

Univerzita Komenského v Bratislave  
Fakulta matematiky, fyziky a informatiky

---

## **Odhadovanie miery samo-rekombinácie v koncoch DNA**



Univerzita Komenského v Bratislave  
Fakulta matematiky, fyziky a informatiky  
Katedra aplikovanej matematiky a štatistiky

---

Evidenčné číslo: ef191c50-ae53-424f-9b61-e9ba1305df1b

# Odhadovanie miery samo-rekombinácie v koncoch DNA

Bakalárska práca

Študijný odbor: 9.1.9. Aplikovaná matematika  
Študijný program: Ekonomická a finančná matematika  
Vedúca bakalárskej práce: Mgr. Katarína Boďová, PhD.



## ZADANIE ZÁVEREČNEJ PRÁCE

**Meno a priezvisko študenta:** Jozef Jakubík  
**Študijný program:** ekonomická a finančná matematika (Jednoodborové štúdium, bakalársky I. st., denná forma)  
**Študijný odbor:** 9.1.9. aplikovaná matematika  
**Typ záverečnej práce:** bakalárska  
**Jazyk záverečnej práce:** slovenský

**Názov :** Odhadovanie miery samo-rekombinácie v koncoch DNA

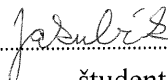
**Cieľ :** Lineárne konce DNA zvyknú vykazovať náhodný pohyb. Ak sa však dostanú do blízkosti tzv hot-spotu (iná časť tej istej DNA, s ktorou sú schopné reagovať), potom nastane ich vzájomná rekombinácia. V práci budeme za pomoci diskretných algoritmov numericky študovať, ako rýchlo táto reakcia nastáva, v závislosti od tuhosti DNA, teploty, resp. vzdialenosti medzi koncom a hot-spotom. Alternatívne sa dá tento problém predstaviť ako pohyb guľôčky v energetickom potenciáli.


**Vedúci :** Mgr. Katarína Bod'ová, PhD.

**Dátum zadania:** 27.10.2010

**Dátum schválenia:** 08.11.2010

.....  
doc. RNDr. Margaréta Halická, CSc.  
garant študijného programu

.....  
  
študent

.....  
  
vedúci práce

Dátum potvrdenia finálnej verzie práce, súhlas s jej odovzdaním (vrátane spôsobu sprístupnenia)

.....  
vedúci práce

## Čestné vyhlásenie

Vyhlasujem, že som bakalársku prácu vypracoval samostatne s použitím uvedenej odbornej literatúry.

Bratislava 2. 6. 2011

.....

*Jozef Jakubík*

## **Podakovanie**

Chcem využiť tento priestor a vysloviť úprimné podakovanie svojej vedúcej záverečnej práce, Mgr. Kataríne Bodovej, PhD. za odborné rady a mnohé podnetné pripomienky k práci.

## **Abstrakt**

Jakubík Jozef, Odhadovanie miery samo-rekombinácie v koncoch DNA, Univerzita Komenského v Bratislave, Fakulta matematiky fyziky a informatiky, Katedra aplikovanej matematiky a štatistiky, vedúca práce Mgr. Katarína Boďová, PhD., Bratislava, 2011, 35 strán.

Konce DNA, nazývané aj teloméry, sú špecifické svojou stavbou a funkciou v bunke. V niektorých prípadoch na nich môže prebiehať proces samo-rekombinácie, ktorý sa podieľa na ich obnovovaní. Cieľom bakalárskej práce je skúmanie vplyvu jednotlivých faktorov na priemerný čas potrebný na samo-rekombináciu koncov DNA. Skúmané faktory sú maximálna vzdialenosť hotspotov, tvar potenciálu, teplota a viskozita prostredia.

## **Kľúčové slová**

teloméra, Fokker-Planckova rovnica, first-exit čas, náhodný pohyb v potenciáli

## **Abstract**

Jakubík Jozef, The Estimation of Rate of Self-Recombination in thr End of DNA , Come-  
nius University, Bratislava, Faculty of Mathematics, Physics and Informatics, Department  
of Applied Mathematics and Statistics, supervisor Mgr. Katarína Boďová, PhD., Bratislava,  
2011, 35 pages.

DNA ends, also known as telomeres, have specific structure and function in the cell. In some cases they may undergo the process of self-recombination, which is involved in their main-  
tenance. The aim of the bachelor thesis is to investigate the impact of individual factors  
on the average time needed for the emergence of self-recombination of DNA ends. Investi-  
gated factors are the maximum distance of hotspots, the shape of potential, temperature  
and viscosity of the environment.

## **Key words**

telomere, Fokker-Planck equation, first-exit time, random walk in potential

## Obsah

<b>Úvod</b>	<b>8</b>
<b>1 Teloméry</b>	<b>9</b>
1.1 Telomerické krúžky . . . . .	10
1.2 Hotspot . . . . .	11
<b>2 Model</b>	<b>12</b>
2.1 Fokker-Planckova rovnica . . . . .	13
2.2 Stochastický algoritmus na výpočet first-exit času . . . . .	14
2.3 Stochastický algoritmus . . . . .	16
<b>3 Analytický výpočet first-exit času</b>	<b>18</b>
<b>4 Závislosť času do rekombinácie od jednotlivých faktorov</b>	<b>20</b>
4.1 Závislosť času do rekombinácie od tvaru potenciálu . . . . .	20
4.2 Závislosť času do rekombinácie od teploty a viskozity prostredia . . . . .	28
4.3 Závislosť času do rekombinácie od maximálnej možnej vzdialenosti hotspotov . . . . .	31
<b>5 Záver</b>	<b>33</b>



## Úvod

Cieľom práce je matematické skúmanie mechanizmov spojených s predlžovaním a skracovaním telomér — nekódujúcich repetícií DNA na koncoch chromozómov. Teloméry [7] sú dôležité pri replikácii DNA, chránia kódujúce úseky DNA a bránia nevhodným DNA opravám. Pri delení buniek sa teloméry skracujú v dôsledku problému replikácie koncov chromozómov. Existuje viacero spôsobov predlžovania telomér. V tejto práci budem skúmať alternatívny mechanizmus ich predlžovania pomocou telomerických krúžkov (t-circles), ktoré vznikajú po oddelení lasovitých a slučkovitých štruktúr z telomér. Telomerické slučky vznikajú, ak sa dve miesta na telomére, ktoré sú náchylné na rekombináciu (tzv. hotspoty), dostanú do dostatočnej blízkosti a zrekombinujú. V práci nebudem skúmať vznik lasovitých štruktúr, ktoré vznikajú odlišným spôsobom. Telomerické krúžky môžu voľne difundovať, replikovať sa a spätne sa pripojiť k telomére, čím ju predĺžia [8, 9, 6].

V práci budem modelovať náhodný pohyb telomerických hotspotov na tom istom reťazci a skúmať, s akou pravdepodobnosťou sa dostanú do dostatočnej blízkosti, aby mohli zrekombinovať. Model teloméry je zjednodušený na dva hotpoty, z ktorých jeden má zafixovanú polohu a druhý mení relatívnu polohu vzhľadom na pohyb prvého hotspotu. Tuhosť telomér je závislá od vzdialenosti hotspotov, krátka teloméra je tuhšia a horšie sa ohýba. Dlhšia teloméra je síce ľahšie ohybná, ale je menšia pravdepodobnosť, že sa hotpoty stretnú. Môžeme si to predstaviť ako náhodný pohyb guľôčky v energetickom potenciáli.

Pri modelovaní použijem stochastický algoritmus [1] založený na diskretizácii Fokker-Plankovej rovnice [4] na konečnom intervale dĺžky rovnej maximálnej nožnej vzdialenosti medzi hotpotmi na telomére. Pravdepodobnosti presunu do susedných bodov závisia od potenciálu, v ktorom sa hotpoty pohybujú.

## 1 Teloméry

Dvojjávitnica DNA nesie genetickú informáciu dôležitú na správnu funkciu a prežitie bunky, ako aj celého organizmu. V DNA sú dva typy sekvencií, kódujúce — úseky, ktoré nesú genetickú informáciu potrebnú na syntézu proteínov nevyhnutných pre život bunky a nekódujúce — úseky s rôznou funkciou, často neznámou. Teloméry [7] sú nekódujúce úseky na koncoch DNA, ktoré sa skladajú z viacerých rovnakých tandemových repetícií nukleotidov. Počet repetícií sa môže líšiť na každej telomére. Počet nukleotidov na jednu repetíciu sa mení v závislosti od organizmu. Teloméry majú tri základné funkcie:

- chránia kódujúce časti DNA pred poškodením enzýmami,
- riešia problém replikácie koncov chromozómov, ktorý vzniká, lebo úsek materskej DNA na mieste, kde sa odbúra primer (úsek RNA, ktorý zohráva dôležitú úlohu pri replikácii DNA) sa už nedosyntetizuje, a preto je dcérska DNA kratšia ako pôvodná materská DNA,
- bránia nevhodným DNA opravám.

Poškodenie spôsobené enzýmami či nevhodnou opravou, prípadne skrátenie kódujúcej časti, môže viesť k nežiadúcej či nebezpečnej genetickej mutácii v bunke. Aby bunka predišla nežiadúcej genetickej mutácii, ktorou by mohla poškodiť celý organizmus, tak sú pre ňu skrátené teloméry signálom, aby podstúpila bezpečnú programovanú bunkovú smrť (apoptózu). Pre niektoré organizmy a typy buniek je ale potrebný vyšší počet bunkových delení ako dovoľuje dĺžka teloméru. Pre tieto bunky je počet delení dôležitejší ako prípadné poškodenie mutáciou, prípadne je u nich znížené riziko výskytu mutácií. Napríklad v ľudských embryonálnych bunkách sa vyskytuje telomeráza, lebo sa potrebujú intenzívne deliť a majú znížené riziko mutácie, pričom v bunkách dospelého jedinca sa už telomeráza nenachádza. V závislosti od organizmu a typu bunky sa teloméry môžu predlžovať rôznymi mechanizmami, prípadne sa môže uplatňovať viac mechanizmov aj v jednej bunke.

Mechanizmy predlžovania:

- telomeráza [7] (enzým syntetizujúci telomerické repetície)

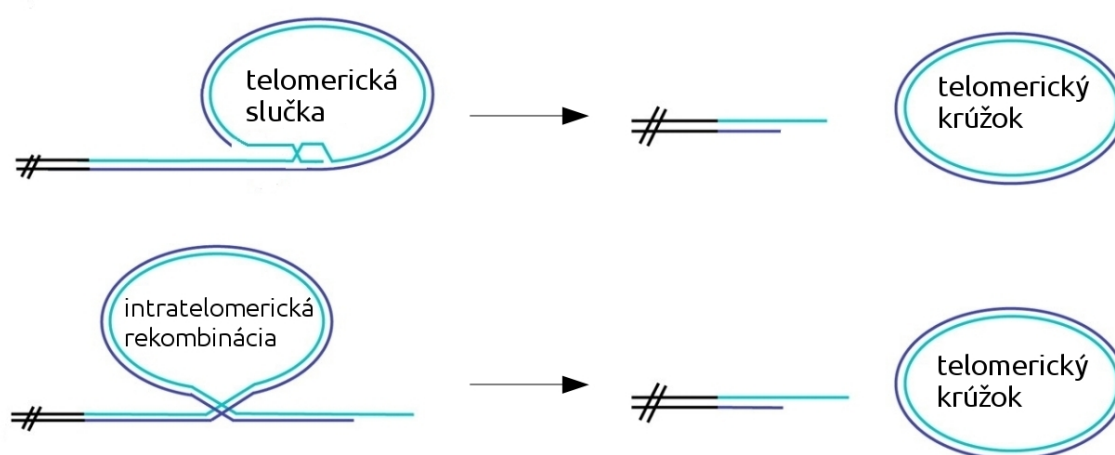
- alternatívne spôsoby [8] (homologická rekombinácia telomér, replikácia valivou kružnicou, replikácia pomocou telomerických krúžkov, ...)

Nie každý druh buniek obsahuje telomerázu, preto je dôležité zaoberať sa aj alternatívnym spôsobom predlžovania telomér. V práci budem študovať jeden z alternatívnych mechanizmov, a to predlžovanie pomocou telomerických krúžkov.

Ako **modelový organizmus** pri výskume telomerických krúžkov sa často používa kvasinka *Candida parapsilosis*, ktorá má lineárnu minochondriálnu DNA. V jej mitochondriách sa nenachádza telomeráza a ich teloméry sa môžu predlžovať iba alternatívnymi spôsobmi. *C. parapsilosis* má do dvanásť repetícií, každú dĺžky 738 bázových párov.

## 1.1 Telomerické krúžky

Telomerický krúžok (t-circle) [8, 6, 9] môže vzniknúť z lineárnej teloméry oddelením slučkovitej štruktúry od pôvodnej teloméry. Takáto slučkovitá štruktúra vzniká intratelomerickou rekombináciou (vo vnútri teloméry) alebo z telomerickéj slučky (t-loop), ktorá sa tvorí na voľnom konci teloméry.



Obr. 1–1 Vznik telomerických krúžkov.

Týmto procesom sa pôvodná teloméra skráti a vzniknutý telomerický krúžok sa môže pripojiť k inej telomére, prípadne sa môže replikovať mechanizmami, ktorými sa nemôže replikovať pôvodná teloméra. Nareplikované telomerické krúžky sa môžu opäť pripojiť k teloméram a rozvinúť sa, čím teloméru predĺžia. Vznik telomerických krúžkov bol pozorovaný v rôznych experimentoch na kvasinkách *Candida parapsilosis*, *Pichia philodendra* a *Candida salmanticensis* s rôznou veľkosťou [9], pričom počet básových párov je vždy celočíselným násobkom počtu básových párov telomerickéj repetície.

## 1.2 Hotspot

Hotspot je miesto na DNA, kde môže prebiehať rekombinácia. Keď sa dostanú dva hotspoty dostatočne blízko k sebe, môže prebehnúť rekombinácia, či už hotspoty prislúchajú k jednému alebo dvom vláknam DNA. Následne sa hotspoty opäť rozdelia, čo môže viesť pri spojení dvoch vlákien DNA k výmene úsekov pred a za hotspotom, čím sa môže zmeniť dĺžka výsledných vlákien. Pri samorekombinácii jedného vlákna môže vzniknúť telomerický krúžok a skrátená teloméra. Predpokladá sa, že hotspot sa vyskytuje na DNA v tých miestach, kde je viac A-T párov ako G-C párov, pretože dvojité vodíkové väzby medzi adenínom a tymínom je slabšia ako trojitá vodíková väzba medzi guanínom a cytozínom. V takomto slabšom mieste sa môže dvojzávitnica ľahšie rozdeliť. Keďže teloméry sa skladajú z viacerých rovnakých repetícií, tak ak je hotspot na niektorej repetícii, potom musí byť na každej repetícii.

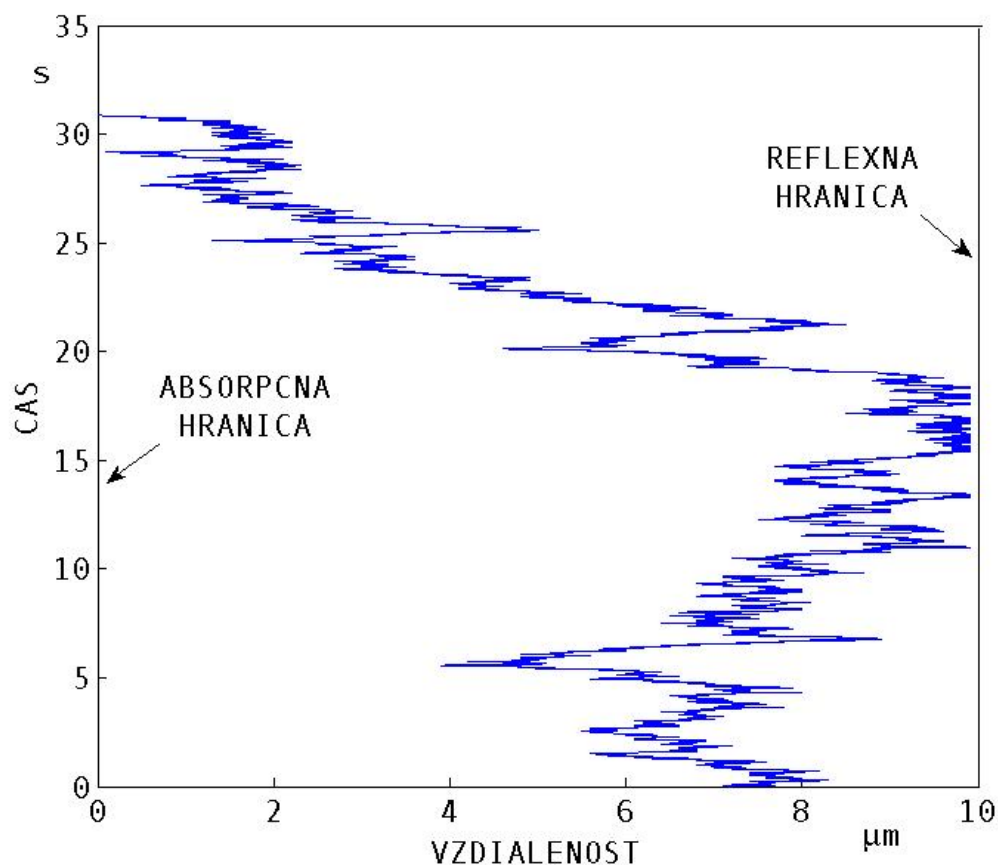
Teloméry sa pohybujú a ohýbajú v cytoplazme náhodným pohybom prostredníctvom difúzie a dôsledkom elastických stresov. Ich pohyb môže byť ovplyvnený proteínmi, naviazanými na kosť DNA, ktoré môžu určovať preferované skrútenia telomér. V súčasnosti sa tomuto problému venuje veľká pozornosť — wormlike chain model a J-faktory [3, 10]. Pohyblivosť telomér taktiež závisí od stavby telomér, teploty a viskozity okolitého prostredia. Ak sa majú dostať dva hotspoty jednej DNA dostatočne blízko na to, aby mohli zrekombinovať, musia tieto hotspoty prekonať určitý potenciál, ktorý je závislý od stavby teloméry a od okolitého prostredia.

## 2 Model

V práci budem modelovať pohyb hotspotov na tej istej telomére. Model bude zjednodušením reality a vo výsledku sa vzdiali od zložitých biologických procesov v bunke. Uvažované zjednodušenia:

- Namiesto pohybu oboch hotspotov budeme uvažovať iba pohyb jedného hotspotu vzhľadom na pohyb druhého zafixovaného hotspotu.
- Teloméra sa nachádza vo voľnom prostredí, teda neinteraguje so žiadnou inou štruktúrou.
- Na pohyb teloméry majú vplyv len difúzny a driftový faktor.
- Odhliadnutie od dimenzionality problému a uvažovanie pohybu len v jednej dimenzii — vzdialenosť hotspotov ( $R$ ). Podobne by sa dal preformulovať aj problém s radiálne symetrickým potenciálom.
- Predpoklad okamžitej rekombinácie po tom, ako sa hotspoty dostanú dost blízko. Dala by sa implementovať aj konečná miera reaktivity.

Voľný hotspot sa môže pohybovať po úsečke, ktorej dĺžka je maximálna možná vzdialenosť oboch hotspotov. Viac sa voľný hotspot vzdialiť nemôže, lebo teloméra sa nemôže predĺžiť a hotspoty sa nemôžu presunúť na iné miesto na telomére. V jednom konci úsečky bude zafixovaný hotspot. Ak sa voľný hotspot dostane do maximálnej vzdialenosti (teda na druhý koniec úsečky), nemôže už pokračovať vo vzdalovaní, a tak sa odrazí späť ( $R = L$ ,  $L$  je maximálna možná vzdialenosť). Ak sa dostane dostatočne blízko k zafixovanému hotspotu ( $R = 0$ ), tak hotspoty zrekombinujú a môže vzniknúť telomerický krúžok. Uvažujeme pohyb v potenciáli po úsečke s absorpčnou okrajovou podmienkou na jednej strane a reflexnou na strane druhej. Voľný hotspot sa bude pohybovať po diskretizovanom priestore, ktorý nahradí spojitý pohyb.



**Obr. 2–2** Vývoj polohy voľného hotspotu v čase, pričom vzdialenosť  $10 \mu\text{m}$  je diskretizovaná 100 deliacimi bodmi a v nultom bode sa nachádza zafixovaný hotspot.

Pravdepodobnosť presunu z jedného bodu do susedného bodu sa určí na základe potenciálu, v ktorom sa hotspoty pohybujú. Vo výsledku nás bude zaujímať, za aký priemerný čas sa dostanú oba hotspoty dostatočne blízko na to, aby mohla prebehnúť rekombinácia (tzv. first-exit time problém) a ako tento čas závisí od rôznych faktorov. Model vychádza z článku [1] a je postavený na Fokker-Planckovej rovnici.

## 2.1 Fokker-Planckova rovnica

Fokker-Planckova [4] rovnica, tiež známa ako Kolmogorova dopredná rovnica, popisuje vývoj hustoty pravdepodobnosti v čase. Je to parciálna diferenciálna rovnica popisujúca proces

zahŕňajúci difúziu a drift. V jednorozmernom prípade má tvar:

$$\begin{aligned}\frac{\partial}{\partial t}P(x, t) &= -\frac{\partial}{\partial x}[V(x, t)P(x, t)] + \frac{\partial^2}{\partial x^2}[D(x, t)P(x, t)] \\ P(x, t_0) &= \delta(x - x_0) \\ x &\in \mathbb{R},\end{aligned}\tag{2.1-1}$$

kde  $P(x, t)$  je hustota pravdepodobnosti,  $V(x, t)$  driftový koeficient a  $D(x, t)$  difúzny koeficient. Používa sa na výpočet funkcie hustoty stochastického procesu popísaného stochastickou diferenciálnou rovnicou

$$dX_t = f(t, X_t)dt + g(t, X_t)dW_t,\tag{2.1-2}$$

pričom  $f$  popisuje drift,  $g$  popisuje difúziu a  $W_t$  je Wienerov proces.

## 2.2 Stochastický algoritmus na výpočet first-exit času

V našom prípade nás bude zaujímať čas, za ktorý sa dostanú dva hotspotsy dostatočne blízko na rekombináciu. Predpokladáme, že voľný hotspot je ovplyvnený difúziou v bunke a náhodne sa pohybuje v potenciáli, ktorý mu predurčuje stavba teloméry.

Budeme vychádzať z Fokker-Planckovej rovnice tvaru

$$\begin{aligned}\frac{\partial \rho}{\partial t} &= \frac{\partial v'(x)\rho}{\partial x} + D\frac{\partial^2 \rho}{\partial x^2} \\ \rho(x, t_0) &= \delta(x - x_0) \\ \rho(0, t) &= 0 \\ J(L, t) &= 0,\end{aligned}\tag{2.2-1}$$

kde  $\rho$  je hustota pravdepodobnosti,  $v(x)$  je potenciál určený stavbou teloméry,  $D$  je difúzny koeficient, pre ktorý platí Einstein-Stokesova rovnica  $D = K_B T / 6\pi\eta r$ , pričom  $K_B$  je Boltzmanova konštanta,  $T$  je absolútna teplota,  $\eta$  je dynamická viskozita prostredia a  $r$  je polomer častice (hotspotu) a  $J = -v'(x)\rho - D\partial\rho/\partial x$  označuje pravdepodobnostný tok. Vypočítať priamo hustotu pravdepodobnosti  $\rho$  je často pre komplikované potenciály  $v(x)$  netriviálne. Na numerický výpočet strednej hodnoty však nepotrebujeme poznať priamo hustotu pravdepodobnosti. Stačí, ak budeme poznať dostatočne veľa realizácií daného procesu, z ktorých

vieme určiť strednú hodnotu skoro presne. Jednotlivé realizácie stochastického procesu, ktorý popisuje rovnica 2.2-1, sú produktom Langevinovej rovnice

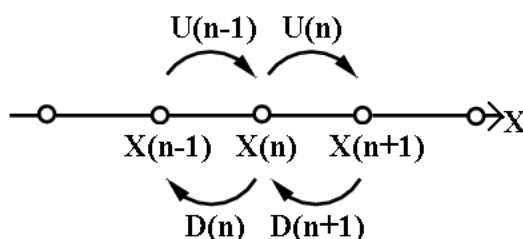
$$\begin{aligned} dx &= -v'(x)dt + (2D)^{1/2}dW \\ x(0) &= x_0, \end{aligned} \quad (2.2-2)$$

kde  $W$  je Wienerov proces. Na reflexnej hranici sa hotspot odrazí a v absorpčnej hranici proces končí. Dobrý prístup, ako numericky simulovať proces, ktorý popisuje rovnica 2.2-2, je diskretizovať čas a použiť Eulerov typ algoritmu na popis napredovania z miesta  $x(t_n)$  do miesta  $x(t_{n+1})$ .

$$x(t_{n+1}) = x(t_n) - \Delta t v'(x_n) + (2D\Delta t)^{1/2} \xi_n, \quad (2.2-3)$$

kde  $\Delta t$  je časový krok a  $\xi_n \sim N(0, \Delta t)$  je náhodná premenná, vyberaná nezávisle pre každý krok.

Namiesto diskretizácie času sa dá odvodiť diskretizácia závislej premennej  $x$ . Tým sme prerobili problém na skokovitý problém.



**Obr. 2–3** Skokový problém aproximujúci jednotlivé realizácie stochastického procesu.

V článku [1] je odvodený čas, za ktorý preskočíme z pozície  $x_n$  do pozície  $x_{n+1}$  alebo  $x_{n-1}$ . Situáciu na obrázku 2–3 popisuje rovnica

$$\frac{dP_n}{dt} = U_{n-1}P_{n-1} - D_nP_n + D_{n+1}P_{n+1} - U_nP_n, \quad (2.2-4)$$

pričom  $P_n$  je pravdepodobnosť, že v čase  $t$  sa nachádzame v bode  $x_n$ ,  $U_n$  je miera, s akou sa častice presúvajú z  $x_n$  do  $x_{n+1}$  a  $D_n$  je miera, s akou sa častice presúvajú z  $x_n$  do  $x_{n-1}$ . Miery prechodu medzi stavmi náhodného procesu ( $P_n$  a  $U_n$ ) sú v [1] zvolené tak, aby bola



Fokker-Planckova rovnica splnená s presnosťou druhého rádu vzhľadom na priestorový krok a aby bola splnená ekvilibriová podmienka tzv. detailed balance. Výsledné pravdepodobnosti miery prechodu sú

$$U_n = \frac{D}{\Delta x^2} e^{-[v(x_{n+1})-v(x_n)]/2D}, \quad (2.2-5)$$

$$D_n = \frac{D}{\Delta x^2} e^{[v(x_n)-v(x_{n-1})]/2D}. \quad (2.2-6)$$

Na základe rovnice 2.2-4 môžeme za pomoci 2.2-5 a 2.2-6 zostrojiť numerický algoritmus na výpočet priemerného času, za ktorý sa dostanú dva hotspoty dostatočne blízko na to, aby mohli zrekombinovať.

### 2.3 Stochastický algoritmus

Stochastický algoritmus aplikovaný v [1] je klasickým algoritmom používaným v problémoch chemickej kinetiky [2], ako aj v iných oblastiach vo fyzike a v aplikovaných vedách. Začneme z bodu  $x_1$ , ktorý je náhodne vybraný z rozdelenia, ktoré generuje daný potenciál (viď nižšie). Čas, ktorý hotspot strávi v bode  $x_n$ , je vybraný náhodne z exponenciálneho rozdelenia, charakterizovaného

$$E(t) = \frac{1}{D_n + U_n}.$$

Na výpočet času použijeme vzťah

$$t = \frac{-1}{D_n + U_n} \ln(X),$$

pričom  $X$  je z rovnomerného rozdelenia  $U([0, 1])$ . Na určenie, či sa hotspot pohne do bodu  $x_{n+1}$  alebo  $x_{n-1}$ , nám posluží iné náhodne vygenerované číslo z  $U([0, 1])$ , ktoré porovnáme s pravdepodobnosťami, že hotspot skočí do  $x_{n+1}$ :

$$P(x_n \rightarrow x_{n+1}) = \frac{U_n}{D_n + U_n},$$

prípadne, že skočí do  $x_{n-1}$ :

$$P(x_n \rightarrow x_{n-1}) = \frac{D_n}{D_n - U_n}.$$

Keď sa voľný hotspot dostane na maximálnu vzdialenosť od fixovaného hotspotu (reflexnú hranicu), odrazí sa späť. Pri dostatočnej blízkosti (absorpčná hranica) hotspot zrekombinuje,

čím celý proces končí. Opakovaním algoritmu dostaneme priemerný čas, za ktorý sa dostane voľný hotspot k fixovanému.

**Štartovací bod**  $x_1$  sa volí z rozdelenia, ktoré vznikne, ak do každého bodu  $x_n$  umiestnime rovnako veľa hotspotov. Každý necháme náhodne sa pohybovať v danom potenciáli, pričom obe hranice budú reflexné, dostatočne dlho. Potom sa pozrieme na to, v akej pozícii sa ustália. Podľa počtu hotspotov, ktoré sa ustália v jednotlivých bodoch, rozdelíme interval  $[0, 1]$ . Potom generujeme náhodné číslo z  $U([0, 1])$  a porovnáme ho s intervalmi vzniknutými našim delením. Kolkému intervalu prislúcha číslo, z toľkého bodu začneme s  $x_1$ .

### 3 Analytický výpočet first-exit času

Problémom výpočtu first-exit času sa zaoberal vo svojej práci [5] Attila Szabo. Vychádzal z rovnakej Fokker-Planckovej rovnice,

$$\begin{aligned}\frac{\partial \rho}{\partial t} &= \frac{\partial v'(x)\rho}{\partial x} + D \frac{\partial^2 \rho}{\partial x^2} \\ \rho(x, t_0) &= \delta(x - x_0) \\ J(0, t) &= \kappa \rho(0, t) \\ J(L, t) &= 0,\end{aligned}\tag{3.0-1}$$

s tým rozdielom, že neuvažoval, že častice zrekombinujú okamžite, ale s určitou mierou efektivity reakcie  $\kappa$ . Keď  $\kappa \rightarrow \infty$ , prvú okrajovú podmienku môžeme nahradiť podmienkou z predchádzajúcej kapitoly,  $\rho(0, t) = 0$ , ktorá sa tiež nazýva Smoluchovského okrajová podmienka. Szabo v článku [5] odvodil všeobecnú formulu pre priemerný čas do reakcie, ak uvažujeme  $\kappa \rightarrow \infty$ :

$$E(t) = \left[ \int_0^L x^{d-1} e^{-\beta v(x)} dx \right]^{-1} \int_0^L \frac{x^{1-d}}{D} e^{\beta v(x)} \left[ \int_x^L y^{d-1} e^{-\beta v(y)} dy \right]^2 dx,\tag{3.0-2}$$

kde  $d$  je dimenzia a  $\beta$  je konštanta odvodená od teploty. S použitím tohoto výsledku vieme implicitne vypočítať, aký je čas do reakcie a porovnať ho s výsledkom numerického výpočtu. V niektorých prípadoch, ak vieme výsledok spočítať všeobecne, môžeme dostať aj závislosť medzi výsledným časom a maximálnou možnou vzdialenosťou hotspotov  $L$ .

Pre konštantný potenciál  $v(x) = k$  a dimenziu  $d = 1$  vieme odvodiť závislosť

$$\begin{aligned}E(t) &= \left[ \int_0^L e^{-\beta k} dx \right]^{-1} \int_0^L \frac{1}{D} e^{\beta k} \left[ \int_x^L e^{-\beta k} dy \right]^2 dx \\ &= \frac{e^{2\beta k}}{e^{2\beta k}} \left[ \int_0^L 1 dx \right]^{-1} \int_0^L \frac{1}{D} \left[ \int_x^L 1 dy \right]^2 dx \\ &= \frac{1}{D} \cdot \frac{1}{L} \int_0^L L^2 - 2Lx + x^2 dx \\ &= \frac{1}{D} \cdot \frac{1}{L} \left[ xL^2 - Lx^2 + \frac{x^3}{3} \right]_0^L \\ &= \frac{L^2}{3D}.\end{aligned}\tag{3.0-3}$$

Z výsledku sa dá usúdiť, že pri rovnakej difúzii a konštantnom potenciáli rastie priemerný čas, potrebný na zrekombinovanie, kvadraticky s maximálnou možnou vzdialenosťou hotspotov  $L$ . Podobne vieme aj pre lineárny potenciál  $v(x) = -\alpha x$  a dimenziu  $d = 1$  odvodiť závislosť

$$\begin{aligned}
E(t) &= \left[ \int_0^L e^{\beta\alpha x} dx \right]^{-1} \int_0^L \frac{1}{D} e^{-\beta\alpha x} \left[ \int_x^L e^{\beta\alpha y} dy \right]^2 dx \\
&= \frac{\alpha\beta}{e^{\alpha\beta L} - 1} \int_0^L \frac{e^{-\alpha\beta x}}{D} \left[ \frac{e^{\alpha\beta L} - e^{\alpha\beta x}}{\alpha\beta} \right]^2 dx \\
&= \frac{1}{\alpha\beta D} \cdot \frac{1}{e^{\alpha\beta L} - 1} \int_0^L e^{-\alpha\beta x} \left( e^{2\alpha\beta L} + e^{2\alpha\beta x} - 2e^{\alpha\beta x} e^{\alpha\beta L} \right) dx \\
&= \frac{1}{\alpha\beta D} \cdot \frac{1}{e^{\alpha\beta L} - 1} \int_0^L e^{-\alpha\beta x} e^{2\alpha\beta L} + e^{\alpha\beta x} - 2e^{\alpha\beta L} dx \tag{3.0-4} \\
&= \frac{1}{\alpha\beta D} \cdot \frac{1}{e^{\alpha\beta L} - 1} \left( -\frac{e^{2\alpha\beta L}}{\alpha\beta e^{\alpha\beta L}} + \frac{e^{2\alpha\beta L}}{\alpha\beta} + \frac{e^{\alpha\beta L} - 1}{\alpha\beta} - 2e^{\alpha\beta L} L \right) \\
&= \frac{1}{\alpha\beta D} \cdot \frac{1}{e^{\alpha\beta L} - 1} \left( \frac{e^{2\alpha\beta L}}{\alpha\beta} - \frac{1}{\alpha\beta} - 2e^{\alpha\beta L} L \right) \\
&= \frac{1}{\alpha^2 \beta^2 D} \cdot \frac{1}{e^{\alpha\beta L} - 1} \left( e^{2\alpha\beta L} - 1 - 2e^{\alpha\beta L} L \alpha \beta \right).
\end{aligned}$$

## 4 Závislosť času do rekombinácie od jednotlivých faktorov

Budeme skúmať závislosť priemerného času do rekombinácie od týchto faktorov:

- tvar potenciálu,
- teplota a viskozita,
- maximálna vzdialenosť hotspotov.

### 4.1 Závislosť času do rekombinácie od tvaru potenciálu

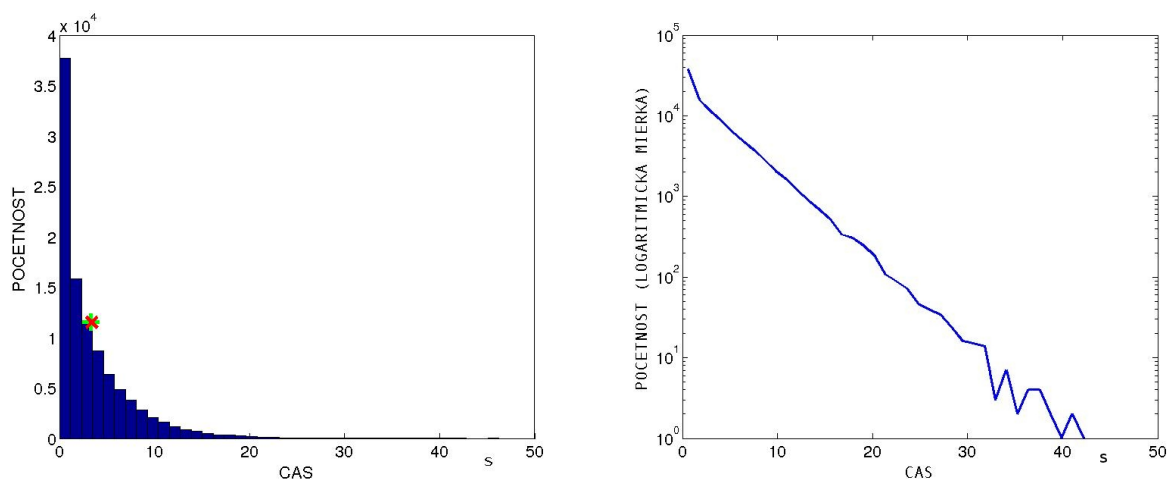
Do úvahy prihádza mnoho tvarov potenciálov. Pozrieme sa na lineárne potenciály a budeme študovať, ako čas závisí od ich sklonu. Taktiež sa pozrieme na periodické potenciály a budeme skúmať, ako závisí čas od počtu periód a rozdielu medzi maximom a minimom na perióde.

Konkrétne:

- konštantný potenciál  $u(x) = k$ ,  $x \in [0, L]$
- lineárny potenciál  $u(x) = \alpha x$ ,  $x \in [0, L]$
- periodický potenciál  $u(x) = \beta \sin(\gamma 2\pi x/L)$ ,  $x \in [0, L]$
- lineárny s periodickými perturbáciami  $u(x) = \alpha x + \beta \sin(\gamma 2\pi x/L)$ ,  $x \in [0, L]$

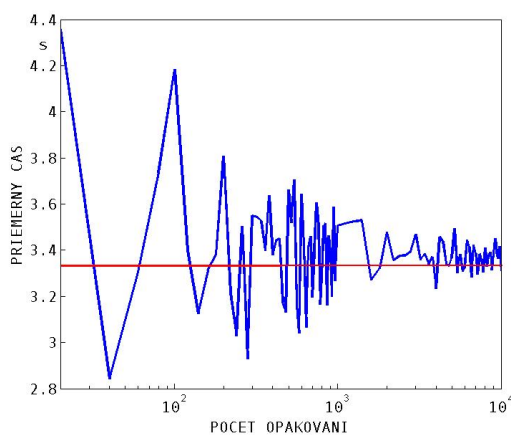
Ako si môžeme všimnúť, v numerickom algoritme nezáleží na posune potenciálu o konštantu, lebo pri výpočte  $U_n$  a  $D_n$  nás zaujíma len rozdiel potenciálov v dvoch po sebe idúcich polohách hotspotu. Všetky simulácie boli robené pri vzdialenosti 10  $\mu\text{m}$  a 100 deliacich bodoch. Dáta sú z 10000 opakovaní algoritmu.

## Konštantný potenciál



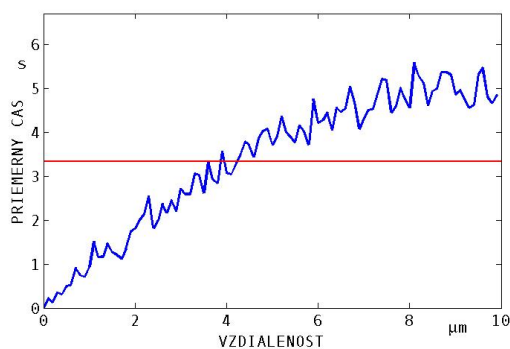
**Obr. 4–4** Histogram časov pre konštantný potenciál. Zelený marker označuje explicitne spočítaný priemerný čas pomocou výsledku z článku [5] a červený marker označuje priemerný čas spočítaný numerickým algoritmom. Takmer lineárny graf vpravo naznačuje, že časy pochádzajú z exponenciálneho rozdelenia.

Ako je vidno vľavo na obrázku 4–4, časy trvania jednotlivých realizácií stochastického procesu sa zdajú byť exponenciálne rozdelené. Toto sa dá overiť na logaritmickom grafe 4–4 vpravo. Ak pochádzajú dáta z exponenciálneho rozdelenia, potom zakreslením početností pomocou logaritmickkej mierky dostaneme lineárny graf. Spriemerovaním jednotlivých realizácií dostávame strednú hodnotu daného exponenciálneho rozdelenia, ako aj priemerný čas do absorpcie. V tomto prípade vieme presnosť numerického riešenia overiť pomocou explicitného výsledku z článku [5]. Ako dobre sa dá v tomto prípade explicitné riešenie odhadnúť numerickou simuláciou, naznačuje konvergenčný diagram 4–5.



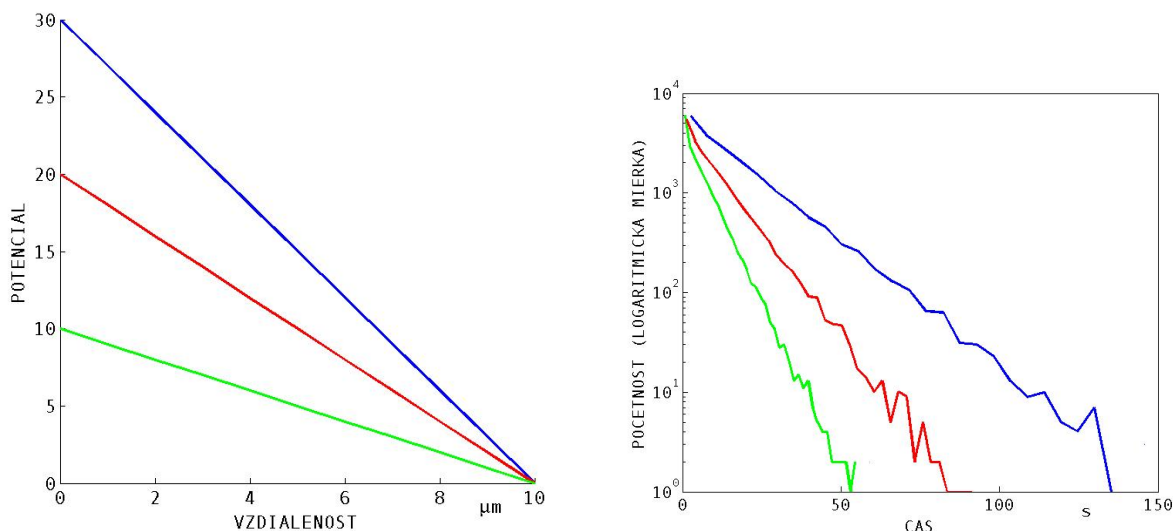
**Obr. 4–5** Konvergenčný diagram znázorňujúci priemerný čas v závislosti od počtu počtu dát na jeho výpočet. Červená hodnota znázorňuje explicitne spočítaný priemerný čas pomocou výsledku zo [5].

Zaujímavé je sledovať závislosť priemerného času do rekombinácie pre jednotlivé polohy štartovacieho bodu (obr. 4–6). Obrázok ukazuje, ako sú na tom štartovacie body z pohľadu rýchlosti, s akou majú šancu dostať v danom potenciáli dostatočne blízko k zafixovanému hotspotu. Vidno, že asi prvých tridsať bodov je na tom približne rovnako, lebo ich priemerný čas do rekombinácie je rovnaký. Z toho sa dá usúdiť, že v určitej vzdialenosti od zafixovaného hotspotu je v podstate jedno, kde začíname, lebo všetky vzdialenejšie body majú rovnakú šancu zrekombinovať po takmer rovnakom čase.



**Obr. 4–6** Závislosť priemerného času od počiatočného štartovacieho bodu. Červená hodnota znázorňuje explicitne spočítaný priemerný čas pomocou výsledku zo [5]. Na výpočet priemerného času bolo pre každý štartovací bolo použitých 1000 simulácií.

## Lineárny potenciál



**Obr. 4–7** Tri typy lineárneho potenciálu ( $-x$  — zelený,  $-2x$  — červený,  $-3x$  — modrý) a k nim vpravo histogramy na logaritmickej osi.

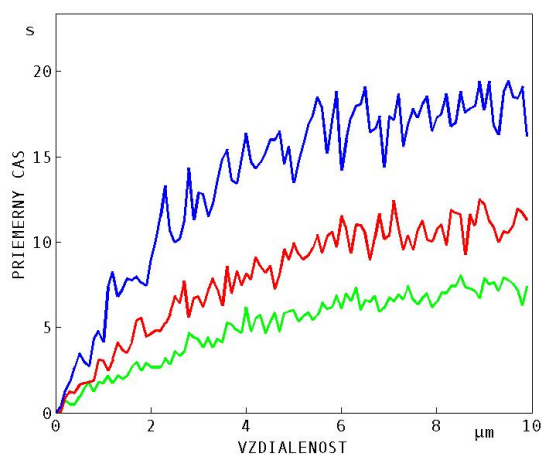
Na základe obrázku 4–7 vpravo sa dá opäť tvrdiť, že jednotlivé realizácie stochastického algoritmu pochádzajú z exponenciálneho rozdelenia. Jednotlivé zmeny strednej hodnoty súvisia so zmenou sklonu grafu na logaritmickej osi. Početnosť malých časov sa nijako zásadne nemení, ale kvôli rozdielnym sklonom sa mení početnosť väčších časov, čo má za dôsledok zmenu výsledného priemerného času do rekombinácie. Z uvedeného jasne vyplýva, že s rastúcim sklonom potenciálu výraznejšie rastie aj priemerný čas do rekombinácie.

Zaujímalo by nás, aký je vzťah medzi sklonom potenciálu a sklonom grafu na logaritmickom grafe. Potom by sme poznali priamo hustotu rozdelenia priemerných časov do rekombinácie, lebo ak predpokladáme  $p_\lambda(x) = 1/\lambda \cdot e^{-\lambda x}$ , z čoho dostávame  $\log(p_\lambda(x)) = \log 1/\lambda - \lambda x$ . Vieme, že  $1/\lambda$  je stredná hodnota exponenciálneho rozdelenia, preto je aproximovanie sklonu rovnako ťažké, ako aproximovanie priemerného času.

Na obrázku 4–8 vidíme, ako sa so zmenou sklonu potenciálu mení čas do rekombinácie pre jednotlivé štartovacie body. Dá sa odpozorovať nielen rastúci priemerný čas do rekombinácie s rastúcim sklonom potenciálu, ale aj zväčšujúca sa oblasť s približne rovnakým priemerným

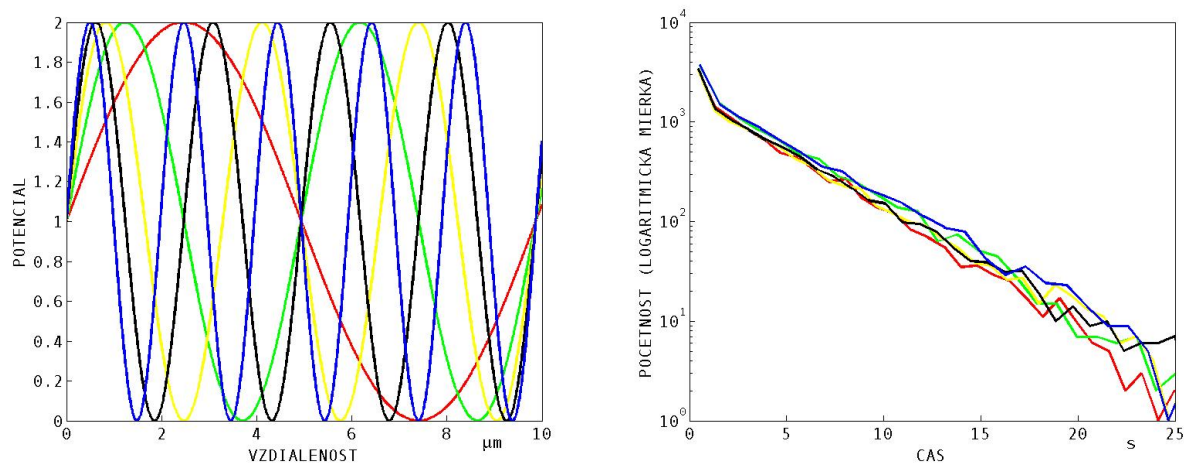


časom do rekombinácie. V prípade modrého grafu sa javí, že do tejto časti patrí už skoro polovica bodov. Je to spôsobené tým, že body začínajúce v strednej časti sa ľahšie a rýchlejšie pohnú v smere spádu potenciálu, čím sa dostanú v podstate na rovnakú vzdialenosť ako body, ktoré v tejto oblasti začínali.



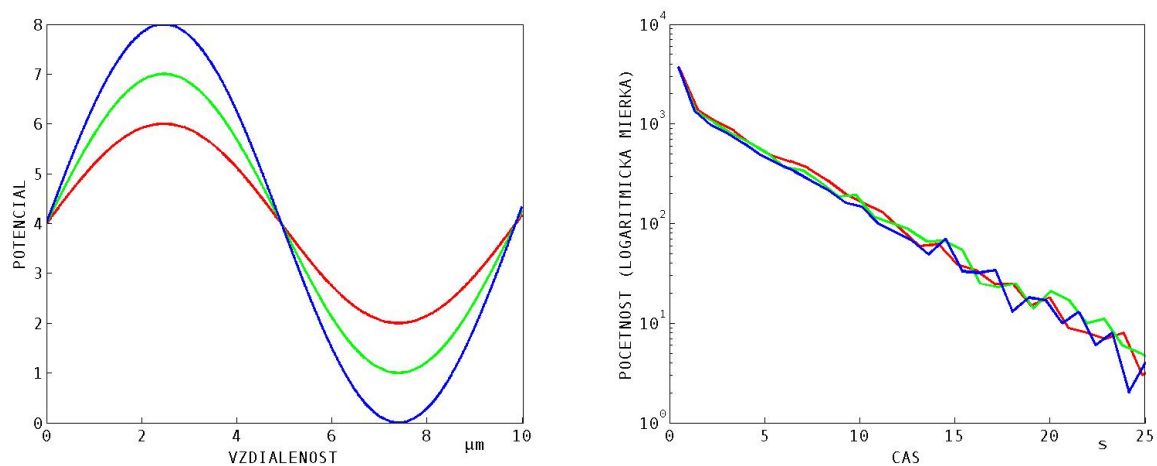
**Obr. 4–8** Závislosť priemerného času od počiatočného štartovacieho bodu. Na výpočet priemerného času bolo pre každý štartovací bod použitých 1000 simulácií. Graf sa dá vyhladiť len zväčšovaním počtu simulácií, čo nie je z technických príčin možné.

## Periodický potenciál



Obr. 4–9 Päť typov potenciálov s rôznym počtom periód. Vpravo histogramy na logaritmickej osi.

Z obrázku 4–9 vyplýva exponenciálne rozdelenie časov pre jednotlivé potenciály, ale o rozdieloch medzi jednotlivými funkciami sa nedá veľa povedať. Z toho sa dá usúdiť, že medzi počtom periód potenciálu a časom do rekombinácie je malá závislosť. Pre uvažované hodnoty pri náraste periód na dvojnásobok nastáva asi len trojpercentný nárast priemerného času.



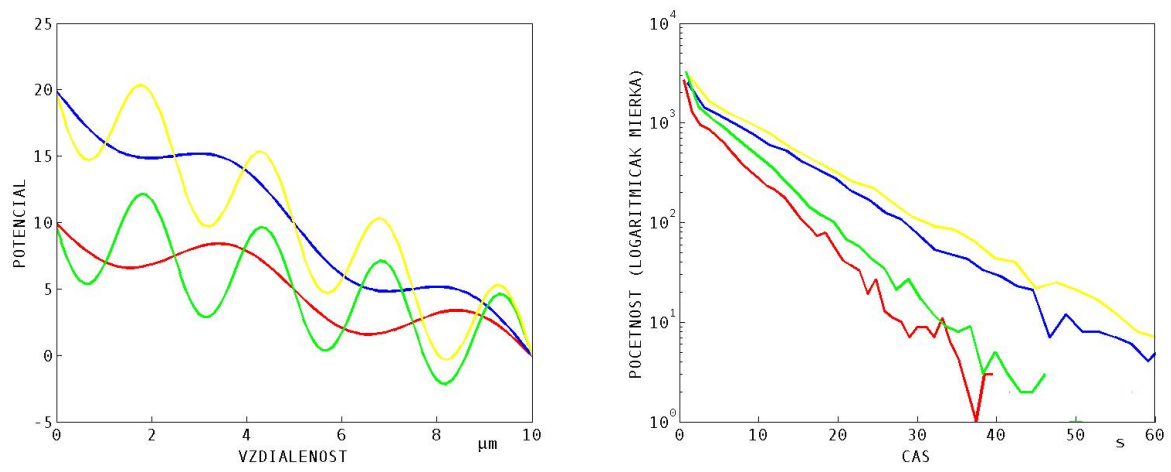
Obr. 4–10 Tri typy potenciálov s rôznou veľkosťou amplitúdy. Vpravo histogramy na logaritmickej osi.

Podobná situácia nastáva aj pri amplitúde periódy, ako naznačuje obrázok 4–10. Pri nami meraných hodnotách spôsobí dvojnásobná zmena amplitúdy asi dvojpercentnú zmenu v priemernom čase.

Zväčšenie amplitúdy môže v špeciálnom prípade spôsobiť aj pokles času. Deje sa to vtedy, ak má potenciál tvar len jednej prevrátenej sínusovej periódy ( $-\sin 2\pi x/L$ ). V takomto prípade sa počiatočná podmienka vygeneruje tak, že veľa bodov sa usadí v potenciálovej jame na sínuse, ktorá je blízko zafixovaného hotspotu a je vysoká pravdepodobnosť, že vyštartujeme z bližšieho bodu. S rastúcou amplitúdou sa prehĺbuje potenciálová jama a stále viac štartovných bodov sa generuje blízko zafixovaného hotspotu.

Predchádzajúce výsledky sú dosť neintuitívne, pretože keby sme pozorovali len jednu časticu, ktorá by bola v potenciálovej jame sínusu, teda vo vzdialenosti  $3L/4$ , tak jej priemerný čas do rekombinácie by rástol s veľkosťou amplitúdy výraznejšie. Tento rozdiel sa ale stráca, lebo nezačíname s časticou len v tomto bode, ale vychádzame z ekvilibriovej počiatočnej podmienky.

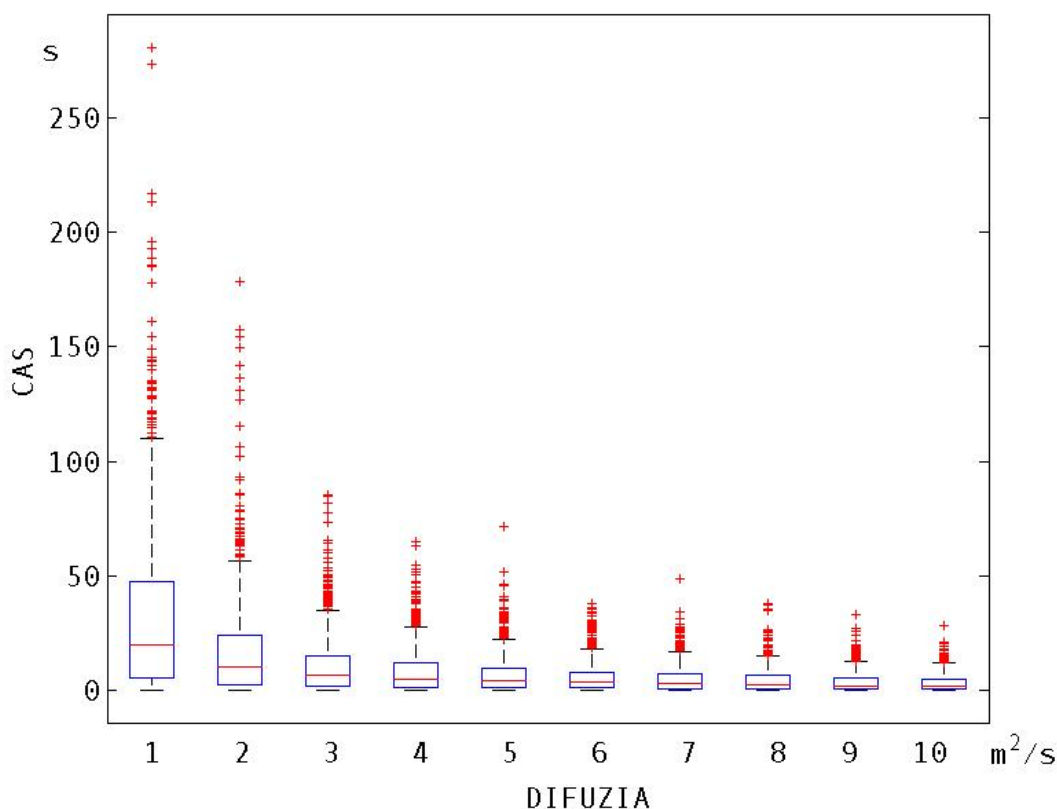
Keďže periodicitu potenciálu má len malý vplyv na priemerný čas do rekombinácie, nemá zmysel uvádzať ďalšie grafy priemerného času pre jednotlivé štartovacie body, ktoré by boli na nerozoznanie od už uvedených. Taktiež nemá zmysel uvažovať rastúci periodický potenciál, lebo zmena vyvolaná periodicitou je v tomto prípade tiež veľmi malá. Nadôvažok jeden príklad na obrázku 4–11, kde v oboch prípadoch dvojnásobný nárast počtu periód, ako aj dvojnásobný nárast amplitúdy vyvolal len asi 2 % zmenu v priemernom čase. V porovnaní s lineárnym potenciálom bola zmena asi 4 %, ale nárast priemerného času dvojnásobnou zmenou sklonu potenciálu bol asi 75 % pri každom tvare periódy.



Obr. 4–11 Malý vplyv periodicity potenciálu na priemerný čas do rekombinácie.

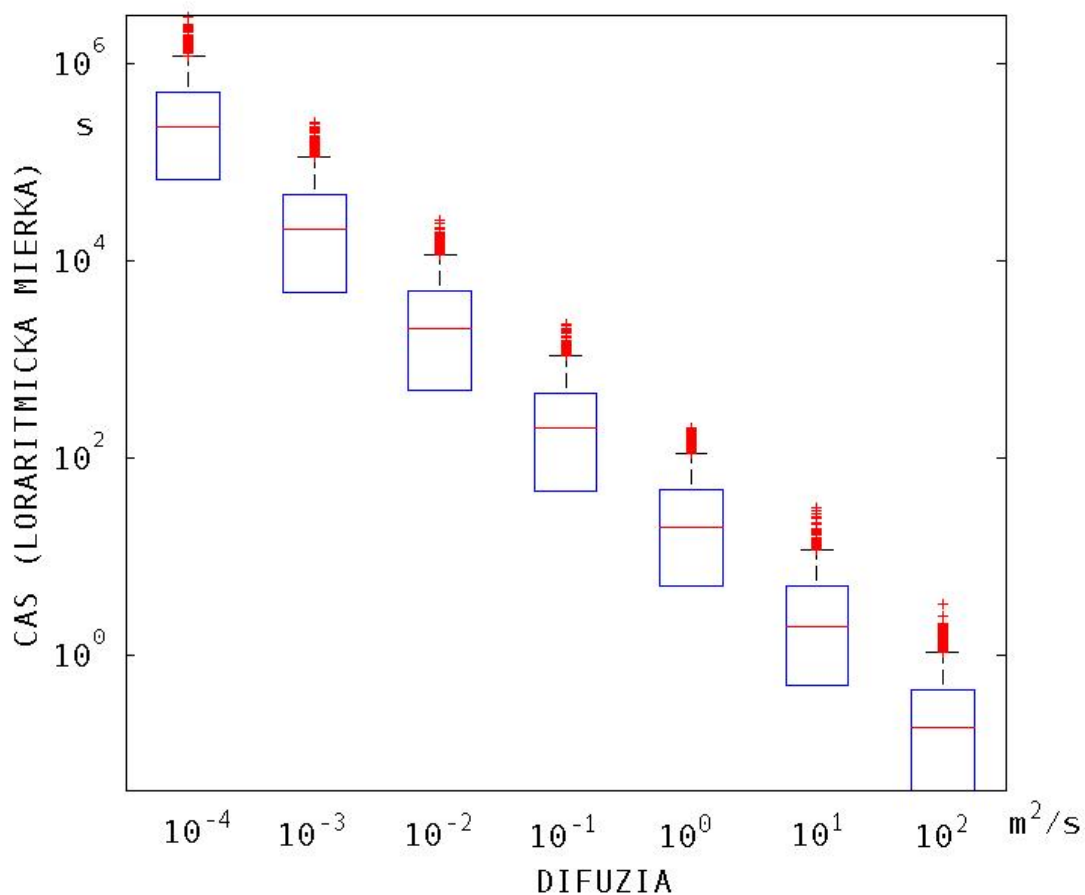
## 4.2 Závislosť času do rekombinácie od teploty a viskozity prostredia

Závislosť času do rekombinácie od teploty a viskozity prostredia sa dá zhrnúť do jedného faktora, lebo obe fyzikálne konštanty vystupujú vo vzorcoch len v difúznej konštante  $D = K_B T / 6\pi\eta r$ . Bude nás teda zaujímať závislosť od difúzneho faktora. Difúzia je jav, kedy častice jednej látky samovoľne prenikajú medzi častice druhej látky. Čím je difúzny koeficient väčší, tým rýchlejšie je prenikanie častice (hotspotu) medzi okolité častice cytoplazmy alebo mitochondriálnej matrix (vnútro mitochondrie).



**Obr. 4–12** Boxplot času pre difúzie od 1 po 10  $m^2/s$ . Maximálna vzdialenosť 10  $\mu m$ , každý boxplot je zostrojený z 1000 opakovaní procesu. Boxplot zachytáva postupne 25 %, 75 % dát a outliery sú vykreslené zvlášť.

Ako ukazuje graf 4–12, v modeli platí závislosť, že s rastúcim difúznym faktorom klesá priemerný čas. Ďalej sa dá odpozorovať aj klesajúci rozptyl časov s rastúcim difúznym faktorom.



**Obr. 4–13** Boxplot časovej závislosti od difúzie. Pri simulácii boli použité rovnaké parametre ako v predchádzajúcom prípade.

Graf 4–13 ukazuje, že medzi difúznym faktorom a priemerným časom je lineárny log-log vzťah. Rovnako to platí aj pre rozptyl časov pre jednotlivé difúzne koeficienty, lebo boxploty si udržiavajú rovnakú veľkosť.

Pre reálne koeficienty ( $T = 25\text{ }^\circ\text{C}$ ,  $\eta = 44.5 \cdot 10^{-4}\text{ g/cm} \cdot \text{s}$ ,  $r = 0.17 \cdot 10^{-9}\text{ m}$ ) je hodnota difúzneho koeficientu  $2,88 \cdot 10^{-10}\text{ m}^2/\text{s}$ . Pri týchto hodnotách aj malá zmena niektorého z

ovplyvniteľných parametrov (teplota, viskozita) vyvolá veľkú zmenu priemerného času do rekombinácie.

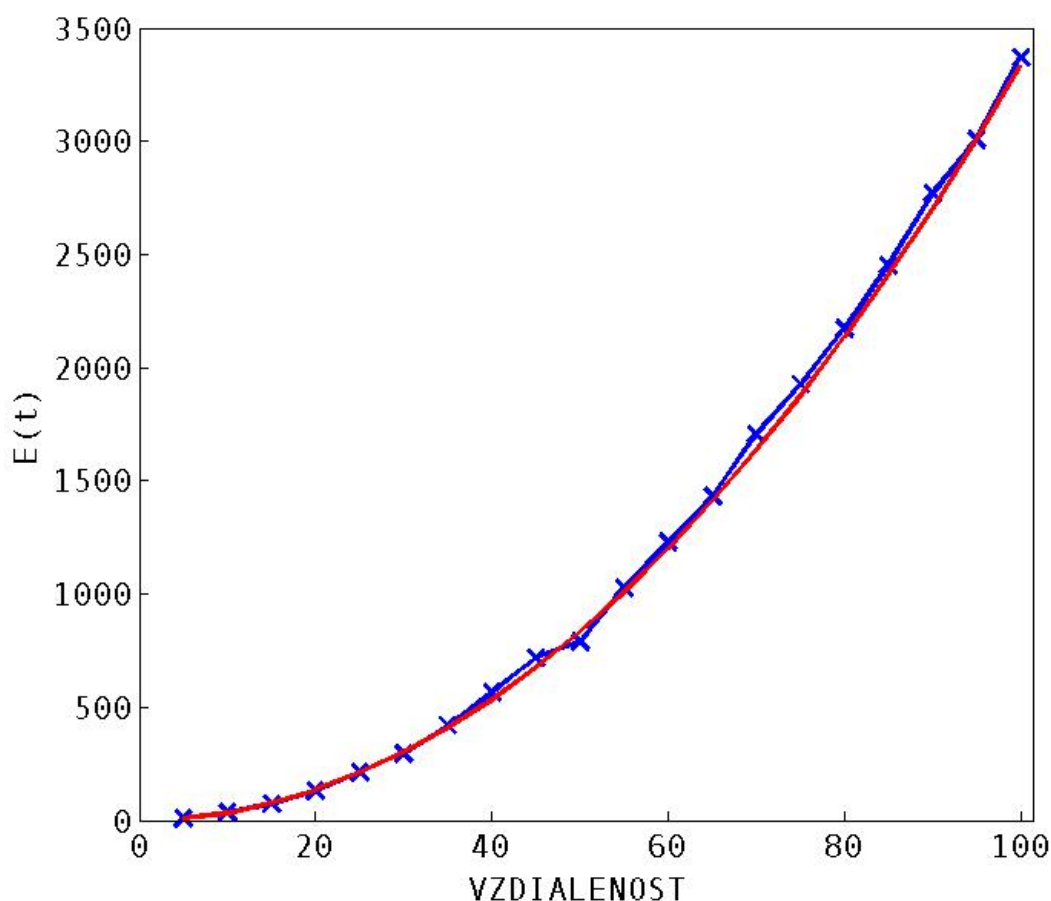
Ako si môžeme všimnúť na obrázku 4–13, tak sklon boxplotov je takmer -1. Teda platí vzťah

$$\log E(t) \approx a - \log(D),$$

z čoho úpravou dostávame  $E(t) \approx A/D$ , čo zodpovedá výsledku, ktorý sme získali z 3.0-3, a taktiež všeobecnému vzorcu 3.0-2, ak predpokladáme, že  $D$  je konštantné.

### 4.3 Závislosť času do rekombinácie od maximálnej možnej vzdialenosti hotspotov

Už sme explicitne pomocou [5] odvodili závislosť medzi maximálnou vzdialenosťou dvoch hotspotov a priemerným časom, potrebným na rekombináciu pre konštantný potenciál 3.0-3. Ako vidno na obrázku 4–14, v tomto prípade sa numerické výsledky zhodujú s explicitným výpočtom.



**Obr. 4–14** Porovnanie priemerných časov pre konštantný potenciál získaných pomocou numerického algoritmu (modrá) s odvodenou funkciou  $L^2/3D$  pre  $D = 1$  (červená). Každý priemerný čas je získaný pomocou 10000 simulácií, s konštantnými ostatnými parametrami.

Pre lineárny potenciál máme explicitný výpočet 3.0-4, ktorý sa nezhoduje s numerickým



výpočtom. Zhodujú sa iba rádovo, čo je najpravdepodobnejšie spôsobené rôznou počiatočnou podmienkou pre numerický výpočet a explicitné riešenie.

V súvislosti s maximálnou možnou vzdialenosťou hotspotov a strmou potenciálu treba byť obozretný pri numerickej simulácii a opatrne vyberať počet deliacich bodov, teda veľkosť kroku, ktorý robíme. Môže sa stať, že v prípade zlej voľby kroku bude algoritmus trvať príliš dlho — príliš dlhý krok (pravdepodobnosti prechodu sú príliš malé) alebo bude dávať nepresné výsledky — príliš malý krok (nastane numerické zaokrúhlenie).

## 5 Záver

Cielom práce bolo použiť stochastický algoritmus z článku [1] na matematické skúmanie alternatívneho mechanizmu spojeného s predlžovaním a skracovaním telomér — telomeric-kých krúžkov. V práci sme popísali pohyb dvoch hotspotov ako stochastický, relatívny pohyb jedného voľného hotspotu vzhľadom na druhý zafixovaný hotspot. Skúmali sme, za aký priemerný čas sa dostanú hotspoty dostatočne blízko k sebe na to, aby mohli zrekombinovať a hlavne, ako na tento proces vplyvajú jednotlivé faktory (teplota a viskozita prostredia, tvar potenciálu). Dospeli sme k prekvapivému záveru, že periodicita potenciálu má pri porovnaní so sklonom potenciálu takmer zanedbateľný vplyv na priemerný čas do rekombinácie. Je to spôsobené tým, že nezačíname z rovnomernej počiatocnej podmienky, ale z ekvilibriovej počiatocnej podmienky, špeciálnej pre každý tvar potenciálu. Ukázalo sa, že asi najvplyvnejší a najľahšie ovplyvniteľný faktor je teplota. Explicitne sme za pomoci výsledku z [5] spočítali, ako závisí priemerný čas do rekombinácie od maximálnej vzdialenosti hotspotov pre konštantný potenciál.

## Literatúra

- [1] T. C. Elston and C. R. Doering. Numerical and analytical studies of nonequilibrium fluctuation-induced transport processes. *Journal of Statistical Physics*, 83(3):359–383, 1996.
- [2] D. T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of computational physics*, 22(4):403–434, 1976.
- [3] L. Ringrose, S. Chabanis, P. O. Angrand, C. Woodroffe, and A. F. Stewart. Quantitative comparison of DNA looping in vitro and in vivo: chromatin increases effective DNA flexibility at short distances. *The EMBO Journal*, 18(23):6630–6641, 1999.
- [4] H. Risken. *The Fokker-Planck equation: Methods of solution and applications*, volume 18. Springer Verlag, 1996.
- [5] A. Szabo, K. Schulten, and Z. Schulten. First passage time approach to diffusion controlled reactions. *The Journal of Chemical Physics*, 72(8):4350, 1980.
- [6] L. Tomaska, M. J. McEachern, and J. Nosek. Alternatives to telomerase: keeping linear chromosomes via telomeric circles. *FEBS letters*, 567(1):142–146, 2004.
- [7] L. Tomaska and J. Nosek. Telomere heterogeneity: Taking advantage of stochastic events. *FEBS letters*, 583(7):1067–1071, 2009.
- [8] L. Tomaska, J. Nosek, J. Kramara, and J. D. Griffith. Telomeric circles: universal players in telomere maintenance&quest. *Nature Structural & Molecular Biology*, 16(10):1010–1015, 2009.
- [9] L. Tomaska, J. Nosek, A. M. Makhov, A. Pastorakova, and J. D. Griffith. Extragenomic double-stranded DNA circles in yeast with linear mitochondrial genomes: potential involvement in telomere maintenance. *Nucleic Acids Research*, 28(22):4479, 2000.

- 
- [10] P. A. Wiggins, T. van der Heijden, F. Moreno-Herrero, A. Spakowitz, R. Phillips, J. Widom, C. Dekker, and P. C. Nelson. High flexibility of DNA on short length scales probed by atomic force microscopy. *Nature Nanotechnology*, 1(2):137–141, 2006.