

OTDK Konferencia
Biológia Szekció

Második hibaküszöb:
a függetlenül replikálódó gének
kompartmentekben fenntartható
maximális száma

Szerző:

Hubai András Gábor

Biológus MSc hallgató
Eötvös Loránd Tudományegyetem
Természettudományi Kar

Témavezető:

Dr. Kun Ádám

tudományos főmunkatárs
Eötvös Loránd Tudományegyetem
Természettudományi Kar
Növényrendszertani, Ökológiai és Elméleti Biológiai Tanszék



Budapest, 2012

Kivonat

Az élet kialakulásának korai szakaszára jellemző (ún. minimális) genomot nem volt lehetséges egyetlen kromoszómán eltárolni: az enzimatisz másolás pontatlansága miatt az információ jelentős része hamar elveszett volna (vö. Eigen Paradoxona). És bár rövid szakaszok már ekkor is kellő pontossággal másolódtak, különálló gének formájában sem problémamentes e minimális genom fenntartása: a szaporodási ráták közötti elkerülhetetlen különbségek kompetitív kizáráshoz, ezáltal pedig információvesztéshez vezetnek. A Sztokasztikus Korrektor Modell (SCM) a protosejtekbe csomagolt gének osztódáskori véletlenszerű szétválása révén teszi lehetővé, hogy az előnyös összetételű protosejtek folyamatosan újra felbukkanjanak, s így a belső versengés ellenére is fennmaradjon a teljes génkészlet.

Egyedalapú modell segítségével kerestük a vezikulaszám, illetve a vezikulán belüli molekulaszám függvényében fenntartható különböző gének maximális számát. Megmutattuk, hogy a sztochasztikus korrekció lehetővé teszi közel 100 gén együttélését, még bizonyos egyenlőtlen szaporodási ráták mellett is. Egy minimális élő sejthez körülbelül 60-100 különböző gén szükséges, így elmondható, hogy kompartmentalizált rendszerünkben az információ integrációja sikeres: elegendő a protosejt működéséhez. Eredményeink felvetnek egy lehetséges evolúciós utat, amely során újabb és újabb gének épülnek be a rendszerbe, annak összeomlása nélkül.

Tartalomjegyzék

Kivonat.....	2
1. Bevezetés.....	4
1.1. Tudománytörténet: néhány szemléletmód	4
1.2. Az információ-fennmaradás problémája	5
1.3. Információ-integrátor rendszerek	8
1.4. Célkitűzés.....	10
2. Modell	11
2.1. Implementáció.....	12
3. Eredmények	14
3.1. Egyensúly kialakulása	15
3.2. Korrekciós exponens (ϵ).....	16
3.3. Vezikulák száma (v).....	17
3.4. Molekulák száma (m)	18
3.5. Replikációs affinitások (a)	20
4. Diskusszió	22
4.1. Többszintű szelekció.....	22
4.2. Sztochasztikus korrekció	24
4.3. Komplementáció	24
4.4. A fenntartható információ becslése.....	25
4.5. A komplexitás növekedése.....	26
5. Köszönetnyilvánítás	28
6. Irodalomjegyzék	29

1. Bevezetés

Az élővilág szembeötlő diverzitása folyamatos inspiráció forrása az emberiség számára, ezzel összhangban minden kornak megvoltak az e sokféleség kialakulását magyarázó elméletei. A mai szekuláris konszenzus a Darwin munkássága nyomán kidolgozott ún. Kiterjesztett Evolúciós Szintézis (Darwin, 1859; Pigliucci, 2009; Pigliucci és Müller, 2010). Ez – többek közt – kiterjed az élőlények küllemének (*fenotípus*) környezethez igazodó evolúciójára és idomulására (*plaszticitás*), jellegeik genetikai (és epigenetikai) öröklődésre és evolvabilitásra, valamint a fajképződésre (*speciáció*). Azonban az élet kialakulását, amely szükségképpen megelőzte a diverzifikációt, nem magyarázza. Nem meglepő tehát, hogy korunk tudósai fokozott megszállottsággal fordulnak az élet keletkezésének problémaköre felé.

1.1. Tudománytörténet: néhány szemléletmód

Persze az eredet kérdését mindig is érdeklődés övezte, csak hogy sokáig triviálisnak gondolták az élet megjelenését. Már az ókorból ismert, és széleskörű elterjedtségnek örvendett a gondolat, miszerint az apró élőlények (pl. tetvek, nyüvek, férgek) képesek az élettelenből való spontán keletkezésre; azonban 1668-ban Redi megmutatta, hogy a legyekről elzárt húson nem jelennek meg nyüvek (Oparin, 1938; Brack, 1998). A spontán keletkezés elmélete mégsem dőlt meg ezzel, csupán az erre képesnek gondolt alanyok köre változott meg: a soksejtűek helyébe az ezidőtájt felfedezett mikrobák léptek. További kétszáz évnek kellett elteltie, míg Pasteur 1861-es sterilizációs kísérletei az elmélet meggyőző cáfolatául szolgáltak (Pasteur, 1878). Felismerésének tükrében Pasteur kimondta a biogenezis törvényét, mely szerint minden élő csakis élőből keletkezik (*omne vivum ex vivo*).

Hasonlóképp az ókorból ismerjük a vitalizmus elméletét, mely szerint az atomok minden értelmet nélkülöző kölcsönhatásai képtelenek az élőlények leírására: e magyarázatban az életerő fogalma megkerülhetetlen (állította többek közt Galenos, II. századi orvos). A kémia (alkímia) fejlődésével azonban egyre nehezebben vált elfogadhatóvá, hogy az orron át belélegzett gázok ilyennel szolgálhatnának. Ám mielőtt az elmélet kudarcot vallott volna, Berzelius, a kiváló svéd kémikus (az általa bevezetett elkülönítés nyomán) a szerves anyagokat javasolta az életerő hordozóinak.

Ennek nyomán tartjuk Wöhler 1828-as kísérletét – amelyben szervesen kiindulóanyagokból szerves ureát állított elő – a vitalizmus elmélet döntő cáfolatának

A spontán keletkezés gondolatának elvetése, a vitalizmus elmélet bukása és a fajok egymásból való leszármazásának ötlete felvetette az élet történeti keletkezésének (az ún. *abiogenezis*) merőben új kérdését. A lyelli uniformizmus elvével azonban ütközött a megfigyelés, hogy míg a laborban élettelen nem alakul élővé, az élet kezdetén ennek meg kellett történnie. E látszólagos ellentmondást Darwin 1871-ben a következő gondolatmenettel igyekezett feloldani: ha valamikor egy „langyos kis tavacskában” ki is alakulhatott az első élőlény, ilyennek mi azért nem lehetünk tanúi, mert az ezen úton „egyre bonyolultabb átalakulásokra kész, kémiai úton kialakuló fehérje vegyületek [...] napjainkban rögvést felfalásra, felszívódásra kerülnének” (F. Darwin, 1887).

Az abiogenezis jelensége így valamelyest körülhatárolható: az élettelen folyamatok által (*abiotikusan*) kialakult szerves anyagok és az első, bizonyára kezdetleges, biológiai evolúcióra képes élőlény közötti átmenet. Az evolúcióra való képesség kritériuma Maynard Smith (1987) megfogalmazásában a sokszorozódás, a jellegek öröklődése, valamint ennek ritka hibája révén új változatok kialakulása. Az élet definíciójának a mai napig nincs hasonlóan bevett formája: ha az önfenntartáshoz igénybe is vesszük az anyagcserét¹, a féligáteresztő burkot (*membrán*) és a géneket (Luisi, 1998), ezek nem működnek a külső környezettől függetlenül, és főleg nem a legelső élőlényekben (Szathmáry, 2007). Az abiogenezis kutatásának legfőbb kérdései így a korai környezet kémiai összetételére (ennek részeként az abiotikus úton keletkező szerves anyagokra), illetve az első élőlények anyagcseréjére, membránjára illetve örökítőanyagára vonatkoznak.

1.2. Az információ-fennmaradás problémája

A tudományos vizsgálódás e területei közül kiemelkedik a géneké, azok ugyanis a legvalószínűbb² alanyai a korai evolúciónak. Ez pedig sajátos felépítésüknek tudható be – különféle alegységek (*modulok*) követik egymást az egyszálú, elágazatlan láncban – mely szerkezet lehetővé teszi a

¹ Anyagcserén (*metabolizmus*) külső tápanyagokból a saját belső komponensek előállítását értjük.

² Az evolúció gondolata egyébként felmerült már az anyagcsere és a membrán kapcsán is, mivel azok összetételükben (*kompozíció*) információ hordozására képesek (Kauffman, 1986; Segré & Lancet, 2000). Az ún. kompozicionális információ azonban csak korlátozottan örökletes, nem evolvábilis (Vasas *et al.*, 2010).

gének *processzív* (modulonkénti) másolódását³. Másolási pontatlanság (hiba) révén az eredeti szekvenciától eltérő (mutáns) változatok is kialakulhatnak; processzív másolódás esetén az egyes modulokban egymástól függetlenül lehetnek hibák, a lehetséges változatok száma így csillagászati⁴. És minthogy a másolatok száma jóval kisebb, mint ahányan az így elérhető változatok vannak, a gének *korlátlan öröklődésű* replikátoroknak minősülnek (Szathmáry és Maynard Smith, 1997); e replikátorok sajátsága, hogy „nyílt végű” evolúcióra képesek.

E gondolatmenetnek azonban kiemelendő egy gyengéje: hogy ti. csak a kellően hosszú gének változatai vannak csillagászati mennyiségben. Az élet keletkezésekor pedig egyáltalán nem biztos, hogy ez adott lett volna. Sőt, mint azt Eigen (1971) felismerte, a hosszú gének fenntartása is problematikus lehet. Ugyanis ha a processzív másolódás során hibázni is modulonként lehet⁵, úgy a hosszabb mesterszekvenciából készült másolatok nagyobb arányban lesznek mutánsok, mint rövidebb társaik. Egy határon, az ún. hibaküszöbön⁶ túl pedig olyan arányban keletkeznek mutánsok, hogy a másolatok felhőjéből (*quasispecies*) kiveszik maga az örökítendő információt hordozó mesterszekvencia (Eigen, 1971; Schuster, 2010).

Hosszú genom nélkül nincsen pontos enzim⁷, pontos enzim nélkül viszont nincs hosszú genom – mondja ki Eigen Paradoxona (Maynard Smith, 1979). A pontos másoláshoz szükséges enzimek ugyanis hosszúak lehettek. De vajon mennyire? És mekkora lehetett a hibaráta az élet kialakulásakor? Talán ez utóbbi becslése könnyebb: a baktériumok hibaráta viszonylag alacsony (nukleotidonként kb. 10^{-8} – 10^{-10} (Schaaper, 1993; Kunkel, 2004)), ám ezt hatalmas enzimkomplexek végezte hibajavító mechanizmusoknak köszönhetik. Az ugyanezen szervezetekben másolódó vírusok hibaráta csupán 10^{-5} körüli, ámde ezen is megmutatkozik a hibajavítás hatása, amely pedig még a legelső élőlényekből is bizonyosan hiányzott. Hogyha tehát realisztikus becslést kívánunk adni, akkor érdemes szemügyre vennünk az *RNS-világ*⁸

³ Erre példa a *templát alapú másolódás*, melynek lényege, hogy az alegységeiből felépülő új gén követni kényszerül egy már meglevő gén alegység-sorrendjét (*mesterszekvencia*): a minta szál egyes alegységei ugyanis fizikai közelségük révén megszabják, hogy milyen alegység épülhet be velük szemben az épülő szálba.

⁴ K féle alegység (pl. nukleotidokra $K = 4$) esetén a lehetséges típusok száma $I_{\max} = K^L$, ahol L a mesterszekvencia hossza. L növelésével K^L hatványosan növekszik, erre vonatkozik a „csillagászati” jelző. Figyeljük meg, hogy mivel a mesterszekvencia egy a lehetséges típusok közül, így a benne hordozott információ maximuma ($I_{\max}$) is éppen ekkora!

⁵ A hibázás – lemásolt alegységekre jutó – rátáját állandónak szokás venni.

⁶ Minthogy dolgozatomban bevezetem a második hibaküszöb fogalmát, erre innentől első hibaküszöb néven utalok.

⁷ A paradoxon természetesen a másolóenzimekre értendő.

⁸ Az RNS-világ egy az élet keletkezése során feltételezett bioszféra, amelynek minden élőlényében RNS (azaz nem DNS, illetve fehérje) alkotta az örökítőanyag és biokatalizátor (enzim) molekulákat. *Riboorganizmusnak* hívjuk az ilyen élőlényeket. A katalitikus aktivitással bíró RNS molekulát pedig *ribozimnak* nevezzük.

forгатókönyve (Gilbert, 1986) nyomán végzett kísérleteket: a mesterségesen előállított⁹ RNS polimerázok eddigi legalacsonyabb hibaráta 8,8 · 10⁻³ (Wochner *et al.*, 2011). Kérdés, hogy ez mekkora génhosszt engedélyez.

Eigen Paradoxonának formalizálása vezet el bennünket az első hibaküszöbhez – mely a gén¹⁰ hossza és a másolási pontosság közötti összefüggés. Tekintsünk meg Maynard Smith (1983) egyszerűsítése nyomán a következő modellt! Feltesszük, hogy az összes mutáns másolódási rátája egységes (α), és alacsonyabb a mesterszekvenciánál (A). Ekkor a kettő arányából számolt $s = A/\alpha$ szelekciós előny állandónak tekinthető. A mesterszekvencia maximális hossza (L) és a legnagyobb, még azt megőrző (ún. kritikus) másolási pontosság (q) összefüggése:

$$L < \frac{\ln(s)}{(1 - q)} \quad [1]$$

ha elhanyagoljuk a mesterszekvenciába való visszamutáció lehetőségét, valamint a hibázást függetlennek tekintjük a beépítendő alegység típusától illetve pozíciójától.

Ha pedig ennél is közelebb akarunk evezni a valósághoz, akkor figyelembe vehetjük, hogy az enzimek aktivitása a térszerkezetüktől függ (Anfinsen, 1973), amelyre pedig a szekvenciát érintő mutációk jelentős hányada (λ) semleges (*neutrális*). Takeuchi és munkatársai (2005) belefoglalták az [1] egyenlőtlenségbe a neutrális változatokat:

$$L < \frac{-\ln(s)}{\ln(q + \lambda - q\lambda)} \quad [2]$$

A [2] képletben szereplő értékeket Kun és munkatársai (2005) a következőnek becsülik: $\lambda \approx 0,24$ és $s \approx 350$. A már említett hibarátaival ($q = 0,9912$) a maximális fenntartható génhossz 872 alegységnek adódik. Ezen a hosszon bár eltárolható egy kellően pontos polimeráz ribozim, egy feltételezett *riboorganizmus*⁸ teljes genomját ez nem képes fenntartani (Kun *et al.*, 2005).

⁹ Az eljárást, amely az evolúció folyamatait kémcsőben szimulálja, találó kifejezéssel *in vitro evolúciónak* mondjuk.

¹⁰ Ezt szokás a genom hosszának is nevezni, merthogy itt egyetlen gén hordozza a rendszer minden információját.

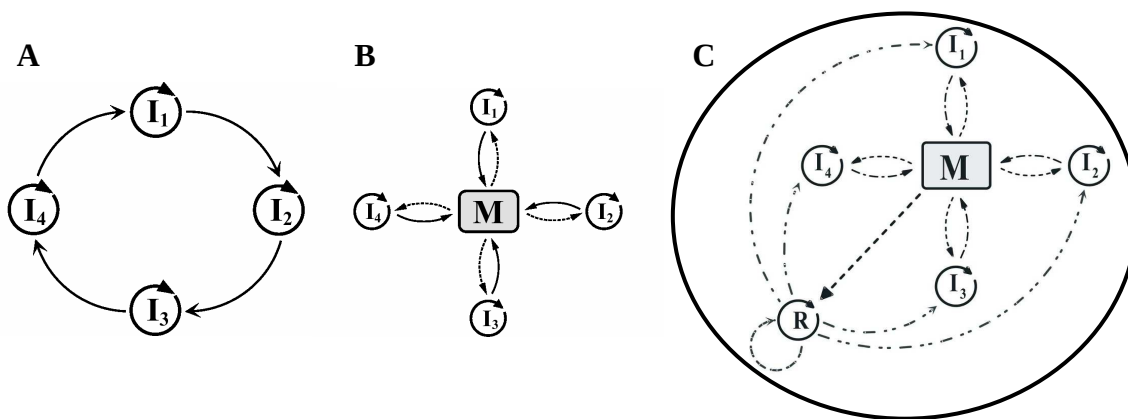
1.3. Információ-integrátor rendszerek

Eigen és Schuster (1977) felveti annak a lehetőségét, hogy amennyiben egy genom nem tartható fenn egyetlen információhordozó molekulán, több kisebb molekula is megfelelhet e célnak – persze ehhez a molekulák együttélése és együtt szaporodása is megvalósítható kell legyen. Az elsőként javasolt, erre képes rendszer a Hiperciklus (1/a. ábra), melyben különböző szekvenciájú (típusú) független replikátorok körkörös topológia¹¹ szerint segítik egymás másolódását (heterokatalízis). A különböző replikátor típusok, azaz a Hiperciklus fennmaradása által biztosított, hogy bármelyik típusú replikátor hiányában megáll (jelentősen lassul) az összes többi molekula másolódása is.

A Hiperciklus információfenntartó képességét azonban alapvető kritika érte: heterokatalitikus rövidzárlatot (*shortcut*) követően csak a rövidebb ciklus információtartalma marad fenn, a levágott rész kiesik; a heterokatalitikus segítséget igénybevevő, de nem továbbító molekulák (paraziták) beépülhetnek a rendszerbe, annak replikációját akadályozva, és akár teljesen meg is állítva (Maynard Smith, 1979). Továbbá a modell nem ad lehetőséget az információ akkumulációjára, a ciklusba nem lehet újabb molekulákat illeszteni annak jelentős megváltoztatása nélkül. A heterokatalitikus segítség olyan altruista viselkedés, amelynek elmaradása nem jár negatív visszacsatolással, a parazita viselkedésnek nincs közvetlen büntetése. E büntetést a diszperzió gátlásával meg lehet valósítani: ha a parazitáknak nincs lehetőségük elhagyni a kizsákmányolt rendszert, akkor az általuk okozott hátrány idővel visszahat rájuk (Szabó *et al.*, 2002).

A rövidzárlat okozta információvesztés elkerülhető, illetve az új típusok beépülése megvalósítható, ha a replikátor molekulák egymástól függetlenül csak a központi rendszerrel (pl. anyagcserével) állnak kapcsolatban, azzal egymásnak kölcsönös segítséget nyújtva: ez a Metabolikus Replikátor Modell (1/b. ábra; Czárán és Szathmáry, 2000). A paraziták jelentette veszély pedig újfent kivédhető a diszperzió csökkentésével, így például a membránok általi kompartmentalizációval: ez a Sztochasztikus Korrektor Modell (SCM; 1/c. ábra; Szathmáry és Demeter, 1987). E modell feltételezi több, a központi anyagcserét segítő molekula, valamint egy aspecifikus replikáz jelenlétét valamennyi kompartmentben (*vezikula*); a replikáz az összes molekulát másolni képes.

¹¹ E topológia alapja, hogy minden típus pontosan egy másikat segít, mindegyikük pontosan egyetlen típustól kap segítséget, és egy típus segítése indirekt módon valamennyi másikhöz is eljut.



1. ábra: A Hiperciklus (A), a Metabolikus Replikátor Modell (B) és a Sztokasztikus Korrektor Modell (C) felépítésének vázlata; I: információhordozó molekula, M: anyagcsere, R: aspecifikus replikáz (Czárán és Szathmáry, 2000; Könnýű *et al.*, 2008 nyomán).

E három információ-integrátor rendszert összehasonlítva szembevetendő, hogy a SCM felépítése a leginkább összetett – mégsem mondhatjuk feltevéseit túlzónak. A membrán általi körbezártság a diszperzió gátlásának könnyen megvalósítható módja; már a korai membránok is átjárhatatlanok lehettek a polimerekkel szemben (Schrum *et al.*, 2010), így védelmül szolgáltak a génvesztés és a paraziták (vezikulák közti) elterjedése ellen. Valószínűsíthető továbbá, hogy az anyagcseréjük által termelt alegységek sem juthattak ki a vezikulából (Yarus, 1999), tehát a kompartmentek közötti szomszédság hatása elhanyagolható. A vezikulák populáción belüli előnye, a rátermettségük elsősorban azon múlik, hogy hányféle segítő enzimet tartalmaznak, melyek az alegységek szintézisét (Unrau és Bartel, 1998) és az azokhoz vezető köztes anyagcseretermékek átalakítását szolgálják (Jensen, 1976); e típusok birtoklása ilyen szempontból esszenciálisnak¹² tekintendő. És ugyan még nem ismerünk aspecifikus replikáz ribozimot (Bartel és Unrau, 1999; Johnston *et al.*, 2001; Zaher és Unrau, 2007; Wochner *et al.*, 2011), ilyen molekula létének nem látjuk elméleti akadályát, mi ezért nem is foglalkoztunk e probléma megválaszolásával.

A vezikulák osztódása kezdetben bizonyosan szabályozatlan folyamat, a molekulák véletlenszerűen (és egymástól függetlenül) jutnak a két utódvezikula egyikébe. Nem garantált, hogy a két vezikula azonos méretű lesz, hát még hogy valamennyi típusú molekulából kerül beléjük. Az esetek többségében a két helyen mégis közel azonos molekuláris összetételre számíthatunk: az utódvezikulákban a molekulák száma binomiális eloszlást¹³ követ, és ez minden molekulatípusra

¹² Elgondolásunk szerint a korai molekulák hatalmas szelekciós nyomásnak lehettek kitéve. Hosszú távon csak azon típusok maradhattak fenn, amelyek a központi anyagcserét jobban serkentették, mint amennyivel a másolásuk azt lassította. Ebből pedig arra következtetünk, hogy a vezikula nem élhette túl, ha bármely ilyen típus kiveszett belőle.

¹³ Véges sok molekula, egymástól függetlenül, és kétféle értéket vehet fel (egyik vagy másik utódvezikulába kerül).

külön-külön is igaz. Ez megkönnyíti az azonos szaporodási ütemű molekulák együttélését. Mi azonban a helyzet, ha a szaporodási ütemek eltérése miatt már a szétszandó vezikula tartalma sem egyenletes a típusokra nézve – ha egyes típusok nagyon kis arányban vannak benne?

1.4. Célkitűzés

A fentebb bemutatott mechanizmusok révén a SCM alkalmas a független replikátorok együttes fenntartására (Zintzaras *et al.*, 2002), ám mindeközéig senki nem határozta meg a koegzisztenciára képes típusok maximális számát. Pedig mennyiségi becslést csak úgy adhatunk a korai rendszerek összetettségére (a független replikátorokban együttesen tárolható maximális információra), ha ismerjük a fenntartható génszámra vonatkozó, szabályozatlan másolódás és osztódás jelentette korlátot: a második hibaküszöböt. A SCM viselkedése (dinamikája) és stabil egyensúlyi állapotai is kellőképp feltáratlanok. Fontanari *et al.* (2006) matematikai modell segítségével kísérli ezt meg, – melyben a vezikulák számát végtelennek feltételezi, – ám megjegyzi, hogy munkája nem helyettesíti egy alapos, a paraméterter feltárása általi vizsgálatot. Jelen dolgozat e hiányosságokat igyekszik pótolni.

2. Modell

Modellünkben az információ hordozói rövid, pontosan másolódó molekulák (gének), melyek enzimatis aktivitással is rendelkeznek (vö. ribozim). Enzimaktivitásuk alapján a molekulák többféle (τ) típusba sorolhatóak. E típusok együttesen a környezet biztosította tápanyagokból saját alegységeik előállítására képesek – ezt a folyamatot nevezzük anyagcserének. Illetve a típusok közül egy az aspecifikus replikáz, mely ezen alegységek felhasználásával készít másolatot a molekulákról. Az egyes típusok replikázhoz való kötődésének erősségétől, a replikációs affinitástól (a_i) függ, hogy a rendelkezésre álló alegységekből mely molekulának sikerül megkettőződnie.

A SCM alapötlete mentén: valamennyi molekula vezikulában található, és ezen elkülönült közösségben verseng a replikáz általi másolódásért; a vezikulák a belső, molekuláris összetétel alapján részesednek a környezet biztosította tápanyagokból, ezáltal növekednek és osztódnak; tehát molekuláris és vezikuláris szinten is hat szelekció, amely együttesen hat a különböző típusú molekulákban hordozott információ együttes fennmaradására.

Minden vezikula $m_i \in [1; m_{\max}]$ molekulát tartalmaz, mely típusonként $m_{i,j}$ darabra oszlik:

$$m_i = \sum_{j=1}^{\tau} m_{i,j} \quad [3]$$

A vezikula rátermettségét (R) anyagcseréjének gyorsasága szabja meg, mely két folyamatból áll: a tápanyagok vezikulába áramlásából, majd azok alegységgé alakításából. Mindkét folyamat sebessége a belső molekulaállomány milyenségén múlik. A tápanyagok beáramlása a vezikula felületével arányos, mely felületet a molekulaszám (m_i) határoz meg – modellünkben lineárisan. Az alegységgé alakítás pedig attól függ, hogy a különböző típusok milyen arányban vannak jelen – legnagyobb sebesség az egyenletes típuseloszláshoz¹⁴ tartozik; az összefüggést a típusonkénti molekulaszámok ($m_{i,j}$) produktuma írja le. És mivel azt szeretnénk, hogy a különböző maximális

¹⁴ A vezikula belső anyagcseréje megkívánja valamennyi különböző típusú molekula jelenlétét. E molekulák típusok szerinti megoszlásának létezik olyan aránya, amely mellett az anyagcsere sebessége maximális. Minthogy semmilyen adatunk nincs arra nézve, hogy az ismeretlen enzimaktivitással bíró molekuláknak mi volna ezen eloszlása, a maximumot önkényesen ahhoz az állapothoz rendeltük, amikor minden típusból ugyanannyi van a vezikulában.

molekula- ($m_{\max}$) és típusszámok (τ) mellett kapott rátermettségeket össze lehessen hasonlítani, ezért úgy kalibráltuk az [4] egyenletet, hogy $0 \leq R \leq 1$ legyen¹⁵:

$$R_i = \frac{m_i}{m_{\max}} \left(\prod_{j=1}^{\tau} \left(\frac{m_{i,j} \cdot \tau}{m_i} \right) \right)^{\varepsilon} \quad [4]$$

Az ε kitevő a két komponens közötti súlyozásra szolgál. Alacsonyabb kitevő esetén a típusok kiegyenlítettége meghatározóbb, míg magasabb kitevő esetén a molekulaszám hatása dominál a rátermettségben.

A szelekció mindkét szintjén nem-determinisztikus algoritmus szabja meg a szaporodást. Minden vezikulának a rátermettségével arányos esélye van, hogy adott körben egy molekulával gazdagodjon – a választás véletlenszerű. Ezt követően a vezikulán belüli molekulák közül, a típusok replikációs affinitásával (a_i) súlyozva, szintén véletlenszerűen választunk egyet másolásra. A vezikulák a maximális molekulaszámot ($m_{\max}$) elérve kettéosztódnak. A bennfoglalt molekulák ilyenkor azonos eséllyel kerülnek a két utódvezikula valamelyikébe, ám egyik leányvezikula sem maradhat üres¹⁶. A konstans populációméret (N) fenntartása érdekében osztódáskor el is pusztul egy vezikula, amire valamennyinek azonos esélye van. A vezikulák dinamikája tehát Moran folyamat: véges és állandó a populációméret, és sztochasztikus hatások mentén változik a rendszer. Vizsgálatunk addig tart, amíg a vezikulák dinamikája egyensúlyba nem kerül, ideértve a rendszer kihalását is – amikor minden vezikula anyagcseréje leáll: $\forall i \in [1; N]: R_i = 0$

2.1. Implementáció

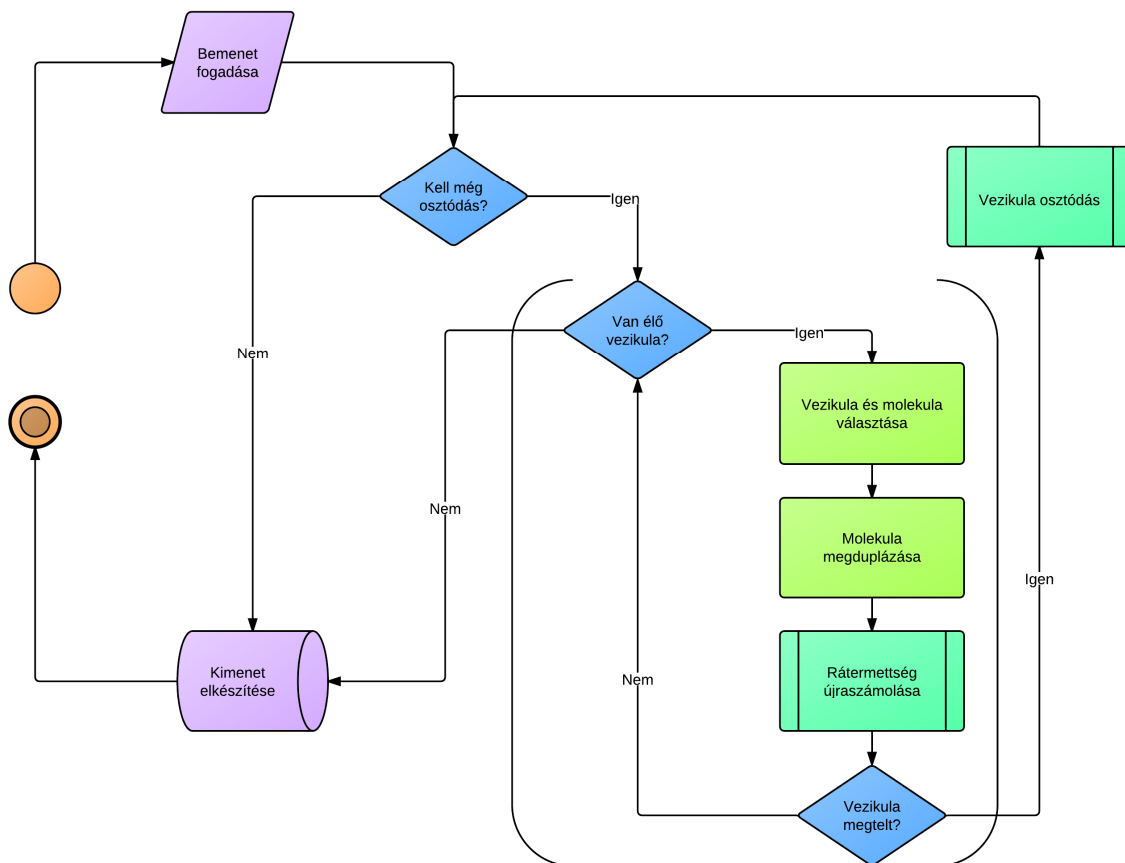
Modellünk viselkedését számítógépes szimulációval vizsgáltuk. A szoftvert Object Pascal programnyelven írtam, algoritmusát a 2. ábra foglalja össze. E program bemeneti paraméterei: az osztódási események („generációk”) maximális száma, g , a vezikulák száma, v , a molekuláris típusok száma, τ , a hozzájuk tartozó replikációs affinitások, a_i , a kópiaszám (minden típusból

¹⁵ A tápanyagok beáramlása kapcsán a maximális molekulaszám reciproka a kalibrálási tag, az alegységgé alakulás kapcsán pedig típusonként a molekula- és típusszám reciproka. A kalibrálási tagok megfigyelhetők a [4] képlet kiterjesztett formájában:

$$R_i = \left(m_i \frac{1}{m_{\max}} \right) \cdot \left(\prod_{j=1}^{\tau} (m_{i,j}) \left(\frac{\tau}{m_i} \right)^{\tau} \right)^{\varepsilon} \quad [4.2]$$

¹⁶ Ha mégis üresen maradna valamelyik vezikula, előlről kezdjük a molekulák kettéosztását.

kezdetben ennyi van a vezikulában), $c = m_{\max} / 2\tau$, valamint a rátermettség függvényének kitevője, ε . A kimeneti adatok – a populáció átlagos rátermettsége, illetve a túlélő (pozitív rátermettséggel bíró) vezikulák száma – a rendszer egyensúlyi állapotára vonatkoznak (ld. 3.1. alfejezet).



2. ábra: A modell vizsgálatához használt algoritmus folyamatábrája. Az alakzatok követik a folyamatábrák bevett jelöléseit¹⁷, a színek alkalmiak. A molekulák másolódásának ciklusa zárójelek között szerepel; a vezikulák osztódásának ciklusa (minden feltételes elágazódásnál az igen opciót követve) ezen kívül nyúlik.

A szoftvert magas futási sebességre és alacsony memórafogyasztásra igyekeztem optimalizálni. A szimulációkat az ELTE IIG *Atlasz* nevű nagyteljesítményű számítógépén¹⁸ végeztük. Szempont volt a paraméterek megválasztásakor, hogy a program ésszerű idő alatt lefusson: a jelenleg elérhető kapacitás tehát limitálta, hogy a paramétertér mekkora részét tudtuk feltérképezni¹⁹. A dolgozat ábráihoz 1800 szimuláció kimenetét használtam fel, ezek össz futásideje kb. 300 nap volt²⁰.

¹⁷ Paralelogramma (lila) jelöli a bemeneti paraméterek (itt fájlból történő) beolvasását, henger (lila) a kimenet kiírását, rombusz (kék) a feltételes elágazódásokat (ha... akkor), téglalap (sárgászöld) az egyszerű eljárásokat, duplafalú téglalap (zöld) pedig a függvények meghívását.

¹⁸ Részletek a <http://www.caesar.elte.hu/hpc/atlasz-hw.html> címen érhetőek el.

¹⁹ A paramétertér feltérképezése így arra korlátozódott, hogy az előzetes vizsgálatok alapján belőtt, lokálisan optimális paraméter-kombináció körül, a paraméterek meghatározta tengelyek mentén külön-külön, „sugárirányban” mozogva vizsgáljuk a modell viselkedését.

²⁰ Ez becsült érték, 4 órás átlagos futásidővel számolva. A szimulációkat párhuzamosan több nóduson futtattuk.

3. Eredmények

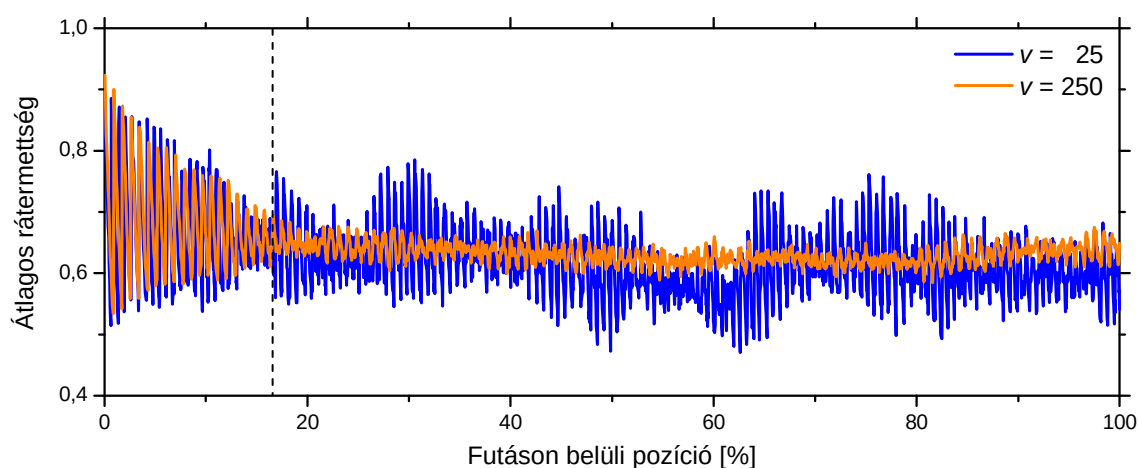
A célkitűzésben foglaltak szerint érdeklődésünk középpontjában a SCM információ-integrációs képessége áll. Az információ fenntartása pedig két tényezőre bontható: a minél pontosabb fenntartásra, melyet a rendszer átlagos rátermettségével, illetve vezikuláinak túlélésével jellemzünk; valamint a minél több információra, amelyet a molekuláris típusok száma ír le. A fenntartható információ mennyiségét azonban alapvetően befolyásolhatják a rendszer és környezetének körülményei; e körülményeket igyekeztünk megragadni az imént ismertetett paraméterekkel (ld. 2.1. alfejezet). A maximális típusszám meghatározásához tehát fel kell tárunk a SCM viselkedését (dinamikáját) a paraméterek különféle értékei által kirajzolt sokdimenziós térben.

Első körben olyan paraméter-kombinációt keresünk, amely lehetővé teszi vezikulák tartós túlélését, azaz információ fennmaradását. Minthogy a realisztikus körülményeket illetően nincs kapaszkodónk, a paramétereket kénytelenek vagyunk ehhez önkényesen megválasztani. Amennyiben találunk fenntartható rendszert, innen indulhatunk neki a paramétertér felderítésének. Ennek során a maximális fenntartható típusszámot, illetve a paraméterek ehhez hozzájáruló, optimális értékét keressük. Minthogy a modell egyik egyensúlyi mutatójára van hatással, másodjára a vezikuláris rátermettséget meghatározó függvény kitevőjét vizsgáljuk; majd sorra a populációméretet (vezikulaszámot) és molekulaszámot.

Idáig, a korábbi kutatások nyomán, feltételezzük a molekuláris típusok azonos eséllyel történő másolódását. Valószínűtlen azonban, hogy ez adott lett volna a szabályozástól mentes, korai rendszerekben. Végezetül tehát kitérünk a típusok eltérő másolódási esélyének (replikázhoz kötődésük különböző affinitásának) vizsgálatára is.

3.1. Egyensúly kialakulása

A különböző molekulatípusok fenntartása egyensúlyt feltételez, első körben ennek kialakulását vizsgáltuk. Pár paraméter-kombinációra részletesen megfigyeltük a modell időbeli lefutását. Eredményeink alapján (egyenletes típuseloszlásból kiindulva) vezikulánként átlagosan 17 osztódás szükséges az egyensúly eléréséhez. A 3. ábra megmutatja, hogy 25 illetve 250 vezikula rendszere (2500 illetve 25000 generáció alatt) miképp közelíti meg, és éri el az egyensúlyt. A megközelítés fázisa során a kezdeti, uniform típuseloszlás veszik ki a rendszerből: itt szabályos ritmusú, csillapodó oszcilláció figyelhető meg. Az egyensúly fázisában az átlagos rátermettség egy konstans érték körül fluktuál. Kisebb vezikulaszám mellett nagyobb a sztochasztikus hatások mértéke, ezt tükrözi a fluktuáció nagyobb szórása.



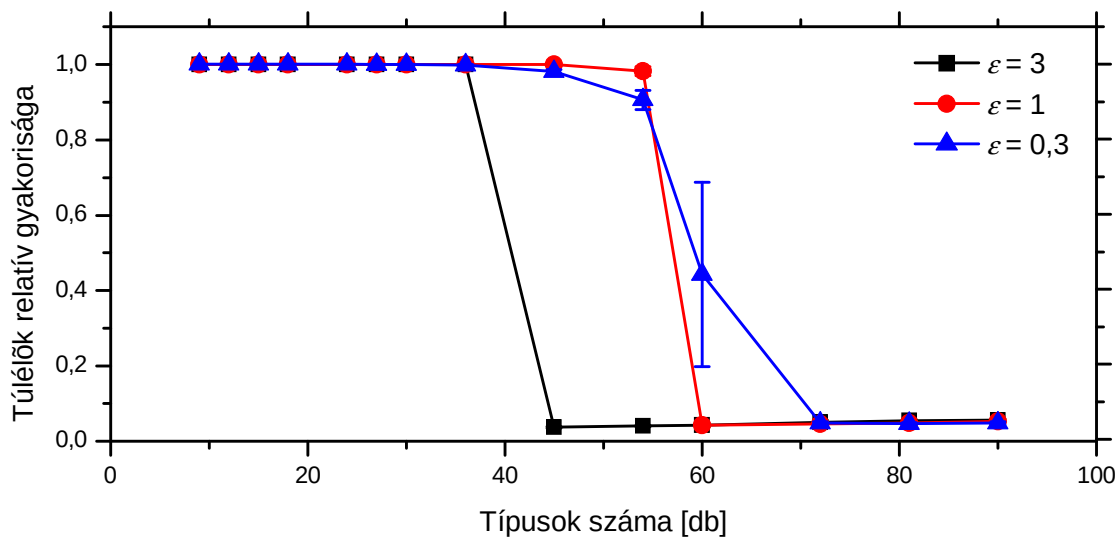
3. ábra: Két szimuláció időbeli lefutása, melyek paramétereit csak a vezikulaszámban térnek el. A futások során markánsan elkülönül két fázis: a kezdeti csillapodó oszcilláció (egyensúly megközelítése), és az ezt követő fluktuáció (egyensúly). A két fázis határát szaggatott vonal jelöli. Kék vonal: $g = 2500$; $v = 25$. Narancs vonal: $g = 25000$; $v = 250$. Közös paraméterek: $m = 1000$; $\tau = 5$; $\varepsilon = 0,3$; $a_i = 1$, ahol $i \in [1; \tau]$.

A továbbiakban bemutatásra kerülő szimulációkat az itt kapott eredményhez igazítottuk: a futás első $17v$ osztódásáig nem is feltételezzük, hogy beállt az egyensúly. Az egyensúlyi mutatók értékét értelemszerűen csak ezt követően kezdjük el számolni – ezen átlagok pontos meghatározásához minél hosszabb futásra van szükség, valamint a kezdeti, nem-egyensúlyi értékek mellőzésére. Előbbit szolgálja, hogy a generációk számát innentől mindig a vezikulaszám 100-szorosának választjuk; utóbbit pedig, hogy háromszoros ráhagyást biztosítunk az egyensúly elérésére: a mutatók kiszámolásából mellőzzük az első ($17v$ helyett) $50v$ generáció értékeit. Az egyensúly beálltát végül a sokgenerációs átlag és az utolsó generáció értékeinek összevetésével ellenőrizzük.

3.2. Korrekciós exponens (ε)

A következő alfejezetekben egy-egy paraméter változtatásával keressük a fenntartható típuszám maximumát, miközben az összes többi paraméter értékét lerögzítjük. Először a vezikulák rátermettségfüggvényének kitevőjét, az ún. korrekciós exponenst vizsgáljuk. E kitevő értéke mutatja meg, hogy az anyagcsere sebességét a tápanyagok vezikulába áramlása, avagy azok alegységgé alakítása szabja-e meg jobban. Ha a tápanyagok beáramlása a jelentősebb, akkor a belső molekulaszám járul inkább hozzá a rátermettséghez ($\varepsilon > 1$); ha pedig az alegységgé alakulás lépése a kritikus, úgy a típuseloszlás egyenletessége segíti a gyors anyagcserét ($\varepsilon < 1$). Megvizsgáltuk a típusok fennmaradását a kitevő elhagyásával is ($\varepsilon = 1$).

Szimulációink eredménye alapján (4. ábra) a rendszer információtartalmának fenntartását az segíti, ha a típuseloszlás egyenletessége járul inkább hozzá a rátermettséghez, $\varepsilon = 0,3$. Látható, hogy jelentősen alacsonyabb típuszámnál esik le a túlélő vezikulák gyakorisága amennyiben $\varepsilon = 3$ ($\tau = 45$), mint amennyiben $\varepsilon = 0,3$ ($\tau = 72$).

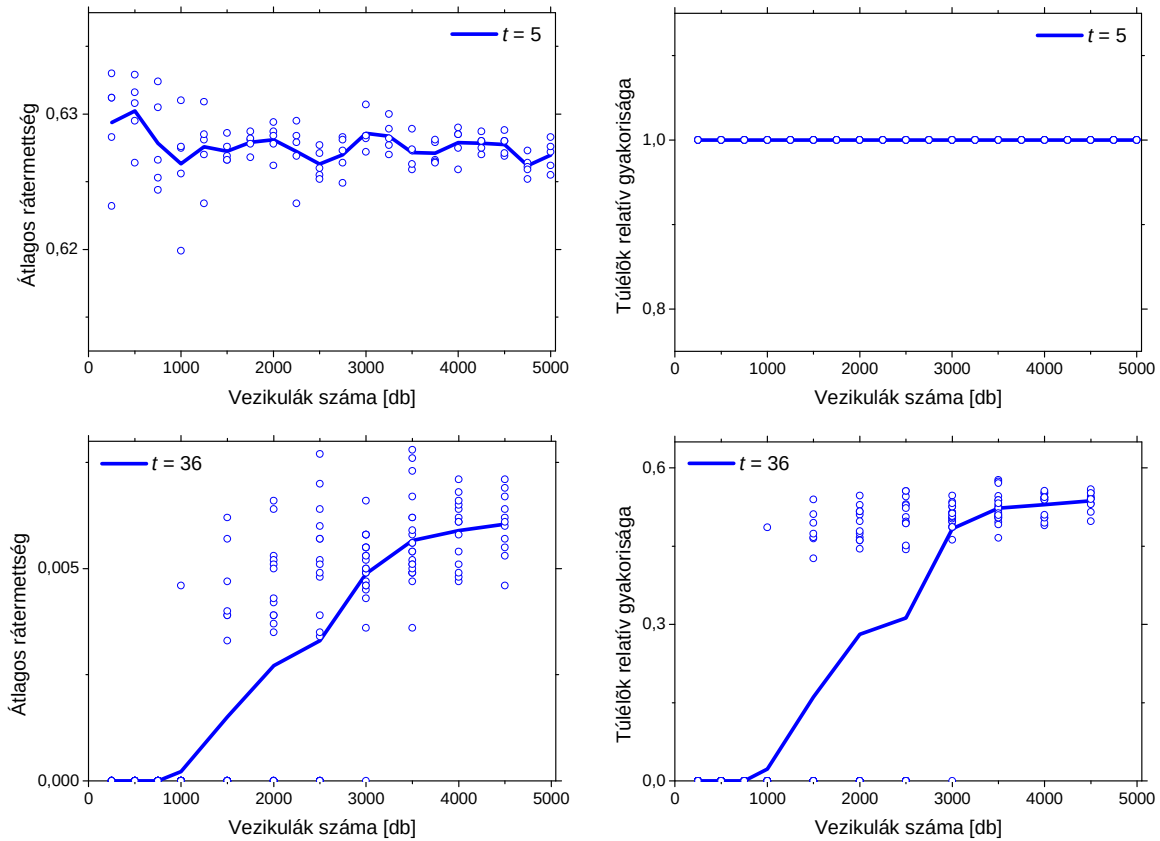


4. ábra: A túlélő (pozitív rátermettséggel rendelkező) vezikulák egyensúlyi gyakorisága a vezikulán belüli molekulatípusok számának függvényében. A grafikon 252 futás kimenetét összegzi; a pontok 6 ismétlés átlagát jelölik. A legtöbb pont eltakarja a hozzá tartozó szórás hibasávot. Közös paraméterek: $g = 100000$; $v = 1000$, $m = 3240$; $a_i = 1$, ahol $i \in [1; \tau]$.

3.3. Vezikulák száma (v)

Már az egyensúly kialakulása kapcsán láttuk (3.1. alfejezet), hogy a vezikulaszám kihat a rátermettség futáson belüli fluktuációjára: előbbi növelésével csökken az utóbbi szórása. Itt a futások egyensúlyi mutatóit, azaz a rátermettség és a túlélők számának átlagait vizsgáljuk – azt tapasztaljuk, hogy ezek is függenek a vezikulaszámtól (5. ábra).

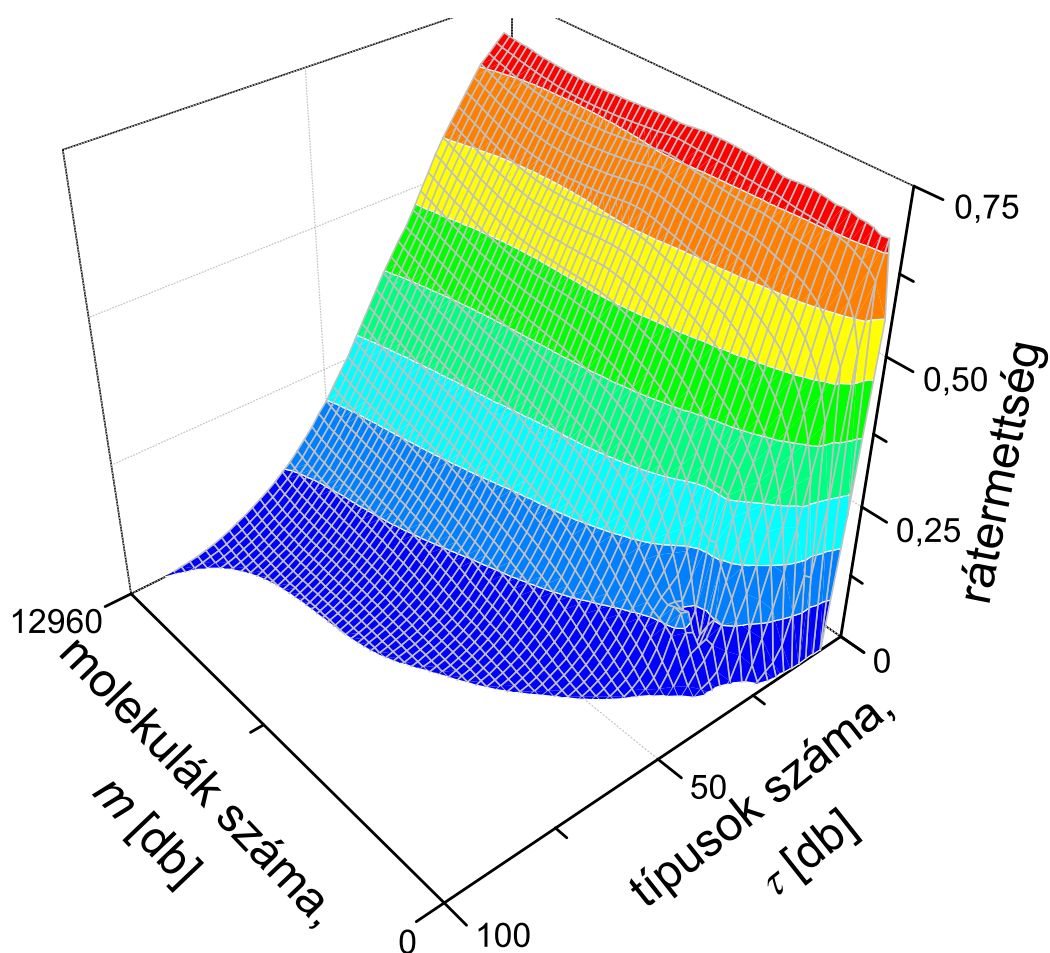
Mindkét egyensúlyi mutató telítési görbét mutat a vezikulaszám növelésével. Ha 250 vezikula elégséges a rendszer fennmaradásához, akkor nem feltétlenül segíti ezt a több vezikula: tehát csak a görbe platóját látjuk. Viszont amennyiben 250 vezikula nem elegendő, és a vezikulák számának növelésével válik fenntarthatóvá a rendszer: ilyenkor látványos az összefüggés telítés jellege. Megfigyelhető, hogy az átlagos rátermettség érzékenyebben tükrözi a sztochasztikus hatásokat, mint a túlélők aránya, amely cserébe (arányaiban) kisebb szórást mutat.



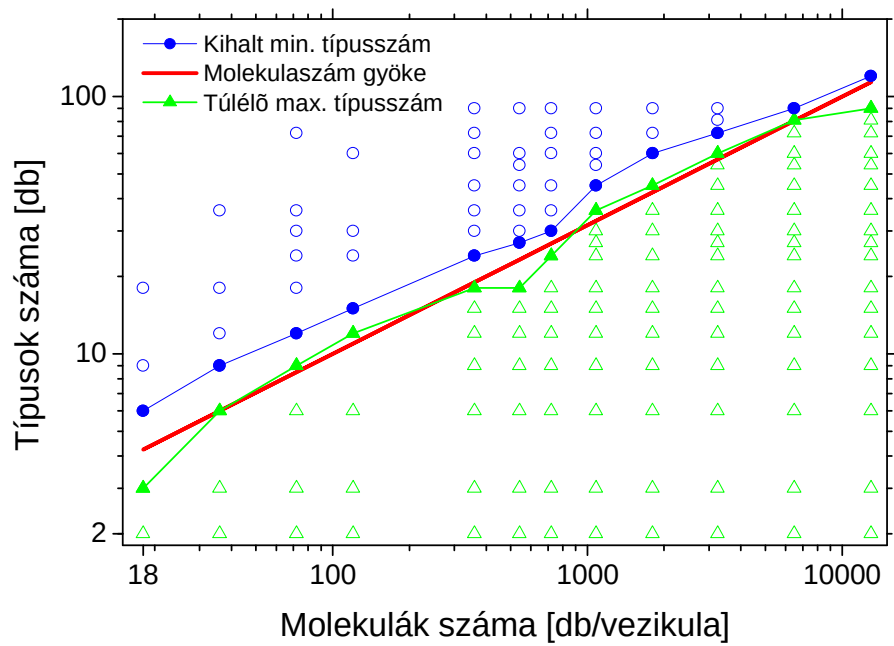
5. ábra: Átlagos rátermettség és túlélők relatív gyakorisága a vezikulaszám függvényében. Míg az alsó grafikonokon szembetűnő a telítési görbe szerinti összefüggés ($\tau_2 = 36$; $c_2 = 30$), a fenti grafikonokon ennek csak a platója látszik ($\tau_1 = 5$; $c_1 = 200$). A fenti grafikonok 100-100, az alsók 214-214 futás kimenetét összegzik; a pontok az egyes futásokat jelölik. Közös paraméterek: $v_{\min} = 250$; $v_{\max} = 5000$; $\varepsilon = 0,3$; $a_i = 1$, ahol $i \in [1; \tau]$. A molekulák száma is közel azonos: $m_1 = 1000$ és $m_2 = 1080$.

3.4. Molekulák száma (m)

Állandó vezikula- és molekulaszám mellett az egyre több típus fenntartása megnehezíti a túlélést: a rendszer átlagos rátermettsége a típuszám növelésével meredeken csökken (6. ábra). A molekulaszám növelésével – a populációméret növeléséhez hasonlóan (ld. 3.3. alfejezet) – telítésszerűen nő a rátermettség, ez segíti a több típus együttlését. $\nu = 1000$ vezikulát tartalmazó rendszerben, $m = 12960$ molekulaszám mellett a legnagyobb fenntartható típuszám $\tau = 90$. A molekulaszám függvényében fenntartható maximális típuszám görbéje (6. ábra, fehér és sötétkék területek határa) gyökös összefüggést sugall; a 7. ábra ezt egyértelműen igazolja. $\nu = 1000$ mellett a legnagyobb fenntartható típuszám közel egyenlő a molekulaszám gyökével.



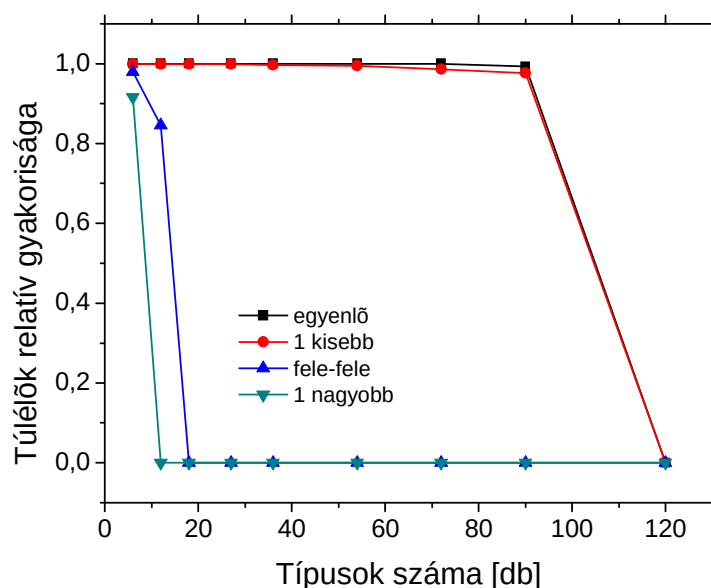
6. ábra: Átlagos rátermettség a molekula- és típuszám függvényében. A grafikon 1031 futás XYZ adatainak extrapolációjával készült – következtetéseinket természetesen a mögöttes adatokból vontuk le. Paraméterek: $g = 100000$; $\nu = 1000$; $\varepsilon = 0,3$; $a_i = 1$, ahol $i \in [1; \tau]$.



7. ábra: A túlélő, azaz a típusok együttélését biztosító, s a kihalást eredményező paraméterteret elválasztó kritikus típusszám a molekulák számának függvényében. Mindkét tengely mentén logaritmikus a skálázás! A paraméterter feltüntetett helyein a simulációt 6-8 ismétléssel futtattuk. A grafikon (a 6. ábrával azonos) 1031 futás kimenetét összegzi. Paraméterek: $g = 100000$; $v = 1000$; $\varepsilon = 0,3$; $a_i = 1$, ahol $i \in [1; \tau]$.

3.5. Replikációs affinitások (a)

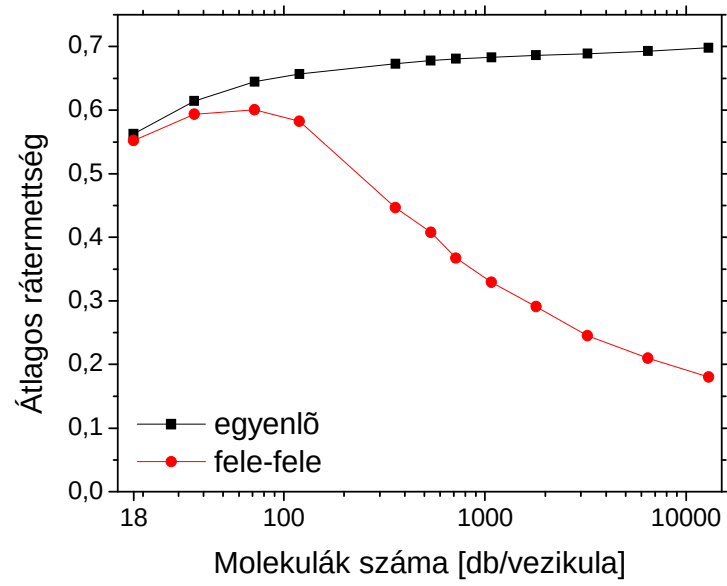
Idáig olyan típusokkal számoltunk, amelyek másolódási esélye azonos, így kevésbé veszélyeztetette fennmaradásukat valamelyikük túlszaporodása. A különböző térszerkezetű molekula típusok azonban eltérő affinitással kötnek az általános replikázhoz, így másolódási rátájuk is különböző lehet. Megvizsgáltuk, hogy eltérő replikációs affinitás értékekkel számolva mennyiben változik a fenntartható típusok száma (8. ábra). Szimulációnkban kétféle replikációs affinitással számoltunk: $a_{\text{nagyobb}} = 1$ és $a_{\text{kisebb}} = 0,9$. Amennyiben csak egy típus affinitása kisebb a többinél, akkor éppen ugyanannyi típus fenntartható, mint egyenlő affinitásokkal ($90 \leq \tau_{\text{max}} < 120$). Egészen eltérő hatású, ha a megkülönböztetett típus affinitása nagyobb a többiekénél: ilyenkor a fenntartható típusok száma hihetetlenül alacsony ($6 \leq \tau_{\text{max}} < 12$). A kisebb és nagyobb affinitású típusok fele-fele arányú megoszlásakor a fenntartható típusok száma szintén alacsony ($12 \leq \tau_{\text{max}} < 18$).



8. ábra: Az „egy kisebb” ($a_{1..(\tau-1)} = 1$ és $a_{\tau} = 0,9$), „fele-fele” ($a_{1..(\tau/2)} = 1$ és $a_{(\tau/2+1).. \tau} = 0,9$) és „egy nagyobb” ($a_1 = 1$ és $a_{2.. \tau} = 0,9$) affinitás-kiosztások hatása a fenntartható típusszámra. A grafikon 235 futás kimenetét összegzi; a fekete négyzet 4-6, a többi pont 7 ismétlés átlagát jelöli.

Közös paraméterek: $g = 100000$; $v = 1000$; $m = 12960$; $\varepsilon = 0,3$.

Megvizsgáltuk továbbá, hogy hogyan változik a molekulaszám növekedésével a rendszer átlagos rátermettsége (9. ábra). Egyenlő affinitások mellett a rátermettség telítésszerűen nőtt, ahogy azt a 3.4. alfejezetben láttuk. Azonban a fele-fele arányban eltérő affinitások fenntartását nem segíti a molekulaszám jelentős növelése: bár kezdetben itt is növekszik a rátermettség, $m = 72$ fölött jelentős (bár egyre lassuló) csökkenés mutatkozik benne.



9. ábra: Két típust tartalmazó rendszer ($\tau = 2$) kétféle affinitás-kiosztással: egyik az „egyenlő” ($a_1 = 1$; $a_2 = 1$), másik a „fele-fele” ($a_1 = 1$; $a_2 = 0,9$). A molekulaszám (m) növelésével eltérő tendenciát mutat a rátermettségük változása. A vízszintes tengely mentén logaritmikus a skálázás!

Közös paraméterek: $g = 100000$; $\nu = 1000$; $\varepsilon = 0,3$.

4. Diszkusszió

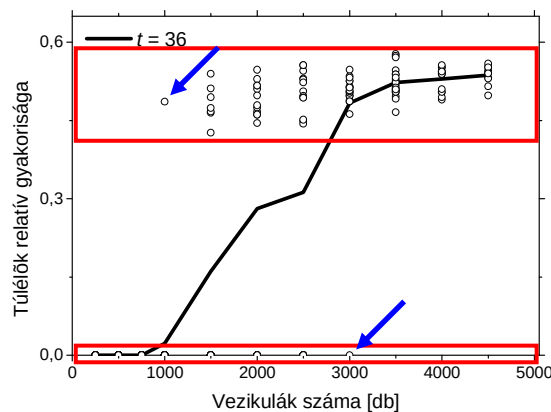
A SCM viselkedését tanulmányozó kutatások érdekességét az adja, hogy eredményeik betekintést nyújtanak az élet keletkezésének világába. Szimulációink értelmezésével mi is ezt célozzuk meg. Kíváncsiak vagyunk az imént megismert, információ fennmaradását szolgáló mechanizmusokra, a sztochasztikus korrekcióra és a többszintű szelekcióra. Érdekel bennünket, hogy ezek maximálisan hány különböző típusú molekula fenntartását teszik lehetővé, azaz hogy mekkora a második hibaküszöb; így adhatunk csak mennyiségi becslést a korai rendszerek összetettségére. És végül szeretnénk tudni, hogy az élet keletkezésének forgatókönyvében milyen szerep juthat a SCM-nek.

4.1. Többszintű szelekció

A molekulák és vezikulák sorsa a SCM-ben számos véletlen hatásnak van kitéve, ennek megfelelően egyes típusok kiveszhetnek, e helyeken pedig jelentősen lelassul az anyagcsere. Mégis, a rendszerek széles paraméter-tartományban képesek a hosszútávú fennmaradásra. És míg futáson belül szembetűnő fluktuációt mutat a pillanatnyi átlagos rátermettség (vö. 3.1. alfejezet, 3. ábra), a független futások sokgenerációs átlagai egymás közvetlen közelébe érkeznek. E dinamikai rendszer paraméterei tehát megszabnak egy stabil egyensúlyi pontot (*vonzó*), amihez bár tart az átlagos rátermettség, a véletlen folyamatok révén annak közeléből rendszeresen kimozdul. A dinamikus egyensúly feltételezi az anyagcsere sebességének, és ezen keresztül a túlélő vezikulák számának szabályozását – ennek csak a kétszintű, molekuláris és vezikuláris szinten ható, szelekció lehet a mechanizmusa. Enélkül elképzelhetetlen volna az információ fennmaradása.

Azonban nem a típusok koegzisztenciája az egyetlen egyensúlyi állapot: a rendszer kihalása is az. Ez utóbbi bár instabil, de egyben statikus egyensúlyi állapot: az ide jutó rendszerek ezt nem tudják elhagyni²¹. Az 5. ábra jobb alsó grafikonján igen látványos, hogy a futások (véges sok, de nem kevés generációt követően) e két állapot valamelyikében találhatóak (10. ábra, piros keretek). Ugyanezen grafikonon szembetűnő még, hogy viszonylag nagy szórás mellett is túlélhet egy rendszer, és alacsony szórás mellett is kihalhat (10. ábra, kék nyilak). Ez sokat elárul a dinamikáról: minél nagyobb szerepet kap az esetlegesség, annál valószínűbb, hogy egyszer végérvényesen kihal a rendszer; ugyanakkor a nagy számok törvénye a túlélés felé mozgatja azt.

²¹ Diszkrét rendszerünk ezen állapota ún. nyelő, az irányított gráfok kapcsán közkeletű kifejezéssel élve.



10. ábra: Az 5. ábra jobb alsó grafikonjának (a túlélők relatív gyakoriságának vezikulaszámtól való függése) értelmezése. Piros keret: a SCM egyensúlyi állapotai; kék nyíl: átlagtól markánsan eltérő adat.

Az információ fennmaradását három tényező is akadályozza: a zaj, a sodródás és a másolódási aszinkronitás²². Az első kettő csak sztochasztikus rendszerekben fordul elő, ezek kiküszöbölésével szolgál a többszintű szelekció. Alacsony molekulaszám a vezikulák osztódásakor okozhat zajt: amikor is egy viszonylag egyenletes típuseloszlású vezikula (és minden osztódó vezikula ilyen, hisz a kellő méret elérése jó anyagcserét igényel) egyenlőtlen leányvezikulákra bomlik. Nem meglepő tehát, hogy a molekulaszám növelésével nő az átlagos rátermettség (vö. 3.4. alfejezet, 6. ábra).

A kis populációméret (alacsony vezikulaszám) is kiszolgáltatja a rendszereket a sztochasztikus hatásoknak, ezt tükrözi az egyensúly körüli fluktuáció nagyobb szórása (vö. 3.1. alfejezet, 3. ábra). Kevés típus fenntartásakor ennek nincs hatása a rendszer túlélésére: az átlagos rátermettség ingadozásait korrigálni képes a SCM (vö. 3.3. alfejezet, 5. ábra, bal felső grafikon). Azonban sok típus fenntartásakor az egyensúlyi átlagos rátermettség alacsonyabb, így már egy kisebb fluktuáció is elegendő a rendszer kihalásához. Az alacsony vezikulaszám zajt és sodródást okozva veszélyezteti az információ fennmaradását (vö. 3.3. alfejezet, 5. ábra, bal alsó grafikon).

A molekula- és vezikulaszám emelésével nő a rendszer túlélésének valószínűsége, ráadásul mindkét esetben telítésszerűen. Ugyanis annak a valószínűségét, hogy a zajjal módosított érték negatív értéket vegyen fel, úgy is csökkenthetjük, hogy a stabil egyensúlyi pont (átlagos rátermettség) értékét növeljük, és úgy is, hogy a zaj szórását csökkentjük.

²² Zajért az ismétlések alacsony száma felelős; jellemzője a megnövekedett szórás, amely nagy eltéréshez vezethet az átlag és a várható érték között. Sodródást az alacsony elemszámmal történő (így kevésbé reprezentatív), és ismételt mintavétel okoz; az egyes mintavételek kihatnak a következő mintavételekben tapasztalható várható értékre, ilyenkor tehát az is eltolódhat. A másolási aszinkronitás ezekhez képest eleve kedvezőtlen várható értéket jelent.

4.2. Sztochasztikus korrekció

Ha a többszintű szelekció megmagyarázza a típusok együttélését, akkor mi szükség sztochasztikus korrekcióra? A válasz megdöbbentően egyszerű: a szelekció csak a létező egyedek közül válogat; és ha a populációban kevés a magas rátermettségű egyed, ez önmagában nem fogja fenntartani az előnyös jellegeket. A sztochasztikus korrekció lényege abban áll, hogy megnövelheti egyes egyedek rátermettségét. A kifejezés neve is erre utal: ahol a determinisztikus körülmények kihaláshoz vezetnek, ott a véletlen folyamatok, – az előnyös összetételű vezikulák folyamatos újratermelése révén, – képesek a rendszer túlélését biztosítani (Szathmáry és Demeter, 1987).

Leginkább tehát akkor van szükség sztochasztikus korrekcióra, ha egyébként egyáltalán nincsenek jól rátermett vezikulák: másolódási aszinkronitás esetén. Ezért a sztochasztikus korrekció, szemben a többszintű szelekcióval, akkor segíti legjobban a rendszerek fennmaradását, ha minél nagyobb az esetlegesség. A 9. ábra (3.5. alfejezet) ezt látványosan példázza: egy nagyobb és egy kisebb replikációs affinitással bíró típus együttélését az alacsony (kópia- és) molekulaszám okozta zajnál is jobban hátráltatja a magas (kópia- és) molekulaszám garantálta egyenlőtlen típuseloszlás. Érdekes volna tudni, hogy e görbe nullához, avagy pozitív értékhez tart! Fontanari *et al.* (2006) modellje, amely végtelen vezikulaszámmal dolgozik, nem vizsgált eltérő másolódási affinitásokat.

A 8. ábra (3.5. alfejezet) állandó (vezikulán belüli) molekulaszám mellett hasonlítja össze az egyre több típust fenntartó rendszereket. Azt látjuk, hogy a „fele-fele” és „egy nagyobb” affinitás-kiosztású rendszerek csak alacsony típusszám, azaz magas kópiaszám mellett élnek túl. Ez azt sugallja, hogy a magas kópiaszám segíti a kisebb affinitású típusok fenntartását, ami ellentmond az előző bekezdés érvelésének. Azonban ha alaposan szemügyre vesszük az ábrát, megállapíthatjuk azt is, hogy azok a rendszerek élnek túl, amelyekben nincs több 6 alacsony affinitású típusnál. Ennek magyarázatára egy harmadik információ-integrációs mechanizmust javaslunk, a komplementációt.

4.3. Komplementáció

Az alacsony replikációs affinitással bíró típusok várhatóan ritkábban másolódnak, mint az ugyanazon vezikulában levő és azonos kópiaszámmal képviselt, magasabb affinitású típusok; ezáltal az alacsony affinitású típusok kópiaszáma idővel lecsökken a vezikulákban. Ez pedig egy önerősítő folyamat, hisz minél kevesebben vannak, annál kevesebb másolat készül róluk a továbbiakban. Mi menti meg ezeket a teljes kiveszéstől? Természetesen az, hogy amelyik

vezikulában nincs (vagy csak kevés van) ezekből a ritka típusokból, annak leáll (illetve lelassul) az anyagcseréje; és ezáltal kiviszi a rendszerből a többi típus fölöslegben termelt, számos kópiáját.

Nézzük meg az esetet, amikor egyetlen ilyen típus van a rendszerben! Ha az osztódás során ennek a ritka típusnak követjük a kópiáit, akkor nyomát sem leljük a vezikulák pusztulásának: e molekulák mindig olyan környezetbe kerülnek, ahol az összes többi, gyakori típus jelen van. Ezt a jelenséget nevezzük komplementációnak: amikor egy típus olyan vezikulába kerül, ahol már az összes többi típus fellelhető, amely vezikulát tehát mindenképp megkíméli a szelekció. A ritka típus garantált komplementációja és a hiányos génkészletű vezikulákra ható szelekció biztosítja az alacsony affinitású típus, és azon keresztül a teljes rendszer fennmaradását. Ez magyarázza, hogy „egy kisebb” affinitás-kiosztás mellett ugyanannyi típus tartható fenn, mint egyenlő esetben (vö. 3.5. alfejezet, 8. ábra).

Azonban ha egy rendszerben – például egy kiugróan magas affinitású típus megjelenésével – egyszerre több típus affinitása lesz hasonlóan alacsony, akkor e típusok ritkává válását követően, az osztódások során már egyikük komplementációja sem lesz garantált. Ilyenkor lehetséges, hogy egy teljes génkészletű vezikula osztódásával létrejövő mindkét leányvezikula hiányos legyen, persze különböző típusokra nézve. Ezért minél több alacsony affinitású típus van egy rendszerben, a komplementáció hatása annál jelentéktelenebb. A szimulációk eredménye alapján úgy néz ki, hogy 6 alacsony affinitású típus fennmaradását – amin a sztochasztikus korrekció nem tudott segíteni – a komplementáció biztosítani tudja (vö. 3.5. alfejezet, 8. ábra).

4.4. A fenntartható információ becslése

Miután áttekintettük a SCM absztrakt információ megtartó mechanizmusait, térjünk rá a korai élő rendszerek tárgyalására! Ezek kapcsán – dolgozatom fő célkitűzéseként – a protosejtekben fenntartható gének maximális számát, a második hibaküszöb értékét keresem. A szimulációk során feltárt maximális fenntartható típusszám (egyenlő, illetve „egy kisebb” affinitás-kiosztások mellett) legalább 90, de legalábbis kisebb, mint 120 ($90 \leq \tau_{\max} < 120$). Ez vajon azt jelentené, az első hibaküszöb által meghatározott maximális molekulahosszal ($L = 873$) számolva, hogy a korai sejtek képesek lehettek egy akár $873 \cdot 90 = 78570$ nukleotidnyi genom fenntartására? A válasz egyértelmű: nem; a hibarátát figyelembevevő modellek nem vizsgálják a fenntartható gének számát, az e dolgozatban bemutatott modell pedig elhanyagolja a hibarátát.

A maximális génszámra vonatkozó $90 \leq \tau_{\max} < 120$ megállapításunk azonban csak a paramétertér egyetlen pontjára jellemző. A szimulációk eredményéből általánosságban csupán annyi szűrhető le, hogy a fenntartható típuszám a (vezikulákon belüli) molekulaszám gyökével arányos:

$$\tau_{\max} \sim \sqrt{m} \quad [5]$$

$\nu = 1000$ mellett e két érték közel egyenlőségét is megállapítottuk (ld. 7. ábra, 3.4. alfejezet), mely esetben a fenntartható típuszámot csak a kópiaszám limitálja. Elképzeléseink szerint az általunk használt, a belátható futásidőhöz választott populációméretek és molekulaszámok a reális értékek töredékei; magasabb vezikula- és molekulaszámok várhatóan még nagyobb típuszámot tesznek fenntarthatóvá. Ám mindezidáig ismeretlen, hogy milyen molekuláris koncentrációviszonyok uralkodhattak a prebiotikumban, illetve egy ebből felhalmozó vezikulában, valamint hogy mekkora lehetett a vezikulák populációmérete – így nem adhatunk pontos becslést a maximális fenntartható típuszámra. Az [5] arányosságban feltárt összefüggést nagyszerűen ki tudná egészíteni egy e kérdésre irányuló kísérleti kutatás.

Eigen (1971) nyomán, ahogy ő azt a(z első) hibaküszöb kapcsán tette, megvizsgálhatjuk hogy a második hibaküszöb gátat állít-e a sejtes élet kialakulásának. Ehhez meg kell becsülnünk, hogy egy sejt, illetve az ahhoz szükséges minimális anyagcsere örökítésére mekkora genom lehet képes. Kun és munkatársai (2005) kalkulációi nyomán a lehetséges legkisebb, minimális anyagcserét fenntartani képes genom méretét vegyük kb. 7000 nukleotid hosszúnak! Ennyi információ fenntartására, – a $\nu = 1000$ mérsékelt becsléssel élve, valamint – $\tau_{\max} = 90$ -t feltételezve, akár $L = 80$ nukleotid hosszú gének is képesek; ez a génhossz az első hibaküszöb figyelembevételével is biztosan fenntartható.

4.5. A komplexitás növekedése

A bevezetőben arról írok, hogy az élet keletkezése egy átmenet, – az abiogenetikusan is kialakuló szerves anyagok és az első egysejtű között; – valamint hogy az ehhez elengedhetetlen információ-integrációra éppen azért képes a SCM, mert az a Hiperciklussal szemben ez lehetőséget teremt a molekulatípusok, tehát az információ felhalmozására. Az összehasonlítást ott azért tettem, merthogy a Hiperciklus kifejezetten akadályozza a típusok beépülését. Most azonban nézzük meg az érme másik oldalát: vajon segíti-e a SCM a komplexitás növekedését?

A szabályzástól mentes, korai rendszerekben – mint már kifejtettem – valószínűtlen, hogy a molekulák különféle típusai azonos affinitással kötődjenek az őket másoló replikáz enzimhez. Sőt, e tekintetben valószínűleg minden két típus különbözött. Az egyenletes affinitás-kiosztás mellett fenntartható típuszámot tehát csak úgy érthette el a rendszer, ha az affinitás mértéke evolvábilis volt. Ez minden bizonnyal így is lehetett, hisz a molekulák kötődési erősségét a térszerkezet, és azon keresztül a szekvencia határozza meg – amiről pedig tudjuk, hogy processzív másolódása révén kifejezetten sokféle változat létrehozására képes (Szathmáry és Maynard Smith, 1997).

Azt is kiderítettük, hogy a sztochasztikus korrekció és a komplementáció mechanizmusai lehetőséget teremtenek a különböző affinitású típusok együttélésére. Ha kialakult egyszer egy kétféle és különböző affinitással bíró típust egyesítő rendszer, akkor az viszonylag alacsony kópiaszám mellett (vö. 3.5. alfejezet, 9. ábra) tartósan fennmaradhatott; mely idő alatt a típusok replikációs affinitása felzárkózhatott egymáshoz²³. Ez bár egy kis típuszámú rendszer, egyenlő az affinitás-kiosztása; és minden ilyen ($\tau < \tau_{\max}$) rendszerre igaz, hogy komplementációja révén képes egy alacsonyabb affinitású típus integrálására, anélkül hogy ezzel csökkenne a túlélő vezikuláinak száma. Azaz az integrációs és felzárkózási lépések során a rendszer újra és újra visszajut az egyenlő affinitás-kiosztású állapotba, így újra és újra képes egy-egy típus integrálására. Ez az evolúciós dinamika lehetővé teszi a genom méretének hosszútávú, fenntartható növekedését, azaz felvet egy utat, amely mentén megkerülhető Eigen Paradoxona.

²³ Önző replikátorokról lévén szó, az affinitás szinte kizárólag csak nőhet. Az alacsonyabb affinitással rendelkező mutáns elterjedésének gátat szab, hogy a vezikulán belül túlszaporodják a vele azonos típusú (aktivitású) molekulák. Ha pedig valamelyik vezikulában egy mutáns molekula affinitása a többieké fölé emelkedne, az magával együtt pusztulásba vinné az őt hordozó vezikulát.

5. Köszönetnyilvánítás

Hálás köszönet illeti témavezetőmet, Kun Ádámot mindazért amivel jelen kutatásomat támogatta: programozási ismereteim elmélyítéséért, az abiogenezis tudományából bemutatott érdekességekért, lelkesedésem fenntartásáért, és nem utolsósorban a dolgozat lektorálásáért.

A kutatásom koncepcióját Szathmáry Eörsnek köszönhetem. Neki, valamint munkatársainak, Mauro Santosnak és Elias Zintzarasnak tartozom köszönettel a vezikulák rátermettség-függvényét illető kritikai észrevételekért.

Köszönöm az ELTE Informatikai Igazgatóságának, hogy felhasználhattam az Atlasz nagyteljesítményű számítógépet a szimulációk futtatására, és kifejezetten is Maulis Ádám rendszergazdának, hogy többször is készségesen a rendelkezésemre állt.

Köszönöm szépen bátyámnak, Hubai Tamásnak, hogy lehetővé tette, hogy távolról hozzáférjek az Atlaszon futó szimulációimhoz, az EISZ-en keresztül tudományos publikációkhoz, valamint hogy segített a programomban felmerült hibák azonosításában.

Hála illeti szüleimet a taníttatásomba való befektetésükért, a belém vetett bizalmukért.

6. Irodalomjegyzék

1. Anfinsen, C. B. (1973) Principles that govern the folding of protein chains. *Science* **181**, 223–230.
2. Bartel, D.P. & Unrau, P.J. (1999) Constructing an RNA world. *Trends Cell. Biol.* **9**, M9-M13.
3. Brack, A. (1998) *Introduction*. In: André Brack (ed.) *The Molecular Origins of Life: Assembling Pieces of the Puzzle*. p. 1. Cambridge: Cambridge University Press.
4. Czárán, T. & Szathmáry, E. (2000) *Coexistence of competitive-mutualist replicators in prebiotic evolution*. In: Dickmann, U., Law, R. and Metz, J.A.J. (eds.) *The Geometry of Ecological Interactions: Simplifying Spatial Complexity*. pp. 116-134. Cambridge: Cambridge University Press
5. Darwin, C. (1959) *The origin of species by means of natural selection*. London: John Murray
6. Darwin, F. (1887) *The life and letters of Charles Darwin, including an autobiographical chapter*. **3**, 18. London: John Murray
7. Eigen, M. (1971) Self organization of matter and the evolution of biological macromolecules. *Naturwissenschaften* **10**, 465 -523.
8. Eigen, M. & Schuster, P. (1977) The Hypercycle, part A: Emergence of the Hypercycle. *Naturwissenschaften* **64**, 541-565.
9. Fontanari, J.F., Santos, M. & Szathmáry, E. (2006) Coexistence and error propagation in pre-biotic vesicle models: A group selection approach. *J. Theor. Biol.* **239**, 247-256.
10. Gilbert, W. (1986) The RNA world. *Nature* **319**, 618.
11. Jensen, R.A. (1976) Enzyme recruitment in evolution of new function. *Annu. Rev. Microbiol.* **30**, 409-425.

12. Johnston, W.K., Unrau, P.J., Lawrence, M.S., Glasner, M.E. & Bartel, D.P. (2001) RNA-catalyzed RNA polymerization: Accurate and general RNA-templated primer extension. *Science* **292**, 1319 -1325.
13. Kauffman, S. A. (1986) Autocatalytic sets of proteins. *J. Theor. Biol.* **119**, 1–24.
14. Könnyű, B., Czárán, T. & Szathmáry, E. (2008) Prebiotic replicase evolution in a surface-bound metabolic system: Parasites as a source of adaptive evolution. *BMC Evol. Biol.* **8**, 267.
15. Kun, Á., Santos, M. & Szathmáry, E. (2005) Real ribozymes suggest a relaxed error threshold. *Nat Genet* **37**, 1008-1011.
16. Kunkel, T. A. (2004) DNA replication fidelity. *J. Biol. Chem.* **279**, 16895–16898.
17. Luisi, P. L. (1998) About various definitions of life. *Orig. Life Evol. Biosph.* **28**, 613–622.
18. Maynard Smith, J. (1983) Models of evolution. *Proc. R. Soc. Lond. B* **219**, 315 -325.
19. Maynard Smith, J. (1987) *How to model evolution*. In: Dupré, J. (ed.) *The Latest on the Best. Essays on Evolution and Optimality*. pp 119–131. Cambridge: MIT Press.
20. Maynard Smith, J. (1979) Hypercycles and the origin of life. *Nature* **280**, 445-446.
21. Oparin, A. I. (1936) *The origin of life*. Moscow: Moscow Worker Publisher.
22. Pasteur, L. (1878) The germ theory and its applications to medicine and surgery. *Comptes Rendus de l' Academie des Sciences*, **LXXXVI.**, 1037-1043.
23. Pigliucci, M. (2009) An extended synthesis for evolutionary biology. *The Year in Evolutionary Biology 2009: Ann. N.Y. Acad. Sci.* **1168**, 218–228.
24. Pigliucci, M., Müller, G.B. (eds.) (2010) *Evolution - the Extended Synthesis*. The MIT Press
25. Schaaper, R. M. (1993) Base selection, proofreading, and mismatch repair during DNA replication in *Escherichia coli*. *J. Biol. Chem.* **268**, 23762–23765.
26. Schrum, J.P., Zhu, T.F. & Szostak, J.W. (2010) The origins of cellular life. *Cold Spring Harbor Perspectives in Biology* **2**.

27. Schuster, P. (2010) Mathematical modeling of evolution. Solved and open problems. *Theory Biosci.* **130**, 71-89.
28. Segré, D. & Lancet, D. (2000) Composing life. *EMBO Reports* **1**, 217–222.
29. Szabó, P., Scheuring, I., Czaran, T. & Szathmáry, E. (2002) In silico simulations reveal that replicators with limited dispersal evolve towards higher efficiency and fidelity. *Nature* **420**, 340-343.
30. Szathmáry, E. & Demeter, L. (1987) Group selection of early replicators and the origin of life. *J. Theor. Biol.* **128**, 463-486.
31. Szathmáry, E. & Maynard Smith, J. (1997) From replicators to reproducers: The first major transitions leading to life. *J. Theor. Biol.* **187**, 555–571.
32. Szathmáry, E. (2007) Coevolution of metabolic networks and membranes: The scenario of progressive sequestration. *Phil. Trans. R. Soc. B* **362**, 1781 –1787.
33. Takeuchi, N., Poorthuis, P. & Hogeweg, P. (2005) Phenotypic error threshold; additivity and epistasis in RNA evolution. *BMC Evol. Biol.* **5**, 9.
34. Unrau, P.J. & Bartel, D.P. (1998) RNA-catalysed nucleotide synthesis. *Nature* **395**, 260-263.
35. Vasas, V., Szathmáry, E. & Santos, M. (2010) Lack of evolvability in self-sustaining autocatalytic networks constraints metabolism-first scenarios for the origin of life. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 1470–1475.
36. Wochner, A., Attwater, J., Coulson, A. & Holliger, P. (2011) Ribozyme-catalyzed transcription of an active ribozyme. *Science* **332**, 209 -212.
37. Yarus, M. (1999) Boundaries for an RNA world. *Curr. Opin. Chem. Biol.* **3**, 260-267.
38. Zaher, H.S. & Unrau, P.J. (2007) Selection of an improved RNA polymerase ribozyme with superior extension and fidelity. *RNA* **13**, 1017-1026.
39. Zintzaras, E., Santos, M. & Szathmáry, E. (2002) „Living” under the challenge of information decay: The Stochastic Corrector Model vs. Hypercycles. *J. Theor. Biol.* **217**, 167-181.