

Molekulárna darwinovská evolúcia *in silico*¹

Vladimír Kvasnička

Katedra matematiky FCHPT STU, Bratislava

email: kvasnic@stuba.sk



1. Úvod

Darwinovská evolúcia patrí medzi štandardné predmety štúdia **umelého života** (AL), hlavný stimul pre túto skutočnosť bol prudký rozvoj evolučných algoritmov na prelome 80-90 rokov minulého storočia. Táto skutočnosť má svoj počiatok u teoretického genetika Seala Wrighta [31], ktorý v r. 1932 publikoval prácu v ktorej zaviedol pojem "adaptive landscape"² a charakterizoval darwinovskú evolúciu ako adaptívny proces (v matematickej terminológii "optimalizačný proces"), kde genotyp populácie je adaptovaný tak, že dosiahne lokálne (snáď aj globálne) optimum na povrchu fitness. O mnoho neskôr, po 40 rokoch, využil informatik John Holland [15] túto metaforu darwinovskej evolúcie pri návrhu genetických algoritmov, ktoré zo súčasného pohľadu môžu byť interpretované ako abstrakcia darwinovskej evolúcie do formy univerzálneho optimalizačného algoritmu^{3,4}.

Cieľom tohto príspevku je prezentácia jednoduchého modelu darwinovskej evolúcie, ktorý postihuje jej elementárne aspekty a procesy vyskytujúce sa na úrovni biomakromolekúl. Diskutovaný model evolúcie poskytuje semikvantitatívnu teóriu, ktorá je schopná interpretovať mnohé jej pojmy a koncepty, tak napr. úlohu neutrálnych mutácií v priebehu adaptačného procesu. Dôležitú úlohu v našom prístupe hrajú metódy **umelej chémie** [2,4,6,8,13,14,22,23,24], ktorá je v súčasnosti považovaná za oblasť úzko súvisiacu s umelým životom. Tento prístup je založený na metafore chemostatu (chemického reaktoru), ktorý obsahuje "molekuly" reprezentované abstraktnými objektami (reťazcami znakov, grafmi, formulami z nejakého jazyku, atď.). Molekuly sú v chemostate transformované pomocou

¹ Prepracovaná a rozšírená verzia práce [24], ktorá bude prezentovaná na "Second Euro-International Symposium on Computational Intelligence", 16. -19. júna 2002, Košice.

² V súčasnosti sa používa angl. termín "fitness landscape". Vychádzajúc zo súčasnej slovenskej (a českej) matematickej terminológie, kde je zaužívaný termín "(hyper)povrch funkcie" (ako preklad angl. "function (hyper)surface"), používame preklad tohto termínu ako "povrch fitness". Týmto sa chceme vyhnúť krkolomným prekladom termínu "landscape".

³ Americký filozof a kognitívny vedec Daniel Dennett vo svojej knihe *Darwin's Dangerous Idea - Evolution and the Meaning of Life* [5] univerzálnosť Darwinovej evolúcie zdôraznil tým, že je charakterizoval ako univerzálny algoritmus.

⁴ Richard Dawkins v eseji *Univesal Darwinism* [7] charakterizoval Darwinovu evolúciu ako univerzálny prístup k spontánnemu vytváraniu adaptívnej zložitosti (adaptive complexity), ktorá je charakteristická pre život v celom našom vesmíre. Problém univerzálnosti darwinovskej evolúcie je široko diskutovaný Dušanom Gálikom v jeho prednáške [32], v ktorej kriticky hodnotí možnosti univerzálneho darwinizmu ako východiska pre zjednotenie prírodných a spoločenských vied.

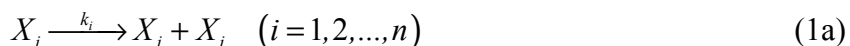
"chemických reakcií" na iné prípustné objekty, pričom pravdepodobnosť týchto transformácií je striktné určená štruktúrou reagujúcich objektov. Kinetika procesov prebiehajúcich v chemostate je popísaná Eigenovými replikátorovými diferenciálnymi rovnicami, ktoré majú svoj „zdravý“ základ vo fyzikálnej chémii (zákon účinných hmotností [1]). Hlavným predmetom záujmu umelej chémie je (i) štúdium formálnych systémov založených na metafore chemostatatu a ktoré sú schopné vykonávať špeciálne vysoko paralelné výpočty a (ii) vytvárať formálne autokatalytické systémy (molekuly sú reprezentované štruktúrovanými objektmi) pre účely *in-silico* simulácií emegencie "živých" systémov.

Newman a Engelhardt [26] ukázali, že použitím Kauffmanovej *KN* funkcie⁵ [17,18,26] môžu byť zreprodukované skoro všetky základné výsledky získane Petrom Schusterom a jeho spolupracovníkmi [12,27,28] o molekulárnej darwinovskej evolúcii, ktoré sú založené na použití "fyzikálneho" modelu RNA a jej foldingu,. V tomto príspevku ukážeme, že použitie tohto jednoduchého modelu povrchu fitness poskytuje výsledky, ktoré sú taktiež formálne veľmi blízke tým, ktoré sú predpovedané replikátorovými diferenciálnymi rovnicami, ktoré boli postulované už pred viac ako 30 rokmi Eigenom [9] ako odozva na Spiegelmanove *in-vitro* experimenty [29] s molekulárnou evolúciou RNA molekuly⁶.

2. Eigenové replikátory

Na prelome 60. a 70. rokov publikoval Manfred Eigen významnú prácu "*Self organization of matter and the evolution of biological macro molecules*" [9], v ktorej študoval vlastnosti chemického systému zloženého z tzv. *replikátorov* (priamo v texte tejto práce uvádza, že bol stimulovaný k jej napísaniu Spiegelmanovými [29] experimentami *in-vitro* darwinovskej evolúcie). Replikátory sú hypotetické biomakromolekuly, ktoré majú schopnosť replikovať sa. Eigen ukázal, že takýto jednoduchý chemicko-kinetický systém formálne simuluje darwinovskú evolúciu na molekulárnej úrovni. Ďalšie zovšeobecnenie modelu replikátorov bolo vykonané Eigenom a Schustrom [10] v sérii prác so spoločným názvom "*The Hypercycles: A Principle of Natural Evolution*", v ktorých sa položili základy teórie *hypercyklov*, ktoré sú v súčasnosti považované za jeden z nemnohých formálnych systémov schopných simulovať nárast zložitosti (komplexity) na rozhraní živých a neživých systémov.

Študujme molekuly - replikátory X_1, X_2, \dots, X_n , ktoré sú schopné nasledujúcich dvoch chemických reakcií



Prvá reakcia (1a) znamená, že molekula X_i sa replikuje s rýchlostnou konštantou k_i a druhá reakcia (1b) znamená, že molekula X_i zaniká s rýchlostnou konštantou ϕ (ktorá sa nazýva *zried'ovací tok* a bude špecifikovaná neskoršie). Ak použijeme k týmto reakciám štandardný fyzikálno-chemický zákon účinných hmotností [1], dynamika kinetického systému obsahujúceho reakcie (1a-b) je určená diferenciálnymi rovnicami

$$\dot{x}_i = x_i (k_i - \phi) \quad (i = 1, 2, \dots, n) \quad (2a)$$

⁵ Táto stochastická funkcia bola vytvorená [17,18] teoretickým biológom Stuartom Kauffmanom pomocou fyzikálnej „metafory“ spinových skiel, je v súčasnosti veľmi populárna v AL a teoretickej biológii, ako jednoduchý prostriedok pre numerickú implementáciu „rozoklaných“ (rugged) povrchov fitness, ktorých členitosť sa dá jednoducho „nastavovať“ pomocou niekoľkých celočíselných parametrov.

⁶ Pozri prednášku E. Tomáška a J. Noseka [30] na seminári Kognitívne vedy III (CogSci2000), CHTF STU, Bratislava. Text prednášky je dostupný na http://math.chtf.stuba.sk/CogSci_2000.htm.

kde $0 \leq x_i \leq 1$ je koncentrácia molekuly X_i . Zried'ovací tok ϕ je chápaný ako "volný parameter", ktorý sa nastaví tak, aby suma časových derivácií koncentrácií bola nulová, $\sum \dot{x}_i = 0$

$$\dot{x}_i = x_i \left(k_i - \sum_{j=1}^n k_j x_j \right) \quad (i=1,2,\dots,n) \quad (2b)$$

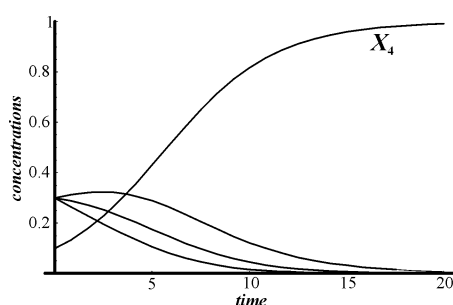
kde podmienka $\sum x_i = 1$ bola použitá bez straty všeobecnosti našich úvah. Analytické riešenie systému diferenciálnych rovníc má tvar [9]

$$x_i(t) = \frac{x_i(0)e^{k_i t}}{\sum_{j=1}^n x_j(0)e^{k_j t}} \quad (3)$$

Toto riešenie vyhovujú asymptotickej podmienke, kde "preživa" len tá molekula, ktorá má maximálnu rýchlostnú konštantu k_{max} , zatiaľ čo ostatné molekuly vymiznú

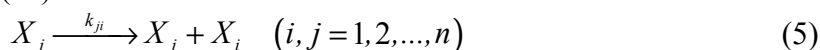
$$\lim_{t \rightarrow \infty} x_i(t) = \begin{cases} 1 & (\text{pre } k_i = k_{max} = \max\{k_1, \dots, k_n\}) \\ 0 & (\text{ostatné prípady}) \end{cases} \quad (4)$$

Každý typ molekuly je formálne chápaný ako zvláštny biologický druh s fitness špecifikovanou rýchlostnou konštantou k , potom prežívajú len tie molekuly - biologické druhy, ktoré sú najlepšie prispôsobené, t.j. majú najväčšiu rýchlostnú konštantu k_{max} ; všetky ostatné molekuly - druhy s rýchlostnými konštantami - fitness menšími než ako k_{max} vymiznú v priebehu časovej evolúcie chemicko-kinetického systému, pozri obr. 1. Podmienka konštantnosti sumy koncentrácie (t.j. $\sum x_i = 1$) zavádza do systému dodatočný "selekčný tlak", replikujúce sa molekuly vytesňujú ostatné molekuly.



Obrázok 1. Časový priebeh koncentrácií 4-zložkového systému s rýchlostnými konštantami $k_1=1$, $k_2=2$, $k_3=3$ a $k_4=4$. Molekuly X_4 víťazia, aj keď ich počiatočná koncentrácia bola najmenšia, v čase $t > 15$ systém už obsahuje skoro výlučne len molekuly X_4 , ostatné vymizli.

Eigenov model replikátorov môže byť jednoducho modifikovaný tak, že sa zavedú mutácie do procesu replikácie (1a)



kde k_{ji} je rýchlostná konštantá tejto reakcie. predpokladá sa, že matica rýchlostných konštant $\mathbf{K} = (k_{ij})$ má dominantné diagonálne elementy, t.j. nediagonálne elementy sú podstatne (rádovo) menšie ako diagonálne elementy ($k_{ij} \ll k_{ii}$, pre $i, j = 1, 2, \dots, n$ a $i \neq j$). Tento dôležitý predpoklad vyplýva priamo zo skutočnosti, že "nepresné" replikácie (5) pre $i \neq j$ sú veľmi vzácne, t.j. produkt X_j je uvažovaný ako mutácia eduktu X_i , $X_j = O_{mut}(X_i)$. Systém diferenciálnych rovníc (2a) má potom tvar

$$\dot{x}_i = x_i (k_{ii} - \phi) + \sum_{j \neq i} k_{ji} x_j \quad (i = 1, 2, \dots, n) \quad (6)$$

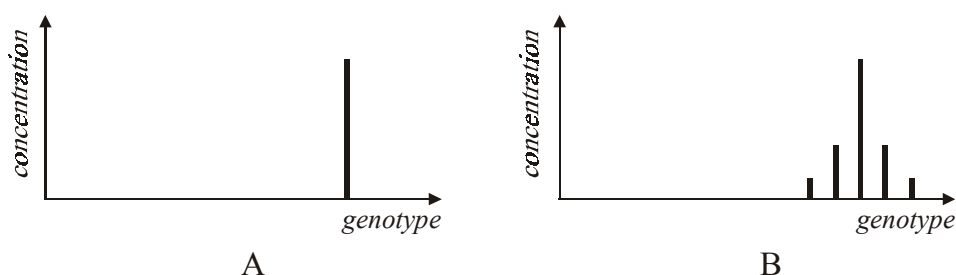
Zried'ovací tok ϕ z (6) je určený tak, aby aj v tomto všeobecnejšom prípade suma časových derivácií koncentrácií bola nulová ($\sum \dot{x}_i = 0$, alebo $\sum x_i = 1$)

$$\phi = \sum_{i,j=1}^n k_{ij} x_j \quad (7)$$

Analytické riešenie (6) má tvar [16]

$$x_i(t) = \frac{\sum_{j=1}^n q_{ij} x_j(0) e^{\lambda_j t}}{\sum_{m,j=1}^n q_{mj} x_j(0) e^{\lambda_j t}} \quad (8)$$

kde $\mathbf{Q} = (q_{ij})$ je nesingulárna matica, ktorá diagonalizuje maticu rýchlostných konštánt \mathbf{K} , $\mathbf{Q}^{-1} \mathbf{K} \mathbf{Q} = \mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$. Pretože bolo postulované, že matica \mathbf{K} má dominantné diagonálne elementy, jej vlastné hodnoty λ sú blízke diagonálnym elementom, $\lambda_i \doteq k_{ii}$, a transformačná matica \mathbf{Q} je blízka jednotkovej matici, $q_{ij} \doteq \delta_{ij}$ (kroneckerovský delta symbol). To znamená, že zavedenie slabých mutácií nemení dramaticky všeobecné vlastnosti pôvodného jednoduchého systému replikátorov (1-4) bez mutácií. Menovite, v asymptotickom čase $t \rightarrow \infty$ systém obsahuje skoro výlučne len molekuly majúce najväčšie rýchlostné konštanty k_{max} , ktoré sú vzácne doprevádzané inými replikátormi tvorenými z víťazného replikátora (s rýchlostnou konštantou k_{max}) pomocou nepresného replikačného procesu (5), pozri obr. 2.



Obrázok 2. Diagram A znázorňuje asymptotickú situáciu pre $t \rightarrow \infty$ pre systém replikátorov popísaných reakciami (1a-b), t.j. ktorí sa presne replikujú. V tomto prípade asymptoticky stabilný stav systému obsahuje len replikátory s maximálnou rýchlostnou konštantou k_{max} . Diagram B znázorňuje asymptoticky stabilný stav systému, kde replikácia nie je už presný proces, kde sa už uvažujú mutácie pri replikačnom procese. Tento stabilný stav je zložený z niekoľkých replikátorov, pričom najväčšia koncentrácia prislúcha replikátoru s maximálnou rýchlostnou konštantou k_{max} , ostatné replikátory sa vyskytujú s podstatne menšou koncentráciou. Replikátory vyskytujúce sa súčasne s hlavným víťazom X_{max} sú tie, ktoré sú z neho tvorené mutáciami.

3. Metafora chemického reaktoru - chemostat

Uvažujme chemický reaktor - *chemostat*, ktorý obsahuje formálne objekty nazývané molekuly, ktoré sú reprezentované reťazcami znakov. Chemostat nie je priestorovo štruktúrovaný, v chémii sa hovorí, že je s dobrým miešaním (well stirred reactor), pozri obr. 3. Akt chemickej reakcie - replikácie spočíva v nasledujúcich troch krokoch:

- (1) náhodný výber replikátora M (ktorý je totožný s X_i) z chemostatu,

- (2) pre takto vybraný replikátor sa uskutoční s pravdepodobnosťou úmernou rýchlostnej konštante k_{ii} proces replikácie, pričom vznik chybného replikátora M' (ktorý je totožný s X_j) je uskutočnený s pravdepodobnosťou úmernou rýchlostnej konštante k_{ij} .
- (3) Vzniklým novým replikátorom M' (bez mutácie alebo s mutáciou) sa nahradí iný náhodne vybraný replikátor z chemostatu (t.j. vo všetkých elementárnych krokoch máme v chemostate konštantný počet replikátorov).

Počítačová simulácia chemostatu a v ňom prebiehajúce replikačné reakcie sa dá pomerne jednoducho implementovať pomocou jednoduchého stochastického prístupu [14]. Tento prístup je založený na skutočnosti, že ak náhodne vyberieme z chemostatu nejaký replikátor, potom pravdepodobnosť toho, že tento replikátor je totožný s nejakým daným replikátorom X sa jednoducho rovná koncentrácii tohto replikátora v chemostate, $0 \leq c(X) \leq 1$. Pravdepodobnosť uskutočnenia replikácie sa rieši tak, že táto pravdepodobnosť je úmerná replikačnej rýchlostnej konštante. Algoritmicky môžeme chemostat implementovať pomocou pseudo-Pascalu týmto jednoduchým algoritmom:

Algoritmus 1

```

chemostat:=a multiset composed of randomly
           generated replicators X;
t:=0;
while t<t_max then
begin t:=t+1;
      X:=O_select(chemostat); {selection}
      if random<prob_rep(X) then
      begin X' := X;           {replication}
            X'' := O_mut(X'); {mutation}
            X''' := O_select(chemostat);
            chemostat := (chemostat + {X''}) - {X'''};
      end;
end;
end;

```

Algoritmus je inicializovaný náhodným vygenerovaním chemostatu. Pravdepodobnosť $\text{prob}_{\text{rep}}(X)$ je určená tak, že je úmerná rýchlostnej konštante replikácie X , pravdepodobnosť mutácie je zahrnutá priamo v operátore $O_{\text{mut}}(X')$.

Všeobecná idea chemostatu budú ilustrované jeho použitím ako optimalizátora binárnej funkcie

$$f : \{0,1\}^N \rightarrow [0,1] \quad (9)$$

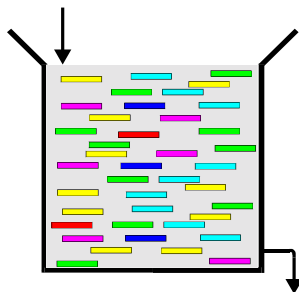
Táto funkcia mapuje binárne reťazce $\mathbf{x} = (x_1, x_2, \dots, x_N) \in \{0,1\}^N$ dĺžky N na reálne čísla z intervalu $\langle 0,1 \rangle$. Naším cieľom bude nájsť optimálne riešenie

$$\mathbf{x}_{\text{opt}} = \arg \max_{\mathbf{x} \in \{0,1\}^N} f(\mathbf{x}) \quad (10)$$

Pretože kardinalita množiny $\{0,1\}^N$ sa rovná 2^N , CPU čas nutný pre riešenie tejto úlohy bude rásť exponenciálne

$$t_{\text{CPU}} \approx 2^N \quad (11)$$

To znamená, že optimalizačný problém (10) patrí do triedy numericky obtiažnych NP-uplných problémov. Toto je jeden z hlavných dôvodov prečo optimalizačné problémy typu (10) sú riešené pomocou evolučných algoritmov [11,15], ktoré reprezentujú veľmi efektívne numerické techniky schopné rýchlo získať dobré suboptimálne riešenie nbinárnych optimalizačných problémov.



Obrázok 3. Chemostat obsahuje homogénnu "zmes" replikátorov. Replikačný akt spočíva v tom, že náhodne vybraný replikátor sa zreplikuje (buď perfektne alebo s malou mutáciou), vzniklý produkt - replikátor vytesní iný náhodne vybraný replikátor. To znamená, že chemostat obsahuje v celej svojej histórii konštantný počet replikátorov.

Nech chemostat obsahuje molekuly, ktoré sú realizované binárnymi reťazcami $\mathbf{x} = (x_1, x_2, \dots, x_N) \in \{0,1\}^N$. Uvažujme nasledujúcu monomolekulárnu reakciu - replikáciu (pozri rov. (5))



kde molekula - reaktant \mathbf{x}' nahradí náhodne vybranú molekulu z chemostatu. Funkčná hodnota $f(\mathbf{x})$ priradená chemickej reakcii je interpretovaná ako pravdepodobnosť (rýchlostná konštanta) reakcie (12). V evolučných algoritmoch selekčný tlak v populácii riešení (chromozómov) je vytvorený pomocou fitness. Chromozómy s väčším fitness majú väčšiu šancu zúčastňovať sa reprodukčného procesu (miera kvality chromozómov), a naopak, chromozómy s malým fitness len vzácné vstupujú do reprodukčného procesu. Tento jednoduchý prejav princípu Darwinovho prirodzeného výberu zabezpečuje postupnú monotónne neklesajúcu gradáciu fitness v priebehu evolúcie populácie. V tomto prístupe zmienený princíp výberu na základe fitness je rozšírený o dodatočný selekčný tlak založený na skutočnosti, že populácia v chemostate má konštantný počet jedincov. Molekula vstupujúca do reakcie je náhodne vybraná z chemostatu, kvázináhodne sa rozhodne, či vstupuje do reakcie alebo nie (s pravdepodobnosťou úmernou jej funkčnej hodnote), vzniklá molekula vytesní z chemostatu inú, náhodne vybranú molekulu. Produkt \mathbf{x}' z pravej strany reakcie (12) je mutácia [15] vstupujúcej molekuly \mathbf{x} , $\mathbf{x}' = O_{mut}(\mathbf{x})$, kde O_{mut} je stochastický operátor mutácie, ktorý zmení jednotlivý bit s pravdepodobnosťou P_{mut} . Pseudo-Pascalovský kód tohto optimalizačného algoritmu dostaneme jednoduchou modifikáciou Algoritmu 1.

Ako ilustračný príklad študujme chemostat pre unimodálnu funkciu určenú nad binárnymi reťazcami dĺžky 6, chemostat je multimnožina

$$P = \{\dots, (110011), \dots\} \subset \{0,1\}^6 \quad (13)$$

Každý binárny vektor \mathbf{x} je ohodnotený racionálnym číslom z intervalu $\langle 0,1 \rangle$

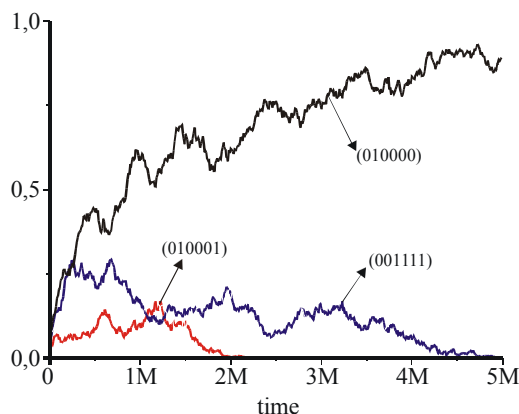
$$real(\mathbf{x}) = \frac{1}{2^6 - 1} int(\mathbf{x}) \quad (14)$$

kde $int(\mathbf{x})$ je nezáporné celé číslo vyjadrujúce dekadickú interpretáciu binárneho reťazca \mathbf{x} . Rýchlostná konštanta k priradená reťazcu

$$k(\alpha) = f(\text{real}(\alpha)) = \frac{1}{2}(1 + \sin(2\pi \cdot \text{real}(\alpha))) \quad (15)$$

s optimálnym riešením $\alpha_{\text{opt}}=(010000)$, kde $\text{real}(x)=16/63$ and $f(16/63)=0.999845$. Chemostat je inicializovaný 1000 náhodne generovanými binárnymi reťazcami a mutačný operátor O_{mut} je špecifikovaný 1-bit pravdepodobnosťou $P_{\text{mut}}=0.01$. Z numerických výsledkov znázornených na obr. 4 vyplýva, že v chemostate spontánne emerguje binárny reťazec, ktorý je optimálnym riešením s racionálnou hodnotou blízkou k presnému riešeniu $x_{\text{opt}}=0.25$. Hlavné výsledky tejto kapitoly môžu byť zhrnuté takto:

- (1) Metafora Eigenových replikátorov poskytuje efektívny stochastický tzv. **replikátorový optimalizačný algoritmus**, kde
- (2) **dôkaz konvergencie** tohto algoritmu ku globálnemu riešeniu bezprostredne vyplýva z existencie jediného asympticky stabilného riešenia s najväčšou rýchlostnou konštantou.
- (3) Tento algoritmus je veľmi **podobný** štandardnému genetickému algoritmu [15], avšak je založený na úplne odlišnej metafore ako GA, metafora darwinovskej evolúcie v GA je nahradená metaforou chemostatu replikátorov.

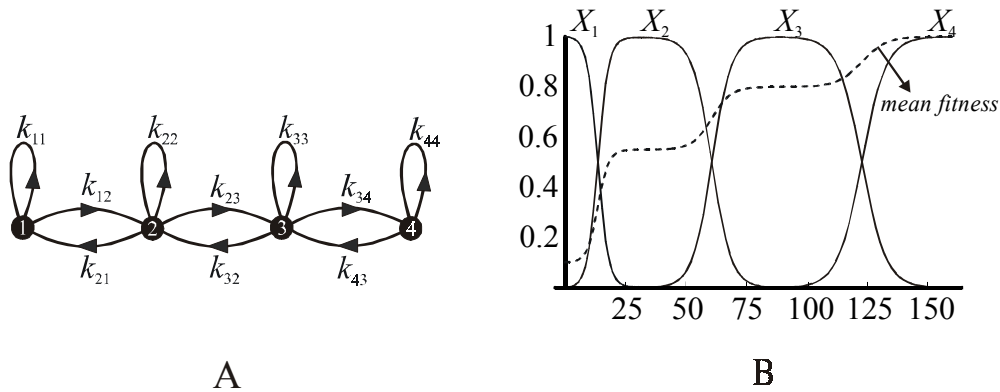


Obrázok 4. Priebeh frekvencií výskytu dominantných binárnych reťazcov dĺžky 6. Chemostat bol inicializovaný náhodne generovanými 1000 binárnymi reťazcami, po 5×10^6 krokoch dominantným riešením je optimálny reťazec $x_{\text{opt}}=(010000)$. V chemostate dlhú dobu preživa ešte iné riešenie $x=(001111)$, $\text{real}(x)=17/63$, ktoré je susedné s optimálnym riešením ale ktoré má od neho veľkú Hammingovu vzdialenosť (v teórii GA táto skutočnosť sa nazýva Hammingov útes).

4. Replikátory ako simulátory darwinovskej evolúcie

Eigenov systém replikátorov s mutáciou (t.j. s nepresnou autoreplikáciou) študovaný v 2. kapitole bude použitý k štúdiu molekúlárnej darwinovskej evolúcie. Uvažujme hypotetický systém obsahujúci 4 replikátory X_1, X_2, X_3 , a X_4 . Tieto replikátory majú tú vlastnosť, že ľubovoľný z nich X_i produkuje pomocou nepresnej autoreplikácie susedné replikátory $X_{i\pm 1}$, pozri obr. 5, diagram A. Ak počiatočná koncentrácia replikátora X_1 je $x_1(0)=1$, potom v priebehu časovej evolúcie systému vznikajú koncentračné vlny, ktoré sú postupne priradené replikátorom X_2, X_3 , a X_4 , pozri obr. 5, diagram B. Táto skutočnosť môže byť jednoducho interpretovaná ako prejav molekúlárnej darwinovskej evolúcie, kde fitness jednotlivých "druhov" je určené diagonálnymi rýchlostnými konštantami k_{ii} . Evolučný proces bol inicializovaný populáciou, ktorá obsahovala len replikátor X_1 . V dôsledku skutočnosti, že v

priebehu jeho replikácie môžu vznikáť nepresné kópie - mutanty (rýchlosť ich vzniku je špecifikovaná rýchlostnou konštantou k_{12}), nasledujúci replikátor X_2 s väčším fitness (rýchlostnou konštantou k_{22}) než akú mal jeho predchodca X_1 ($k_{11} < k_{22}$) v ďalšej etape evolúcie zvíťazí (t.j. jeho koncentrácia narastie na skoro jednotkovú hodnotu, zatiaľ čo koncentrácia X_1 klesne na veľmi malú hodnotu). Tento proces sa opakuje pre replikátor X_2 , ktorý je teraz považovaný za počiatočný replikátor (hrá rovnakú úlohu ako replikátor X_1 v predchádzajúcej etape evolúcie). Pretože replikácia X_2 nie je bezchybná, s malou pravdepodobnosťou je produkovaný replikátor X_3 s väčším fitness než aký mal jeho predchodca X_2 ($k_{33} > k_{22}$), t.j. nový "druh" X_3 dočasne zvíťazí. Tento proces je ukončený vtedy, keď posledný replikátor X_4 sa začne vyskytovať ako dôsledok nepresnej replikácie X_3 , jeho koncentrácia postupne narastie na skoro jednotkovú hodnotu a potom už zostáva časovo nemenná (asymptoticky stabilný stav).



Obrázok 5. Diagram (A) reprezentuje systém 4 replikátorov, kde replikátor X_i produkuje pomocou nepresných replikácií (mutácií) susedné replikátory X_{i-1} a X_{i+1} . Hrany grafu sú ohodnotené rýchlostnými konštantami, ktorých numerické hodnoty sú určené maticou \mathbf{K} (18). Diagonálne rýchlostné konštanty sú medzi sebou dobre odseparované a sú o mnoho väčšie ako nediagonálne rýchlostné konštanty. Diagram (B) znázorňuje priebehy koncentrácií replikátorov, ktoré tvoria postupnosť "koncentračných vln". Tento diagram tiež obsahuje priebeh priemerného fitness systému špecifikovaného výrazom $\bar{k} = k_{11}x_1 + \dots + k_{44}x_4$, ktorý tvorí typickú neklesajúcu "schodovú" funkciu.

Tieto úvahy môžeme zformalizovať na semikvantitatívnej úrovni. Predpokladajme, že replikátorový systém v čase t_0 je v takom tranzitnom stave, kde koncentrácia X_i je skoro jednotková, zatiaľ čo koncentrácie susedných replikátorov X_{i-1} a X_{i+1} sú zanedbateľne malé, t.j. $x_i(t_0) = 1 - 2\delta$, $x_{i-1}(t_0) = x_{i+1}(t_0) = \delta$, a $x_j(t_0) = 0$ pre zostávajúce koncentrácie, pričom δ je malé kladné číslo. Zried'ovací tok $\phi(t)$ (7) je určený $\phi(t_0) = \delta(k_{11} + k_{12}) + (1 - 2\delta)(k_{22} + k_{21} + k_{23}) + \delta(k_{33} + k_{32} + k_{34})$. Potom diferenciálna rovnica (5) pre $i=2$ má tvar

$$\dot{x}_2 = x_2(k_{22} - \phi) + x_1k_{12} + x_3k_{32} \quad (16)$$

Ak do tejto diferenciálnej rovnice dosadíme koncentrácie v čase t_0 a použijeme predpoklad, že $\dot{x}_2(t_0) = 0$ (koncentrácia replikátora X_2 v čase t_0 dosiahla maximum), dostaneme $0 = -(k_{21} + k_{23}) + \delta(k_{12} + k_{32})$, alebo

$$\delta = \frac{k_{21} + k_{23}}{k_{12} + k_{32}} \rightarrow 0 \Rightarrow k_{21} + k_{23} \ll k_{12} + k_{32} \quad (17)$$

Z predpokladu dobrej separovateľnosti koncentračných vln vyplývajú silné podmienky pre rýchlostné konštanty replikátorov, ktoré sú susedné s replikátorom X_2 ; môžeme povedať, že suma "vychádzajúcich" rýchlostných konštant musí byť podstatne menšia než ako suma "prichádzajúcich" rýchlostných konštant. Inými slovami povedané, "**pravdepodobnosť**"

vzniku daného replikátora v dôsledku nepresných autoreplikácií susedných replikátorov musí byť o mnoho väčšia než ako "pravdepodobnosť" jeho zániku v dôsledku jeho nepresnej autoreplikácie.

Tieto semikvantitatívne úvahy boli numericky verifikované pre 4-replikátorový systém s maticou rýchlostných konštánt

$$\mathbf{K} = \begin{pmatrix} 0.1 & 10^{-3} & 0 & 0 \\ 10^{-7} & 0.55 & 10^{-7} & 0 \\ 0 & 10^{-11} & 0.8 & 10^{-11} \\ 0 & 0 & 10^{-7} & 1.0 \end{pmatrix} \quad (18)$$

Táto matica vyhovuje podmienkam: (i) Jej diagonálne maticové elementy sú o mnoho väčšie než ako nediagonálne elementy, a (ii) nediagonálne rýchlostné konštanty vyhovujú nerovnosiam (17). Z predchádzajúcej diskusie vyplýva, že v tomto prípade môžeme očakávať darwinovské správanie sa systému. Toto očakávanie pekne súhlasí s numerickými výsledkami z obr. 5, diagram B, kde sú ukázané koncentračné profily, ktoré sú typické pre darwinovskú evolúciu. Môžeme teda na záver tejto kapitoly konštatovať, že **Eigenova fenomenologická teória replikátorov tvorí vhodný teoretický rámec pre počítačové simulačné štúdiá molekulárnej darwinovskej evolúcie** (t.j. biomakromolekúl, ktoré sú schopné replikačných proces, akými sú RNA alebo DNA).

5. Zovšeobecnená Kauffmanova funkcia KN

Nech \mathbf{g} je reťazec obsahujúci N celých čísel z $\{0, 1, \dots, p-1\}$

$$\mathbf{g} = (g_1 g_2 \dots g_N) \in \{0, 1, \dots, p-1\}^N \quad (19)$$

Každý index $1 \leq i \leq N$ je ohodnotený podmnožinou obsahujúcou $K+1$ náhodne vybraných indexov (včítane i) z $\{1, 2, \dots, n\}$

$$\Gamma(i) = \{j_1 < j_2 < \dots < j_{K+1}\} \subseteq \{1, 2, \dots, N\} \quad (20)$$

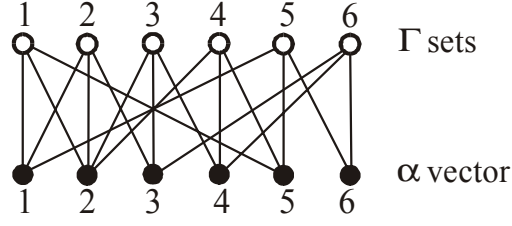
Zovšeobecnená Kauffmanova funkcia zobrazuje p -nárne vektory dĺžky N na reálne čísla z intervalu $\langle 0, 1 \rangle$

$$F(\mathbf{g}) = \frac{1}{N(F-1)} \sum_{i=1}^N \varphi(\langle \mathbf{g}, \Gamma(i) \rangle) = \frac{1}{N(F-1)} \sum_{i=1}^N \varphi(g_{j_1}, g_{j_2}, \dots, g_{j_{K+1}}) \quad (21)$$

kde $F-1$ je maximálne kladné celé číslo priradené pomocou funkcie φ špecificovanej pomocou náhodného generátora

$$\varphi(g_{j_1}, g_{j_2}, \dots, g_{j_{K+1}}) = \text{random} \left(\underbrace{\text{int}(g_{j_1}, g_{j_2}, \dots, g_{j_{K+1}})}_{\text{RandSeed}}, F \right) \quad (22)$$

kde celé číslo $\text{int}(g_{j_1}, g_{j_2}, \dots, g_{j_{K+1}}) = \sum_{l=0}^K g_{j_{K+1-l}} p^l$ je použité ako iniciátor (RandSeed) daného náhodného generátora celých čísel z $\{0, 1, \dots, F-1\}$.



Obrázok 6. Príklad Kauffmanovej funkcie špecifikovanej pre $N=6$ a $K=2$, podmnožiny Γ sú určené pomocou hrán medzi dolnými a hornými vrcholmi takto: $\Gamma(1)=\{1,2,5\}$, $\Gamma(2)=\{1,2,3\}$, $\Gamma(3)=\{2,3,4\}$, $\Gamma(4)=\{2,4,5\}$, $\Gamma(5)=\{1,5,6\}$, $\Gamma(6)=\{3,4,6\}$.

Pre lepšie pochopenie definície zovšeobecnenej Kauffmanovej funkcie, študujme jej konkrétnu realizáciu pre podmnožiny Γ z obrázku 5 a pre kvaternárny reťazec $\mathbf{g}=(022311)$ (t.j. $p=4$)

$$\begin{aligned}
 F(\mathbf{g}) &= \frac{1}{6(F-1)} (\varphi(021) + \varphi(022) + \varphi(223) + \varphi(231) + \varphi(011) + \varphi(231)) \\
 &= \frac{1}{6(F-1)} (\text{random}(9, F) + \text{random}(10, F) + \text{random}(43, F) \\
 &\quad + \text{random}(45, F) + \text{random}(5, F) + \text{random}(45, F))
 \end{aligned} \tag{23}$$

kde napr. výraz $\text{random}(9, F)$ vyjadruje nezáporné celé číslo generované náhodným generátorom inicializovaným $\text{RandSeed}=9$, ktoré je menšie ako F .

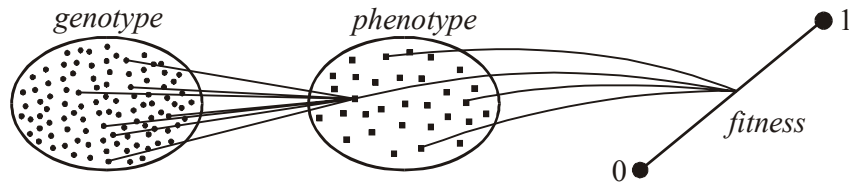
Reťazec $\mathbf{g}=(g_1g_2\dots g_N) \in \{0,1,\dots,p-1\}^N$ bude v našich ďalších úvahách interpretovaný ako genotyp špecifikujúci "genetickú informáciu" hypotetického organizmu a ktorý je predmetom darwinovskej evolúcie. Jednotlivé zložky fenotypu $\mathbf{ph}(\mathbf{g})=(ph_1ph_2\dots ph_N) \in Ph=\{0,1,\dots,F-1\}^N$ hypotetického organizmu sú určené

$$ph_i = \varphi(\langle \mathbf{g}, \Gamma(i) \rangle) = \varphi(g_{j_1}, g_{j_2}, \dots, g_{j_{k+1}}) \tag{24}$$

Konečne, fitness genotypu $\mathbf{g}=(g_1g_2\dots g_N)$ je určený pomocou priradeného fenotypu $\mathbf{ph}(\mathbf{g})=(ph_1ph_2\dots ph_N)$ ako suma jeho jednotlivých komponent (pozri rov. (21))

$$\text{fitness}(\mathbf{g}) = f(\mathbf{g}) = \frac{1}{N(F-1)} \sum_{i=1}^N ph_i \tag{25}$$

To znamená, že v danom prípade tak genotyp, ako aj fenotyp sú určené celočíselnými vektormi rovnakej dĺžky N , aktuálne hodnoty fenotypu sú špecifikované nielen podmnožinami Γ , ale taktiež aj implementáciou náhodného generátora celých čísel $\text{random}(\text{RandSeed}, F)$.



Obrázok 7. Znáznornenie zloženého zobrazenia (26). Obe čiastkové zobrazenia sú typu *mnoho-na-jeden*, t.j. mnoho genotypových reťazcov je zobrazených na jeden reťazec reprezentujúci fenotyp, a podobne, mnoho fenotypových reťazcov je zobrazených na jednu hodnotu fitness. Táto vlastnosť silnej redundantnosti kódovania genotypu je veľmi dôležitá pre existenciu neutrálnych epoch (stázií) v Darwinovej evolučnej teórii.

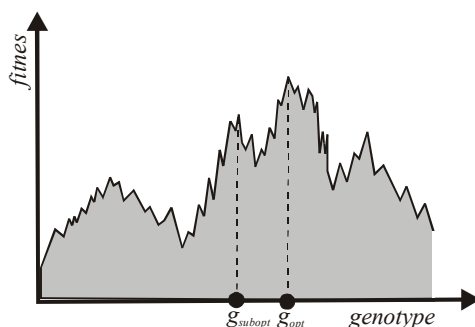
Vzájomný vzťah medzi elementami "triády" genotyp - fenotyp - fitness je formálne vyjadrený pomocou sekvencie dvoch zobrazení (pozri obr. see Fig. 7)

$$G \xrightarrow{ph} Ph \xrightarrow{fitness} [0,1] \quad (26)$$

To znamená, že najprv je genotyp zobrazovaný na fenotyp, a potom je fenotyp zobrazovaný na fitness. Skrátaná forma tohto zobrazovania môže byť tiež vyjadrená ako jedno zložené zobrazovanie

$$G \xrightarrow{f = fitness \circ ph} [0,1] \quad (27)$$

Toto nové zobrazovanie priamo mapuje genotyp na fitness bez nutnosti explicitného uvažovania fenotypu ako medzivýsledku. Pri tejto súvislosti môžeme si položiť otázku či je potrebné v našich teoretických štúdiách darwinovskej evolúcie explicitne uvažovať fenotyp ako sprostredkovateľa medzi genotypom a fitness. Z čisto matematického pohľadu fenotyp nie je potrebný explicitne uvažovať, genotyp môže byť zobrazovaný priamo na fitness. Avšak, z pohľadu názornej interpretácie darwinovskej evolúcie, fenotyp poskytuje výbornú heuristiku pre názorné pochopenie a interpretáciu jej mechanizmov a výsledkov. Tak napríklad, konkrétny fenotyp môže byť postulovaný ako evolučný cieľ, potom evolúcia je interpretovaná ako sekvencia fenotypov blížiac sa postupne k tomuto požadovanému fenotypu.



Obrázok 8. Povrch fitness bol pôvodne postulovaný v teórii darwinovskej evolúcie Sealom Wrightom [31], ktorý charakterizoval evolúciu ako adaptačný (optimalizačný) proces. V ideálnom prípade, výsledný genotyp populácie je totožný s optimálnym riešením g_{opt} s maximálnym fitness (globálne maximum na povrchu fitness funkcie), avšak môže nastať situácia, keď evolúcia dosiahne len suboptimálne riešenie g_{subopt} , ktorého fitness je menšie ako fitness optimálneho riešenia.

Ako už bolo poznamenané v predchádzajúcej časti tejto prednášky, Wright charakterizoval darwinovskú evolúciu ako adaptačný (optimalizačný) proces na povrchu fitness, ktorý predstavuje priame zobrazovanie (27) genotypu na fitness (pozri obr. 8). Populácia v priebehu adaptácie pomocou darwinovskej evolúcie hľadá globálne maximum na povrchu fitness

$$g_{opt} = \arg \max f(g) \quad (28)$$

Tento zložitý optimalizačný problém bude v nasledujúcej časti prednášky riešený pomocou metód umelej chémie, ktoré sú založené na metafore chemostatu a Eigenových replikátorov. Bude ukázané, že prístup chemostat poskytuje efektívny prístup k riešeniu optimalizačného problému (28), t.j. "replikátorové" *metódy umelej chémie sú mimoriadne vhodné k simulovaniu molekulovej darwinovskej evolúcie.*

6. Vlastnosti povrchu fitness určeného zovšeobecnenou Kauffmanovou K/N funkciou

Jeden zo základných konceptov teórie molekulárnej darwinovskej teórie sú neutrálne grafy [12,27,28], ktoré sú určené ako podgrafy tzv. *genotypového* grafu (pozri obr. 8)

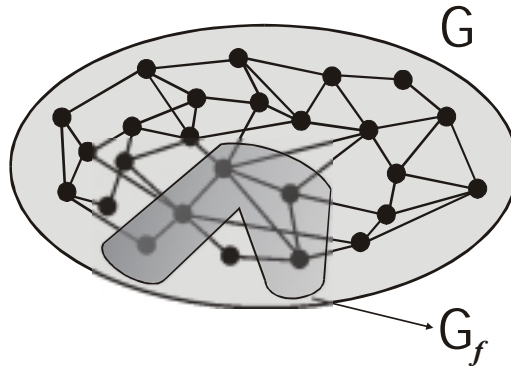
$$G = (V = \{0,1,\dots,p-1\}^N, E) \quad (29)$$

Jeho vrcholy sú reťazce - genotypy, pričom dva vrcholy sú spojené hranou, ak Hammingova vzdialenosť medzi nimi je jednotlová (t.j. jeden reťazec môže byť pretransformovaný na druhý reťazec pomocou 1-bodovej mutácie)

$$(g, g') \in E \Leftrightarrow (g' = O_{1-mut}(g)) \wedge (g = O_{1-mut}(g')) \quad (30)$$

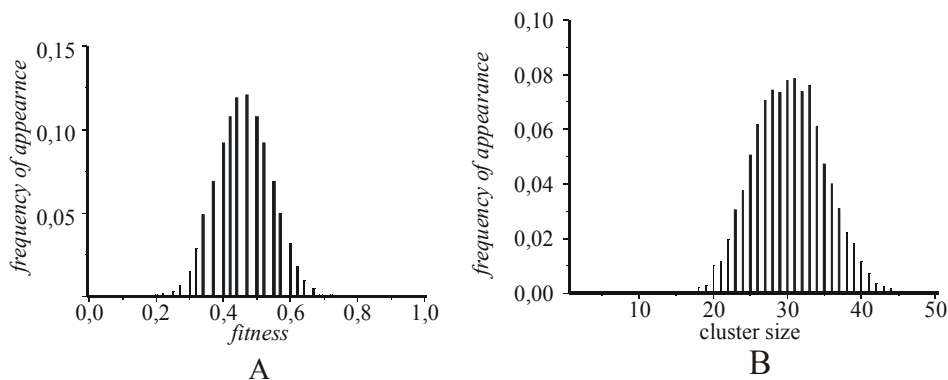
Genotypový graf G má $|V| = p^N$ vrcholov a $|E| = p^N \cdot (p-1)^N / 2$ hrán. Tento graf môže byť rozložený na tzv. *neutrálne grafy*, $G_f = (V_f, E_f)$, ktoré sú indukované vrcholmi (reťazcami) s rovnakým fitness (pozri obr. 8)

$$V_f = \{g \in V; fitness(g) = f\} \quad (31)$$

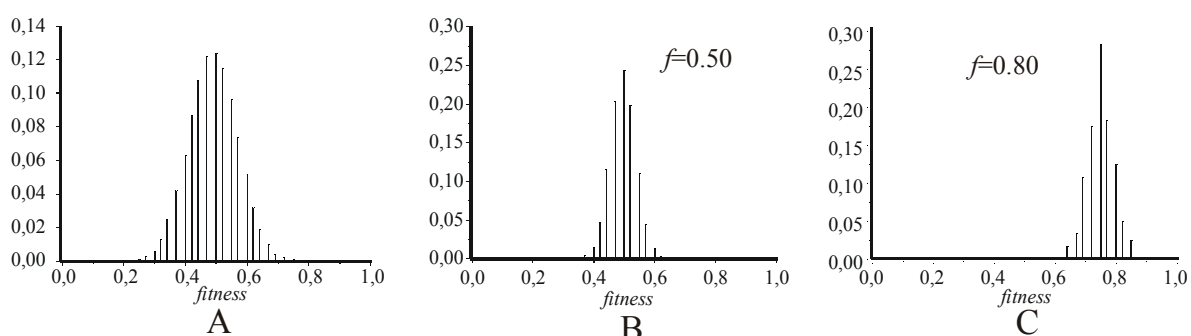


Obrázok 9. Genotypový graf G obsahuje ako vrcholy všetky možné genotypy - reťazce, dva vrcholy sú spojené hranou vtedy, ak ich Hammingova vzdialenosť sa rovná jednej. Neutrálny (sub)graf $G_f \subset G$ je indukovaný tými vrcholmi G , ktoré sú ohodnotené rovnakým fitness f .

Distribúcia fitness v genotypovom grafe G pre špecifické hodnoty parametrov K , N , p , a F je ukázaná na obr. 9. Pre dostatočne veľké hodnoty parametra K (počet susedov v indexových podmnožinách Γ) náhodne generované kvaterné reťazce ($p=4$) majú fitness s najväčšou pravdepodobnosťou výskytu okolo 0.5. Ako ukazujú naše simulačné výpočty, pre menšie hodnoty K , maximálna frekvencia výskytu fitness náhodne generovaných reťazcov je podstatne posunutá k menším hodnotám ako je 0.5, a navyše, začína byť silne závislá na aktuálnej špecifikácii pomocnej funkcie φ , pozri rov. (22). Táto závislosť φ na aktuálnej špecifikácii je potlačená zväčšovaním hodnôt parametra K . Obrázok 9 tiež ukazuje histogram veľkosti klastrov neutrálnych grafov G_f , kde pod pojmom "*klastery*" rozumieme podmnožinu reťazcov - vrcholov, ktorá indukuje súvislý graf podgraf. To znamená, že daný neutrálny graf G_f môže byť rozložený na súvislé podgrafy nazývané *klastery*. Vo vnútri daného klastra môžeme "cestovať" pomocou 1-bodových mutácií, avšak medzi rôznymi klastrami neexistujú priame spoje - hrany (to znamená, že pomocou 1-bodovej mutácie nemôžeme preskočiť z jedného do druhého klastra).



Obrázok 10. (A) Histogram vyjadruje frekvenciu výskytu fitness náhodne generovaných reťazcov pre parametre $N=40$, $p=4$, $F=2$, a $K=3$. (B) Histogram vyjadruje výskytu veľkosti klastrov pre náhodne generované reťazce (s rovnakými parametrami ako v predchádzajúcom histograme).



Obrázok 11. Numerické experimenty s efektom 1-bodových mutácií na zmenu fitness náhodne generovaných reťazcov pre parametre $N=40$, $p=4$, $F=2$, and $K=4$. Diagram (A) ukazuje histogram fitness reťazcov, ktoré vznikli 1-bodovou mutáciou z náhodne vygenerovaného reťazca. Diagramy (B-C) znázorňujú histogramy fitness vytvorené 1-bodovou mutáciou z náhodne generovaných reťazcov, ktoré majú fixovaný fitness pre $f=0.50$ a $f=0.80$. V prípade väčšieho pôvodného fitness (diagram C) vytvorené reťazce 1-bodovou mutáciou majú distribúciu fitness posunutú k menším hodnotám. To znamená, že pre reťazce s vysokým fitness pravdepodobnosť vzniku nových reťazcov s väčším fitness ako pôvodným je už podstatne menšia, ako v prípade malého pôvodného fitness.

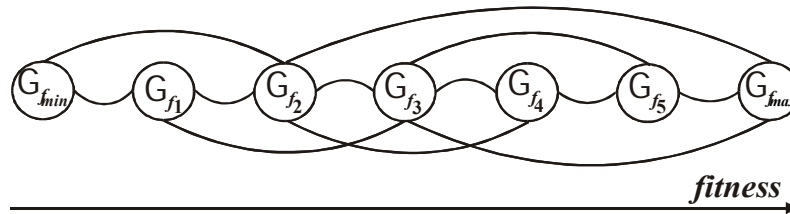
Efekt 1-bodových mutácií na zmenu fitness reťazcov je veľmi dôležitý pre lepšie pochopenie podrobností mechanizmu predneseného modelu molekulárnej darwinovskej evolúcie. Menovite, je potrebné popísať možnosti 1-bodových mutáčnych "skokov" z jedného neutrálneho grafu G_f na druhý neutrálny graf $G_{f'}$, pričom $f' > f$. Obrázok 11 obsahuje histogramy našich numerických experimentov s pôsobením 1-bodových mutácií na zmenu fitness. Najdôležitejší výsledok je znázornený na histogramoch B a C, ktoré ukazujú podstatnú zmenu rozdelenia výskytu fitness reťazcov, ktoré vznikli 1-bodovou mutáciou z reťazcov s prepísaným fitness. Ak pôvodné reťazce majú už vysoké fitness, potom z nich vytvorené reťazce 1-bodovou mutáciou majú distribúciu fitness podstatne posunutú k menším hodnotám ako bola pôvodná hodnota. Z tohto pozorovania vyplýva, že na záver evolúcie sa rýchlosť zmeny fitness k väčším hodnotám podstatne znižuje.

Genotypový graf môže byť vyjadrený v kondenzovanej reprezentácii tak, že jednotlivé neutrálne grafy sú priradené vrcholom kondenzovaného grafu, pričom jednotlivé vrcholy sú spojené hranou ak medzi pôvodnými neutrálnymi grafmi existovala aspoň jedna hrana, pozri obr. 12. Každá hrana z kondenzovaného grafu môže byť ohodnotená pravdepodobnosťou

prechodu z jedného neutrálneho grafu G_f na druhý neutrálny graf $G_{f'}$, $Pr(G_f \rightarrow G_{f'})$, pozri obr. 13, [12,27,28]⁷

$$P(G_f \rightarrow G_{f'}) = \frac{|V_{f'} \cap \partial V_f|}{|\partial V_f|} \quad (32)$$

kde ∂V_f je okolie neutrálnej podmnožiny zostrojenej pomocou 1-bodových mutácií. Numerická hodnota tejto pravdepodobnosti je určená ako podiel dvoch kardinalít $|V_{f'} \cap \partial V_f|$ a $|\partial V_f|$; pretože $|V_{f'} \cap \partial V_f| < |\partial V_f|$, potom $0 < P(G_f \rightarrow G_{f'}) < 1$. Ak platí $|V_{f'} \cap \partial V_f| \ll |\partial V_f|$, potom táto pravdepodobnosť je veľmi malé číslo, obrazne povedané, môžeme povedať, že prechod z jedného neutrálneho grafu G_f na druhý neutrálny graf $G_{f'}$ je veľmi vzácna udalosť. Diagramy B a C na obr. 11 poskytujú tieto pravdepodobnosti prechodu (sú úmerné výške jednotlivých stĺpcov). Tak napríklad z grafu D môžeme vydedukovať, že pravdepodobnosť $Pr(G_{0,8} \rightarrow G_{0,9})$ je veľmi malé číslo, t.j. „preskoku“ pomocou 1-bodovej mutácie z neutrálneho grafu $G_{0,8}$ do neutrálneho grafu $G_{0,9}$ je extrémne vzácna udalosť.



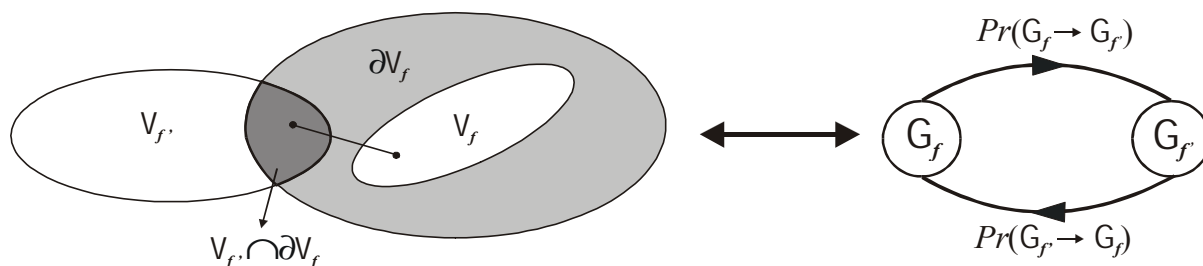
Obrázok 12. Kondenzovaná forma genotypového grafu G vznikne tak, že príslušné neutrálne grafy G_f sú lineárne usporiadané podľa rastúceho fitness, pričom dva vrcholy – neutrálne grafy sú spojené hranou vtedy, ak v pôvodnom fenotypovom grafe existovala aspoň jedna hrana medzi hranami patriacimi k týmto rôznym neutrálnym grafom. To znamená, že ak nejaký reťazec patriaci do neutrálneho grafu G_f môže byť transformovaný pomocou 1-bodovej mutácie na iný reťazec z neutrálneho grafu $G_{f'}$, kde $f' \neq f$, len vtedy, ak oba neutrálne grafy sú spojené hranou. Pomocou ohodnotenia jednotlivých hrán pravdepodobnosťami môžeme kvantitatívne odhadnúť 1-bodové mutačné skoky z jedného neutrálneho grafu do druhého neutrálneho grafu.

Vráťme sa na moment k nerovnosti (17), ktorá vyplýva z Eigenovej teórie replikátorov a ktorej splnenie je podmienkou dobrej separovateľnosti koncentračných profilov v priebehu evolúcie. K tomu, aby koncentrácia reťazcov z neutrálneho grafu G_f bola dobre separovaná od koncentrácie reťazcov zo susedných neutrálnych grafov $G_{f-previous}$ respkt. $G_{f-later}$, je potrebné, aby pravdepodobnosti vyhovovali podmienke

$$Pr(G_{f-previous} \rightarrow G_f) \gg Pr(G_f \rightarrow G_{f-later}) \quad (33)$$

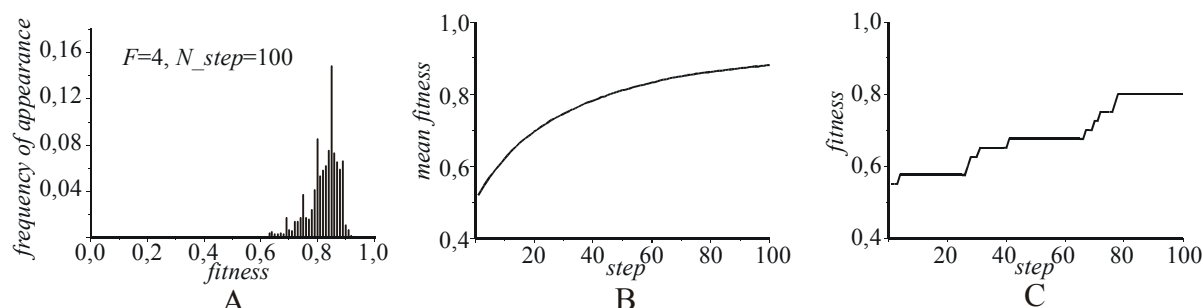
t.j. klesajúcu postupnosť v priebehu celej evolúcie. Ako vyplýva z našich simulačných výsledkov (pozri obr. 11) táto podmienka dobrej separovateľnosti koncentračných profilov reťazcov patriacich do daného neutrálneho grafu je dobre splnená menovite v terminálnej fáze evolúcie. Vtedy pre reťazce s vyšším fitness pomocou 1-bodovej mutácie len vzácne vytvoríme reťazec s ešte vyšším fitness.

⁷ J. Pospíchal ma upozornil, že táto definícia pravdepodobnosti $Pr(G_f \rightarrow G_{f'})$ nie je „kóšer“, pretože kardinalita okolia ∂V_f by mala obsahovať počet hrán, ktoré mieria z V_f von do okolia, a nie len počet vrcholov dosiahnutých z množiny V_f pomocou 1-bodových mutácií.



Obrázok 13. Množinovo-teoretický znázornenie 1-bodových mutačných transformácií z neutrálneho grafu G_f na iný neutrálny graf $G_{f'}$. Pretože tieto pravdepodobnosti nie sú symetrické, $Pr(G_f \rightarrow G_{f'}) \neq Pr(G_{f'} \rightarrow G_f)$, každá hrana v kondenzovanom genotypovom grafe musí byť nahradená dvoma opačne orientovanými grafmi, pričom každá z nich je potom ohodnotená danou pravdepodobnosťou.

Na záver tejto kapitoly obrátíme našu pozornosť na možnosť optimalizácie samotných reťazcov metódou lokálneho prehľadávania⁸ (hill climbing). Táto jednoduchá optimalizačná metóda spočíva v opakovanom aplikovaní nasledujúcej jednoduchšej prehľadávacej procedúry: Pre daný reťazec zostrojíme všetky možné reťazce pomocou 1-bodových mutácií. V tomto okolí nájdeme reťazec s najvyšším fitness, ktorý v nasledujúcom prehľadávacej procedure bude použitý ako východzí reťazec (stred). Tento jednoduchý lokálny prehľadávaci proces je opakovaný predpísaný počet-krát (v našich výpočtoch 100 krát). Na obr. 14 sú prezentované typické výsledky získané pomocou optimalizácie lokálneho prehľadávania. Diagram A znázorňuje distribúciu fitness výsledného reťazca, ktorý bol získaný použitím metódy lokálneho prehľadávania aplikovanej na náhodne generovaný reťazec. najčastejšie sa vyskytujúce fitness je okolo 0.9. To znamená, že metóda lokálneho prehľadávania s najväčšou pravdepodobnosťou neposkytuje optimálne riešenie s jednotkovým fitness.

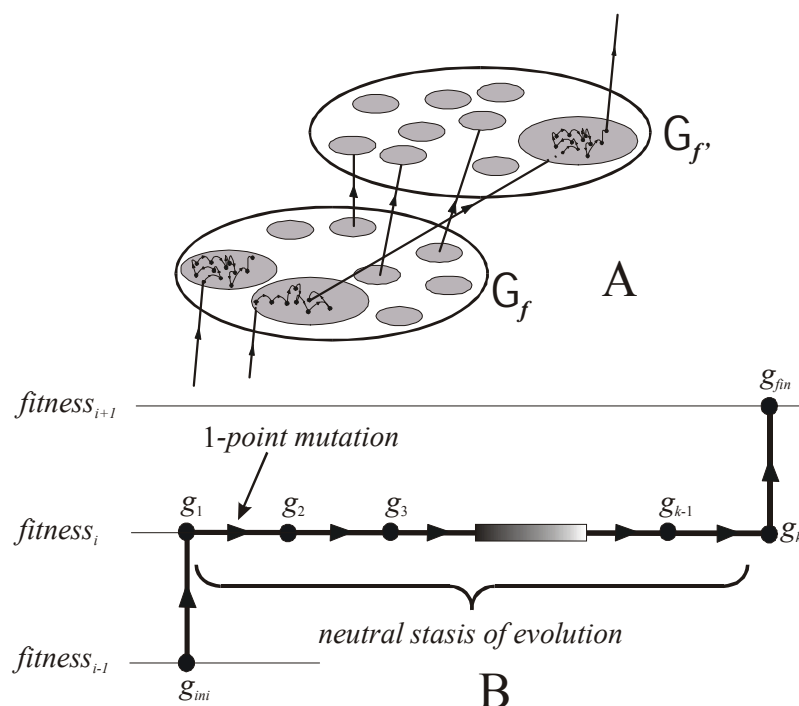


Obrázok 14. Optimalizácia pomocou lokálneho prehľadávania (hill-climbing optimization) náhodne vygenerovaného reťazca pre parametre $N=40$, $p=4$, $F=2$, and $K=3$. Diagram A znázorňuje histogram fitness reťazcov rezultujúcich z lokálnych optimalizácií, ktoré boli inicializované náhodne generovanými reťazcami. Diagram B znázorňuje priemerné fitness pre jednotlivé kroky lokálnych optimalizácií. Oba diagramy A a B boli vytvorené pomocou 10^4 optimalizácií lokálneho prehľadávania. Diagram C je ilustračný príklad náhodne zvolenej jednotlivej lokálnej optimalizácie, priebeh fitness pre jednotlivé optimalizačné kroky vytvára typický „schodový“ priebeh, ktorý je typický pre molekulárnu darwinovskú evolúciu [29,30].

Toto pozorovanie môže byť jednoducho vysvetlená skutočnosťou, že pre klastre s vysokým fitness je len malá šanca k uskutočneniu „skoku“ pomocou 1-bodovej mutácie do iného klastru s vyšším fitness. Diagram B znázorňuje priemerné fitness jednotlivých krokov optimalizačnej metódy lokálneho prehľadávania. Priebeh tohto grafu znázorňuje monotónny nárast fitness, ktorý na záver optimalizácie dosiahne hodnotu priemerného fitness približne 0.9. Diagram C znázorňuje typický priebeh fitness náhodne vybraného jednej optimalizácie.

⁸ V učebnici [25] o evolučných algoritmoch sme použili „úsmevný“ preklad termínu “hill-climbing optimization” ako “horolezecký optimalizačný algoritmus”.

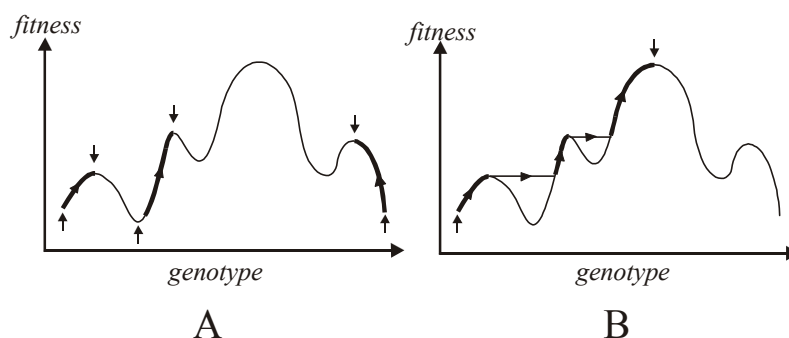
Tento diagram znázorňuje typický „monotónne-schodový“ priebeh fitness pre molekulárnu darwinovskú evolúciu, kde existujú relatívne dlhé neutrálne epochy, v ktorých sa fitness nemení. Táto skutočnosť znamená existenciu dlhej série neutrálnych mutácií, v priebehu ktorých optimalizačná metóda lokálneho prehľadávania hľadá pomocou týchto neutrálnych mutácií šancu pre „skok“ do nasledujúceho neutrálneho grafu s väčším fitness (pozri obr. 12). Obrazne povedané, dĺžka neutrálnej etapy je úmerná veľkosti neutrálneho klastra v ktorom sa optimalizačná metóda práve nachádza (pozri obr. 15). Na záver môžeme konštatovať, že jednoduchá optimalizačná metóda lokálneho prehľadávania v podstate simuluje priebeh molekulárnej darwinovskej evolúcie, jej hlavné obmedzenie je (na rozdiel od štandardných evolučných algoritmov založených na populácii riešení) len vzácné dosiahne globálne optimum (nie je schopná prekonať bariéru pascí existencie neutrálnych klastrov s vyšším fitness).



Obrázok 15. (A) Schematické znázornenie neutrálnych epoch medzi dvoma „skokmi“ z jedného neutrálneho grafu G_f do druhého neutrálneho grafu $G_{f'}$. V počiatočnej fáze lokálneho prehľadávania vzniká sekvencia reťazcov vo vnútri klastra, prvé riešenie v sekvencii odovpovedá výsledku posledného „skoku“ a posledné riešenie vyjadruje „skok“ do nasledujúceho neutrálneho grafu $G_{f'}$, potom neutrálna epocha sa znovu opakuje až dovtedy, kým sa neobjaví vhodný reťazec vo vnútri daného klastra, ktorý umožní ďalší „skok“. ľavej dolnej časti tohto diagramu je znázornený klaster (v rámci neutrálneho grafu G_f), z ktorého neexistuje „skok“ do neutrálneho grafu s vyšším fitness, tento klaster vyjadruje „pasce“ pre metódu lokálneho prehľadávania. (B) Podrobnejšie znázornenie jednej neutrálnej etapy v rámci jedného klastra neutrálneho grafu.

Niekoľko všeobecných poznámok k významu neutrálnych mutácií pre molekulárnu darwinovskú evolúciu. Ako už bolo povedané v kapitole 5, existencia alebo neexistencia neutrálnych mutácií závisí od redundantnosti zobrazenia genotypu na fitness. V prípade, že táto redundantnosť je potlačená na minimum (zobrazenie genotypu na fitness je "skoro" jedno-jednoznačné, každý genotyp je ohodnotený fitness vzájomne jednoznačne), potom neutrálne neexistujú. Preto ak chceme mať zabezpečenú existenciu neutrálnych mutácií musíme postulovať silnú redundantnosť zobrazenia genotypu na fitness (toto zobrazenie je typu mnoho-na-jeden, pozri obr. 6). Aj keď neutrálne mutácie môžu byť extrémne náročné na čas (dlhé neutrálne epochy), ich existencia umožňuje relatívne jednoduchým stochastickým (slepým) spôsobom prekonať "pasce" lokálnych miním na povrchu fitness (pozri obr. 16). Táto výhodná potenciálna možnosť úniku lokálnych miním sa ešte umocní vtedy, keď

rovnaká lokálna optimalizácia sa vykonáva vysoko paralelným spôsobom v populácii. Prípadne genotypy, ktoré "zamrzli" v lokálnych minimách, po určitom čase začnú byť nevýhodné (s menším fitness ako priemerným) a preto začnú byť z populácie vytesňované. Táto kombinácia neutrálnych mutácií a vysokou paralelnosťou lokálnych optimalizácií v populácii vytvára veľmi efektívnu "evolučnú stratégiu" riešenia zložitého optimalizačného problému typu (28) tak, aby bolo získané globálne optimálne riešenie.



Obrázok 16. Schematické znázornenie priebehu molekulovej darwinovskej evolúcie v prípade, že neexistujú neutrálne mutácie (diagram A), alebo existujú neutrálne mutácie (diagram B). V prvom prípade evolučná adaptácia (optimalizácia) skončí na najbližšom vrchole povrchu fitness (v tomto prípade môžeme povedať, že molekulárna darwinovská evolúcia je lokálny optimalizátor - fitness genotypu populácie "lezie" po povrchu fitness do najbližšieho maxima). V prípade, že uvažujeme neutrálne mutácie (povrchu fitness obsahuje mnohé neutrálne "platá" s konštantnou hodnotou fitness pre rôzne genotypy), lokálny charakter optimalizácie darwinovskej evolúcie môže byť preklopený práve neutrálnymi oblasťami, ktoré umožňujú efektívne prekonať "údolia" s nižšou hodnotou fitness (diagram B). Týmto spôsobom darwinovská evolúcia má veľkú šancu dostať sa do globálneho maxima na povrchu fitness.

7. Molekulárna darwinovská evolúcia

V predchádzajúcej kapitole bolo ukázané, že jednoduchá optimalizačná metóda lokálneho prehľadávania poskytuje výsledky, ktoré sú veľmi podobné očakávaným výsledkom zo všeobecnej molekulárnej darwinovskej evolúcie, pozri obr. 13. Tento zjednodušený prístup k simulácii *in-silico* darwinovskej evolúcie bude teraz nahradený podstatne všeobecnejším prístupom založeným na metafore chemostatu. Predpokladajme, že chemostat - populácia je multimimiožina, ktorá obsahuje p.nárne reťazce dĺžky N

$$P = \{ \dots, (001\dots 0\dots 4), \dots \} \in \{0, 1, \dots, p-1\}^N \quad (33)$$

Zloženie chemostatu bude charakterizované dvoma parametrami: *homogenitou* a *entropiou*. Ich hlavný význam spočíva v tom, že umožňujú detailný pohľad a interpretáciu molekulárnej darwinovskej evolúcie. Nech i -tý reťazec populácie P je označený $\mathbf{g} = (g_1 g_2 \dots g_N) \in \{0, 1, \dots, p-1\}^N$, kde $0 \leq g_j^{(i)} \leq p-1$. Pravdepodobnosti výskytu hodnoty $0 \leq g \leq p-1$ v celej populácii na i -tom mieste v reťazcoch tvoria p -dimenzionálny vektor

$$\boldsymbol{\omega}_i = (\omega_0^{(i)}, \omega_1^{(i)}, \dots, \omega_{p-1}^{(i)}) \quad (34a)$$

$$\omega_j^{(i)} = \frac{1}{|P|} \sum_{k=1}^{|P|} \delta(j, g_j^{(k)}) \quad (34b)$$

kde Kroneckerovo delta je $\delta(i, g) = 1$ (pre $i = g$) a $\delta(i, g) = 0$ (pre $i \neq g$). Ak všetky elementy $\boldsymbol{\omega}_i$ sú rovné $1/p$, potom i -tá hodnota v populácii reťazcov je plne náhodná. Homogenita i -tého elemntu je určená takto

$$0 \leq \chi_i(P) = \frac{P}{p-1} \sum_{j=0}^{p-1} \left(\omega_j^{(i)} - \frac{1}{p} \right)^2 \leq 1 \quad (35)$$

Ak $\chi_i=0$ ($\chi_i=1$), potom i -tý element v celej populácii sú úplne náhodné (jednoznačné). Konečne, populačná homogenita je určená ako aritmetický priemer všetkých parciálnych homogenít (35)

$$\chi(P) = \frac{1}{N} \sum_{i=1}^N \chi_i(P) \quad (36)$$

Entropia populácie je určená pomocou distribúcie pravdepodobnosti $0 \leq w_f \leq 1$ (kde $\sum_f w_f = 1$) (veľičina w_f popisuje pravdepodobnosť výskytu reťazcov s fitness f)

$$\sigma(P) = - \sum_f w_f \ln w_f \quad (37)$$

Ak všetky reťazce populácie P majú rovnaké fitness, potom entropia populácie je nulová $\sigma(P)=0$, v opačnom prípade, ak populácie obsahuje reťazce s rôznym fitness, potom entropia je kladná, $\sigma(P)>0$ (jej veľkosť určuje mieru populačnej heterogenity vzhľadom k fitness).

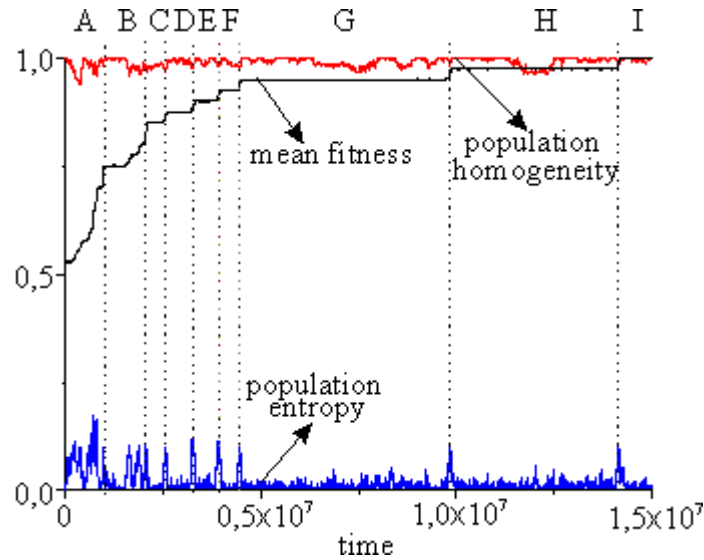


Fig. 17. Numerické výsledky simulácie molekularnej darwinovskej evolúcie pomocou metafory chemostatu. Dolný priebeh (modrý) znázorňuje entropiu chemostatu - populácie, prostredný priebeh (čierny) znázorňuje priemerný fitness populácie a vrchný priebeh (červený) znázorňuje homogenitu populácie. V prvej etape evolúcie (epochy A a B) v populácii existuje prudký nárast priemerného fitness, epochy C - F možno charakterizovať ako relatívne krátkodobé neutrálne epochy, a konečne, v posledné dve epochy G - H sú relatívne dlhodobé neutrálne epochy, v ktorých sa priemerný fitness dlhodobo nemení. Z priebehu homogenity v priebehu celej evolúcie vyplýva, že pre zvolenú hodnotu pravdepodobnosti 1-bodovej mutácie $P_{mut}=10^{-6}$, populácia je vysoko homogénna, čo podstatne "rehabilituje" 1-reťazovú aproximáciu molekularnej darwinovskej evolúcie pomocou lokálneho prehľadávania (pozri obr. 13).

Predpokladajme, že v chemostate P prebieha jednoduchá replikácia (ktorá nie je presná)

$$\mathbf{g} \xrightarrow{f(\mathbf{g})} \mathbf{g} + \mathbf{g}' \quad (38)$$

kde $\mathbf{g}' = O_{mut}(\mathbf{g})$ je stochastická mutácia reťazca \mathbf{g} , pričom mutácia je určená 1-bitovou pravdepodobnosťou P_{mut} . Algoritmus chemostatu bol v kapitole 3 špecifikovaný v pseudo-Pascale vo forme Algoritmu 1. Fitness reťazcov bol určený pomocou zovšeobecnenej Kauffmanovej funkcie s parametrami $N=40$, $p=4$, $F=2$, and $K=3$, pričom chemostat obsahuje 1000 reťazcov a $P_{mut}=10^{-6}$. Chemostat bol inicializovaný populáciou reťazcov náhodne

generovaných. Numerické výsledky sú znázornené na obr. Numerické výsledky sú znázornené na obr. 16. Hlavný záver z týchto výsledkov je, že priebeh priemerného fitness tvorí "schodovú" neklesajúcu funkciu, ktorá je typickou vlastnosťou molekulárnej darwinovskej evolúcie. Podobný výsledok bol získaný aj v predchádzajúcej 6. kapitole použitím jednoduchej optimalizačnej metódy lokálneho prehľadávania aplikovanej k jednému reťazcu (pozri obr. 13). Priebeh entropie na obr. 16 je neustále skoro nulový, len v oblastiach prudkého zvýšenia fitness dochádza ku vzniku entropických pík, ktoré charakterizujú populáciu ako "skokovú", v ktorej sa krátkodobo súčasne vyskytujú reťazce s rôznymi fitness. Faktor homogenity populácie je neustále skoro. Z tohto priebehu bezprostredne vyplýva, že populácia v priebehu celej evolúcie vykazuje vysoký stupeň homogenity. Vyskytujúce sa stochastické fluktuácie homogenity môžu byť vysvetlené ako prejav neutrálnych mutácií v priebehu dlhodobých neutrálnych epoch. Ako už bolo uvedené v legende k obr. 14, tieto neutrálne mutácie sú nutné k príprave populácie k spontánnej emergencii mutačného "skoku" do iného neutrálneho grafu s vyšším fitness, pričom dĺžka týchto neutrálnych epoch je úmerná pravdepodobnostiam $Pr(G_f \rightarrow G_{f'})$ definovanými v 6. kapitole. V prípade, že daná pravdepodobnosť je veľmi malá, potom časové trvanie neutrálnej epochy môže byť mimoriadne dlhé.

8. Súhrn a diskusia

Hlavným cieľom tejto prednášky je ukázať, že *in-silico* simulácie molekulárnej darwinovskej evolúcia poskytujú detailný pohľad na jej priebeh a mechanizmy. Autoreplikujúce biomakromolekuly v našich simuláciách boli reprezentované pomocou p -nárnych reťazcov, pričom ich fitness bol počítaný pomocou zovšeobecnenej Kauffmanovej funkcie. Tento prístup k *in-silico* simulácii darwinovskej evolúcie bol prednedávnom použitý Newmanom a Engelhardtom [26], ktorí na základe svojich výsledkov dospeli k veľmi dôležitému záveru, že tento jednoduchý „formálny“ prístup simuluje v podstate všetky výsledky získane pomocou fyzikálneho modelu RNA a jej foldingu na sekundárnu štruktúru, ktoré boli dosiahnuté teoretickým chemikom (tiež aj teoretickým biológom) Petrom Schusterom so spolupracovníkmi [12,27,28]. Naše výsledky, pozorovania a zovšeobecnenia môžu byť sumarizované takto:

- (1) **Eigenova teória replikátorov** (2. kapitola) tvorí solídny fyzikálno-chemický základ molekulárnej darwinovskej evolúcie. Riešenie jej diferenciálnych rovníc špecifikovaných plauzibilným výberom rýchlostných konštánt poskytuje časové priebehy relatívnych koncentracii replikátorov a priemerného fitness v tvaroch, ktorý je je obvykle hypoteticky predpokladaný pre darwinovskú evolúciu. (pozri obr. 5).
- (2) Darwinovská evolúcia je komplikovaná interakcia medzi **monodovskou náhodou a nutnosťou**, medzi deterministickými a stochastickými procesmi, ktoré sú jej integrálnou súčasťou. Obsahuje časti, ktoré sú plne deterministické a predikovateľné (napr. zobrazenie genotypu na fenotyp), ale tiež obsahuje aj časti, ktoré majú silný stochastický charakter a môžu byť špecifikované len použitím štatistickej terminológie (napr. mutácie).
- (3) Idea teoretického genetika Seala Wrighta o **povrchu fitness** (adaptive landscape) má veľký heuristický význam pre interpretáciu darwinovskej evolúcie a môže byť považovaná za jeden z najväčších teoretických prínosov neodarwinizmu. Podľa tejto idey darwinovská evolúcia môže byť interpretovaná ako špeciálna forma **evolučného**

algoritmu [11,15] pre riešenie extrémne komplikovaných kombinatoriálnych optimalizačných problémov.

- (4) Použitý model poskytuje jednoduchý mechanizmus pre vysvetlenie významu **neutrálnych mutácií** pre darwinovskú evolúciu (pozri obr. 15 a 16). Neutrálne mutácie na povrchu fitness sú dôležité pre prekonanie „evolučných pascí“ lokálnych maxim. Darwinovská evolúcia obsahuje neutrálne epochy, kde priemerné fitness zostáva dlhodobo konštantné, ale zloženie reťazcov populácie sa pomaly mení smerom k reťazcom, ktoré sú schopné „skoku“ 1-bitovej mutácie do neutrálneho grafu s vyšším fitness.
- (5) **Časová orientácia** darwinovskej evolúcie je jednoznačná a jej základným prejavom je existencia neklesajúceho priemerného fitness populácie. Reťazce s väčším fitness sú reprodukované častejšie ako reťazce s malým fitness, „silnejšie“ reťazce v priebehu evolúcie majú evolučnú výhodu vzhľadom k iným reťazcom. Inými slovami povedané, darwinovská evolúcia je progresívna zmena populačného genotypu, kde priemerný fitness je neklesajúci v priebehu celej evolúcie.
- (6) Semikvantitatívna analýza Eigenových diferenciálnych rovníc (pozri 4. kapitolu) poskytuje podmienky (nerovnosti) pre rýchlostné konštanty replikátorov, tak aby **koncentračné profily jednotlivých replikátorov boli dobre separované** v priebehu časovej evolúcie replikátorového systému (pozri obr. 5, diagram B). Pomocou numerických simulácií bolo ukázané, že tieto podmienky sú taktiež splnené aj pre pravdepodobnosti prechodu z jedného neutrálneho grafu na iný neutrálny graf tak, aby sa fitness zväčšovalo. Táto skutočnosť je potvrdením nášho predpokladu, že Kauffmanova *KN* funkcia vhodne aproximuje povrch fitness biomakromolekulárnych replikátorov tak, že použitie chemostatu v podstate dáva formálne identické výsledky s Eigenovou teóriou replikátorov použitou ako fenomenologický prístup k molekulárnej darwinovskej evolúcii.
- (7) Dve rôzne časové škály môžu byť rozlíšené v darwinovskej evolúcii, a to adaptívny stav a neutrálny stav. **Adaptívny stav** je priradený náhlym zmenám priemerného fitness (pozri obr. 17), kde koexistujú rôzne fenotypy. Starý fenotyp má menší fitness ako novovznikajúci fenotyp. Pretože pravdepodobnosť replikácie reťazca je úmerná jeho fitness, reťazce odpovedajúce novému fitness sa častejšie replikujú ako staré. Nové reťazce s väčším fitness zvíťazia v priebehu niekoľkých evolučných krokov, t.j. adaptívny stav, v ktorom sa stare fenotypy vymieňajú za nové fenotypy, sa môže zdať externému pozorovateľovi ako extrémne rýchla evolučný proces. **Neutrálny stav** darwinovskej evolúcie spočíva v dlhodobej neutrálnej etape (stáze), kde výskyt neutrálnych mutácií je významný pre stochastickú prípravu nasledujúceho krátkodobého adaptívneho stavu.
- (8) Aké sú **limity prezentovaného modelu** darwinovskej evolúcie? Použitý tvar genotypu a jeho zobrazenia na fenotyp sú extrémne jednoduché, je vhodný pre štúdium evolúcie u vírusov, bakteriofágov a najjednoduchších prokaryotických baktérii. Pre zložitejšie organizmy musí byť model genotypu (a jeho zobrazenia na genotyp) o mnoho zložitejší, musí zahŕňať také koncepty, akými sú premenlivá dĺžka genotypu, hierarchická modulárna štruktúra, iné mutácie než aké sú 1-bodové mutácie, atď. Jeden z najdôležitejších problémov súčasnej teórie darwinovskej evolúcie je vysvetlenie a emergencia tzv. **neredukovateľných zložítostí (irreducible complexities)** [3]. Súčasná teoretická modely darwinovskej evolúcie nie sú schopné vysvetliť tento problém ani na elementárnej úrovni. Súčasná aktivity v umelom živote sa okrem iného usmerňujú aj na oblasť, ktoré by mohli mať význam pre vypracovanie všeobecnej teórie darwinovskej evolúcie.. V súčasnosti sú veľmi intenzívne študované modulárne aspekty genotypu [20,21] a problém symbiózy [19], kde oba problémy vyžadujú zložitejšie genotypy ako ten, ktorý bol študovaný v tejto prednáške.

Pod'akovanie. Autor ďakuje Jirkovi Pospíchalovi za trpezlivosť pri sledovaní a rozvíjaní mojich "endless" kontemplácií o darwinovskej evolúcii. Autor taktiež ďakuje Dušanovi Gálikovi za kritické prečítanie a zhodnotenie tejto prednášky a taktiež upozornenie na "neprijemné" nepresnosti jej textu. Táto práca bola podporená grantami 1/7336/20 a 1/8107/01 Vedeckej grantovej agentúry SR.

Literatúra

1. Atkins P. W. (1998). *Physical Chemistry*. Oxford University Press, Oxford.
2. Banzhaf W., Dittrich P., Eller B. (1999). Topological Interactions in a Binary String System. *Physica D* **125**, 85.
3. Behe, M. J. (1996). *Darwin's black box*. Simon & Schuster, New York.
4. Benatre J.-P., Le Metayer D. (1990). The Gamma Model and its Discipline of Programming. *Sci. Comp. Progr.* **15**, 55.
5. Dennett D. C. (1995). *Darwin's Dangerous Idea - Evolution and the Meaning of Life*. Penguin Press, London.
6. Berry G., Boudol G. (1992). The Chemical Abstract Machine. *Theoret. Comp. Sci.* **96**, 217.
7. Dawkins R. (1982). Universal Darwinism. In Bendall D. S. (ed.) *Evolution from Molecules to Men*, Cambridge University Press, London, p. 403..
8. Dittrich P. (1999). Artificial Chemistries (tutorial material). A tutorial held at ECAL'99, 13-17 September 1999, Lausanne, Switzerland.
9. Eigen M. (1971). Self organization of matter and the evolution of biological macro molecules. *Naturwissenschaften* **58**, 465
10. Eigen M., Schuster P. (1977). The Hypercycles: A Principle of Natural Evolution. *Naturwissenschaften* **64**, 541; **65**, 7; **65**, 341
11. Fogel, D. B. (1995). *Evolutionary Computation. Toward a New Philosophy of Machine Intelligence*. The IEEE Press, New York.
12. Fontana W. and Schuster P. (1998). Continuity in Evolution: On the Nature of Transitions. *Science*, **280**, 1451.
13. Fontana W. (1991). Algorithmic Chemistry. In: Langton C.G. (ed.) *Artificial Life II* Addison Wesley, Reading, MA, p. 159.
14. Gillespie D. T. (1977). Exact Stochastic Simulation of Coupled Chemical Reactions. *J. Phys. Chem.* **81**, 2340.
15. Holland J. H. (1975). *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor.
16. Jones B.L., Enns R.H., Rangnekar S.S. (1976). On the theory of selection of coupled macromolecular systems. *Bulletin of Mathematical Biology* **38**, 15
17. Kauffman S.A. (1993). *The Origins of Order: Self-Organization and Selection in Evolution*. Oxford University Press, New York

18. Kauffman S.A. and Hohnsen S. (1991). Coevolution to the edge of chaos: Coupled fitness landscape, poised states, and coevolutionary avalanches. *J. Theor. Biology* **149**, 467.
19. Kvasnička V. (2000). An evolutionary model of symbiosis. In: P. Sinčák, J.Vasčák (eds.): *Quo Vadis Computational Intelligence?* Physica-Verlag, Heidelberg, p. 293.
20. Kvasnička V. (2001). An evolutionary simulation of modularity emergence of genotype-phenotype mappings. *Neural Network World* **5**, 473.
21. Kvasnička V. (2002). A modularity emergence of genotype-phenotype mappings. *Artificial Life* (sent for publication).
22. Kvasnička V., Pospíchal J. (2001). Autoreplicators and Hypercycles in Typogenetics, *Journal of Molecular Structure (Theochem)* **547**, 119.
23. Kvasnička V., Pospíchal J. (2001). A study of autoreplicators and hypercycles by typogenetics. In: J. Kelemen, P. Sosik: *Advances in Artificial Life, ECAL 2001, LNAI 2159*, Springer 2001, p. 37.
24. Kvasnička V., Pospíchal J. (2002). Artificial Chemistry, Replicators, and Molecular Darwinian Evolution *In Silico*. In: P. Sinčák, J.Vasčák (eds.): *Quo Vadis Computational Intelligence?* Physica-Verlag, Heidelberg (to be published).
25. Kvasnička V., Pospíchal J., Tiňo J. (2000). *Evolučné algoritmy*, Vydavateľstvo STU, Bratislava.
26. Newman M.E.J. and Engelhardt R. (1998). Effects of neutral selection on the evolution of molecular species. *Proc. Roy. Soc. London B*, **265**, 1333.
27. Schuster P. (2002). Molecular Insight into Evolution of Phenotypes. In *Evolutionary Dynamics - Exploring the Interplay of Accident, Selection, Neutrality, and Function*, ed. by Crutchfield J.P. and Schuster P., Oxford University Press,
28. Schuster P. and Fontana W. (1999). Chance and Necessity in Evolution: Lessons from RNA. *Physica D*, **133**, 427.
29. Spiegelman S. (1971). An Approach to the Experimental Analysis of Precellular Evolution. *Quart. Rev. Biophysics*, **4**, 213.
1. Tomáška E., Nosek J. (2000). Molekulárna darwinovská evolúcia: In vivo, in vitro a in silico. Zborník *Kognitívne vedy III (CogSci 2000)*, Katedra matematiky CHTF STU, Bratislava (prednáška je dostupná na internetovskej adrese http://math.chtf.stuba.sk/CogSci_2000.htm).
31. Wright S. (1931). Evolution in Mendelian Populations. *Genetics* **16**, 97.
32. Gálik D. (2000). Univezálny darvinizmus (Môže byť darvinizmus teóriou všetkého) Zborník *Kognitívne vedy III (CogSci 2000)*, Katedra matematiky CHTF STU, Bratislava (prednáška je dostupná na internetovskej adrese http://math.chtf.stuba.sk/CogSci_2000.htm).