrů. Vidíme, že oba modely se chovají podobně a vykazují oscilace s periodou přibližně 24 hodin. Trajektorie stochastického modelu vypadají jen jako trochu zašuměné deterministické trajektorie. Také si můžeme všimnout, že průměrná perioda stochastického modelu je o něco málo kratší než perioda deterministického modelu.

V tomto kontextu je zajímavé sledovat, jak se oba modely chovají, pokud měníme některý z parametrů. Pro příklad si zvolíme parametr $\delta_R = 0.2$ a budeme ho postupně zmenšovat. Zjistíme, že perioda oscilací obou systémů se tím bude postupně zvětšovat až do jisté kritické hodnoty $\delta_R^* \approx 0.087$, kdy deterministický model oscilovat přestane. Pro hodnotu parametru $\delta_R = \delta_R^*$ nastává v deterministickém systému tzv. bifurkace, při které se oscilující řešení mění na řešení vyvíjející se do stacionárního stavu. Poznamenejme, že v [2] můžeme najít analýzu podobného případu, ale pro jiný typ bifurkace.

Překvapující je, že stochastický model na přechod přes kritickou hodnotu δ_R^* nijak nereaguje. Nadále si zachovává oscilující řešení a perioda těchto oscilací se plynule zvětšuje se zmenšujícím se δ_R . Situace pro $\delta_R = 0.05$ je zachycena na obr. 2 dole. Deterministický model neosciluje a jeho řešení se blíží stacionárnímu stavu, zatímco stochastický model osciluje s průměrnou periodou přibližně 70 hodin. Je tedy zřejmé, že oscilace stochastického modelu nejsou způsobené příslušnou chemickou kinetikou, protože tu dobře zachycuje i deterministický model a ten neosciluje. Jediná možnost tedy je, že stochastický model osciluje jen díky náhodnému šumu. Jeho oscilace jsou indukované šumem a jde o čistě stochastický efekt, který deterministický model nemůže postihnout.



Obr. 3. Průměrná perioda oscilací v závislosti na parametru δ_R pro stochastický a deterministický model



Obr. 4. Průměrná perioda oscilací v závislosti na velikosti systému pro stochastický model a $\delta_R = 0.05$

Pro lepší představu o chování obou systémů při měnícím se parametru δ_R uvádíme obr. 3, na kterém vidíme závislost periody oscilací na parametru δ_R . Pro deterministický model je tato perioda dobře definovaná. Ve stochastickém případě ji počítáme jako průměrnou hodnotu z dlouhé stochastické simulace (simulujeme alespoň 500 period). Vidíme, že stochastický model si zachovává oscilační charakter pro širokou škálu hodnot parametru δ_R , zatímco deterministický model osciluje jen pro $\delta_R > \delta_R^*$. Z tohoto hlediska jsou oscilace stochastického modelu výrazně robustnější než oscilace deterministického modelu. Proto si můžeme dovolit neintuitivní tvrzení, že náhodný šum může zvyšovat robustnost biochemických systémů.

Na závěr se ještě podíváme, jak průměrná perioda oscilací ve stochastickém modelu závisí na velikosti systému. Intuitivně je zřejmé, že pokud budeme stochastický systém zvětšovat, tj. budeme zvětšovat objem reaktoru a tím zvyšovat počet interagujících molekul, tak budou stochastické efekty slábnout a chování stochastického modelu se bude blížit deterministickému. Toto tvrzení lze i rigorózně dokázat, protože deterministický model je limitou mikroskopického stochastického popisu. Proto bychom očekávali, že v případě $\delta_R = 0.05$, kdy deterministický systém neosciluje a jeho perioda je tedy nekonečná, bude perioda stochastického systému se zvětšující se velikostí systému růst k nekonečnu. Obr. 4 ilustruje přesně tuto situaci a potvrzuje naše očekávání.

4. Závěr

Představili jsme stochastický model cirkadiánních rytmů, který osciluje jen díky náhodným fluktuacím. A navíc díky jim jsou tyto oscilace robustní vzhledem ke změnám parametrů. Tento model ukazuje, jak významnou roli mohou stochastické jevy hrát v biochemických procesech uvnitř živých buněk, a v důsledku i v celých organizmech.

Je tedy život jen náhoda? Ano i ne. Biochemické pochody jsou sice ovlivňovány náhodným šumem a jeho projevy mohou být významné, nicméně i tyto jevy je možné matematicky popsat. Jejich analýzou můžeme pochopit, jak fungují, a zjistit, jak je ovlivňovat a regulovat. Biochemické procesy v živých organizmech jsou náhodné, ovšem náhoda je to popsatelná a předvídatelná.

Tento článek je jen krátkou ukázkou matematických metod pro studium rozsáhlých a komplexních biochemických systémů. Je to relativně nová oblast aplikované matematiky, která má potenciál v kombinaci s moderními výpočetními prostředky odhalit tajemství biochemických pochodů v živých organizmech, což by mělo nesmírně pozitivní důsledky především v lékařství. Pak bychom totiž mohli navrhovat nové léky cíleně a s velkou jistotou, že nemají vedlejší účinky. Byl by to dramatický pokrok při hledání nových léků a znamenalo by to revoluci v medicíně. Představa je to nesmírně vzrušující, nicméně úkol je to velmi náročný a přes veškerý pokrok, který jsme za poslední roky udělali, jsme stále ještě na začátku. Zásadní objevy v této oblasti nás teprve čekají.

Poděkování. Autorův výzkum byl podpořen z prostředků programu People (Marie Curie Actions) sedmého rámcového programu Evropské unie (FP7/2007-2013), grant č. 328008, a také z projektu RVO 67985840.

L i t e r a t u r a

- [1] DAVIDSON, E., LEVIN, M.: *Gene regulatory networks*. Proc. Natl. Acad. Sci. USA *102* (14) (2005), 4935–4935.
- [2] ERBAN, R., CHAPMAN, S. J., KEVREKIDIS, I. G., VEJCHODSKÝ, T.: *Analysis of a stochastic chemical system close to a SNIPER bifurcation of its mean-field model*. SIAM J. Appl. Math. *70* (2009), 984–1016.
- [3] ERBAN, R., CHAPMAN, S. J., MAINI, P. K.: *A practical guide to stochastic simulations of reaction-diffusion processes*. Preprint ArXiv:0704.1908 (2007), dostupné z: <http://arxiv.org/abs/0704.1908>
- [4] GILLESPIE, D.: *Exact stochastic simulation of coupled chemical reactions*. J. Phys. Chem. *81* (1977), 2340–2361.
- [5] GILLESPIE, D.: *The chemical Langevin equation*. J. Chem. Phys. *113* (2000), 297–306.
- [6] GILLESPIE, D.: *Approximate accelerated stochastic simulation of chemically reacting systems*. J. Chem. Phys. *115* (2001), 1716–1733.
- [7] LANDER, E. S., et al.: *Initial sequencing and analysis of the human genome*. Nature *409* (6822) (2001), 860–921.
- [8] VENTER, J. C., et al.: *The sequence of the human genome*. Science *291* (5507) (2001), 1304–1351.
- [9] VILAR, J. M. G., KUEH, H. Y., BARKAI, N., LEIBLER, S.: *Mechanisms of noise-resistance in genetic oscillators*. PNAS *99* (2002), 5988–5992.