Az anyagcsere szerkezetének hatása a genetikai interakciókra és a genomszerveződésre

Ph.D. értekezés

Kovács Károly

Témavezető: Papp Balázs

Biológia Doktori Iskola

MTA Szegedi Biológiai Kutatóközpont Biokémiai Intézet

SZTE TTIK

Szeged, 2012

Tartalomjegyzék

1 Bevezetés

A genotípus és fenotípus közötti kapcsolat feltérképezése a biológia alapvető fontosságú problémái közé tartozik. Habár a probléma régi, a nukleotidszekvenciáktól a fenotípusos jellegekig vezető út szisztematikus vizsgálata és realisztikus matematikai modellek megalkotása csak az utóbbi években kezdődhetett el. Ezt az évtizedek alatt felhalmozódott molekuláris biológiai tudás mellett a nagy áteresztőképességű (high-troughput) technológiák megjelenése tette lehetővé (Westerhoff and Palsson, 2004). Először a 90-es évek végétől - a szekvenáló technikák fejlődése révén - a genotípusokra vonatkozó ismereteink nőttek meg robbanásszerűen. A nagyszámú ismertté vált genomszekvencia lehetővé tette a genom szerkezete, funkciója és evolúciója vizsgálatát, az így életre kelt tudomány a genomika. A további nagy áteresztőképességű technikák megjelenésével újabb és újabb molekuláris alkotóra, sejtes alrendszerre váltak elérhetővé szisztematikus, szervezet szintű, kvantitatív adatok, újabb és újabb „omikák" megjelenését eredményezve. Végül a genomikai és nagy léptékű fenotípusos adatok matematikai modellbe integrálásával elsőként vált lehetővé a genotípus-fenotípus kapcsolat szisztematikus, nagy léptékű feltérképezése. Az így létrejött új tudományterület a rendszerbiológia (Bruggeman and Westerhoff, 2007). Disszertációm kérdésfelvetései a genom anatómiájának vizsgálatával egyrészt a genomikához, másrészt az anyagcserehálózatok rendszerszintű vizsgálatával a rendszerbiológia területéhez kapcsolódnak.

Az anyagcsere talán a legrészletesebben ismert sejtes alrendszer, ami a legkézenfekvőbb vizsgálati területté teszi, ha a genotípus-fenotípus kapcsolat jobb megértését tűzzük ki célul. De vajon valójában mennyire használható mai tudásunk a mutációktól a rátermettségig vezető út megértésében, előrejelzésében? Disszertációm második fejezetében ezt a kérdést elemzem a nullmutációk (teljes funkcióvesztéssel járó mutációk) közötti kölcsönhatások speciális esetében (haploid élesztő törzsek rétermettségét telepméreteik alapján becsülve). Másrészt az adatok lehetővé teszik az anyagcserehálózat működésében mutatkozó főbb szabályszerűségek vizsgálatát. A rendszerbiológia egyik általános kérdése, hogy a talált mintázatok mennyiben adaptívak, tükröznek „mérnöki elveket" (design principles) (Alon, 2003; Alon, 2006; Papp et al. 2009) Például mikrobákban egy biomassza-alkotó molekula előállításért felelős útvonal sebességének emelkedése hozzájárulhat a rátermettség növeléséhez.Vajon ez mennyire tükröződik az anyagcsereútvonalak időbeli működésének szabályozottságában? Disszertációm

harmadik fejezetében a fenti kérdést a szabályozottság genomszerveződésre tett lehetséges hatásának tükrében vizsgálom.

1.1 Az anyagcsere rendszerszintű vizsgálata

A következőkben röviden áttekintem az anyagcsere mai, rendszerszintű vizsgálati lehetőségeit, illetve az oda vezető utat (Papin et al., 2003; Westerhoff and Palsson, 2004). Az anyagcsereútvonalak koncepciója és lépésenkénti feltérképezésük feladata a molekuláris biológa történetében tradicionálisan fontos szerepet tölt be (Papin et al., 2003). Ugyanakkor az útvonalleírásokkal párhuzamosan már a 60-as évek végétől megjelentek az anyagcsere működésének megértését célzó számítógépes kinetikai modellek (Garfinkel et al., 1970). Az egyes útvonalak modelljeit egyre nagyobb metabolikus hálózatokéi követték, a 80-as évek végére eljutva az emberi vörösvértest kinetikai modelljéig (Joshi and Palsson, 1989; Westerhoff and Palsson, 2004). A 70-es évekre a metabolikus kontroll analízissel az anyagcsereútvonalak rendszerszintű, ún. emergens[1] tulajdonságainak elméleti vizsgálata is megjelent (Kacser & Burns 1973; Heinrich & Rapoport 1974). A teljes anyagcserehálózat tulajdonságainak vizsgálatára azonban csak a nagy áteresztőképességű technikák megjelenése után kerülhetett sor. A legtöbb genomikai és posztgenomikai adattal két tradicionális modellélőlény emelkedett a genomikának és a rendszerbiológiának is legfontosabb vizsgálati alanyává: prokarióták közül az *Escherichia coli* (Mori 2004), eukarióták közül pedig a sörélesztő *Saccharomyces cerevisiae* (Castrillo and Oliver, 2004). Disszertációmban én is erre a két szervezetre koncentrálok. A 90-es évek közepétől jelentek meg a szervezetszintű genomikai és posztgenomikai adatot tartalmazó metabolikus adatbázisok, mint pl. a MetaCyc, KEGG (Ogata et al., 1999; Caspi et al., 2011; Karp and Caspi, 2011). Ezzel egyrészről lehetővé vált a tradicionális anyagcsereútvonalak genomszintű elemzése, különböző fajok útvonalainak összehasonlítása (Papin et al., 2003), olyan, a teljes alrendszerre jellemző tulajdonságok, mint pl. a szubsztrátok reakciónkénti számának vagy az egyes enzimek által katalizált reakciók száma eloszlásának statisztikai vizsgálata (Ouzounis & Karp 2000). Másrészről a sejtszintű adatok alapján a teljes anyagcserehálózatok egységes matematikai modellbe foglalt rekonstrukciója is elkezdődhetett (Covert et al., 2001).

[1] Emergens: egy rendszer egyes komponensei által külön-külön nem mutatott, azok interakcióiból fakadó tulajdonság.

A modelleket három nagy családba sorolhatjuk: topológiai, kényszer-alapú és kinetikai modellek. Ezek közül a legegyszerűbb az anyagcserehálózat topológiai vizsgálatát jelenti (Jeong et al., 2000). A hálózatok szerkezetét leíró topológiai modellek előnye, hogy valóban genomléptékűek, így a hálózat globális szerkezetének egyes szabályszerűségei feltárhatók. Például a metabolitok kapcsolódási száma hatványfüggvény eloszlást mutat, ami azt jelenti, hogy néhány anyagcseretermléknek nagyon sok kapcsolata van, míg a legtöbbnek csak néhány (Barabási and Oltvai, 2004). Emellett a hálózat hierarchikus modularitást mutat, vagyis kisebb egymás között sűrűn összekötött modulok nagyobb, lazábban kapcsolt csoportokba szerveződnek, valamint a topológiai szintű modularitás a funkcionális modularitást (specifikus biológiai funkciót ellátó alegység) is tükrözi (Ravasz et al., 2002). Ugyanakkor egy topológiai modell nem használ fel semmilyen arra vonatkozó információt, hogy az adott hálózat biokémiai reakciókból áll, így annak funkciójára és evolúciójára vonatkozó következtetések levonására csak korlátozottan alkalmas (Stelling, 2004; Steuer and Junker, 2008). Például míg egy fehérje-fehérje fizikai interakciós hálózatban a fehérje kapcsolatainak száma pozitívan korrelál a kódoló gén esszencialitásával (Jeong et al., 2001) (de lásd (Yu et al., 2008)), egy anyagcserehálózatban ennek megfelelő korrelációt nem találunk a metabolit reakció-összeköttetéseinek száma és az esszenciális reakciók aránya között (Mahadevan and Palsson, 2005). Ennek oka akkor válik érthetővé, ha figyelembe vesszük a hálózat éleinek jelentését: például egy esszenciális terméket termelő tökéletesen lineáris útvonalban szereplő metabolitnak csak két kapcsolata van, de bármelyik eltávolítása letális.

A három modelltípus közül a legtöbb biológiai információt az ún. kinetikai modellek hordozzák a hálózat minden reakcióját enzimkinetikai mechanizmusokkal írva le, lehetővé téve a rendszer dinamikus modellezését. Azonban a kinetikai modellek nagyobb léptékű kiterjesztését, az újabb előrelépések (Jamshidi and Palsson, 2008; Smallbone et al., 2010) ellenére, az enzimkinetikai adatok kis száma jelenleg is akadályozza (Steuer & Junker 2008).

E két modellcsalád között helyezkednek el információtartalomban az ún. kényszer-alapú (constraint-based) modellek, melyek a hálózat topológiáján túl olyan fizikokémiai és biológiai kényszereket is magukban foglalnak, mint a reakciók sztöchiometriája és reverzibilitása, vagy a környezetből felvehető tápanyagok listája (Price et al. 2004). Elnevezésük onnan ered, hogy a kinetikai modellekkel szemben a modell változóinak becslése a lehetséges megoldások terének szűkítésével történik. A kényszer-alapú modellek is genomléptékű adatokat integrálnak (600-1300 gén), ugyanakkor alkalmasak az anyagcsere génjeit adott környezetben a biomasszatermelés optimális mértékére mint fenotípusra

leképezni, illetve arra kvantitatív előrejelzést tenni (Edwards et al., 2001). Mindezek miatt a genotípus-fenotípus leképezés vizsgálatában, ill. általánosabban az evolúciós rendszerbiológiai alkalmazásokban kulcsszerepet játszik (Oberhardt et al. 2009; Papp et al. 2011). Nevében a kényszer a fizikokokémiai kényszereket jelöl, amelyek segítségével a lehetséges fenotípusok terét leszűkíthetjük. Az anyagcseremodellekben legalább két ilyen kényszert alkalmaznak: a reakciók sztöchiometriájából fakadóak (tömegegyensúly) illetve enzimkapacitásbeli korlátok, utóbbi például azt jelenti, hogy ismert irányú reakció esetében a fluxus csak az egyik irányban folyhat. A kényszerek működésének feltétele a steady-state állapot, vagyis minden metabolit azonos rátával termelődik és használódik fel. A környezetben levő tápanyagok a hozzájuk tartozó transzportreakciók fluxusértékének beállításával szimulálhatók. A lehetséges fluxusok terében az FBA (flux balance analysis) módszer egy optimalitási függvény alapján, ami lehet akár egy adott termék maximális termelése, vagy legtöbbször a biomassza termelés maximális mértéke, megkeresi az optimális fluxuseloszlást. Egyszerűsége és előfeltevései ellenére számos különböző területen alkalmazták sikerrel (Oberhardt et al. 2009), mint pl. gén esszencialitás predikciója (Förster et al., 2003; Duarte et al., 2004; Kuepfer et al., 2005), laborbeli szelekciós kísérlet során elért optimális metabolikus állapot megjóslása (Ibarra et al. 2002), vagy akár 200 millió év alatt lezajlott genomredukció szimulálása endoszimbionta baktériumokban (Pál et al. 2006). Fő hátrányai közé sorolható, hogy a steady-state feltétel nem teszi lehetővé a dinamika vizsgálatát, illetve a modell nem tartalmaz a szabályzásra vonatkozó explicit mechanizmusokat, vagyis optimális szabályzást feltételez. Ugyanakkor történtek fontos előrelépések e korlátok kiküszöbölésére (Min Lee et al., 2008; Jamshidi and Palsson, 2010).

Munkám során mind tradicionális útvonal adatbázis adatait, genomléptékű anyagcserehálózat rekonstrukciót, kényszer-alapú modellezés predikcióit és kinetikai modellt is felhasználtam az anyagcsere tulajdonságainak elemzésében.

1.2 Anyagcsere és evolúció: vizsgálandó kérdések

A kísérletes és elméleti háttér gazdagsága különösen kedvező feltételeket teremt a metabolikus hálózatok evolúciójának vizsgálatához. Ezzel kapcsolatban kétféle irányban is tehetünk fel kérdéseket.

Egyrészről milyen evolúciós erők hatására alakultak ki a metabolikus hálózatok egyes tulajdonságai? Megmagyarázhatók-e a neutrális nullmodell alapján, vagy csak adaptív hipotéziseket megfogalmazva? Vagyis általánosságban ezek a tulajdonságok mennyire finomhangoltak a természetes szelekció által?

Másrészről milyen további evolúciós következményei vannak a gének anyagcserehálózatba való szerveződésének? Két mutáció genetikai interakcióban (episztázis) van, ha a kettős mutáns fenotípusa eltér attól, amit az egyes mutánsok alapján várnánk. Vajon mennyiben határozza meg két gén funkcionális kapcsolata a mutációik közti genetikai interakciót, ezen keresztül pedig a lehetséges evolúciós útvonalakat (Poelwijk et al., 2007)? Milyen hatása van az anyagcsere funkcionális szerveződésének a genom szerveződésére? Vajon lehet-e hatása két enzim útvonalbeli sorrendjének génjeik kromoszómán való elhelyezkedésére?

Mint bármilyen fenotípusos mintázatnak, az anyagcsere jellegzetességeinek is különböző evolúciós okai lehetnek: lehet sodródás, mutációs nyomás eredménye, a természetes szelekció révén rögzült adaptív jelleg, vagy más adaptív jellegekre ható szelekció melléktermék (Lynch, 2007; Papp et al., 2009). Például az anyagcsere már említett topológiai jellegzeteségeit (kapcsolatainak hatványfüggvény eloszlása, modularitás) sokan adaptív jellegként értelmezték. Például a modularitás lehet a mutációkkal szembeni robosztusságra ható szelekció (Wagner, 2000) vagy a változatos környezethez való alkalmazkodás ereménye (Parter et al., 2007). Ugyanakkor számítógépes szimuláció alapján az anyagcserehálózat a kapcsolatok számának eloszlása a gyors növekedésre irányuló szelekció melléktermékeként is megjelenhet (Pfeiffer et al., 2005). Másrészt az anyagcserehálózat szerkezetének atmoszférikus reakcióhálózatokkal mutatott hasonlósága a topológia neutrális jellegét támogatja (Holme et al., 2011). Más jellegeknél az optimális („mérnöki") modelleknek való megfelelés szelekciós hatás jelenlétére utalhat (Papp et al. 2009). Például a hagyományos anyagcsereútvonalak előrejelezhetők egy, a reakciók számát minimalizáló és maradék ATP-t maximalizáló modellel (Beasley and Planes, 2007).

Vajon az anyagcsere expressziós szabályzása mennyire a szelekció által finomhangolt? A génexpresszió különböző aspektusainak optimalizálására ható szelekció létét modellek és

kísérleti eredmények is alátámasztják. Először is, a génexpresszió szintje szelekció alatt áll: a *lac* operon génjei kifejeződésének mértéke néhány száz generációs kísérleti evolúció alatt képes elérni az adott környezetben optimális mennyiséget (Dekel and Alon, 2005). Másodszor, annak is lehet funkcionális magyarázata, hogy egy gén aktivátor vagy represszor által szabályozódik-e. Kísérleti eredmények szerint a szabályzás módja azzal függ össze, hogy a gén termékének az idő mekkora hányadában kell jelen lennie megfelelő mennyiségben a baktérium természetes környezetében. Amelyik géntermékre az idő nagy részében szükség van, az többnyire pozitív, amelyikre ritkán, az többnyire negatív szabályzás alatt áll (Savageau, 1977). Ennek eredménye, hogy a regulátor kötőhelye mindkét esetben az idő nagy részében foglalt. A mintázat oka a téves regulátorok kötődéséből adódó hibák minimalizálására ható szelekciós nyomás lehet (Shinar et al., 2006; Sasson et al., 2012). Végezetül léteznek modellek a génexpresszió időbeli szabályozásának optimális módjára is, pontosabban arra, hogy miképpen lehet a legalacsonyabb össz-enzimkoncentráció felhasználásával a leggyorsabban bekapcsolni egy metabolikus útvonalat (Klipp et al., 2002; Zaslaver et al., 2004). A modellek által előrejelzett „tervezési elv" működését kísérletes eredmény is alátámasztja (Zaslaver et al. 2004). Különböző operonok promoter aktivitásának összehasonlítása azt mutatja, hogy minél közelebb van az enzim egy adott metabolikus útvonal kezdetéhez, annál gyorsabban aktiválódik és annál magasabb expressziós szintet ér el. Például az arginin bioszintézis egymás után következő reakciólépéseihez szükséges operonok kb. 10 perc különbséggel aktiválódnak. Disszertációm 3. fejezetében azt vizsgálom, vajon az azonos operonban elhelyezkedő gének szerveződése hozzájárulhat-e a génexpresszió optimális időbeli szabályzásához.

1.3 Célkitűzések

Disszertációm a továbbiakban két nagy fejezetre oszlik, melyeken belül külön bevezető, módszer, eredmények és diszkusszió találhatók. A különválasztásukat az eltérő kérdésfeltevés indokolja. Ami közös bennük, hogy mindkettő az anyagcserében ható génpárok relatív viszonyainak genetikai és evolúciós következményeit vizsgálja.

A második fejezetben azt kérdezem, hogy ha ismerjük két gén relatív pozícióját az anyagcserehálózatban, meg tudjuk-e magyarázni/jósolni a két gén között várható genetikai interakciót? Azaz a kérdés, hogy két gén együttes nullmutációja hogyan fogja befolyásolja a rátermettséget az egyedi mutációk alapján várt kombinált hatáshoz képest? Itt tehát két gén

nullmutációjának a rátermettségre tett azonnali hatását elemzem statisztikai és adatbányászati eszközökkel, közvetlen kvantitatív kísérletes adatok felhasználásával.

A harmadik fejezetben két gén relatív anyagcsereútvonalbeli pozíciójának hosszú távú evolúciós hatását vizsgálom. Befolyásolhatja-e ez a genom szerveződését, konkrétabban a gének operonon belüli sorrendjét? A talált genomi mintázat adaptív jellegét matematikai modellel támasztom alá.

2 Genetikai interakciók modularitása és prediktálhatósága a sörélesztő (*Saccharomyces cerevisiae*) anyagcserehálózatában

2.1 Bevezetés

Milyen mértékben jelezhetők előre illetve érthetők meg az anyagcsere génjei közt fellépő genetikai interakciók a metabolikus hálózatra vonatkozó ismeretek segítségével?

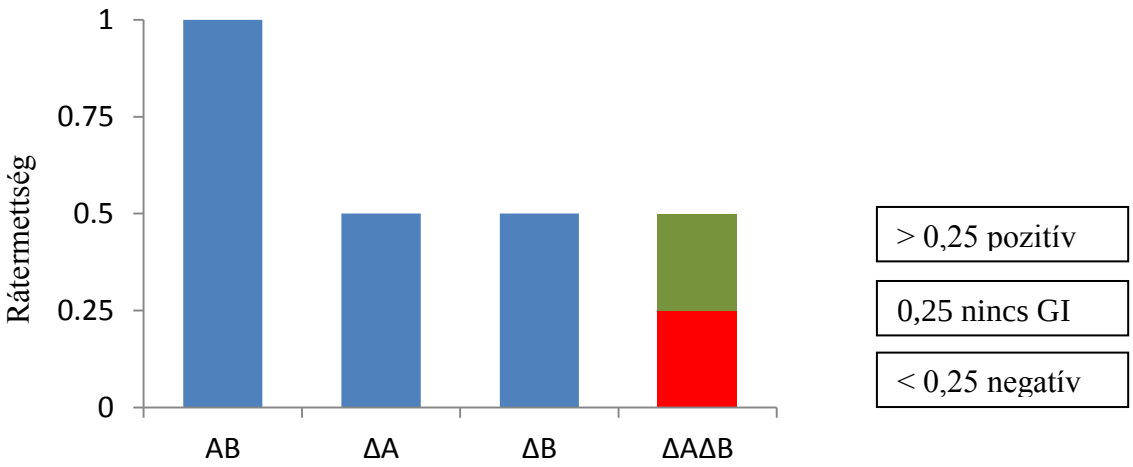
Genetikai interakciónak (GI, episztázis[2]) azt nevezzük, ha egy mutáció fenotípusos hatása nem független egy másik mutáció jelenlététől. A továbbiakban a fenotípusos hatásokon belül a rátermettségre gyakorolt hatást vizsgálom, annak evolúciós jelentősége miatt. Negatív GI-nál a kettős mutáns rátermettsége alacsonyabb, mint azt az egyes mutációk külön-külön vett hatása alapján várnánk, pozitív GI esetén pedig magasabb. Káros mutációknál ez azt jelenti, hogy a pozitív GI esetében a vártnál kisebb a rátermettség csökkenése, negatív GI-nál pedig nagyobb. A negatív GI esetén kapott fenotípusokat synthetic sick/lethal-nak (SS/SL) is nevezzük, utalva arra, hogy két, egyenként nem letális mutáció együttesen letális fenotípust eredményezhet[3].

De mi is pontosan az egyes mutációk hatása alapján várt kombinált hatás, amely alapján a genetikai kölcsönhatásokat definiáljuk? Bár a válasz nem triviális (Mani et al., 2008), rátermettség vizsgálata esetén a legszélesebb körben alkalmazott modell a multiplikatív (Dixon et al., 2009). Eszerint a várt érték az egyes mutánsok relatív rátermettségének szorzata, a negatív és pozitív GI-t pedig az ettől való negatív illetve pozitív eltérésként definiáljuk (*1. ábra*). Az GI mértéke a fenti definíció alapján kiszámolható:

$$\varepsilon = f_{12} - f_1 \times f_2$$

[2] Az episztázis kifejezésnek a fentitől eltérő értelmezései is vannak (Cordell, 2002), ezért az egyértelműség kedvéért disszertációmban végig a genetikai interakció (GI) terminust használom.

[3] A káros mutációk negatív illetve pozitív GI-jainak számos elterjedt alternatív megnevezése is előfordul a szakirodalomban. Negatív: szinergisztikus, synthetic enhancer, aggravating. Pozitív: antagonisztikus, buffering, csökkenő hozam, alleviating.

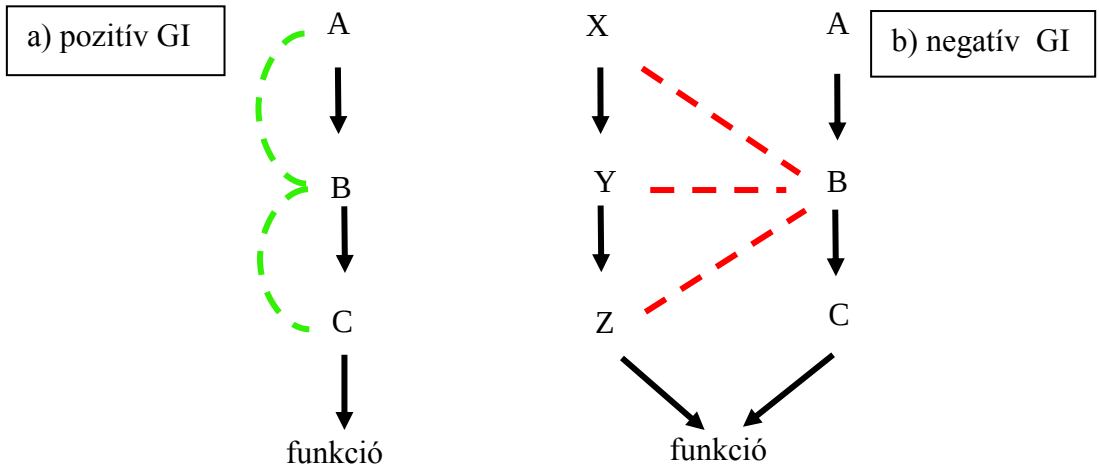


1 .ábra: Példa GI meghatározására a multiplikatív modell alapján.

A vad típus rátermettsége 1, míg ΔA és ΔB nullmutánsoké 0,5. A kétszeres nullmutáns várt értéke 0,25, ha a tapasztalt érték ennél nagyobb, pozitív (zölddel jelölve), ha kisebb, negatív GI-ról beszélünk (piros).

2.1.1 A genetikai interakciók lehetséges intuitív értelmezési módjai

Kutatásomban két gén nullmutációja közötti GI-kat vizsgálom. Vajon ezekben az esetekben milyen intuitív mechanisztikus értelmezései lehetségesek a genetikai kölcsönhatásoknak? Ezekre láthatunk két egyszerű példa modellt a *2. ábrán*. Az első példában, egy lineáris anyagcsereút esetén, bármelyik gén kiesése az útvonal funkciójának teljes elvesztéséhez vezet, így hasonló rátermettség csökkenést eredményez. Ugyanakkor egy második gén kiütése ugyanabban az útvonalban már nem csökkenti tovább a rátermettséget, tehát a kettős mutáns rátermettsége magasabb, mint amit az egyenkénti mutációk alapján várnánk, így pozitív GI-ról beszélhetünk. Ezzel analóg példát lehet elképzelni pozitív GI-ra fehérjekomplexeken belüli alegységek esetén (azaz bármely alegység elvesztése a teljes fehérjekomplex funkciójának kieséséhez vezet). A második példában, ha egy funkciót két párhuzamos anyagcsereút is képes betölteni, akkor egyetlen gén kiütését a redundáns útvonal működése kompenzálni képes, így nincs funkciókiesésből fakadó rátermettségcsökkenés. Ellenben ha a párhuzamos útvonalból is kiütünk egy gént, a funkció megszűnik, a kettős mutánsban jóval nagyobb rátermettségcsökkenést tapasztalunk, mint az egyes deléciók alapján várnánk, vagyis negatív GI-ról beszélhetünk. E két modellt gyakran tekintik kiindulási alapnak a GI-k értelmezésénél (Dixon et al., 2009), és ezek alapján azt várjuk, hogy útvonalon belül a pozitív GI-k, míg útvonalak között a negatív GI-k vannak túlsúlyban.



2. ábra: pozitív és negatív GI-k egyszerű példa modelljei.

Pozitív GI esetén egyetlen lineáris anyagcsereútat (a), negatív GI esetén két redundáns anyagcsereutat (b) láthatunk. A szaggatott vonalak a B gén pozitív illetve negatív GI-it jelzik deléciók esetén. Részletes magyarázat a főszövegben.

2.1.2 A genetikai interakciók jelentősége

Miért fontos feladat a GI vizsgálata? A GI-k gyakorlati és elvi szempontból is fontosak. Már puszta létük is azt jelenti, hogy sem a genetikai rendszerek működését, sem evolúcióját nem érthetjük meg egy-egy gént külön-külön vizsgálva.

Feltehetően a legtöbb betegség több gén mutációja által befolyásolt, így genetikai hátterük felderítésében a GI-k meghatározása fontos szerepet tölthet be (Lehner, 2007; Maxwell et al., 2008; Cordell, 2009). Az utóbbi évek humán genetikájának egyik nagy megoldatlan problémája, hogy a legtöbb többgénes emberi jellegnél a genomléptékű génasszociációs vizsgálatok a jellegek tapasztalt heritabilitásának csak csekély (legfeljebb 20-50%) részét tudják megmagyarázni. A GI-k jelenléte a hiányzó heritabilitás egyik magyarázatául szolgálhat (Zuk et al., 2012). Másrészről a már felderített GI-k felhasználhatók a géntermékek közötti funkcionális kapcsolatok előrejelzésére, illetve ismeretlen funkciójú gének esetében funkció prediktálására (Boone et al., 2007).

A GI evolúcióbiológiai vonatkozásai is sokrétűek, olyan folyamatok magyarázatában szerepel, mint a fajképződés (Dettman et al., 2007) vagy a szexuális szaporodás evolúciója (Visser and Elena, 2007). A kompenzáló negatív GI-k szerepet játszhatnak a genetikai változatosság fenntartásában (Hartman et al., 2001). A GI jelenléte hatással lehet arra, hogyan változik a rátermettség az allélkombinációk függvényében (milyen az adaptív tájkép alakja (Wright, 1932)), ezen keresztül pedig befolyásolhatja a lehetséges evolúciós utakat (Poelwijk et al., 2007; Carneiro and Hartl, 2009) és az adaptáció sebességét (Chou et al., 2011; Khan et al., 2011).

2.1.3 A genetikai interakciók rendszerszintű vizsgálata

A GI több szempontból is kapcsolódik a rendszerbiológiához (Moore and Williams, 2005). Egyrészt fogalmi átfedés miatt: a komponensek interakcióiból fakadó jelenségeket vizsgáló rendszerbiológia bizonyos mértékig a GI-k vizsgálataként is definiálható. Másrészt gyakorlati okból: a nagyléptékű GI adatok redszerbiológiai eszközökkel történő analízise új lehetőségeket nyit a GI-k szerveződésének és mechanisztikus okainak megértésében.

A rendszerszintű vizsgálatokhoz szükséges adatmennyiség csak nemrégen vált elérhetővé. Egyrészt megjelentek azok az adatbázisok, melyek összegzik az irodalomban fellelhető GI adatokat (pl. BIOGRID (Stark et al., 2006)). Másrészről a GI-k szisztematikus kísérletes feltérképezése is lehetségessé vált, ami egy kiválasztott génhalmazon belül az egyszeres és a

lehetséges összes kettős mutáns kombináció összehasonlítását jelenti azonos genetikai háttéren (Dixon et al., 2009). Míg a 90-es években ez egy-egy útvonal génjeinek vizsgálatát jelentette (Hartman et al., 2001), a kétezres évektől százas, majd ezres nagyságrendű génnel képzett génpár vizsgálatát. Ezt egyrészt a genomskálájú génkiütéses könyvtárak létrehozása tette lehetővé (Winzeler et al., 1999; Giaever et al., 2002), másrészt a növekedési ráta, mint könnyen, automatizáltan tesztelhető fenotípus használata, mely egyben a rátermettség közvetlen becslésére is alkalmas (Dixon et al., 2009). A vizsgálatok zöme nem esszenciális gének között fellépő GI-k feltérképezését jelenti, ugyanis itt a gének teljes kiütésére is mód van (Tong et al. 2001; Pan et al. 2004). A vizsgálatok kondicionálisan letális vagy hypomorph allélek előállításával esszenciális génekre is kiterjeszthetők (Tong et al. 2004; Davierwala et al. 2005; Michael Costanzo et al. 2010a). Az interakciós hálózatra vonatkozó legtöbb tudásunkat a a sörélesztővel végzett kutatásoknak köszönhetjük (Boone et al., 2007).[4] A kettős nullmutánsok közötti lehetséges kombinációk körülbelül egyharmadára (1700 × 3900 génre) érhető el kvantitatív GI adat (Koh et al., 2009). A összes sejtfunkciót reprezentálni szándékozó globális vizsgálatokat (Tong et al. 2004; Costanzo et al. 2010) a sejt egyes alrendszereire fókuszáló GI térképek vizsgálata egészítik ki, mint pl. a kromoszómával összefüggő funkciójú (Collins et al., 2007) vagy az endoplazmatikus retikulum működésében és a szekréció folyamatában részt vevő gének (Schuldiner et al., 2005) térképei. Habár az anyagcserehálózatok biokémiai jellemzése előrehaladott, az anyagcsere gének közötti GI-k szisztematikus feltérképezése eddig elhanyagolt területnek számított, így annak statisztikai – bioinformatikai elemzése sem volt lehetséges. Jelen dolgozatban egy új adatsorra támaszkodva (lásd *Módszerek 2.2.1*) elsőként vállalkozunk az anyagcserehálózat genetikai interakcióinak általános jellemzésére.

2.1.4 Az SGA módszer

A GI-k feltérképezésében forradalmi lépésnek számított az ún. SGA módszer bevezetése. A sörélesztő eddigi legnagyobb skálájú GI térképének elkészítése és a vizsgálataim alapjául szolgáló adatok előállítása is az SGA (syntethic genetic array) módszer továbbfejlesztett változatával történt, mely mára automatizált, kvantitatív GI becslésre képes (Tong et al., 2001; Collins et al., 2006; Tong and Boone, 2006; Baryshnikova et al., 2010). A módszer két

[4] Kisebb vagy nagyobb léptékű szisztematikus gén inaktivációs vizsgálatokat végeztek *Schizosaccharomyces pombe*-ben (Dixon et al., 2008; Roguev et al., 2008) *C.elegans*-ban (Lehner et al., 2006, 2006), *Drosophila melanogaster*-ben (Horn et al., 2011) és emlős sejttenyészeten (Yang and Stockwell, 2008) is.

különböző, egyszeres génkiütött törzseket tartalmazó haploid könyvtárból indul ki, melyek eltérő domináns szelekciós markereket tartalmaznak. A haploid sejtek párosodását, majd a meiózist és a marker szelekciókat mindig új szilárd agar felszínekre történő átoltás előzi meg, a protokol végeredménye pedig a kívánt kettős mutáns (szintén haploid állapotban). A vizsgálatok nagy számát a telepek párhuzamosan és automatizáltan, nagy sűrűségű lemezek használatával történő átoltása teszi lehetővé (a teljes nem esszenciális génkollekció elfér 14 darab 386 lyukú lemezen). A kolóniaméretekből történik a rátermettség becslése. Az SGA egy számos statisztikai módszert felhasználó protokoll alkalmazásával a kis skálájú módszerekhez hasonló mértékű torzítatlanság és precízitás elérésére képes (Baryshnikova et al., 2010). Ennek kulcsfontosságú része a szisztematikus hibákra való normalizálás (pl. lemez-specifikus hatás, telep pozíció hatások, tápanyag kompetíciós hatás, stb. kiküszöbölése). A replikált kétszeres és egyszeres génkiütések szórása alapján pedig GI értékhez p-érték is rendelhető (Costanzo et al., 2010). A becslések pontosságát és általánosíthatóságát mutatja, hogy az így nyert GI értékek nagyon jó korrelációt (r =0,89) mutatnak folyadékkultúrában, kompetitív rátermettség-adatokból becsült GI adatokkal (Onge et al., 2007; Costanzo et al., 2010).

2.1.5 A genetikai interakciós hálózatok tulajdonságai

Milyen információkat nyerhetünk a szisztematikus GI vizsgálatokból? Az így kapott nagyszámú GI hálózatba szervezhető, ahol az egyes gének a csúcsok és a gének közötti GI-k az élek (Tong et al. 2001). Egy-egy funkcionális alrendszer génjeinek vizsgálata lehetővé teszi egy-egy alrendszeren belüli kapcsolatok feltérképezését, reprezentatív genom szintű vizsgálatok esetében pedig a szervezet GI-hálózata nagyléptékű szerveződésének vizsgálatát.

Vajon milyen gyakori a két inaktivált gén között GI? A GI ritka, az eddigi legnagyobb skálájú, sörélesztőben végzett vizsgálatban az összes génpár 3%-a mutat GI-t, ezek kétharmada negatív, egyharmada pozitív interakció (Costanzo et al. 2010). A génenkénti interakciók számának hatványfüggveny eloszlást mutat: a legtöbb génnek kevés GI-ja van, míg néhánynak sok (Tong et al. 2004). Adott gén (mind pozitív mind negatív) GI-inak száma erős pozitív korrelációt mutat azzal, hogy egymagában kiütve mennyire csökkenti a rátermettséget (összesítve r=0,73 (Costanzo et al., 2010)). Ennek lehetséges oka, hogy a rátermettségre jelentős hatással levő gének pleiotrópiájának mértéke is nagyobb, azaz több különböző funkciót és génműködést befolyásolnak ezek a mutációk, s így több más gén befolyásolhatja ezen mutációk hatását (Costanzo et al., 2010; Szappanos et al., 2011).

2.1.6 Genetikai interakció és a funkcionális modulok

Mit tudhatunk meg a GI-k alapján a sejt funkcionális szerveződéséről? A GI-t más funkcionális kapcsolatra utaló jellemzőkkel is összehasonlíthatjuk, vizsgálva a funkcionális modulokkal való átfedését. Ha a sejtműködést 17 nagy funkcionális csoportra osztjuk, a GI-k kb. ötödét találjuk a csoportokon belül (Costanzo et al., 2010). A csoportok közti interakciók számában kiemelkedik a kromatin működés és a vezikuláris transzport, ezek a legnagyobb konnektivitású géneken keresztül a többi funkció között is hídként funkcionálnak (Costanzo et al. 2010; Lehner et al. 2006). Nagyobb felbontásban vizsgálva a sejt működési egységeit a fehérjekomplexek mintegy 60%-a mutat GI-ban való szignifikáns feldúsulást, ami lehet pozitív vagy negatív GI is (Baryshnikova et al., 2010). A fizikai kapcsolatban levő fehérjék génjeinek 10-20%-a van GI-ban is, negatív és pozitív GI-k hasonló arányban (Costanzo et al., 2010). A paralóg gének között várható funkcionális átfedéssel egybevágó módon a duplikált párok harmada mutat negatív GI-t, ami erős feldúsulást jelent (Musso et al., 2008; VanderSluis et al., 2010).

A közvetlen GI kapcsolaton túl, két gén funkcionális viszonyára több információt nyerhetünk, ha azt vizsgáljuk, milyen más génekkel vannak GI-ban. Az azonos sejtfunkcióban, útvonalban, vagy fehérjekomplexben szereplő gének pozitív és negatív GI partnerei átfednek, vagyis hasonló a GI profiljuk (Tong et al., 2004). A GI profilok hierarchikus osztályozását több esetben sikerrel használták fel biokémiai útvonalak, fehérjekomplexek előrejelzésére és ismeretlen funkciójú génekhez funkció rendeléséhez (Tong et al. 2004; Michael Costanzo et al. 2010; Collins et al. 2007; Ye et al. 2005; Decourty et al. 2008). Egyakorlati alkalmazások mellett a GI profilok vizsgálata egy új géncsoportosítási (modularizációs) elvhez is vezetett: monokromatikus csoportosításnak nevezzük, ha egy adott géncsoport vagy két csoport génjei kizárólag negatív vagy pozitív GI-val kapcsolódnak egymáshoz (Segre et al., 2005). Ez esetben a GI nem csak egy-egy gén, hanem egy-egy nagyobb géncsoport, modul tulajdonságának tekinthető. Így pl. a *2b ábra* két hipotetikus lineáris útvonala negatív GI-ban van egymással, bármelyik génpárt is választjuk. Az a megfigyelés, hogy találhatunk olyan modulokat, amelyek között az egyik típusú interakció feldúsul (Bandyopadhyay et al., 2008), azt jelzi, hogy a monokromitás létező jelenség. Kérdés, hogy vajon mennyire általános szervezőelv a GI hálózatokban.

Bár az eddigi szisztematikus vizsgálatok számos információt szolgáltattak a GI hálózatok szerveződésével kapcsolatban, több fontos kérdés megválaszolatlan maradt. Továbbra is keveset tudunk arról, milyen molekuláris mechanizmusok eredményezik a GI-kat, illetve milyen mechanizmusok állnak a szisztematikus vizsgálatokban feltárt mintázatok mögött. Bár

mint láttuk, egy adott sejtfunkción, fehérjekomplexen vagy fehérjeinterakciókon belül gyakrabban fordul elő GI a véletlenszerűen vártnál, például a GI-knak csak kevesebb, mint 1-2%-át tudjuk megmagyarázni paralógiával vagy közvetlen fehérje-fehérje kölcsönhatással (Tong et al., 2004; Costanzo et al., 2010). Az anyagcseregének közötti GI-k vizsgálata ezekben a kérdésekben járulhat hozzá tudásunk növeléséhez. Egyrészt az egyik legjobban feltérképezett sejtes alrendszer, másrészt genomskálájú anyagcseremodellek segíhetik a funkcionális kapcsolatok feltérképezését és a GI-k előrejelzését. Harmadrészt segítségével mechanisztikus hipotéziseket tudunk alkotni a GI-k megmagyarázására. A fejezet elején említettem két ilyen lehetséges magyarázatot a pozitív és a negatív GI-kra (*2. ábra*). De vajon leírható-e a GI-k többsége ilyen egyszerű modellek segítségével?

2.1.7 Az anyagcsere genetikai interakcióinak modellezése és predikciója

Az anyagcsere és a GI-k kapcsolatának elméleti vizsgálatára már korábban is volt példa. Az enzimaktivitások és egy lineáris anyagcsereútvonal végtermék termelésének rátája (fluxusa) közötti kapcsolat a metabolikus kontroll elmélet segítségével becsülhető meg (Kacser and Burns, 1973). Az ennek alapján született első általános matematikai modell (Szathmáry, 1993) az útvonalon belüli kis hatású mutációk GI-it vizsgálta, különböző paraméterek optimalizálására ható szelekció esetében. Azonban ezen elméleti konklúziók empirikus tesztelésére adatok hiányában még nem kerülhetett sor.

A genomskálájú kényszer-alapú anyagcseremodellek megjelenése új lehetőséget nyitott a GI-k immár realisztikus modellezésében (Deutscher et al. 2006; He et al. 2010; Harrison et al. 2007; Segre et al. 2005). Az FBA modellek a biomassza termelésének rátáját maximalizálják. Egy gént eltávolítva a modellből, a biomassza-termelés relatív változása alapján becsülhetjük a nullmutáns relatív rátermettségét. Az így becsült rátermettség alapján a modell a gének esszenciális voltát 83-90%-os pontossággal képes becsülni (Förster et al., 2003; Duarte et al., 2004; Kuepfer et al., 2005). A szisztematikus kísérletekkel analóg módon a modellben az egyes és kettős génkiütések biomassza-termelésre gyakorolt hatásából becsülhető a GI. Az első elméleti munka, amely FBA modell segítségével prediktált GI-kat és vizsgálta azok szerveződését, számos megállapítást tett (Segre et al. 2005): A GI párok adott funkcionális csoporton belül fel vannak dúsulva, bár a párok 80%-a ezen csoportok között található (Segre et al., 2005). Ugyanakkor a csoportok közötti interakciók nagymértékű monokromitást mutatnak, a kevés kivétel pedig a többfunkciós géneknek tulajdonítható (Segre et al., 2005). A monokrom kapcsolatok sokszor intuitiv módon jól értelmezhetők. Például az ATP-képződés fermentáció és légzés során is végbemehet, a glikolízis és a légzési lánc egymást kompenzálni

képes redundáns funkciónak megfelelően közöttük negatív GI-t látunk. Ugyanakkor a légzés során történő ATP szintézishez mind az ATP szintetáz, mind a légzési lánc is szükséges, így ezek között pozitív GI-t találunk. De vajon mennyiben lesz igaz a monokrom szerveződés a kísérletek alapján kapott GI hálózatra?

Mint láttuk, a leginkább feltérképezett sörélesztő esetében is csak a génpárok egyharmadára van GI adatunk, az összes többi élőlény esetében pedig jóval kevesebbre, ugyanakkor a GI-k gyakorisága igen alacsony. Ezért a GI-k megbízható előrejelzésének gyakorlati haszna nagy, így egyre növekvő számú módszer szolgál a negatív vagy mindkét irányú GI-k előrejelzése. Ezek legtöbbször a GI-k moduláris szerveződését kihasználva a már ismert GI adatokat használják fel. A módszerek egy része kizárólag GI adatokon alapul (Qi et al., 2008; Ryan et al., 2010), vagy azokat különböző funkcionális adatsorokkal ötvözi (Wong et al., 2004; Chipman and Singh, 2009; Ulitsky et al., 2009). Azonban az utóbbi esetben is a különböző adatok közül a legerősebben prediktáló tulajdonságok maguk a GI-hálózatból kinyert adatok (Wong et al. 2004; Ulitsky et al. 2009). Azonban koncepcionálisan, és mivel a legtöbb fajra nincs GI adatunk ezért gyakorlati szempontból is kulcsfontosságú kérdés, hogy mennyire vagyunk képesek megjósolni két gén között fellépő GI-t a génpár egyéb tulajdonságai alapján? Azaz, mennyire vagyunk képesek a GI-k statisztikus predikciójára kizárólag az anyagcserehálózat jellemzői és a GI-tól különböző genomikai adatsorok alkalmazásával? És vajon az anyagcserehálózat egyszerű biokémiai modellje, az FBA mennyire tesz pontos predikciókat? Bár kisebb empirikus adatsorral összehasonlítva kiderült, hogy az FBA modell által prediktált GI-k a véletlennél jóval gyakrabban képesek előrejelezni a valódi GI-kat (Harrison et al., 2007), mindeddig nem volt elég adat a modell és a valós GI adatok szisztematikus összevetésére.

2.1.8 Célkitűzések

Kutatásom kérdései, hogy az anyagcsere GI-i mennyire kapcsolódnak a hálózat hagyományosan ill. matematikai alapon definiált funkcionális moduljaihoz és mennyire jellemző a GI-k monokróm szerveződése. Ezután azt a kérdést vizsgálom, hogy az FBA mint egyszerű biokémiai modell, illetve anyagcserehálózati és funkcionális genomikai adatok felhasználásával épített statisztikai / adatbányászati modellek mennyire képesek előrejelezni az anyagcseregének GI-it, illetve a különböző módszerek predikciói hogyan viszonyulnak egymáshoz. Eredményeim egy nagyobb projekt részét képezve kerültek közlésre (Szappanos et al., 2011).

Módszerek

2.1.9 A GI adatsor összeállítása

A kísérletes adatokat Charles Boone laboratóriuma (Toronto, Kanada) bocsátotta rendelkezésünkre egy kollaboráció keretében. Kollaborátoraink az SGA protokolt használva (Baryshnikova et al., 2010) a sörélesztő anyagcsere génpárjain szisztematikus kvantitatív GI vizsgálatot végeztek. A gének a sörélesztő 904 gént és 1412 reakciót tartalmazó anyagcserehálózat rekonstrukciója alapján lettek kiválasztva (Mo et al., 2009). A végleges adatsor az új adatok és egy ugyanazon módszerrel végzett korábbi nagyléptékű adatsor (Costanzo et al., 2010) egyesítéséből származik és 652 nem esszenciális anyagcsere enzimet kódoló gént, illetve 176 821 génpárt tartalmaz, amelyből 2668 mutat negatív, 1415 pedig pozitív GI-t.

Azt, hogy két gén GI-ban van-e egymással, egy korábban optimalizált (Costanzo et al., 2010)[5] küszöbérték alapján definiálták: $|\varepsilon| > 0.08$ és $p < 0.05$.[6] Emellett egy megbízhatóbb GI adatsort is meghatároztunk, ami a replikált kísérleti eredmények ismételhetősége alapján csökkenti a hamisan jelzett GI-k számát. Ez olyan génpárokat tartalmaz, amiket legalább két független méréssorozatban vizsgáltak. Ilyenkor legalább az egyik mérés esetén igaz, hogy $|\varepsilon| > 0.08$ és $p < 0.05$, valamint a többi mérésnél is azonos irányú az ε, $p < 0.05$ mellett. Azok a párok, amelyeknél egy mérésre sem igaz, hogy $|\varepsilon| > 0.08$ és $p < 0.05$, nincsenek interakcióban. Az egyik csoporba sem tartozó gének ki lettek hagyva az adatsorból, ami így 122 875 génpárból áll 529 negatív és 194 pozitív GI-val.

2.1.10 A funkcionális hasonlóság és a GI-k kapcsolatának vizsgálata

Kétféle funkcionális modul viszonyát vizsgáljuk a GI-hoz: hagyományos funkcionális csoportok és a hálózat teljes szerkezetét figyelembe vévő ún. fluxus kapcsolt csoportok. A hagyományos funkcionális csoportok esetében a metabolikus rekonstrukcióban közölt csoportosítást használtuk (Mo et al., 2009), amely követi az útvonalak hagyományos biokémiai felosztását (pl. glikolízis - glükoneogenezis, citromsav-ciklus, stb.). A fluxus kapcsoltság azt jelenti, hogy az egyik reakció aktivitása egyben a másik reakció aktivitásával

[5] Becslések szerint az így definiált negatív és pozitív GI-knak 63% illetve 59%-a valós, és lefedi a valós GI-k 35% illetve 18%-át (Costanzo et al. 2010).

[6] Az egyes GI-k statisztikai megbízhatóságát jelző p-értékeket a kísérletes replikátumok és az egyszeres mutánsok hibaeloszlásai alapján számolták (Costanzo et al. 2010; Baryshnikova et al. 2010).

is jár, ami lehet egyirányú (irányított kapcsoltság) vagy kölcsönös (teljes kapcsoltság). A vizsgálathoz Szappanos Balázs (MTA SZBK, Biokémia Intézet) által implementált algoritmus segítségével (Burgard et al. 2004) feltérképezett fluxus kapcsolt génpárokat használtam fel (1491 génpár).

A GI-k funkcionális modulokon és fluxus kapcsolt párokon belüli feldúsulás teszteléséhez randomizációs tesztet használtunk. A negatív vagy pozitív GI-ban levő párok közötti kapcsolatokat randomizálva, az egyes gének GI-nak számát megőriztük. A közös csoportba tartozó génpárok tapasztalt számához a randomizációval kapott eloszlás alapján várt valószínűséget rendeltünk. $p = (R+1)/(N+1)$, ahol R azon esetek száma, ahol a közös csoportba tartozó génpárok száma a random esetben nagyobb vagy egyenlő, mint amennyi az adatokban talált érték, N pedig a randomizáció száma, általában 10 000. Több csoportba is tartozó génpár esetén azt néztük, hogy legalább egy csoport átfed-e egymással. Paralógiára vagy fizikai interakcióra való kontrollálás esetén az ezeket a tulajdonságokat is mutató génpárokat nem vettük figyelembe az átfedő párok számolásánál.

A fehérje-fehérje fizikai interakciós adatokat a BioGrid 2.0.58 adatbázisából gyűjtöttük (Breitkreutz et al., 2008). A paralóg párokat minden élesztő génnek minden más élesztő génnel szembeni BLASTP (Altschul, 1997) keresésével azonosítottuk. Két gént paralógnak definiáltunk, ha i) az E-score $<10^{-8}$, ii) illesztett szekvencia 100 aminosavnál hosszabb, iii) szekvencia hasonlóság > 30% és iv) nem transzpozonok.

2.1.11 Monokromitás vizsgálat

A GI-k funkcionális csoportpárok közötti monokromatikusság vizsgálatához egy monokromatikussági értéket (MC) definiáltunk. Az indexet úgy alkottuk meg, hogy figyelembe vegye a pozitív és negatív GI-k nem egyenlő előfordulási gyakoriságát, és a pozitív vagy negatív irányba való eltérés mértéke ezen háttérértéknél legyen 0, a monokromitás mértéke pedig az ettól való relatív eltéréssel arányosan nőjön vagy csökkenjen 1-ig vagy -1-ig, a teljes monokromitás értékéig. A jelölések: pr_{ij} a pozitív/összes GI arány i és j csoportok között és bpr pozitív/összes GI arány az összes génre (háttér arány).

$$\text{if } pr_{ij} > bpr, \text{MC}_{ij} = (pr_{ij} - bpr)/(1 - bpr)$$

$$\text{if } pr_{ij} = bpr, \text{MC}_{ij} = 0$$

$$\text{if } pr_{ij} < bpr, \text{MC}_{ij} = (pr_{ij} - bpr)/bpr$$

Egy funkcionális csoportpár, amelynek kizárólag pozitív (vagy negatív) GI-ja van egymás között, az MC értéke +1 (vagy −1), míg ha a csoportpár a háttér interakciós arányt tükrözi, az MC értéke 0. Az MC értékek számolásához csak egyetlen funkcionális csoporthoz tartozó géneket veszünk figyelembe. Egy funkcionális csoportpárt akkor definiáltunk monokromatikusnak, ha $|MC_{ij}| > 0.5$.

A monokromitás statisztikai szignifikanciájának meghatározásához randomizációs tesztet végeztünk, a kísérletes adatokból nyert monochromitás értéket az így nyert nulleloszláshoz hasonlítva. A nulleloszlást a GI-k irányának 10 000-szeres randomizálásával kaptuk, a pozitív és negatív GI-k számát és az génpárokhoz tartozó funkcionális csoportosítást megőrizve. Vizsgálatunkat a legalább két vagy legalább három egymás közötti interakciót mutató párokon végeztük (*Függelék: 1.táblázat*).

2.1.12 Génpár tulajdonságok összeállítása a GI prediktálásához

A génpár jellemzőket korábbi munkákat (Wong et al. 2004; Ulitsky et al. 2009) követve állítottuk össze, ugyanakkor kihagyva minden GI-ra vonatkozó információt tartalmazó tulajdonságot. Az egyes deléciók rátermettségre tett hatását a két egyszeres deléciós mutáns rátermettségének átlagával és abszolút különbségével vettük figyelembe. A paralóg párokat a 2.2.2 fejezetben leírtak alapján definiáltuk. Ez alapján minden génpárt három kategória egyikébe soroltunk: nincs paralóg párja, a pár tagjai egymás kizárólagos paralógjai (paralóg géncsalád mérete 2), vagy több génnel állnak paralóg kapcsolatban (paralog géncsalád mérete >2). A fehérje-fehérje interakciós adatokat (PPI) a BioGrid adatbázisból nyertük (B.-J. Breitkreutz et al. 2008). A fehérjékhez géneket rendelve az igraph R csomaggal (Csardi, G., & Nepusz, T. 2006.) meghatároztuk a két gén közötti legrövidebb PPI távolságot. Ha két fehérje hasonló más fehérjékkel van PPI-ban, azzal a két fehérje közötti interakció is prediktálható, ennek becsléséhez négy szomszédsági megfelelést mérő indexet használtunk (mutual clustering coefficients, C_{vw}.) (Goldberg & Roth 2003). $N(x)$ az x él (két gén közötti kapcsolat) szomszédainak számát jelenti. Azonos szomszéd szám esetén, a koefficiensek mindegyike növekszik a szomszédságok közti átfedés mértékével. Két él v és w esetében a következő indexeket definiáltuk:

$C_{vw} = |N(v) \cap N(w)| \,/\, |N(v) \cup N(w)|.$ Jaccard index

$C_{vw} = |N(v) \cap N(w)| \,/\, \min(|N(v)|, |N(w)|).$ Meet/min

$C_{vw} = |N(v) \cap N(w)|^2 \,/\, (|N(v)| \cdot |N(w)|).$ Geometrikus

Hipergeometrikus

$$C_{vw} = -\log \sum_{i=|N(v)\cap N(w)|}^{\min(|N(v)|,|N(w)|)} \frac{\binom{|N(v)|}{i} \cdot \binom{\text{Total}-|N(v)|}{|N(w)|-i}}{\binom{\text{Total}}{|N(w)|}}$$

A „2hop” jellemzők egy génpár és egy harmadik gén közötti specifikus kacsolatot jelölnek (Wong et al. 2004). Például ha A fehérje fizikai interakcióban van C fehérjével és C-nek közös transzkripciós faktora van B-vel, akkor A-B génpár „2hop PPI – Regulator” kapcsolatban van. A kvantitatív fenotípus korreláció az egyes gének deléciós törzseinek 51 stresszkörnyezet esetén folyadékkultúrában mért relatív gyakoriságaiból (kompetitív rátermettség) képzett profiljainak korrelációja (Brown et al., 2006).

Specifikusan az anyagcserehálózatra jellemző tulajdonságokat (pl. metabolikus hálózati távolság) is definiáltunk. A metabolikus hálózatrekonstrukcióból (Mo et al., 2009) a reakciók közös metabolitjai alapján Szappanos Balázs meghatározta a reakciók szomszédsági viszonyait: két reakció szomszédos, ha van közös metabolitjuk (kizárva a kofaktorokat, pl. ATP). Ez alapján a géneket a reakciókhoz rendelve az igraph R csomaggal (Csardi, G., & Nepusz, T. 2006.) meghatároztuk a két gén közötti legrövidebb távolságot (a legkisebb számú metabolit, amelyen keresztül el lehet jutni egyik géntől a másikig). Az anyagcsereút szomszédos reakciói különböző lokális funkcionális kapcsolatban lehetnek egymással, így lehetnek egymás utáni reakciók („chain”), versenghetnek azonos szubsztrátért („forks”), előállíthatják ugyanazt a terméket („OR funnel”), vagy termelhetnek kooperáló szubsztrátokat („AND funnel”) (Chechik et al., 2008). A gének hálózatrekonstrukcióban definiált funkcionális csoportosítása alapján (Mo et al., 2009) meghatároztuk, hogy a génpárok azonos csoportba sorolhatók-e (több csoportba tartozás esetén legalább egy csoport átfed-e egymással).

1. táblázat

Génpár tulajdonságok csoportosítása	Tulajdon-ságok száma	Tulajdonság típusa	Forrás és referencia
Egyszeres deléciós rátermettség (a két egyszeres deléciós mutáns rátermettségének átlaga és abszolút különbsége)	2	számszerű	jelenlegi munka
Paralógia (nincs paralóg pár; paralóg géncsalád mérete 2; paralog géncsalád mérete >2)	1	kategórikus	jelenlegi munka
Legrövidebb útvonal a metabolikus hálózatban	1	számszerű	jelenlegi munka
Közös metabolikus funkcionális csoport	1	kategórikus	(Mo et al., 2009)
Előfoldulás egy adott metabolikus funkcionális csoportban*	30	kategórikus	(Mo et al. 2009)
A metabolikus hálózat helyi algráfjai (nem szomszédos, chain, fork, OR funnel, AND funnel)	1	kategórikus	jelenlegi munka, (Chechik et al., 2008)
Fluxus kapcsoltság (nem kapcsolt, irányítottan, teljesen kapcsolt)	1	kategórikus	jelenlegi munka, (Burgard et al., 2004)
Előfordulás fehérjekomplexek között	1	kategórikus	(Pu et al., 2009)
Előfordulás egy specifikus fehérjekomplexben*	5	kategórikus	(Pu et al. 2009)
Fizikai interakció (PPI,minden fizikai interakció a BioGrid-ből)	1	kategórikus	BioGrid (Breitkreutz et al., 2008)
PPI hálózati kapcsolatok száma (kapcsolatok számának átlaga és abszolút különbsége)	2	számszerű	BioGrid (B.-J. Breitkreutz et al. 2008)
Legrövidebb út a PPI hálózatban	1	számszerű	BioGrid (B.-J. Breitkreutz et al. 2008)
Mutual clustering koefficiensek a PPI hálózatban	4	számszerű	BioGrid (B.-J. Breitkreutz et al. 2008), (Goldberg & Roth 2003)
Közös transzkripciós faktor	1	kategórikus	(Balaji et al. 2006)
2hop PPI – PPI	1	kategórikus	BioGrid (B.-J. Breitkreutz et al. 2008)
2hop PPI – Regulátor	1	kategórikus	
2hop PPI – Paralógia	1	kategórikus	
2hop Regulator – Paralógia	1	kategórikus	
mRNA expresszió korreláció	1	számszerű	(Huttenhower et al. 2006)
Közös MIPS fenotípus	1	kategórikus	(Mewes 2004).
Egy specifikus MIPS fenotípus*	14	kategórikus	(Mewes 2004).
Kvantitatív fenotípus korreláció	1	számszerű	(Brown et al., 2006)
Közös sejten belüli lokalizáció	1	kategórikus	élesztő GO slim (Christie et al., 2004)
Előfordulás egy specifikus sejtkompartmentben*	13	kategórikus	élesztő GO slim (Christie et al. 2004)

*Metabolikus csoportok / komplexek / fenotípusok / kompartmentek, amelyek kevesebb mint négy gént tartalmaznak a vizsgált adatsorunkból, kimaradtak.

2.1.13 A genetikai interakciót prediktáló módszerek kiértékelése

Az 1. táblázat génpár jellemzőit felhasználva prediktáltuk a negatív illetve pozitív GI-kat logisztikus regresszió és random forest (Breiman, 2001) módszereket használva az R statisztikai környezetet alkalmazva (Liaw and Wiener, 2002; R Development Core Team, 2009). A szigorúbb GI definíciót alkalmazva (lásd *2.2.1*) adatsorunk 325 negatív és 116 pozitív GI-t tartalmazott 67 517 génpár között. A random forest egy új, döntési fák együttesén alapuló nem-parametrikus adatbányászati módszer (Breiman, 2001). Előnye, hogy gyors, hatékonyan prediktáló (túlillesztésre kevésbé hajlamos) statisztikai modelleket generál, és sok irreleváns prediktor változó esetén is megbízhatóan működik. Balanced random forest módszert használtunk a GI-ban levő és maradék csoport számában való nagy eltérés (imbalance) miatt, és 5000 döntési fát építettünk. Logisztikus regresszió esetén a predikció sikerét 5-szörös kereszt-validációval teszteltük, ami azt jelenti, hogy a predikcióhoz az összes génpár véletlenszerűen kiválasztott 80%-át használtuk fel és a tesztelés sikerét a maradék 20%-on vizsgáltuk, a folyamatot 10-szer megismételve. A random forest esetén a predikciós sikert hasonló módon, out-of-bag (a predikció során nem felhasznált) mintákon teszteltük. A precision-recall görbéket az ROCR (Sing et al., 2005) vizualizációs R csomaggal készítettük.

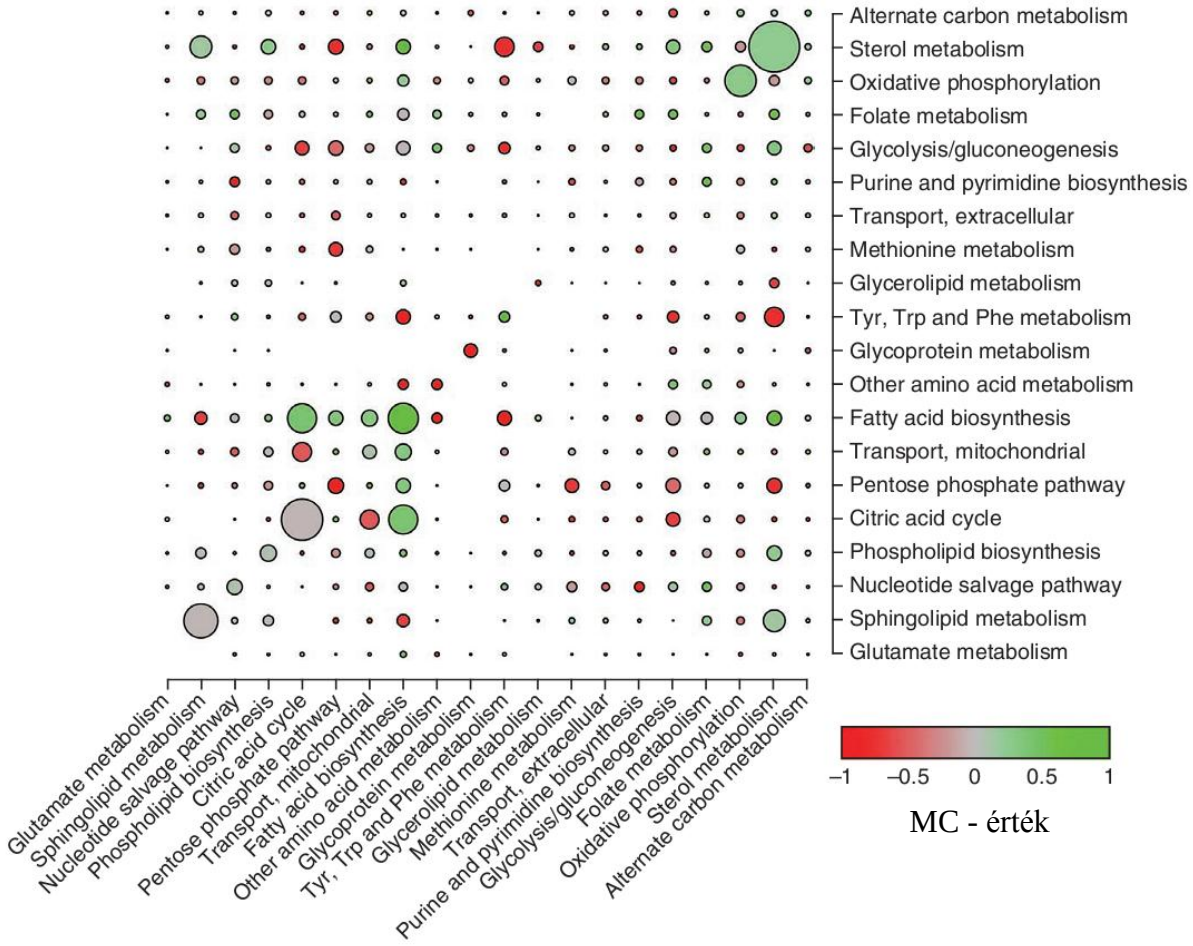
2.2 Eredmények

Az anyagcsere szisztematikus GI adatainak birtokában először egy korábbi genomléptékű anyagcseremodellezési munka a GI és a funkcionális modulok kapcsolatára vonatkozó két előrejelzését teszteltük (Segre et al., 2005). (1.) A GI-k feldúsulnak az anyagcserén belüli funkcionális csoportokon belül. (2.) Az egyes funkcionális csoportok között többségében kizárólag pozitív vagy kizárólag negatív GI-k fordulnak elő (monokromitás). Ezután az FBA modell, illetve anyagcserehálózati és funkcionális genomikai adatok felhasználásával épített statisztikai / adatbányászati modellek GI predikcióinak sikerességét vizsgáltuk.

2.2.1 Genetikai interakciók és funkcionális modularitás

A modell első predikciójával összhangban az anyagcserehálózat klasszikus módon definiált (Mo et al., 2009) funkcionális moduljain belül mind a negatív (1,4-szeres, $p < 10^{-4}$), mind a pozitív GI-k (2,2-szeres, $p < 10^{-4}$) szignifikáns, de a randomizációval kapott átlaghoz képest nem túl nagy mértékű feldúsulását találtuk. Például a lipidanyagcsere különösen fel van dúsulva GI-kban, ezen belül szterol anyagcsere és zsírsav bioszintézis elsősorban pozitív GI-t tartalmaz, míg a szfingolipid bioszintézis mindkét típust *(3. ábra)*. A szigorúbban definiált GI adatsort használva a feldúsulás erősebbé válik (negatív: 2,9-szeres, $p < 10^{-4}$, pozitív: 5,2-szeres, $p < 10^{-4}$).

Vajon megmagyarázható-e ez a mintázat azzal, hogy a funkcionális csoportokon belül olyan, génpárok lehetnek feldúsulva, melyek eleve nagyobb valószínűséggel vannak GI-ban, mint a fizikai interakcióban levő génpárok vagy a génduplikációk? Ha kontrollálunk ezekre a változókra, a feldúsulás mindkét esetben szignifikáns marad, vagyis a funkcionális csoporthoz tartozás plusz információt hordoz (negatív: 1,23-szoros, $p = 0{,}004$; pozitív 2-szeres, $p < 10^{-4}$). Ugyanakkor, ahogy a *3. ábrán* is látszik, a GI-k többsége különböző funkciók között fordul elő (negatív 93%, pozitív 90%; szigorú GI adatsorral 86% és 73%). A GI-ban leginkább feldúsult funkcionális csoportok, mint például a zsírsav bioszintézis génjeinek is számos GI-ja van más csoportok génjeivel is, ami az anyagcserén belüli funkciók közti nagymértékű pleiotrópiára utal.



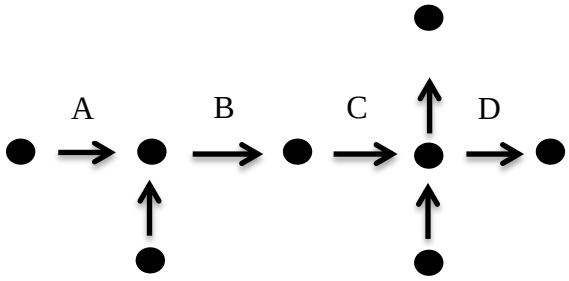
3. ábra: Genetikai interakciók megoszlása és monokromitása funkcionális csoportok között.

A körök sugara az adott funkcionális csoporton belül, vagy csoportok között a kísérletesen tesztelt génpároknak azt a hányadát jelöli, amelyek GI-t mutatnak (például a szterol metabolizmuson belül a legnagyobb a GI –ban levő génpárok aránya 0.225). Az átlóban szemmel látható a GI-k funkcionális csoportokon belüli feldúsulása. A körök színe a monokromatikusság MC-értékét tükrözi, ami a pozitív/összes GI pár normalizált aránya (Módszerek 2.2.3). Azoknak a funkcionális csoportoknak amelyeknek kizárólag pozitív GI-i vannak egymással az MC-értéke +1 (zöld), kizárólag negatív GI-k esetén pedig az MC-érték -1 (piros). A pozitív/összes GI teljes adatsorban tapasztalt háttér aránya (0,348) felel meg 0 MC-értéknek (szürke). Az ábrán csak a 20 legtöbb génpárral tesztelt funkcionális csoport látható és az egyetlen funkcionális csoporthoz sorolható géneken való vizsgálatokon alapul.

Vajon a különböző funkcionális csoportok közti interakciók monokromatikusak-e? Aszámítógépes modell (Segre et al., 2005) első predikciójával összhangban a funkcionális csoportpárok között a monokromitás mértéke szignifikánsan nagyobb, mint azt a randomizációval kapott GI eloszlás alapján várnánk, a monokromitást |MC| > 0.5-nél definiálva (lásd *Módszer 2.2.2.*). Például míg a szterol bioszintézis szinte kizárólag negatív GI-t mutat a tirozin, triptofán és fenilalanin anyagcserével, döntően pozitív interakcióban van a zsírsav bioszintézissel (*3. ábra*). A modell által prediktált GI adatok alapján az egymással legalább 2 illetve legalább 3 GI-t mutató csoportpárok 65% illetve 56% monokróm, ami a véletlen várthoz képest 2,6 illetve 3,6-szoros feldúsulást jelent (*Függelék 1. táblázat*). Az empirikus adatok feldúsulása ennél jóval kisebb mértékű, 24 illetve 34%-kal több monokromatikus génpárt találunk a randomizációval kapott nulleloszlás alapján vártnál, és a szigorú GI adatsoron is hasonló eredményt kapunk (12%,13% többlet). Az alacsony monokromitás oka nem a több funkcióban is részt vevő gének jelenléte, mivel azokat eleve kizártuk a vizsgálatból. Mindez azt sugallja, hogy a valóságban tapasztalt GI-k kisebb hányadára jellemző a monokromatikus szerveződés mint azt az anyagcserehálózat szerkezete alapján várnánk. [7]

Vajon eltérő mintázatokat látunk-e, ha egy szubjektivitástól mentes moduldefiníciót használunk? A funkcionális kapcsolatoknak egy, a klasszikus anyagcsereutakkal alternatív leírási módja a reakciók összefüggő használatán alapuló ún. fluxus kapcsoltság (Papin et al., 2004; Price et al., 2004). A kapcsolt fluxus azt jelenti, hogy az egyik reakció aktivitása minden alkalommal a másik reakció aktivitásával is jár, ami lehet egyirányú (irányított kapcsoltság) vagy kölcsönös (teljes kapcsoltság). A fluxus kapcsoltság összefügg a reakcióutak topológiájával: a teljes kapcsoltság elágazásmentes útvonalszakasznál, a részleges kapcsoltság pl. több bemenő és egy kimenő élt tartalmazó elágazás esetén jöhet létre (*4. ábra*). A modell enzimszabályzást nem tartalmaz, így szabályzási kapcsolatokat nem foglal magába. A fluxuskapcsoltság a funkcionális kapcsolat biokémiailag pontosan értelmezett definícióját adja (Burgard et al., 2004), korábbi munkák pedig rávilágítottak fiziológiai és evolúciós relevanciájukra is jelentős (Pal et al., 2005; Bundy et al., 2007; Notebaart et al., 2008).

[7] A legszigorúbb monokromitás definícióval (|MC|=1) a feldúsulás mértéke alacsonyabb:15% illetve 23%, de így is szignifikáns (p = 0.016, p = 0.026).



4. ábra: Hipotetikus anyagcserehálózat, ahol a metabolitok a csúcsok, reakciók az élek.

Teljes kapcsoltság: ha B reakció aktív, akkor C is és fordítva. Irányított kapcsoltság:ha A reakció aktív, akkor B is, de fordítva nem.

A vizsgálathoz a Szappanos Balázs által implementált algoritmus segítségével (Burgard et al., 2004) feltérképezett fluxus kapcsolt génpárokat használtam fel (1491 génpár). Intuitív módon fluxus kapcsoltság esetén pozitív GI-t várnánk, hiszen ilyenkor definíció szerint az egyik gén kiütésével lenullázott fluxus a másik génhez tartozó reakcióaktivitás megszűnésével jár, ha az nincs is kiütve. Ezzel szemben a GI-k feldúsulását vizsgálva a hagyományos annotációs csoportokhoz hasonló eredményt kaptunk. Mind a negatív (1,4-szeres, $p=0,02$), mind a pozitív (2,3-szoros, $p<10^{-4}$) GI-k fel vannak dúsulva fluxus kapcsolt párokban. Az interakciók döntő többsége (>97%) pedig a nem kapcsolt génpárok között fordul elő, ahogyan szigorú GI adatsorban is (>93%). A kismértékű átfedés visszafele is igaz, a fluxus kapcsolt párok közül mindössze 2% mutat pozitív GI-t és 3% negatív GI-t. Vajon monokromatikus-e az interakciók eloszlása fluxus kapcsolt csoportok között? A teljesen kapcsolt fluxusokból képzett modulok közötti kapcsolatok többsége vegyesen negatív és pozitív, a monokromitás feldúsulás mértéke a klasszikus funkcionális csoportosításhoz hasonlóan csekély, ez esetben ez csak marginális vagy nem szignifikáns különbséget jelent (*Függelék 2. táblázat*).

Összefoglalva, a funkcionális modul definíciójától függetlenül a legtöbb GI különböző funkciókat köt össze. Ugyanakkor mind a negatív, mind a pozitív GI-k fel vannak dúsulva, adott funkcionális csoporton és fluxus kapcsolt párokon belül is. Bár a monokromitás valóban feldúsul a vizsgált funkcionális csoportok között, a legtöbb csoport közötti kapcsolat az elméleti predikcióval ellentétben nem monokróm. Bár a pozitív GI és a fluxus kapcsoltság

közötti összefüggés statisztikailag szignifikáns, az *1a* ábrához hasonló, a reakciók együttes használatából fakadó intuitív magyarázat a pozitív GI-knak csak néhány százalékára alkalmazható.

2.2.2 A legtöbb genetikai interakció nem jelezhető előre

Miután vizsgálataink szerint a GI-k mechanizmusának megértéséhez az anyagcserehálózat modularitásának, vagy funkcionális függőségeinek ismerete csak korlátozott mértékben járul hozzá, kíváncsiak voltunk, hogy a magyarázattól eltekintve a jelenleg az anyagcseregénekről elérhető információk, az FBA modellel együtt, mennyiben hasznosíthatók a GI-k prediktálására.

Számos statisztikai módszer született GI-k jóslására, részben különböző funkcionális genomikai adatok alapján (pl. a géntermékek közötti fizikai interakció, gének koexpressziója stb.), ugyanakkor ezek a meglevő GI-adatokat is felhasználják a predikcióhoz (Wong et al., 2004; Paladugu et al., 2008; Ulitsky et al., 2009). Mi azonban arra vagyunk kíváncsiak, hogy a GI adatokat kizárva, a génpárokra vonatkozó egyéb, funkcionális genomikai, és az anyagcserehálózatra vonatkozó ismereteink birtokában milyen mértékben vagyunk képesek a GI-k előrejelzésére. Második kérdésünk, hogy az anyagcsere FBA modellje milyen mértékben képes a GI prediktálására. Szintén vizsgáltuk, hogy a különböző módszerek predikcióinak sikeressége hogyan viszonyul egymáshoz, illetve mennyiben rejtenek ezek a módszerek kiaknázható komplementer információt.

Az FBA modell előrejelzéseit a modell által becsült egyes és kettős génkiütések növekedésre tett hatását számolva kapjuk (az adatsor generálását a projekt keretén belül Szappanos Balázs végezte). A szigorú GI definíció alapján és a modellben gyengén karakterizált hálózati részeket (blokkolt reakciók (Papin et al., 2004)) kizárva 67 517 génpárból 325 negatív és 116 pozitív empirikus GI-t próbáltunk prediktálni. A statisztikus modellezéshez egyrészt genomszintű génpár jellemzőket használtunk korábbi munkákat követve (Wong et al., 2004; Ulitsky et al., 2009), másrészt anyagcsere hálózati jellemzőket, (1ásd *1. táblázat*). A genomi génpár jellemzők tipukus példái a koexpresszió (Huttenhower et al., 2006), közös transzkripciós faktor (Balaji et al., 2006), vagy azonos fenotípusos kategóriába tartozás a MIPS adatbázisban (Mewes, 2004). Az anyagcserehálózati génpár-jellemzők között olyanokat

definiáltunk, mint a két gén azonos funkcionális csoportba tartozása (Mo et al. 2009), vagy a kódolt reakciók legrövidebb távolsága a metabolikus hálózatban (Mo et al. 2009 alapján).[8] A jellemzők összeállításkor elsődleges célunk az colt, hogy minél több lehetséges tulajdonságot vonjunk be a vizsgálatba, így nem mindegyik prediktív, vagy áll a bevétele mögött konkrét biológiai hipotézis.

Ezután a fenti jellemzők alapján klasszikus statisztikai (logisztikus regresszió) és egy újabb, döntési fák együttesén alapuló (ensemble) adatbányászati módszert (random forest (Breiman, 2001)) alkalmazva osztályoztuk GI adatainkat (a szigorú GI definíció alapján). A predikció sikeressége úgynevezett precision-recall ábrán értékelhető, ami a pontosságot (precision: a prediktált GI-k empirikusan alátámasztott aránya) a lefedettség (recall: true positive rate, az empirikus GI-k előrejelzett aránya) függvényében ábrázolja *(5. ábra)*. Az egyes pontok egy-egy küszöbértékhez tartoznak, ami azt mondja meg, hogy a modellek kvantitatív predikciói alapján mikortól tekintünk valamit pozitív vagy negatív GI-nak.

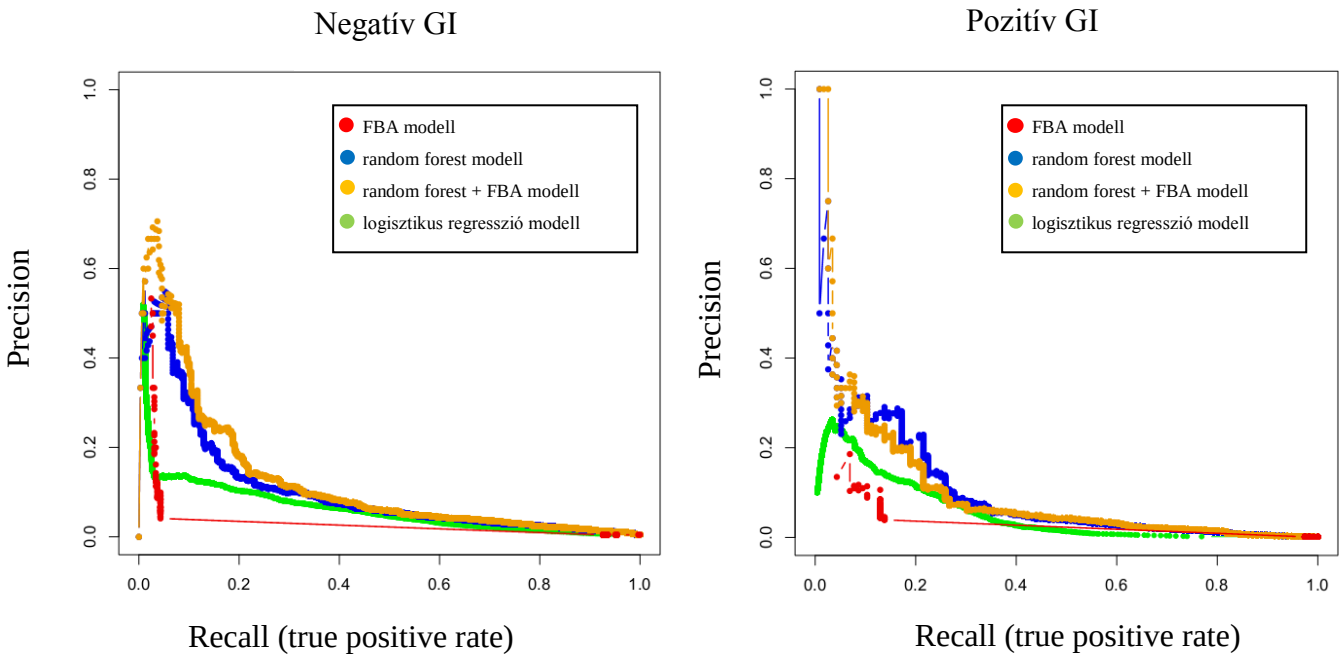
Az FBA modell esetében a prediktált erős interakciók (magas küszöbérték) esetén a kísértetesen mért GI-k nagymértékű feldúsulását tapasztaljuk: negatív GI értékeknél 100-szoros, pozitívaknál 60-szoros feldúsulás (a precision értéke 50%, illetve 11%). Bár ez az eredmény alátámasztja a legerősebbnek előrejelzett GI-k fiziológiás jelentőségét, ugyanakkor azonos küszöbértékeknél a modell az empirikus GI adatoknak csak igen alacsony arányát jelzi előre (recall: 2.8%, 12.9%).

A genomikai és anyagcserehálózati adatokon alapuló statisztikai modellezés, elsősorban a random forest módszer az FBA predikcióinál legtöbb esetben nagyobb precision-t és recall-t eredményezett. Bár ez a módszer az *in vivo* interakcióknak az FBA predikcióhoz képest nagyobb hányadát találja meg, ez 10% precision felett még mindig csak a kísérletesen negatív GI-k 30% -át, illetve a pozitív GI-k 25%-át jelenti.

A becslés jósága akkor sem változik jelentősen, ha az FBA modell által prediktált adatokat (rátermettség és GI értékek) hozzáadjuk a statisztikai modellhez, ugyanakkor negatív GI esetében növeli a maximális precision értékét. Ez arra utal, hogy az anyagcseremodellben található olyan komplementer információ, ami a funkcionális genomikai és hálózati tulajdonságokból közvetlenül nem nyerhető ki. Összefoglalva, a GI-k többségét sem a

[8] A negatív és pozitív GI is rövidebb hálózati távolságnál dúsul fel(logisztikus regresszió $p < 10^{-16}$ illetve $p = 10^{-6}$)

biokémiai modellel, sem a funkcionális genomikai adatsorok és anyagcserehálózati adatok adatbányászati integrálásával nem tudjuk megbízható pontossággal megjósolni.



5. ábra A negatív illetve pozitív GI-t prediktáló módszereink sikerességének kiértékelése

A precision-recall ábrán az x tengelyen azt látjuk, hogy a különböző módszerek által prediktált pozitív vagy negatív GI-k mekkora hányadát jelzik előre a kísérletesen talált GI-knak (lefedés). Az y tengely azt mutatja, hogy a prediktált negatív vagy pozitív GI-knak mekkora hányada GI a kísérletes adatok szerint is (pontosság). Modelleink kimeneti predikciói folytonos értékek, a kategorizába sorolás különböző küszöbértékeknél más lefedettséget és pontosságot eredményez, az ábra egy pontja egy küszöbértékhez tartozó lefedettséget és pontosságot mutat. Mint látható a két érték trade-offban van egymással, vagyis minél pontosabban jóslunk, annál több valódi GI marad ki a megjósolt párok közül. Például a random forest módszerrel ha a prediktált negatív GI-k kb. 50%-a a kísérlet szerint is az, a modell a kísérletes negatív GI-k csak 2,8%-át jelzi előre. A random forest és logisztikus regressziós modellhez tartozó görbékhez az 1. táblázat adatait használtuk fel. Az FBA modell a modell által prediktált GI-értékek, a random forest + FBA az 1. táblázat adatait és az FBA modell GI valamint az egyszeres és kétszeres génkiütés esetén becsült rátermettség értékeit tartalmazza.

2.3 Diszkusszió

Kutatásaink során az anyagcsere modulok és GI közti kapcsolatot vizsgáltuk, valamint azt, hogy mennyire tudjuk előrejelezni a GI-kat genomikai, anyagcserehálózati adatok és egy anyagcseremodell predikciói alapján. Az empirikus adatok részben megerősítették az anyagcseremodell által tett predikciókat: a GI-k feldúsulnak a funkcionális csoportokon belül és a csoportok közti GI-k a véletlenszerűen vártnál gyakrabban monokromatikusak. Azonban egyik hatás sem bizonyult erősnek: a legtöbb GI funkcionális csoportok között fordul elő és a legtöbb csoportpár közti kapcsolat nem monokromatikus. Az *1.a ábra* alapján pozitív GI-kat az egymástól függő aktivitású reakciókat kódoló gének között várunk, azonban a prediktált fluxus kapcsoltság alapján a GI-k csak nagyon kis részét tudtuk megmagyarázni.

Habár funkcionális magyarázatot nem tudunk a legtöbb GI-ra adni, a vele korreláló genomikai adatsorok, anyagcserehálózati információk és az FBA modell felhasználásával elvileg képesek lehetünk a GI-k előrejelzésére. Eredményeink szerint azonban legtöbb GI-t továbbra sem tudjuk hatékonyan prediktálni. Az adatok statisztikai elemzésén alapuló módszereket és az FBA modellt összehasonlítva az előbbiek jobban jelzik előre a GI-t, de az FBA modell is tartalmaz plusz, a predikcióhoz hasznosítható információt.

Összefoglalva, a GI-k többségét továbbra sem értjük jól, sem a biokémiai funkció, sem statisztikai asszociációk szintjén.

A gyenge prediktálhatóság és a funkcionális magyarázatok hiánya részben fakadhatnak technikai hiányosságokból. A funkcionális modulok vizsgálatánál ez jelentheti a hagyományos funkcionális csoportosítás tökéletlen voltát és a modell hibás fluxus kapcsoltság predikcióját (Marashi and Bockmayr, 2011). Az FBA modell alacsony GI találati arányának okai között lehet az egyes mutánsok rátermettségének hibás becslése, a szabályzásra vonatkozó információk hiánya, vagy a hálózat annotációjának hibái (Szappanos et al., 2011).[9] A GI-k jelentős része környezetfüggő (Harrison et al., 2007; Bandyopadhyay et al., 2010), így a kísérletes környezet nem tökéletes szimulálása is a predikciós hibák oka lehet.

Ugyanakkor lehetséges, hogy a GI-k nagy részét tényleg nem lehet az 1. ábrán bemutatotthoz hasonló egyszerű mechanisztikus modellel leírni, azok az eddig vizsgált közvetlen funkcionális vagy hálózati kapcsolatoknál összetettebb összefüggések eredményei. Egy

[9] Ugyanakkor a hibás GI predikciók biokémiai hátterének elemzése és korrigálása a modell továbbfejlesztésének eszköze lehet (Szappanos et al., 2011).

alternatív irány lehet az egyes GI-k mögött álló biológiai összefüggések vizsgálatára a rátermettséget befolyásoló fenotípusos jellegek (rátermettség komponensek, „köztes fenotípusok") közötti függvénykapcsolat vizsgálata (Chiu et al., 2012).

Egy példa a lehetséges függvénykapcsolatra a költség-nyereség viszony, amely gyakori modellje a rátermettség komponensek közti kapcsolatnak. (Tănase-Nicola and ten Wolde, 2008; Wessely et al., 2011). Például mikrobákban egy anyagcsereútvonal működésére ható mutációknak a növekedési rátával becsült rátermettségre tett hatása az útvonal aktivitásával járó biomassza növekedés nyeresége mínusz a fehérjeexpresszió költségeként modellezhető (Chou et al., 2011).[10] A modell paramétereit kísérletesen becsülve a fenti egyszerű költség-nyereség modell tesztelhető. A fenti modell *Methylobacterium extorquens* törzsbe mesterségesen bevitt új anyagcsereútvonalhoz tartozó gének esetében képes volt a laboratóriumi evolúció során fixálódott új mutációk rátermettségre tett hatásának és a közöttük levő negatív GI-k mértékének előrejelzésére (Chou et al., 2011).

A fenti példán túlmutatva, általánosságban is elemezhető, hogy amennyiben a rátermettséget két komponens függvényeként modellezzük, mely függvénykapcsolatok esetében várunk a mindkét komponensre ható mutációk közötti GI-t (Chiu et al., 2012). A köztes fenotípusok közti különböző kapcsolatok eltérő valószínűséggel eredményezhetnek bizonyos típusú GI-kat, így befolyásolhatják a GI-k típusának eloszlását, pl. a költség-nyereség függvény gyakrabban eredményez negatív GI-t, mind káros, mind előnyös mutációk között (Chiu et al., 2012). A rátermettség ilyen köztes fenotípusokra való visszavezetése, a modellek elméleti és kísérletes vizsgálata a jövőben alternatív megközelítést jelenthet az ismeretlen molekuláris hátterű GI-k megmagyarázására.

A GI-k funkcionális alapjainak megismerésében és predikciójában a különböző genomskálájú funkcionális adatsorok eddig is döntő szerepet játszottak. Újabb nagyléptékű vizsgálatok mindkét feladatban további előrelépést jelenthetnek. Ilyen például az egyes génkiütött törzsekben a genom többi génjének expresszióját mérő adatsor („deleteome"). Az egyes GI-ban levő génpárok esetében az esetleges expressziós változás segíthet a molekuláris háttér felvázolásában, például az egyik gén kiütésekor a másik gén expressziós szintjének növekedése a kompenzáló negatív GI eszköze lehet. Egy alternatív magyarázat lehet a

[10] A modellben vad típus esetén a rátermettség (W_0, növekedési rátával becsülve) két komponens költég-nyereség függvényeként írható fel: $W_0 = b_0 - c_0$, ahol b_0 az anyagcserútvonal fluxusa és c_0 a fehérjeexpresszió költsége.
Két az útvonalat érintő mutáció (i,j) esetén, ha a nyereségre illetve a költségre tett multiplikatív hatásuk λ illetve θ: $W_{ij} = \lambda_i\lambda_j b_0 - \theta_i\theta_j c_0$. A GI mértéke ebből: $\varepsilon = b_0\, c_0\, (\lambda_i - \theta_i)(\lambda_j - \theta_j)$

közvetlen funkcionális kapcsolat hiányára a GI-k esetében, hogy a GI-nak a kiütött gén hiánya csak indirekt oka, közvetlen oka a gén kiütése után fel- illetve leexpresszálódó génekkel való interakció (Tucker and Fields, 2003). Ebben az esetben a molekuláris szintű magyarázatot is a génkiütés után fel- illetve leexpresszálódó génekkel való funkcionális kapcsolatában kell keresni. Az elképzeléssel egybevág, hogy egy adott gén GI-inak száma és a kiütésekor fel- vagy leregulálódó gének száma pozitívan korrelál (Spearman r = 0,57, $p<10^{-16}$)[11], de a hipotézist a jövőben konkrétabb vizsgálatoknak kell tesztelniük.

[11] A génkiütés hatására fel- és leregulálódó génekre vonatkozó adatsort Frank Holstege (Utrecht-i Egyetem) bocsátotta rendelkezésünkre kollaboráció keretében.

3 Az anyagcsereutak felépítésének hatása az operonális génsorrendre *E. coli*-ban

3.1 Bevezetés

A következő fejezetben azt vizsgálom, vajon hatással lehet-e az anyagcsere szerkezete a genom szerkezetének evolúciójára. Konkrétabban: vajon befolyásolhatja-e a gének anyagcsereútbeli helyzete a gének operonbeli pozíciójának evolúcióját? Mivel a kérdés a bakteriális génsorrend általános kérdésfelvetéséhez kapcsolódik, ezért először áttekintem az ebben a témakörben végzett eddigi kutatások eredményeit.[12]

A genom szerveződésének egyik legkézenfekvőbb jellegzetessége az egyes gének elhelyezkedési sorrendje, szomszédsági viszonyaik. Vajon független-e egymástól az egyes gének kromoszómális pozíciója? Hasonlóan a genom más olyan fenotípusos jellemzőihez, mint a gének példányszáma, bázisösszetétel, stb., kvantitatív, statisztikai jellegű tulajdonságról beszélünk. Ezért annak eldöntéséhez, hogy a sorrend véletlen folyamatok eredménye vagy a természetes szelekció által létrehozott mintázat, statisztikai módszereket kell segítségül hívnunk (Hurst et al., 2004). A nullmodell szerint a gének sorrendjére nem hat szelekció, de a gének átrendeződést eredményező mutációinak (pl. tandem duplikáció, transzpozíció, stb. rátái fajok vagy kromoszómák között különbözhetnek) lehetnek mintázatképző hatásai, így ezekre kontrollálni kell (Hurst et al., 2004). Az operonon belüli génsorrend esetében nem tudunk olyan mutációs mechanizmusról, amely a génsorrend relatív irányát befolyásolná az anyagcserútbeli helyzethez képest, ezért nullmodellünk szerint az egyes génpárok egyforma eséllyel vannak az enzimek útvonalbeli sorrendjével azonos, vagy azzal ellentétes irányban.

[12] A bakteriális gének kromoszomális elhelyezkedése a szomszédsági viszonyoktól megkülönböztetendő, de azokkal esetleg összefüggő mintázatokat is mutat (Rocha 2008). Például a magasan expresszált a replikációs origóhoz közelebb (Sharp 1989), az esszenciális gének pedig nagyobb valószínűséggel a DNS vezető szálán helyezkednek el (Rocha 2003).

3.1.1 A bakteriális génsorrend evolúciója

Már az 1950-es években, az első genetikai térképezések idején feltűnt, hogy a bakteriális gének kiosztása nem véletlenszerű, gyakran találtak egymáshoz közel elhelyezkedő hasonló funkciójú géneket (például Demerec 1964). A hasonló funkciójú géncsoportok (cluster-ek)[13] rejtélyére látszólag az operonok felfedezése nyújtott magyarázatot (Jacob and Monod, 1961). Később nyilvánvalóvá vált, hogy a géncsoportok nem minden esetben feleltethetőek meg egy-egy operonnak, ugyanakkor az együtt átíródó és szabályozódó egységek eredete maga is számos kérdést vet fel (Lawrence, 2003). Az eukariótákban kisebb számban, de szintén találtak hasonló funkciójú génekből álló csoportokat (Lawrence and Roth, 1996), a genomszintű vizsgálatok pedig mára felfedték általános elterjedtségüket (Hurst et al., 2004; Osbourn and Field, 2009).

A 90-es években elkezdődő genomléptékű vizsgálatok összehasonlító elemzései alapján egy nagymértékben plasztikus bakteriális genom képe rajzolódott ki, amelyben a gének pozíciója nagyban változékony és a génsorrendbeli egyezések néhány kivételtől eltekintve két-három génhossznyira korlátozódnak (Mushegian and Koonin, 1996; Kolsto, 1997; Watanabe et al., 1997; Wolf et al., 2001; Rogozin et al., 2002). A leghosszabb és a legnagyobb filogenetikai távolságokon keresztül megőrződött régió a legtöbb fajban riboszomális fehérjéket kódol (Watanabe et al., 1997; Wolf et al., 2001). A génsorrend evolúciójának sebessége a komparatív adatok alapján jóval meghaladja a fehérjeszekvencia evolúciójáét. Más, genomszintű tulajdonságok evolúciójával összehasonlítva az átrendeződés sebessége nagyobb, mint az ortológ gének elvesztéséé, de kisebb, mint a szabályozó elemeket tartalmazó intergénikus régiók divergálódási sebessége (Huynen and Bork, 1998).

A gének konzerválódott szomszédsági viszonya funkcionális információt hordozhat. A legerősebben konzervált szomszédok gyakran egymással fizikai interakcióban levő fehérjéket kódolnak (Dandekar et al., 1998), a kisebb fokú konzerváltság pedig egyéb funkcionális kapcsolatot jelezhet a géntermékek között (Tamames et al., 1997; Overbeek et al., 1999).[14] Ha az egymás melletti gének nem is őrzik meg szomszédságukat, az ortológok egy másik fajban való együttes előfordulásának valószínűsége genombeli pozíciójuktól függetlenül nagyobb a

[13] Disszertációmban a „géncsoport" kifejezést az angol „cluster" értelmében használom, azaz géneknek a kromoszómán egymás közelében elhelyezkedő csoportja, nem követelve meg az esetleges közös szabályzottságot.

[14] Ezek alapján a gének több fajban is megőrződött szomszédsági viszonyaiból következtethetünk még nem annotált fehérjék lehetséges funkcióira (pl. Wolf et al. 2001; Martín et al. 2003)

véletlenszerűen vártnál, ami szintén a géntermékek között levő funkcionális kapcsolat fontosságára utal (Huynen & Bork 1998).

A géncsoportok fogalmát szabadabban értelmezve megfigyelhetjük, hogy a génsorrend sokszor az operonoknál nagyobb genomi kiterjedtségben mutat konzerváltságot. A hasonló funkciójú fehérjéket kódoló gének, ha nem is pontosan azonos sorrendben maradnak meg, de a kromoszómán csoportokba tömörülve, egymás mellett, vagy közelében fordulnak elő (Rogozin et al., 2002; Yang and Sze, 2008; Ling et al., 2009). Ezt a nagyobb léptékben konzervált szerveződést über-operon[15]-nak is nevezik (Lathe et al., 2000)

Hasonló funkciójú gének operonokon felüli csoportokba rendeződésére legismertebb egyedi példák a riboszomális fehérjék szuperoperonja (Reams and Neidle, 2004) vagy a horizontális transzferhez köthető „genomi szigetek" (Hacker and Carniel, 2001), de ezeknél általánosabb mintázatokat is találhatunk. Egyfajta operonon felüli rendezettséget jelent, és mechanisztikus magyarázatul is szolgálhat rá, hogy a koregulált, vagy egymást szabályzó operonok sokkal közelebb vannak egymáshoz a véletlenszerűen vártnál (Warren and ten Wolde, 2004; Zhang et al., 2012). Az egymást szabályzó géncsoportok közelsége a lokális, kevés gént szabályzó transzkripciós faktorok és célgénjeik között jellemző (Kolesov et al., 2007; Janga et al., 2009). Azt, hogy az ilyen géncsoportok léte a transzkripciós faktor hatékonyabb célbajuttatására ható szelekció eredménye lehet (ún. „rapid search hypothesis"), alátámasztja az is, hogy a transzkripciós faktor génje és a célgén egyirányú orientációt mutat, amely minimalizálja a fizikai távolságot a képződő transzkripciós faktor és a célgén promótere között (Kolesov et al., 2007).

Ezen kutatások alapján úgy tűnik, hogy a részleteiben változékony génsorrend magasabb szinten mutat konzervativizmust és esetleg funkcionális jelentőséggel is bírhat. De vajon találhatunk-e bármiféle rendezettséget az operonokon belül a gének sorrendjében?

3.1.2 Az operonbeli génsorrend evolúciója

A legkorábbról ismert géncsoportok, az operonok, különböző prokarióta genomokban (felső becsléssel) a gének 50-90%-át foglalják magukba (Wolf et al., 2001). *E. coli*-ban az ismert operonok kb. a genom 50%-át alkotják (Okuda et al., 2007). Az operonbeli gének funkcionális hasonlósága jól ismert, például az operonbeli szomszédos génpárok 80%-a

[15] A szerzők explicit módon nem tárgyalják a koncepció operonokhoz való viszonyát.

azonos funkcionális csoportba sorolható (Salgado et al., 2000), 60%-a pedig azonos útvonalba (Okuda et al. 2007).

Az operonok konzerváltságának első összehasonlító vizsgálatai nagymértékű plaszticitást mutattak, például az *Escherichia coli* és a *Haemophilus influenzae* operonjainak csaknem fele átrendeződött, ezért egyes szerzők felvetették, hogy az operonon belüli génsorrend változása általában hosszú távon neutrális (Mushegian and Koonin, 1996; Itoh et al., 1999). Ez azonban azt jelentené, hogy csupán a genomátrendeződések rátája és az eltelt idő szabja meg a fajok génsorrendje közötti eltérést. Ez az eltérés valójában sok nagyságrenddel kisebb, mint azt az átrendeződési mutációk gyakorisága alapján várnánk, a különbség pedig az egy operonban lévő gének esetében kiugró (9 nagyságrend eltérés) (Rocha, 2006), ami azt sugallja, hogy stabilizáló szelekció folyhat az operonon belüli génsorrendre.

Az operonokon belüli gének gyakrabban maradnak egymás szomszédságában, mint az operonok közöttiek (Moreno-Hagelsieb et al., 2001; Rocha, 2006; Yang and Sze, 2008) és a konzerválódott génsorozatok többsége operonokhoz köthető (Wolf et al. 2001)[16]. Legtöbb esetben ha az operon konzerválódott, a génsorrend is változatlan marad (Q. Yang & Sze 2008).

Vajon milyen hatások eredményeként tapasztalhatunk a génsorrendre ható szelekciót? Az operont érintő kromoszómaátrendeződések többsége nagy valószínűséggel káros. Az inzerció révén megszakított operon egyik fele elveszíti eredeti szabályozó szekvenciáit. (Ez egyfajta magyarázatot jelenthet a magasan expresszálódó és esszenciális funkciójú riboszomális fehérjéket kódoló géncsoport konzerváltságára (Itoh et al., 1999)). Egy operonon belüli inverzió hatására pedig az invertált gének a DNS másik szálára helyeződnek át, lehetetlenné téve az együttes átíródást. Amennyiben kizárólag a káros átrendeződéseket tekintjük meghatározónak, akkor az eredetileg kialakult génsorrend határozza meg a konzervált operonok génsorrendjét. Eszerint a forgatókönyv szerint az operonok génsorrendje befagyott véletlen (Itoh et al., 1999).

Ugyanakkor elképzelhető, hogy többfajta működőképes operont eredményező génsorrend is létrejöhet a genomi átrendeződések révén, és ha a különböző génsorrendű operonok eltérő rátermettséget eredményeznek, az operonon belüli génsorrendre is folyhat szelekció. Például a trp és his operonok szerveződését különböző baktériumfajokban összehasonlítva, egy-egy

[16] Ezért a szomszédos gének konzerváltságát az operonok előrejelzésére is alkalmazzák (pl. Omelchenko, 2003)

fajban mindkét operonra érvényes hasonlóságokat találtak (például ismeretlen funkciójú gének inzerciója az operonokba). Ez a fajok között eltérő, míg fajon belül a különböző operonokra egyformán érvényes szelekciós hatásokra utalhat (Xie et al., 2003). Az egyes baktériumcsoportok alakja (például coccus, bacillus) és a sejtfal felépítésében közreműködő géncsoport génsorrendje közötti kapcsolat is szelekciós hatás jelenlétét sugallhatja (Tamames et al., 2001). A befagyott véletlennek ellentmond a „xenológ helyettesítések" megfigyelése is, amelyek során egy operonon belül egy gént horizontális transzferrel érkező homológja hasonló pozícióban vált fel,úgy, hogy a homológ rekombináció a filogenetikai távolság miatt kizárható (Omelchenko et al., 2003).

Vajon milyen funkcionális következményei lehetnek a különböző génsorrendeknek, miért eredményezhetnek különböző rátermettséget? A fizikai interakcióban lévő fehérjék egymáshoz közeli elhelyezkedését már említettük. Mivel az egyazon operonban kódolt gének egy mRNS molekulává íródnak át, ezért a genomi közelség azt eredményezi, hogy a kódolt fehérjék egymás fizikai közelségében transzlálódnak. Így például a fehérjekomplexek alegységeit kódoló gének esetében a genomi távolság csökkentése mérsékelheti a monomerek szabad jelenlétének káros hatását (Papp et al., 2003) gyorsítva a komplex kialakulását (Hurst et al., 2004). Ugyancsak lehetővé válhat a közel elhelyezkedő génekről átírt fehérjék számára a kotranszlációs feltekeredés (Dandekar et al., 1998). Termofil környezethez való alkalmazkodásként a gének közelsége a nem hőstabilis fehérjealegységek és az (esetenként szintén káros) hőlabilis köztitermékek átalakulását gyorsíthatja (Glansdorff, 1999).
Mindemellett az enzimaktivitás alloszterikus szabályozásában is szerepe lehet a génsorrendnek. Azonos anyagcsereútvonalban az egyik enzim terméke egy másik enzim aktivitását növelheti vagy csökkentheti. Az egymást anyagcseretermékeken keresztül szabályozó gének térbeli közelsége hatékonyabb szabályzást tehet lehetővé. E hipotézis alapján azt várjuk, hogy ha mindkét fehérje génje azonos operonon belül található, akkor a gének közelebb helyezkedjenek el egymáshoz a véletlenszerűen vártnál (Fani et al., 2005).

3.1.3 A kolinearitás mintázata

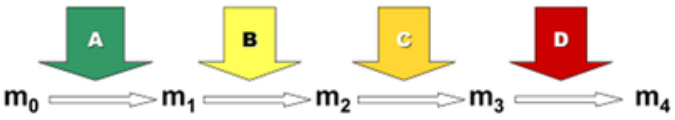
Disszertációmban kolineárisnak nevezem az operonban elhelyezkedő és a metabolikus útvonalban enzimeket kódoló géneket, ha az operon génjeinek genomi sorrendje az általuk kódolt enzimek útvonalbeli sorrendjét tükrözi (*5. ábra*). Már az első géncsoportok, a *Salmonella typhimurium* trp és his operonjai térképezésekor észrevették, hogy esetenként a

gének sorrendje és az általuk kódolt enzimek katalizálta metabolikus lépések sorrendje megfeleltethetőek egymásnak (Demerec, 1964; Lawrence and Roth, 1996). Azonban az anekdotikus megfigyeléseken túl, azóta sem történt meg annak vizsgálata, hogy mennyiben lehet általánosnak tekinteni a fenti mintázatot az operonok között.

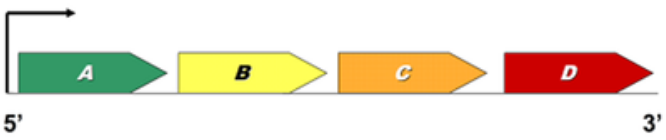
3.1.4 Célkitűzések

A továbbiakban *E. coli*-ban elsőként szisztematikus vizsgálattal összehasonlítom a metabolikus útvonalakat és a hozzájuk tartozó operonokat, hogy teszteljem a kolinearitás általános jelenlétét. Mivel a kolinearitás általános érvényű mintázatnak tűnik, ezért a következőkben a lehetséges funkcionális magyarázatát vizsgálom. A kolinearitás magyarázatára 3 hipotézist mutatok be, azokra általános, de realisztikus paramétereket tartalmazó operon expressziós és enzimkinetikai modellen végzett szimulációk segítségével predikciókat teszek, majd a predikciókat empirikus adatokon tesztelem. Végül tesztelem azt a hipotézist is, hogy a metabolit-szintű alloszterikus reguláció hatással van az operonbeli génsorrendre.

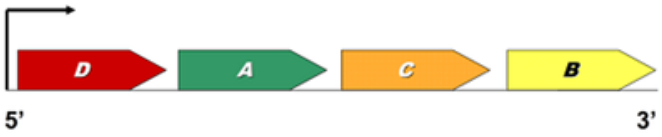
Enzimsorrend az anyagcsereútvonalban



Operon kolineáris génsorrenddel



Operon nem kolineáris génsorrenddel



5. ábra: A kolinearitás jelentése egy hipotetikus operon génjei és az anyagcserútvonalban elfoglalt enzimatikus lépései között

A felső operon génjeinek elrendeződése tökéletesen kolineáris, míg a második esetben a hat operonon belül képezhető génpárból kettőnek van kolineáris sorrendje (A-C és A-B), így a kolinearitás mértéke 1/3. A géneket és az általuk kódolt enzimeket A,B,C,D; a metabolitokat m_0, m_1, m_3, m_4 jelöli.

3.2 Módszerek

3.2.1 Operon expresszió és anyagcsere útvonal modellezése

Modellünk egy négy enzimből álló egyirányú lineáris anyagcsereutat tartalmaz, amelynek enzimeit egy operon kódolja. Az enzimek standard Michaelis-Menten kinetika szerint működnek. Mindegyik enzimnek a katalitikus konstansa (k_{cat}), Michaelis konstansa (K_M) azonos, az aszpartát kináz I enzim kísérletesen mért értékei alapján (*Függelék 4. táblázat*). Egy sejtgeneráció időtartama 60 perc, és minden metabolit és enzim eszerint (D) hígul (ez közel esik az *E. coli* osztódási idejéhez glükóz minimál táptalajon). Kezdetben minden metabolit koncentrációja 0, kivéve az első enzim szubsztrátjáét, aminek koncentrációját 1 mM-ban rögzítettük. Az operon génexpresszióját Swain read-through operon modellje alapján modelleztük (Swain, 2004), sémáját lásd a *6. ábrán* és a *Függelékben*. A matematikai modellezéshez a Copasi 4.4.28 verziót használtuk (Hoops et al., 2006). A sztochasztikus szimulációkat a Copasi hibrid determinisztikus-sztochasztikus szimulációs algoritmusával végeztük ("Hybrid Runge-Kutta", standard beállítások kivéve Runge-Kutta step size 0.1)

3.2.2 Az operonokra vonatkozó adatsorok összeállítása

A metabolikus útvonalak és az operonok megfelelő génsorrendű listájának létrehozásához az Ecocyc adatbázist használtuk (10.5 verzió) (Keseler et al., 2009). Az adatbázisból kigyűjtöttük az átfedő géneket tartalmazó operonokat és anyagcsereútvonalakat (legalább két, különböző reakciólépésekhez tartozó génnel). A kolinearitás mértékének számszerűsítéséhez a vizsgált génekből az enzimatikus és génsorrend szerint rendezett párokat állítottunk elő, a nem egyértelmű útvonalbeli sorrendeket (pl. ciklikus anyagcsereút) kizárva.

A génpárok előállításának menete: (i) Az Ecocyc-ben jelölt operonok egy része hierarchikusan átfed, ilyenkor a legtöbb vizsgált gént tartalmazó operont vettük figyelembe. (ii) Az útvonalak is teljes egészében átfedő szuperútvonalakba csoportosulhatnak, ilyenkor a hosszabb útvonal teljes egészében tartalmazza a rövidebbet. Ebben az esetben a legrövidebb, az összes vizsgált gén által kódolt enzimet tartalmazó útvonalat vettük figyelembe. Nem hierarchikus viszonyban levő, részben vagy teljesen átfedő reakciólépéseket tartalmazó útvonalakkal esetén az átfedő géneket csak egyszer vettük figyelembe, és csak akkor ha az anyagcsereútvonalakban a sorrend azonos volt. (iii) Elágazó útvonalaknál a lineáris

szakaszokat eltérő útvonalakként kezeltük és esetleges átfedés esetén az átfedő génpárokat csak egyszer vettük figyelembe. A továbbiakban ezekre az elágazásmentes lineáris szakaszokra is útvonalakként hivatkozunk. vi) Egy enzim több reakciót is katalizálhat egy útvonalon belül. Amennyiben az enzim egymás utáni reakciólépésekben vesz részt, illetve a két lépés között nincsen másik vizsgált gén, az összes génpár sorrendje egyértelműen meghatározható volt. A kétszer szereplő génnel képzett párokat nem redundáns módon vettük figyelembe. Amennyiben az enzim által katalizált két reakciólépés között egy harmadik vizsgált enzim is szerepel, közöttük nem képezhetők egyértelmű sorrendű párok, ezért ezeket a párokat kihagytuk. A többi, egyértelmű génpárt nem redundáns módon vettük fel az adatsorba. v) Egy reakciólépéshez több enzim is tartozhat, például enzimkomplex alegységek vagy izoenzimek esetén. Ezekben az esetekben az összes alegységet, illetve izoenzimet kódoló génnek mindegyikével külön képeztünk párokat. vi) Az Ecocyc adatbázis útvonalai legtöbb esetben nem tartalmazták az első metabolithoz tartozó esetleges transzporter fehérjéket. Ezért a transzporterek génjeit a kigyűjtött operonok korábban párba nem állított génjei alapján azonosítottuk és adtuk hozzá a listához. A fenti eljárás 321 génpár listáját eredményezete 70 operonban és 73 útvonalban.

A *B. subtilis* operon struktúra adatait a BioCyc (Caspi et al. 2011) és DBTBS (Sierro et al., 2008) adatbázisokból gyűjtöttük. Az ortológok és kromoszomális pozíciók fajok közötti összehasonlításához az EcoCyc adatbázist (Keseler et al., 2009) használtuk.

3.2.3 A kolinearitás mértékének számítása

A kolinearitás mértéke az operonok egy vizsgált csoportjában a kolineáris génpárok száma osztva az összes génpárral. Egy adott génpárt kolineárisnak tekintettünk, ha az 5’ véghez közelebbi gén által kódolt enzim az anyagcsereútvonalban előbb szerepel, mint a 3’ véghez közelebb kódolt enzim (*5. ábra*). A p-értéket randomizációs teszttel becsültük. Megszámoltuk, hogy a génpárok mekkora hányada kolineáris, majd a kapott értéket a vizsgált gének sorrendjének operonon belüli randomizálásával kiszámított értékekből kapott eloszláshoz hasonlítottuk. A p-érték kiszámítási módja: $(R+1)/(N+1)$, ahol R azon randomizálások száma, amikor a vizsgált gének sorrendjét operonon belül véletlenszerűen megváltoztatva ugyanannyi vagy több kolineáris párt kapunk, mint amennyi az általunk az adatokból ismert érték. N az összes randomizálás száma. A p-érték megadja, hogy véletlenszerű génsorrend esetén mekkora valószínűséggel kapnánk az adatokból ismert vagy annál nagyobb értéket. Minden esetben 100 000-szer randomizáltunk.

3.2.4 mRNS abundancia vizsgálatok

Vizsgálatainkhoz egy publikált Affymetrix microarray génexpressziós adatsort használtunk fel (Covert et al., 2004). Ez $\log_2$-transzformált normalizált expressziós profilokat tartalmaz vad típusú *E. coli* K-12 MG1655 törzsre M9 glükóz táptalajon aerob és anaerob körülmények között. Mindegyik génre kiszámoltuk az átlagos expressziós értéket három (aerob) illetve négy (anaerob) adatpontból (ismételt mérések). Az operonok mRNS-szintjét az alkotó gének expressziójának átlagaként számoltuk. Hogy megvizsgáljuk, vajon az operonok csökkenő mRNS szintet mutatnak-e 5'-3' irányban, kigyűjtöttük az *E. coli* genom mindig együtt átíródó transzkripciós egységeit. Ez 386 transzkripciós egységet eredményezett (2199 egységen belüli génpárral) legalább két expresziós adatot tartalmazó génnel. Azoknak az eseteknek a tapasztalati számát, ahol a génpár 5' tagjának nagyobb az expressziós szintje, mint a 3' tagnak (aerob környezet: 1274 pár, anaerob: 1293 pár) összehasonlítottuk a gének transzkripciós egységeken belüli randomizálával kapott ugyanezen érték eloszlásával ($p<10^{-6}$ mindkét környezetben). Hasonló elemzést végeztünk a metabolikus operonokon (270 génpár 65 operonban, $p = 0{,}0295$ aerob és $p = 0{,}0224$ anaerob glükóz táp környezetben).

Mindegyik operon esetében külön-külön is meghatároztuk, hogy az mRNS abundancia profilja szignifikánsan csökkenő tendenciát mutat-e 5'-3' irányban. A monoton csökkenő abundanciát lineáris trend analízissel teszteltük (Quinn and Keough, 2002), a 'gmodels' R csomaggal (Warnes et al., 2006). A változás irányát Spearman rang korrelációval állapítottuk meg. Bonferroni korrekció után 26 (aerob) illetve 23 (anaerob) szignifikánsan csökkenő mRNS expressziót mutató operont találtunk.

A különböző környezetek közötti expressziós variabilitás számszerűsítéséhez egy 213 különböző környezeti feltételek közt mértexpressziós adatokat tartalmazó adatsort használtunk (Price et al., 2006). Kiszámoltuk a $\log_2$ transzformált expressziós értékek szórását, ami invariáns multiplikatív transzformációval szemben (Lewontin, 1966), tehát egy variácós koefficienst eredményez. Az operonok expressziós variabilitását az alkotó gének szórásának átlagaként definiáltuk. Ahhoz, hogy megvizsgáljuk, az operonok expressziós variabilitása korrelál-e a kolinearitás mértékével az mRNS abundanciák mértékére kontrollálva, a szórás és az abundancia közötti lineáris regresszió reziduális értékei alapján osztályoztuk az operonokat, az átlag értéknél nagyobb, vagy kisebb reziduális expressziós varianciájú csoportokba osztva. Randomizációs teszttel vizsgáltuk, hogy a kolinearitás eltér-e a két csoportban. Ezzel analóg módszerrel vizsgáltuk, hogy az operonok mRNS abundancia összefügg-e a kolinearitással, ha az expressziós variabilitásra kontrollálunk.

3.2.5 Metabolit-szintű enzimregulációs adatsor összeállítása

Az EcoCyc (Ingrid M Keseler et al. 2009) adatbázist és egy, a BRENDA (Schomburg et al. 2004) adatbázison alapuló adatsort (Gutteridge et al. 2007) felhasználva, összeállítottunk egy adatsort a metabolit-szintű útvonalon belüli szabályzási kapcsolatokról azonos operonban kódolt enzimek között. Vagyis olyan eseteket gyűjtöttünk ki, amikor az egyik enzim terméke a másik enzimet aktiválja vagy gátolja, kivéve, ha az adott metabolit a szabályozott enzim szubsztátja is egyben (tehát alloszterikus kapcsolatokra fókuszáltunk). 19 génpárra találtunk ilyen interakciót 11 operonban. Annak teszteléséhez, hogy az átlagos géntávolság az interakcióban levő gének esetében eltér-e a véletlenszerűen várttól, az útvonalban részt vevő gének pozícióját randomizáltuk (a többi gén pozícióját megőrizve). A megfigyelt átlagos géntávolság nem különbözött szignifikánsan a randomizáció során kapott véletlenszerű eloszlástól. (1.84 vs 2.07, $p=0,234$; a szomszédos gének közötti távolság 1-nek lett definiálva). Annak megvizsgálásához, hogy az operonon belüli metabolit-szintű szabályzási kapcsolatoknak van-e hatása a kolinearitás mértékére, összehasonlítottuk a kolinearitás mértékét az ismert regulációs kapcsolatokat tartalmazó operonokban (11 operon) az adatsor többi tagjával (59 operon). A különbség szignifikanciáját az operonok ugyanilyen arányú véletlenszerű csoportokba sorolásával kapott kolinearitás értékek eloszlásával való összehasonlítással határoztuk meg. Mivel a metabolit-szintű regulációs kapcsolatokat tartalmazó operonok mRNS szintje magasabb, mint a többi operoné, és az expressziós szint korrelál a kolinearitás mértékével, a fenti vizsgálatot az mRNS szintre korrelálva is elvégeztük. Ehhez alegnagyobb olyan alcsoportot definiáltuk az 59 operon közül, amely expressziós szintje nem különbözik szignifikánsan a regulációt tartalmazó operontól, 35 illetve 36 operont választva ki anaerob illetve aerob környezetben ($p=0,0504$ és $p=0,0516$ értékek alapján). Végül az így kapott csoportok kolinearitását hasonlítottuk össze a metabolikus regulációs interakciókat tartalmazó operonokéval.

3.3 Eredmények

3.3.1 A metabolikus operonok kolineáritásának mértéke nagyobb a véletlenszerűen vártnál

Vajon a metabolikus operonok génsorrendje a kódolt enzimek funkcionális sorrendjét tükrözi? A kérdés vizsgálatához az *E.coli*-t választottuk, a jó minőségű és lefedettségű operonszerkezeti és anyagcsereútvonal-adatok miatt. Legalább két, azonos anyagcsereútvonalba tartozó enzimet kódoló operonokról gyűjtöttünk adatokat az EcoCyc (Keseler et al., 2009) adatbázisból, 70 operont és 321 operonon belüli génpárt eredményezve (lásd *Módszerek 3.2.2*.). Minden egyes operonális génpárra megvizsgáltuk, hogy a relatív pozíciója megegyezik-e a kódolt enzimek funkcionális sorrendjével, vagyis kolineáris-e. Körülbelül a 321 génpár 60%-a mutatott kolinearitást, ami szignifikánsan nagyobb a véletlen alapján várt 50%-nál (p=0,0011, randomizációs teszt).

3.3.2 Hipotézisek a kolinearitás funkcionális magyarázatára

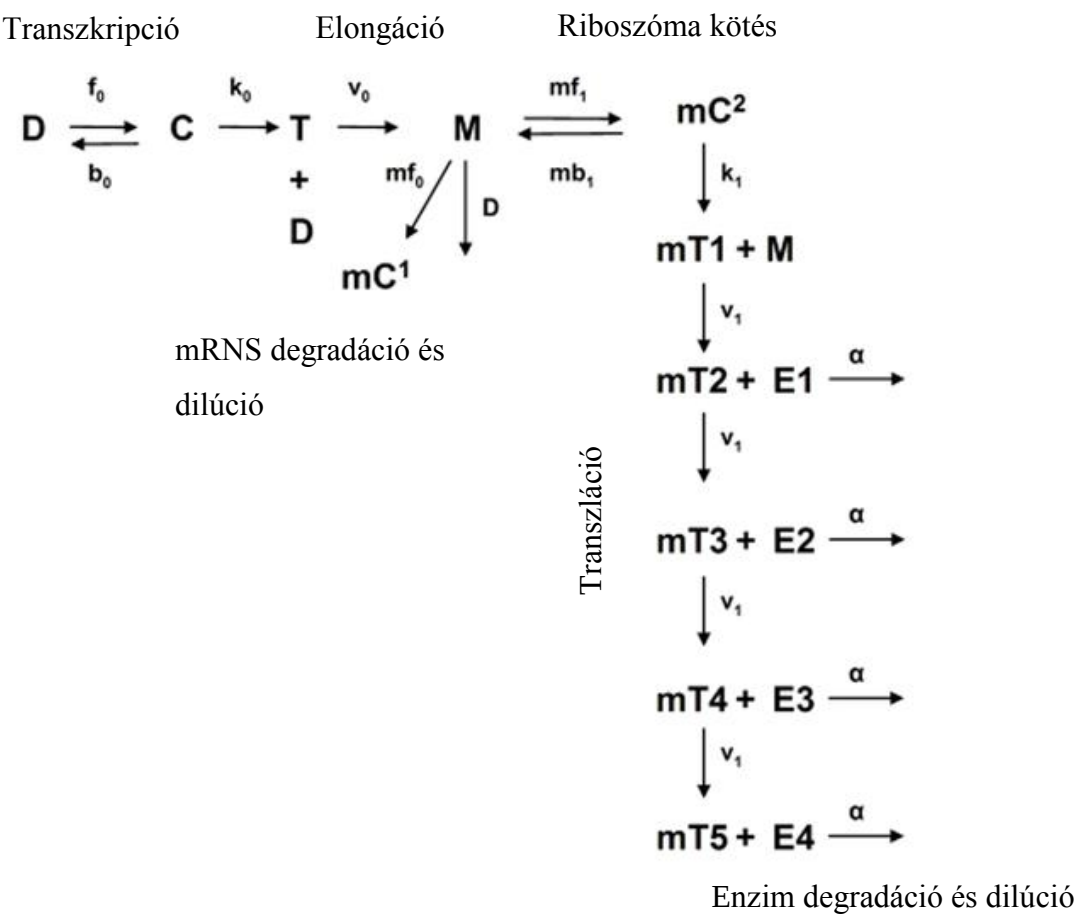
A fenti eredmény meglepő, mivel a legegyszerűbb esetet, az enzimek steady-state állapotát feltételezve a génsorrend az anyagcsereutak produktivitására nincs hatással (0. hipotézis). Ennek alátámasztására korábbi munkák alapján (Swain, 2004; Zaslaver et al., 2004) általános matematikai modelleket építettünk, amelyek egy négy enzimet (E_1, E_2, E_3, E_4) expresszáló operonból és az enzimek által katalizált lineáris anyagcsereútvonalból állnak. Realisztikus modell paramétereket használtunk (lásd *Módszerek 3.2.1* és *Függelék 3. táblázat*), standard Michaelis-Menten kinetikát és egyforma enzimkinetikai paramétereket mind a négy enzimre. Az első három termék metabolitkoncentrációinak időbeli változását az alábbi egyenlet írja le (i = 1,2,3):

$$\frac{dS_i}{dt} = k_{cat} \cdot E_i \cdot \frac{S_{i-1}}{S_{i-1} + K_m} - k_{cat} \cdot E_{i+1} \cdot \frac{S_i}{S_i + K_m} - D \cdot S_i \qquad \textit{1. egyenlet}$$

ahol k_{cat} a katalitikus konstans, D dilúciós ráta (sejt növekedési rátája), K_m a Michaelis konstans (értékek: *Függelék 3. táblázat*). Az első szubsztrát koncentrációja (S_0) 1 mM-ra lett beállítva, míg a többi metabolit kezdeti koncentrációja 0. Az anyagcsereútvonal produktivitása az operon indukciója után adott idő alatt megtermelt végtermék összmennyiségét jelenti. A végtermék (S_4) termelése a 2. egyenlettel írható le (az S_4 összesített termelt mennyisége érdekes számunkra, ezért a végtermék dilúciója nem szerepel):

$$\frac{dS_4}{dt} = k_{cat} \cdot E_4 \cdot \frac{S_3}{S_3 + K_m} \qquad \textit{2.egyenlet}$$

Az operon expresszióját a „read-through” modell (Swain, 2004) alapján modelleztük, ebben a riboszómák közvetlenül haladnak egyik génről a következőre, így a transzlációs események teljes mértékben korreláltak az operon génjei között. A transzlációs ráta (v_1) úgy lett finomhangolva, hogy a modellben az egymásutáni géntermékek (E_i) megjelenése között eltelt idő a kísérletesen megfigyelt értéknek, átlagosan 60 másodpercnek feleljen meg (Alpers and Tomkins, 1965, 1966).



6. ábra: Az operon génexpressziós modelljének reakciósémája. *A modell egy négy génes operonból és egy lineáris anyagcsereútból áll, ami 5 metabolitot és 4 enzimet tartalmaz, melyeket az operon génjei kódolnak. Az RNS-polimeráz promóterhez (D) való reverzibilis kötődéset f_0 asszociációs és b_0 disszociációs rátákkal modellezzük. A nyílt iniciációs komplex és az transzkcripció iniciációja egyirányú folyamatként szerepel k_0 rátával. Csak az mRNS leader régiója (M) szerpel a modellben, melyet a transzkripciót végző RNS polimeráz (T) v_0 rátával ír át. Az mRNS mf_0 rátával bomlik le és D rátával hígul. A riboszómák a degradoszómákkal versengenek a leader mRNS-hez való reverzibilis kötődésért (mf_1 asszociációs and mb1 disszociációs rátával). A transzláció az mC2 állapotból veszi kezdetét k_1 rátával ami után az M mRNS szakasz szabaddá válik a további riboszóma vagy degradoszóma kötéshez. Az enzimek transzlációja az mT állapotban v_1 rátával zajlik, lebomlása és hígulása pedig α rátával (α = D+kdegr). A „read-through" operon modellben csak az első génnek van riboszóma kötő helye így a transzlációt végző riboszóma, mT2 az E1 enzim megtermelése után írja át a következő enzimet (mT3 állapotban). A modell egyenletei megtalálhatók a Függelékben.*

Szimulációnk alátámasztja, hogy steady-state állapotban az anyagcsereútvonalon átmenő fluxus (steady-state reakciósebesség) független a génsorrendtől (*Függelék: 4. táblázat*). A kolinearitás jelenléte azonban azt jelzi, hogy a fenti modellből hiányzik valami. Az alábbiakban három hipotézist vázolunk fel, egy újat és két régit, amelyek a kolinearitás magyarázataként szolgálhatnak. A 3 hipotézis felvázolása után megkíséreljük azok tesztelését meglévő funkcionális genomikai adatok alapján. A három hipotézis a kolinearitást a következő okok miatt tekinti előnyösnek: (1.) az operon 5' véghez közelebbi gének magasabb expressziós szintet érnek el, és emiatt a kolineáris operon magasabb produktivitást eredményez (Nishizaki et al., 2007), (2.) a kolinearitás átmeneti előnyt jelent közvetlenül az operon bekapcsolása után, (3.) a kolinearitás az enzimek véletlenszerű elfogyását követő útvonal-leállás idejét minimalizálja.

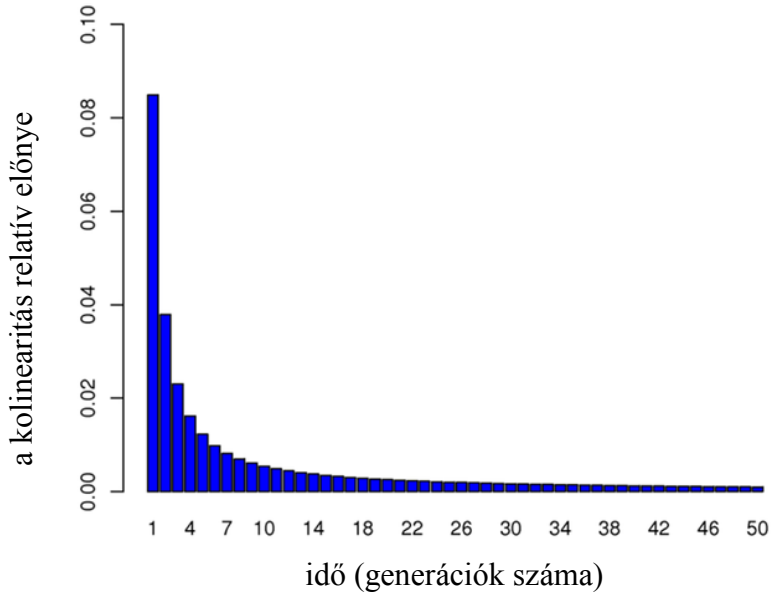
1. hipotézis: a steady-state fluxus növekedése az 5' közeli gének nagyobb expressziós szintje miatt

Az első hipotézis (Nishizaki et al., 2007) a kolineáris elrendezés megnövekedett produktivitását az operonon belüli monoton csökkenő expressziónak, az úgynevezett polaritásnak (Ullmann et al., 1979) tulajdonítja. Azt, hogy az útvonal első enzimének magasabb expressziója a végtermék termelését megnövelheti, kísérletesen manipulált operonon igazolták (Nishizaki et al., 2007). Elméleti modellek szerint, ha az útvonalon belüli enzimmenyiség állandó (felső korlátja van az enzimkoncentrációnak), akkor lineáris anyagcsereutak esetében az egymás után következő enzimek (azonos katalitikus hatékonyságú enzimeket feltételezve) koncentrációjának monoton csökkenése maximalizálhatja a steady-state fluxust (Heinrich and Klipp, 1996). Így, amennyiben az operonon belüli gének között expressziós gradiens van jelen, a kolinearitás előnyt jelenthet. Operon-expressziós modellünk, amennyiben a polaritást beleépítjük, megerősíti a fentieket (lásd *Függelék 5. táblázat*). E hipotézis alapján a csökkenő mRNS abundancia profilú operonok esetében magasabb kolinearitást várunk.

2. hipotézis: az operon bekapcsolása utáni átmeneti előny

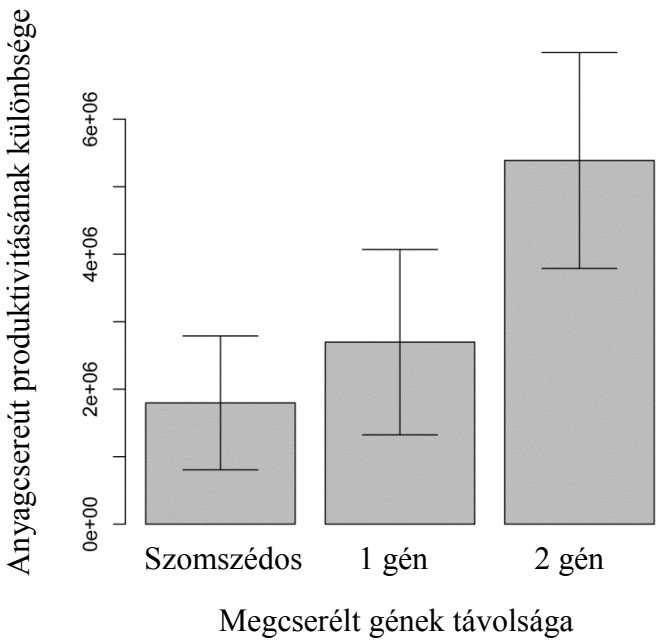
A második elképzelés szerint a kolinearitás átmeneti előnyt jelenthet az operon expressziójának kezdetén, a steady-state enzimszintek beállása előtt. Egy két enzimet (A,B) tartalmazó anyagcsereút esetén, ha az operonbeli génsorrend AB, A enzim működésbe léphet a szubsztrátján, míg a B enzim megjelenik (t.i. először az operon első génterméke jelenik

meg, majd kb. 1 perces késlekedéssel jelenik meg a következő, stb.). Megfordítva, BA sorrend esetén az útvonal a második enzim szintézise előtt nem kezdhet el működni, így lassabbá téve az végtermék első megjelenését. Szimulációnk alátámasztja a fenti lehetőséget (*7. ábra, Függelék 5. táblázat*). Különböző génsorrendű operonok esetén a kolinearitás az operon aktivációját követő egy sejtgenerációnyi idő alatt 8.49%-kal volt képes növelni az útvonal produktivitását (a legkolineárisabb: ABCD és a legkevésbé kolineáris DCBA sorrendeket összehasonlítva). Ennek oka az egymást követő géntermékek megjelenése közötti időkülönbség (Alpers and Tomkins, 1965, 1966), ami az összenzim koncentráció limitációja esetén csökkenti a metabolitok átalakulási rátáját (Klipp et al., 2002; Zaslaver et al., 2004). Ezenkívül szimulációinkban a gének felcserélésének hatása a köztük levő távolságtól függ: átlagosan a szomszédos gének felcserélésének van a legkisebb hatása (*8. ábra*).



7. ábra A kolinearitás relatív előnye csökken az az operon aktivációja után eltelt idővel

Az oszlopok a végtermék relatív többletét mutatják kolineáris (ABCD) operon esetében anti-kolineáris (DCBA) génsorrenddel összehasonlítva. Az operon t=0 időpontban indukálódik. (Az eredmények különböző paraméter értékek esetében Függelék 6. táblázatában találhatók.)



8. ábra Két gén sorrendje felcserélésének hatása az útvonal produktivitására függ a fizikai távolságuktól. *Egy négygénes operon minden lehetséges génsorrendje esetében meghatároztuk az anyagcsereút produktivitását determinisztikus szimulációval, és három csoportot definiáltunk a felcserélt gének fizikai távolsága alapján (azokat a génsorrendeket hasonlítottuk össze párosával, amelyek egy génpár felcserélésével egymásba alakíthatók). Az anyagcsere produktivitása az operon indukció után egy generációs idő alatt előállított végtermék mennyiségeként lett definiálva. Átlagokat és 95%-os konfidencia intervallumokat ábrázoltunk. Randomizációs tesztet alkalmaztunk a 2. és 1. illetve a 3. és 2. csoport átlaga közti eltérés szignifikanciájának vizsgálatára (p= 0,0001 az egyedi produktivitás különbségek 100 000 permutációja alapján).*

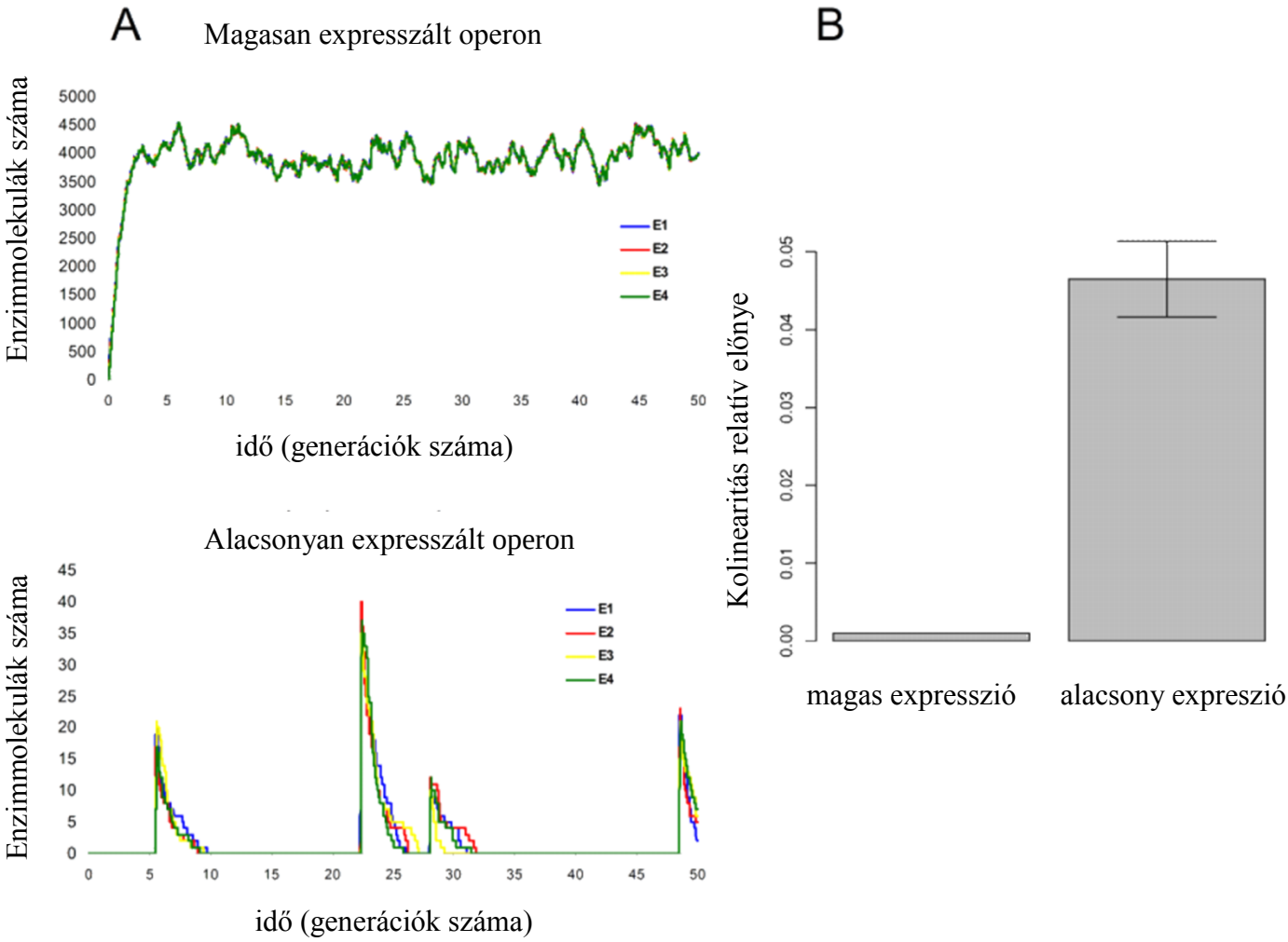
3. hipotézis: az anyagcsere sztochasztikus leállásának minimalizálása

A fenti determinisztikus modellek nem törődnek azzal a megfigyeléssel, hogy a génexpressziós folyamatában kis számban jelen levő molekulák vesznek részt, ami a fehérjemennyiség jelentős mértékű sztochasztikusságához vezethet (Elowitz et al. 2002). Míg az indukció után a magasan expresszált operonok által kódolt enzimek valószínűleg mindig jelen vannak a sejtben, az alacsonyan expresszált operonok esetében az enzimek degradálódhatnak vagy sejtosztódás következtében kihígulhatnak két expressziós esemény között (Cai et al., 2006) visszatérő módon megakasztva az anyagcserefolyamatok működését. A kolinearitás minimalizálhatja a sztochasztikus enzimvesztések hatását, ugyanis felgyorsítja a leállt anyagcserefolyamatok újraindulását, hasonló módon, mint az operonok indukciója után tapasztalt átmeneti előny esetében (2. hipotézis).

Operon-expressziós modellünkön sztochasztikus szimulációt hajtottunk végre, hogy megvizsgáljuk, hogyan befolyásolja a génsorrend az útvonal produktivitását az expressziós szint függvényében. Különböző expressziós szinteket szimuláltunk az RNS polimeráz DNS disszociációs rátáját változtatva *(Függelék: 3. táblázat)*.

Először is szimulációink azt mutatják, hogy bár az enzimszint mind a magas, mind az alacsony expressziós szint mellett is fluktuál, az enzimek ismétlődő eltűnése csak alacsony expressziós szint mellett jellemző (*9A ábra*). Ilyenkor jellemzően mind a négy enzim elveszik az expressziós események között, ami az útvonal teljes leállásához vezet. Másodszor, szimuláltuk két hipotetikus lineáris útvonal működését melyeket ugyanazon operon kódolt és az egyik kolineáris, míg a másik anti-kolineáris volt vele. Így az anyagcsere produktivitása a két elrendezés esetén közvetlenül összehasonlíthatóvá vált a szimulációk sztochasztikussága ellenére. Az útvonal produktivitását az indukció utáni 50 sejtgenerációnyi idő alatt követtük. Elemzésünk azt mutatta, hogy míg a kolinearitás egy nagyon alacsonyan expresszált operonban (átlag ±standard deviáció [SD] fehérje / sejt = 2,4±6,1) az útvonal produktivitását 4,65%-kal tudta emelni, ez a mennyiség 0,1%-ra esik magasan expresszált operonok esetében (3959±232 fehérjeszám/sejt). A hatás a sztochasztikusságnak köszönhető, mivel a kolinearitás előnye lecsökken, ha az alacsony expressziójú szimulációt determinisztikusan futtatjuk (az átlagos fehérjeszintre kontrollálunk) és a magas expresszójú sztochasztikus szimulációéhoz hasonló értéket: 0,07%-os előnyt kapunk. Figyelemreméltó, hogy az alacsony expressziójú sztochasztikus szimuláció esetén nem csak a tökéletesen, hanem már a közepesen kolineáris elrendezések is érzékelhető előnyt jelentenek (2,42%).

Így a fenti szimulációk alapján azt várjuk, hogy a kolinearirás elsősorban az alacsonyan expresszált operonok, és operonon belül az egymástól távolabbi génpárok tulajdonsága. (A szimulációk paraméterek változtatására nézve robosztusak. Lásd *Függelék 7. táblázat*.) A predikció olyan szempontból is érdekes, ill. nem intuitív, hogy az általános vélekedés szerint a magasabban expresszált génekre hat erősebb stabilizáló szelekció, például ez magyarázhatja azok lassabb evolúciós tempóját (Pál et al., 2001; Drummond et al., 2005).



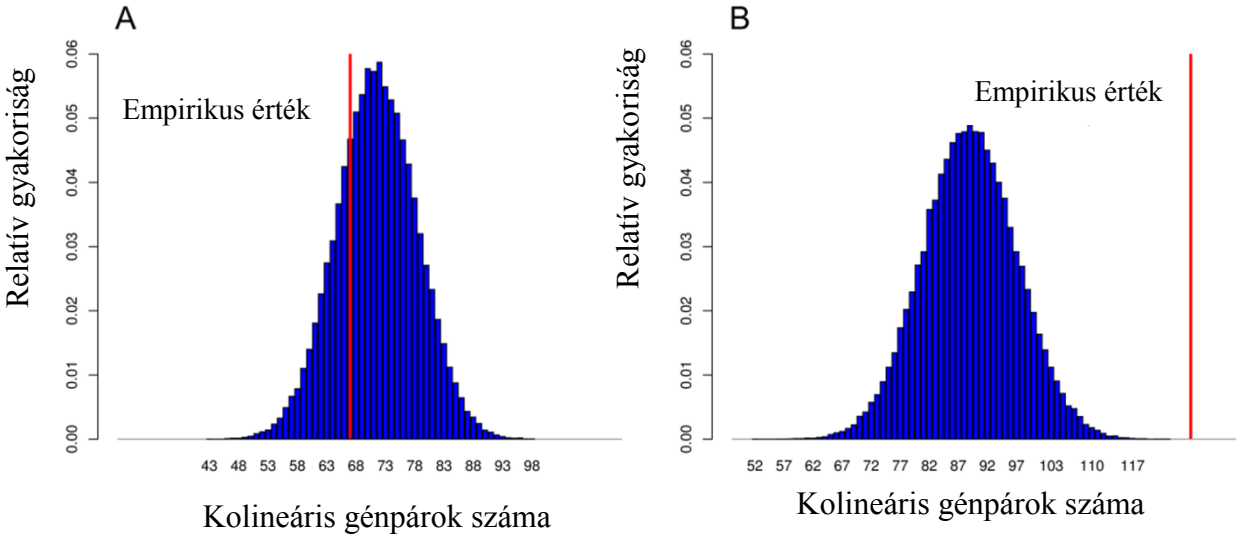
9. ábra: A sztochasztikus szimuláció eredményei

(A) Enzimek molekulaszámának időbeli fluktuációi magasan és alacsonyan expresszált operonban, a modell sztochasztikus szimulációinak eredményei. (modell paraméterek: Függelék 1. táblázat) (B) A kolinearitás átlagos relatív előnye magasan és alacsonyan expresszált operonok esetén 50 sejtgeneráció után (180 000 mp) a modell sztochasztikus szimulációi alapján: 1000 ismételt szimuláció átlaga és 95%-os konfidencia-intervallum. Alacsonyan expresszált operonban a kolinearitásnak szignifikánsan magasabb az előnye, mint magasan expresszált operonban ($p<2.2\times10^{-16}$, Brunner-Munzel teszt: rang alapú módszer eltérő szórású csoportok összehasonlítására (Wilcox, 2005)). Az ábrán a 95%-os konfidencia-intervallum látható.

3.3.3 A modell predikcióinak empirikus tesztelése

A három, a fenti modellek alapján működőképes hipotézist az alapján tesztelhetjük, hogy megvizsgáljuk milyen típusú operonok mutatnak kolinearitást. Az 1. hipotézis teszteléséhez az irodalomból microarray expressziós adatokat gyűjtöttünk glükóz minimál táptalajon növő vad típusú *E. coli* törzsekről aerob és anaerob körülmények között mérve (lásd *Módszerek 3.2.4*). Először megvizsgáltuk, hogy megtalálható-e az operonok esetében az 5'-3' irányban csökkenő mRNS abundancia trendje (polaritás), ahogyan azt korábban mesterséges (Nishizaki et al., 2007) és natív operonok esetében (Ullmann et al., 1979) is tapasztalták. A fentiekkel összhangban az operonon belüli génpárok közül a csökkenő expressziós trendet mutató párok szignifikáns többségét találtuk ($p<10^{-6}$; lásd *Módszerek 3.2.4*), bár számos olyan operon is van amelynél nem látható ilyen trend. A mintázat csak anyagcserében résztvevő operonok esetében is jelen van ($p<0,03$). A fenti hipotézis alapján kizárólag a polaritást mutató operonok esetében várunk kolinearitást. Ezért megvizsgáltuk, hogy a kolinearitás mértéke eltér-e a poláris és nem poláris operonok között. Lineáris trend analízist (Quinn and Keough, 2002) alkalmazva, 26 illetve 23 operont találtunk, amelyek szignifikánsan csökkenő mRNS abundancia profilt mutattak. A hipotézis predikciójával ellentétben nem találtunk nagyobb mértékű kolinearitást ezekben az operonokban akár az aerob (p = 0.97) vagy anaerob (p = 0.36) körülményekre vonatkozó mRNS abundancia adatot használtuk.

A „sztochasztikus leállás" hipotézis predikciója, hogy kolinearitást elsősorban alacsonyan expresszált operonok esetében kell látnunk. Ezt tesztelendő, a metabolikus operonokat két csoportra osztottuk az mRNS expressziós-szintjük alapján, az expressziós-szint logaritmikus skálán vett átlágánál kettéválasztva. A predikcióval egybevágóan azt találtuk, hogy az alacsonyan expresszált operonok esetében a kolinearitás mértéke szignifikánsan magasabb, mint a magasan expresszált operonok esetében (*10. ábra*; aerob körülmények között: p = 0,0011; kolinearitás mértéke: 44,8% vs. 71.6%; aerob körülmények között: p = 0,0006; kolinearitás mértéke: 45% vs. 70.8%, lásd *Módszerek*). A magasan expresszált operonok esetében a kolinearitás mértéke nem tér el a véletlenszerűen várttól (*10. ábra*). A vizsgálat feltételezi, hogy a populáció szinten mért alacsony mRNS szint esetében nagyobb az esély a fehérjék eltűnésére. A genomszintű expressziós adatsort *E.coli* fehérje zaj adatokkal (Taniguchi et al., 2010) összehasonlítva a számunkra releváns nagy zajú ($\mu^2/\sigma^2 > 1$; ahol μ a sejtenkénti fehérje darabszám átlaga, σ pedig a szórása) fehérjék génjeinek kb. 80%-a az átlagosnál alacsonyabb mRNS expressziójú csoportba tartozik aerob és anaerob környezetben egyaránt (Fisher-teszt az mRNS expresszió és zaj szerinti csoportosítás között: $p=10^{-15}$).



10. ábra A kolineáris génpárok számának eloszlása randomizált mintákban magas (A) és alacsony (B) expressziójú operonok esetében.

A piros vonal a kolinearis génpárok E. coli genomban megfigyelhető számát mutatja ([A] 67/143, [B] 127/178 génpár). A magasan és alacsonyan expresszált operonok csoportjait glükóz mininál táptalajon, aerob környezetben mért mRNS szintek alapján határoztuk meg. Az operonon belüli génsorrendet 100 000-szer randomizáltuk.

A mintázat, hogy a kolinearitás az alacsony expressziójú operonokra jellemző, az 1. és 2. hipotézissel is kompatibilis lehet. Az első hipotézist feltételezve elképzelhető, hogy a kolinearitás előnye polaritás esetén alacsonyan expresszált operonok között jobban érvényesülhet. Determinisztikus szimulációnk azonban ennek ellenkezőjét jósolja: a polaritás és a kolinearitás kombinációja magas expresszió mellett jelent nagyobb előnyt (*Függelék, 5. táblázat*). A második hipotézis esetén a kolinearitást az expresszió környezetenként eltérő mértéke erősítheti, az operonok környezetváltozáskor történő gyakori fel- és leregulálása az indukció utáni átmeneti előnyt megsokszorozhatja. Az átlagosan magas expressziójú gének nagyobb valószínűséggel expresszálódhatnak konstitutívan, így magyarázva a kolinearitás expressziós szinttel való (indirekt) összefüggését. Ebben az esetben azonban önmagában az expresszió varianciája ami számít és nem a mértéke. A relatív génexpressziós variabilitás becsléséhez 213 környezetben mért mRNS szint relatív varianciáját használtuk (lásd *Módszerek 3.2.4*). A várakozásokkal ellentétben azt találtuk, hogy az operonok génexpressziós variabilitása és átlagos expressziós szintje gyenge pozitív korrelációt mutat ($r = 0{,}347$, $p = 0{,}004$ és $r = 0{,}267$, $p = 0{,}0268$). A 2. hipotézis predikciójával ellentétben

amikor expressziós szintre kontrolláltunk a magas expressziós varianciájú operonok között nem találtunk nagyobb mértékű kolinearitást, inkább ellenkező irányú trendet tapasztaltunk (p = 0,057 and p = 0,128; lásd *Módszerek 3.2.4*). Ugyanakkor az expressziós szint hatása a kolinearitás mértékére szignifikáns marad az expressziós varianciára történő kontrollálás után is (p = 0,01 és p = 0,016 aerob és anaerob környezetben). Következtetésünk szerint az operonok környezetspecikusságában tapasztalt eltérések nem tudja megmagyarázni az alacsonyan expresszált operonokban látott nagyobb mértékű kolinearitást.

3.3.4 A kolinearitás mértéke a gének fizikai távolságával nő

Operon-modellünkön végzett szimulációink azt is prediktálják, hogy a kolineáris elrendezés előnye átlagosan magasabb akkor, ha a gének közötti távolság nagyobb (*9. ábra*). Az alacsonyan expresszálódó operonokban ellenőrizve a predikciót azt találtuk, hogy a több mint egy génhossz távolságra elhelyezkedő génpárok nagyobb mértékű kolinearitást mutatnak, mint az egymáshoz közelebbiek (Fisher-egzakt teszt, p<0.005; adatsorunkban a medián génhossz 1070 bázispár). Tehát a kolinearitás az operonon belül távolabbi gének esetében nagyobb mértékű, ami megerősíti a modell predikcióját.

3.3.5 Az operonon belüli metabolit-szintű szabályzás kolinearitásra gyakorolt hatását adataink nem támasztják alá

Az eddigi vizsgálataink mind azt feltételezték, hogy az operonon belüli legelőnyösebb génsorrend az, amelyik tükrözi az anyagcsereút enzimsorrendjét. Ugyanakkor elképzelhetőek másféle előnyös elrendezések is, például ha az egyik enzim terméke egy másik, azonos anyagcsereúton levő, enzim aktivitását alloszterikusan szabályozza. Az egymást anyagcseretermékeken keresztül szabályozó gének térbeli közelsége hatékonyabb szabályzást tehet lehetővé a géntermékek szintjén, ami a gének operonon belüli közeli pozíciójára szelektálhat (Fani et al., 2005), akkor is ha egyes esetekben az a kolineárissal ellentétes mintázatot eredményezhet. Ha például a metabolit-szintű szabályzás gyakoribb lenne magasan expresszálódó operonokban, az alternatív magyarázat lehet a kolinearitás hiányára ebben a csoportban.

A hipotézis teszteléséhez útvonalon belüli, metabolit-szintű enzimatikus szabályozókapcsolatokat gyűjtöttünk az EcoCyc (Keseler et al., 2009) adatbázisból és egy korábban publikált adatsorból (Gutteridge et al., 2007), amely a BRENDA (Schomburg et al.,

2004) adatbázison alapul. A hipotézis predikciójával ellentétben a megfigyelt átlagos géntávolság a szabályozási interakcióban lévő génpárok között nem tér el szignifikánsan a véletlenszerűen várttól (p = 0,234, n = 19 génpár), ami a hatás hiányát sugallhatja. Ezután megvizsgáltuk, hogy vajon a kolinearitással potenciálisan ellentétes szelekciós hatású szabályzási interakciók befolyásolják-e a kolinearitás mértékét. Összehasonlítottuk a kolinearitás mértékét az ismert metabolit-szintű regulációs kapcsolatokat tartalmazó operonok és a többi operon esetében (11 illetve 59 operon). A kolinearitás mértékét randomizációs teszttel nem találtuk alacsonyabbnak a regulációs kapcsolatokat tartalmazó operonok esetében (p=0,089, lásd Módszerek). A két csoport közti expressziós szintbeli különbségekre kontrollálva is hasonló eredményt kaptunk (p = 0,35; aerob és anaerob környezetben is). Összefoglalva, nem találtunk bizonyítékot arra, hogy a metabolit-szintű szabályzási kapcsolatok befolyásolnák a génsorrend mértékét vagy csökkentenék a kolinearitás mértékét.

3.4 Diszkusszió

Eredményeink nyújtják az első szisztematikus bizonyítékot arra, hogy a metabolikus operonon belüli génsorrend nem véletlenszerű, hanem a kódolt enzimek funkcionális sorrendjével korrelál. A fenti mintázat azonban csak alacsonyan expresszálódó operonok esetében érvényes. Vizsgálataink nem támasztják alá a következő alternatív hipotéziseket: (1.) a kolinearitás az operonbeli gének 3' vég felé csökkenő expressziójával együtt magasabb steady-state fluxust eredményező adaptív jelleg; (2.) a kolinearitás a környezeti változások hatására gyakran ki- és bekapcsolódó operonok számára a bekapcsolás után átmeneti (nem steady-state) előnyt nyújt. Eredményeink ezekkel szemben támogatják azt a hipotézist, miszerint a kolinearitás az enzimek véletlenszerű elvesztéséből fakadó útvonalleállás idejét minimializálja. A konstitutívan, de alacsony szinten expresszált operonok erősebb szelekció alatt állnak, mivel az expresszió diszkrét, sztochasztikus esemény (Cai et al., 2006) és az előző expressziós eseményből származó enzimek elveszhetnek degradáció vagy sejtosztódás következtében, leállítva az anyagcsereútvonal működését. Az operonok kolineáris génsorrendje minimalizálhatja a leállás idejét. Ez az eredmény alátámasztja a sztochasztikus folyamatok fontosságát a sejtműködés szempontjából (Raj and Oudenaarden, 2008) , másrészt egy újabb példáját nyújtja annak, hogy egy génsorrendbeli mintázat a sztochasztikus folyamatokkal szembeni védekezés eredménye lehet (Batada and Hurst, 2007).

Az, hogy az operonon belüli génsorrend a zaj elleni adaptáció lehet, némileg ellentmondásban van az operonon belüli génsorrend magas fokú konzerváltságával, ami a sorrenden ható erős stabilizáló szelekció létét jelezheti (Rocha, 2006; Yang and Sze, 2008). Az *E. coli* metabolikus operonjait *Bacillus subtilis* operonszerkezeti és ortológia információival összevetve (lásd *Módszerek*) azt találtuk, hogy az *E. coli* operonok 70%-a nem feleltethető meg konzervált *B. subtilis* operonoknak, vagy az ortológ gének hiánya miatt, vagy azért, mert az ortológ gének nem voltak ugyanabban az operonban. Az operonok további 22%-ában, az ortológok relatív sorrendje teljesen megőrződött és csak öt operont (8%) találtunk, ahol a gének relatív pozíciója eltért. A ellentmondás feloldható lehet, ha a szelekció nem közvetlenül az operonok génsorrendjére, hanem az operonok kialakulására hat. Ha két azonos útvonalban levő enzimatikus gén egy mutáció révén operont alkot, az operon megtartására irányuló szelekció erősebb lehet, ha a gének sorrendje kolineáris. Az operon kialakulásakor ható „szelekciós szűrő" konzisztens a kialakulás utáni konzervált sorrenddel.

A fenti eredményeinket bemutató közlemény (Kovács et al., 2009) óta mások újabb, operonon belüli génsorrend-mintázatok létét is felvetették. 510 genom adatait összesítve, az operonokon belüli kisebb géntávolság alapján prediktált operonokban (Moreno-Hagelsieb and Collado-Vides, 2002) mutattak ki szignifikáns trendeket. Eszerint az esszenciális gének az operonok 5'-véghez közelebbi felében (5362 vs 5874 gén), míg a pszeudogének a 3'-véghez közelebbi félben helyezkednek el nagyobb gyakorisággal (Muro et al., 2011), amit az 5'-3' irányban csökkenő expressziós trenddel hoznak összefüggésbe. Bár a metaboliton keresztüli szabályzás hatását nem tudtuk kimutatni metabolikus operonainkon, egy ennél általánosabb regulációs hatást mutattak ki nemrég *E. coli*-ban és *Bacillus subtilis*-ben (Rubinstein et al., 2011): a saját operonját gátló (autorepresszor) transzkripciós faktor az 5'-vég első pozíciójában fordul elő szignifikánsan nagyobb arányban (pl. az *E. coli* esetében 51 esetből 27-szer). A negatív autoreguláció ilyen módon való hatékonyabbá tétele a traszkripciós reguláció optimalizálásának újabb példája lehet. A fenti példák alapján elmondható, hogy egyre pontosabb kép kezd kirajzolódni az operonon belüli szerveződésről. Annak megfejtéséhez, hogy miként zajlik ennek evolúciója, további összehasonlító vizsgálatok szükségesek.

4 Összegzés és kitekintés

Kutatásainkban az anyagcserehálózat génjeivel foglalkoztunk, kihasználva, hogy az anyagcsere talán a legrégebb óta kutatott és legrészletesebben feltérképezett sejtes alrendszer.
A 2. fejezetben az anyagcserehálózat, mint lehetséges eszköz jelent meg, ami segíthet a genotípus és a fenotípus közötti kapcsolat jobb megértésében. Munkánk során elsőként vizsgáltuk az anyagcsere nagyléptékű GI-térképét (Szappanos et al., 2011). Teszteltünk korábbi elméleti predikciókat a GI-k funkcionális modulokon belüli és közötti megoszlásával kapcsolatban és a modulok korlátozott monokromitását mutattuk ki. A negatív és pozitív GI-k egyaránt feldúsulnak a tradicionális funkcionális modulokban és a biokémiailag definiált kapcsolatokban, de az interakciók többsége ezeken kívül helyeszkedik el. Elsőként tesztelhettük nem csak egy nagyléptékű biokémiai modell (Szappanos et al. 2011), hanem az anyagcserehálózati és funkcionális genomikai adatokat integráló statisztikai modellek alkalmazhatóságát a GI-k predikciójára. Eredményeink azt mutatják, hogy jelenleg a GI-k egy jelentős részét sem az anyagcserehálózat szerkezetével megmagyarázni, sem más genomi tulajdonságok alapján nem tudjuk előrejelezni.
A 3. fejezetben egy adaptív hipotézist vizsgáltunk meg az anyagcsere működésére vonatkozóan, illetve az anyagcserére ható szelekció hatását a genom szerveződésére. Elsőként mutattuk ki szisztematikusan az operonon belüli génsorrend nem véletlenszerű mintázatát: az alacsonyan expresszálódó metabolikus operonokban a gének a kódolt enzimek reakciósorrendjét tükrözik (Kovács et al., 2009). Több alternatív adaptív hipotézist matematikai modellel vizsgáltunk, majd azok predikcióit empirikus *E. coli* adatokon teszteltük. Konklúziónk szerint a kolinearitás oka az alacsony expressziójú operonok esetében jelentkező sztochasztikus útvonalleállás idejének minimalizálása.
A nyitott kérdések megválaszolásában illetve a felállított hipotézisek közvetlen tesztelésében a kísérletes és modellezési technikák további fejlődése nyújthat segítséget. A továbblépési lehetőségek közül a következőkben három nagy témakört emelek ki: (i) új fenotípusos jellemzők kvantitatív és nagyléptékű kísérletes vizsgálhatósága(ii) összehasonlító vizsgálatok a mérések több fajra való kiterjesztésével (iii) a genommanipuláció fejlődése a szintetikus biológia eszközeinek alkalmazásával.
A molekuláris biológiai és analitikai módszerek fejlődésével egyre többféle fenotípusos bélyeg válik nagyléptékben és kvantitatív módon mérhetővé. Újabb funkcionális genomikai adatsorok segítséget jelenthetnek mind a GI-k jobb funkcionális megértésében, mind

predikciójában, például a 2. fejezet végén említett, a gének delécióját követő expressziós változások vizsgálatában (van Wageningen et al., 2010). A rátermettséget komponensei függvényeként kezelve, a mutációk egyes komponensekre tett hatása alapján a GI-k irányára és mértékére elméleti predikciókat tehetünk (Chiu et al. 2012). A jóslás azonban kvantitatív fenotípusos méréseket igényel: a komponensek rátermettségre tett hatását mutációk hiányában és az egyes mutációk hatását a komponensekre. Az utóbbi években már néhány esetben volt példa egyes és kettős deléciók a rátermettségtől különböző fenotípusra tett hatásának szisztematikus vizsgálatára, vagyis az adott fenotípus szintjén található GI-k feltérképezésére. FBA modellben az egyes és kettős deléciók hatását a modellben található összes reakció fluxusára (Snitkin & Segrè 2011), kísérletesen pedig az expressziós szintre (G. W. Carter et al. 2007), filamentáris növekedésre (Drees et al. 2005), fehérje-feltekeredés mértékére (Jonikas et al. 2009) és az endocitózisra (Burston et al. 2009) vizsgáltak. A különböző szinten mért GI-k általában csak kis átfedést mutatnak, sőt az irányuk is gyakran eltérhet, de funkcionális jelentőségüket jelzi, hogy gén funkció prediktálásához plusz információt hordoznak a rátermettség szintjén mért GI-hoz képest (Michaut & Gary D. Bader 2012). A jövőben a deléciós GI-k mellett másfajta mutációk közötti GI-k nagyléptékű vizsgálataira is számítani lehet, például overexpressziós könyvtárak (Winzeler et al. 1999) felhasználásával. Érdekes kérdés, hogy milyen mértékben és milyen funkcionális kapcsolatok esetén szolgálnak majd információval egymásról a funkcióvesztéses és nyeréses mutációk közti GI-k. A jövőre vár a regulációs kapcsolatok szerepének jobb megismerése is. A komplex regulációs hálózatokban végbemenő mutációk nem-intuitív válaszokat eredményezhetnek (Gjuvsland et al. 2007), lehetséges, hogy a GI-k egy része a regulációs hálózatok nem-lináris működésére vezethető vissza (Lehner 2011). A regulációs adatok FBA modellbe történő integrálása a modell fejlesztésének is az egyik legfontosabb iránya (Blazier and Papin, 2012). Fontos kísérletes áttörést jelent a sejtszintű mérések megjelenése (Taniguchi et al. 2010; Silander et al. 2012). A sejtszintű fehérje vagy mRNS abundancia-adatok közvetlenebbül teszik majd tesztelhetővé a génsorrend és a zaj összefüggését. Szintén a sejtszintű vizsgálatokkal lehet majd eldönteni, hogy vajon azonos genetikai és környezeti háttér mellett az eddig populáció szinten mért GI-k között vannak-e sejt-sejt szintű sztochasztikus különbségek, és milyen ennek mértéke (Burga et al., 2011).

A két modellélőlényben alkalmazott technológiákat más fajokra is kiterjesztve összehasonlító vizsgálatok válnak lehetségessé, ami segíti az operonsorrend és a GI evolúciójának jobb megértését. Több faj operonszerkezetének ismeretében az operon szerkezeti változásai

nagyobb felbontásban is vizsgálhatók lesznek, az ökológiai tényezők szerepével együtt. A negatív GI-ban levő génpárok együttes nyerése vagy vesztése az evolúció során véletlenszerű, ami azt támasztja alá, hogy legalább részben eltérő funkciót töltenek be és a kompenzáló interakció nem adaptív jelleg, hanem melléktermék (Harrison et al., 2007). Megfordítva, több faj nagyléptékű GI analízisével az evolúció során megőrződött génpárok közötti GI-k konzerváltságának vizsgálata is lehetővé vált (Dixon et al., 2008; Roguev et al., 2008; Tischler et al., 2008; Koch et al., 2012). A kb. 500 millió éve evolúciósan elkülönült *Schizosaccharomyces pombe* –vel összehasonlítva kiderül, hogy a pozitív és negatív GI-k annál konzerváltabbak, minél szorosabb közöttük a funkcionális kapcsolat: így ez míg a különböző sejtfunkciók esetén a GI-k kb. 20%-át jelenti, azonos komplexbe tartozó gének között 70%-ot (Ryan et al., 2012). Egy, a jövőben tesztelendő kérdés, hogy a konzerváltság mennyire függ össze a prediktálhatósággal.

Az anyagcsere génjeinek kísérletes manipulációja adhatja a legközvetlenebb választ a funkcionális összefüggésekre vonatkozó kérdésekre. Bár az adaptív hipotézisek tesztelésében az optimalizációs modellek és az összehasonlító módszer fontos szerepet játszanak, az ok-okozati összefüggést ebben az esetben is kísérletes beavatkozásokkal lehetne tesztelni. A szintetikus biológia eszközeinek alkalmazása egész anyagcserehálózatok kísérletes manipulálását teszi lehetővé (Boyle and Silver, 2012; Lee et al., 2012). A szintetikus biológia célja adott feladat végrehajtására alkalmas biológiai modulok, illetve ezekből összeálló komplex rendszerek mérnöki konstruálása vagy átalakítása (Andrianantoandro et al., 2006). Ez gyakran anyagcsereútvonalak *de novo* előállítását vagy optimalizálását jelenti adott metabolitok kontrollált szintű termelésére (Holtz and Keasling, 2010; Medema et al., 2012). A gének elhelyezkedésének az anyagcserére gyakorolt lehetséges hatásainak vizsgála ezért gyakorlati szempontból is fontossá vált. Például *Escherichia coli*-ban egy transzkripciós faktorból és célgénjéből álló szintetikus szabályzási kört használtak arra, hogy egymástól függetlenül tesztelhessék a két gén kromoszómális pozíciójának, orientációjának és egymástól való távolságának hatását az expressziós szintre, egyedül a kromoszomális helyzet esetében mérve szignifikáns hatást (Block et al., 2012). Egy másik vizsgálatban fluoreszcens fehérjéket kódoló operonokat konstruálva pedig nem csak az 5’ végtől való távolsággal csökkenő expressziós szintet mérték ki, hanem azt találták, hogy önmagában az operon méretének növelése képes megnöveli az első gén expresszióját, ami a transzkripció közben végbemenő transzláció nagyobb sebességére utalhat (Lim et al., 2011). A rendszerbiológia és a szintetikus biológia közötti kapcsolat kétirányú. Mint láttuk, a szintetikus genetikai áramkörök és

anyagcserehálózatok segítségével az elméleti modellek közvetlenül tesztelhetők (Lu et al., 2009), ugyanakkor a modellek predikciói irányt mutatnak a tervezésben és felvázolhatják a várható korlátokat (Oyarzún and Stan, 2012).

A modern biológia egyik nagy kihívása, hogy a biológiai rendszerek összetevőiből megjósolja az egész működését és evolúcióját. Vajon utóbbiak mennyire vezethetők vissza az egyes komponensek tulajdonságaira, illetve azok interakcióira? A rendszerbiológia egyik célja az molekuláris hálózatok emergens tulajdonságainak kutatása. Hangsúlyos kutatási program az adaptív „mérnöki elvek" (design principles) feltárása, melyek a szintetikus biológiában nyerhetnek gyakorlati alakalmazást (Alon, 2003; Alon, 2006). Munkánk második részében arra vázoltunk fel egy forgatókönyvet, hogyan járulhatott hozzá a genomszerveződés evolúciója az anyagcsereútak időbeli működésének „finomhangolásához". A két gén közötti genetikai interakció az egyik legrégebb óta ismert és kutatott emergens tulajdonság, de csak nemrég vált lehetővé a szisztematikus feltérképezése. Mostanra az is nyilvánvalóvá vált, hogy az egyszerű intuitív modellek csak az interakciók kis részére nyújtanak magyarázatot, és az interakciók komplexitását csak további kíséletes és elméleti innovációk segítségével lesz esélyünk megérteni.

5 Köszönetnyílvánítás

Elsősorban szeretnék köszönetet mondani témavezetőmnek, Papp Balázsnak, szakmai irányításáért, a munka minden fázisában nyújtott folyamatos segítségéért és támogatásáért.

Hálával tartozom Pál Csabának a kolinearitás ötletének felvetéséért és a doktori évek alatt nyújtott támogatásáért.

Köszönettel tartozom az Evolúciós Rendszerbiológia Csoport tagjainak, akikkel a doktori évek alatt együtt dolgozhattam.

A 2. fejezet eredményei egy nagyobb projekt részét képezve kerültek közlésre (Szappanos et al. 2011). Ezzel kapcsolatban köszönettel tartozom:
Charles Boone laboratóriumának (Toronto, Kanada) a kísérletes adatokért,
Szappanos Balázsnak (MTA SZBK, Biokémia Intézet) az FBA modellel végzett predikciókért és a közös munkáért,
Honti Ferencnek és Szamecz Bélának a közös munkáért

A 3. fejezet eredményei három szerző munkáján alapulnak (Kovács et al. 2009).
Ennek kapcsán köszönetet mondok Laurence Hurst-nek (Bath University) az együttműködésért, innovatív ötleteiért, és a sztochaszticitás szerepének felvetéséért.

6 Irodalomjegyzék

Alon, U., 2003. Biological networks: the tinkerer as an engineer. Science,301(5641), 1866-1867.

Alon, U., 2006. An introduction to systems biology: design principles of biological circuits (Vol. 10). Chapman & Hall/CRC.

Alpers, D.H., Tomkins, G.M., 1965. The order of induction and deinduction of the enzymes of the lactose operon in E. coli. Proc. Natl. Acad. Sci. U.S.A 53, 797–802.

Alpers, D.H., Tomkins, G.M., 1966. Sequential transcription of the genes of the lactose operon and its regulation by protein synthesis. J Biol Chem 241, 4434–43.

Altschul, S., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research 25, 3389–3402.

Andrianantoandro, E., Basu, S., Karig, D.K., Weiss, R., 2006. Synthetic biology: new engineering rules for an emerging discipline. Molecular Systems Biology 2.

Balaji, S., Babu, M.M., Iyer, L.M., Luscombe, N.M., Aravind, L., 2006. Comprehensive Analysis of Combinatorial Regulation using the Transcriptional Regulatory Network of Yeast. Journal of Molecular Biology 360, 213–227.

Bandyopadhyay, S., Kelley, R., Krogan, N.J., Ideker, T., 2008. Functional Maps of Protein Complexes from Quantitative Genetic Interaction Data. PLoS Comput Biol 4, e1000065.

Bandyopadhyay, S., Mehta, M., Kuo, D., Sung, M.-K., Chuang, R., Jaehnig, E.J., Bodenmiller, B., Licon, K., Copeland, W., Shales, M., Fiedler, D., Dutkowski, J., Guénolé, A., van Attikum, H., Shokat, K.M., Kolodner, R.D., Huh, W.-K., Aebersold, R., Keogh, M.-C., Krogan, N.J., Ideker, T., 2010. Rewiring of genetic networks in response to DNA damage. Science 330, 1385–1389.

Barabási, A.-L., Oltvai, Z.N., 2004. Network biology: understanding the cell's functional organization. Nature Reviews Genetics 5, 101–113.

Baryshnikova, A., Costanzo, M., Kim, Y., Ding, H., Koh, J., Toufighi, K., Youn, J.-Y., Ou, J., Luis, B.-J.S., Bandyopadhyay, S., Hibbs, M., Hess, D., Gingras, A.-C., Bader, G.D., Troyanskaya, O.G., Brown, G.W., Andrews, B., Boone, C., Myers, C.L., 2010. Quantitative analysis of fitness and genetic interactions in yeast on a genome scale. Nature Methods 7, 1017–1024.

Batada, N.N., Hurst, L.D., 2007. Evolution of chromosome organization driven by selection for reduced gene expression noise. Nature Genetics 39, 945–949.

Beasley, J.E., Planes, F.J., 2007. Recovering metabolic pathways via optimization. Bioinformatics 23, 92–98.

Blazier, A.S., Papin, J.A., 2012. Integration of expression data in genome-scale metabolic network reconstructions. Front Physiol 3, 299.

Block, D.H.S., Hussein, R., Liang, L.W., Lim, H.N., 2012. Regulatory consequences of gene translocation in bacteria. Nucl. Acids Res.

Boone, C., Bussey, H., Andrews, B.J., 2007. Exploring genetic interactions and networks with yeast. Nature Reviews Genetics 8, 437–449.

Boyle, P.M., Silver, P.A., 2012. Parts plus pipes: synthetic biology approaches to metabolic engineering. Metab. Eng. 14, 223–232.

Breiman, L., 2001. Random forests. Machine learning 45, 5–32.

Breitkreutz, B.-J., Stark, C., Reguly, T., Boucher, L., Breitkreutz, A., Livstone, M., Oughtred, R., Lackner, D.H., Bähler, J., Wood, V., Dolinski, K., Tyers, M., 2008. The BioGRID Interaction Database: 2008 update. Nucleic Acids Res 36, D637–640.

Brown, J.A., Sherlock, G., Myers, C.L., Burrows, N.M., Deng, C., Wu, H.I., McCann, K.E., Troyanskaya, O.G., Brown, J.M., 2006. Global analysis of gene function in yeast by quantitative phenotypic profiling. Mol Syst Biol 2, 2006.0001.

Bruggeman, F.J., Westerhoff, H.V., 2007. The nature of systems biology. Trends in Microbiology 15, 45–50.

Bundy, J.G., Papp, B., Harmston, R., Browne, R.A., Clayson, E.M., Burton, N., Reece, R.J., Oliver, S.G., Brindle, K.M., 2007. Evaluation of predicted network modules in yeast metabolism using NMR-based metabolite profiling. Genome Res. 17, 510–519.

Burga, A., Casanueva, M.O., Lehner, B., 2011. Predicting mutation outcome from early stochastic variation in genetic interaction partners. Nature 480, 250–253.

Burgard, A.P., Nikolaev, E.V., Schilling, C.H., Maranas, C.D., 2004. Flux coupling analysis of genome-scale metabolic network reconstructions. Genome Res 14, 301–312.

Cai, L., Friedman, N., Xie, X.S., 2006. Stochastic protein expression in individual cells at the single molecule level. Nature 440, 358–362.

Carneiro, M., Hartl, D.L., 2009. Colloquium Paper: Adaptive landscapes and protein evolution. Proceedings of the National Academy of Sciences 107, 1747–1751.

Caspi, R., Altman, T., Dreher, K., Fulcher, C.A., Subhraveti, P., Keseler, I.M., Kothari, A., Krummenacker, M., Latendresse, M., Mueller, L.A., Ong, Q., Paley, S., Pujar, A., Shearer, A.G., Travers, M., Weerasinghe, D., Zhang, P., Karp, P.D., 2011. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Research 40, D742–D753.

Castrillo, J.I., Oliver, S.G., 2004. Yeast as a touchstone in post-genomic research: strategies for integrative analysis in functional genomics. J. Biochem. Mol. Biol. 37, 93–106.

Chassagnole, C., Fell, D.A., Raïs, B., Kudla, B., Mazat, J.P., 2001. Control of the threonine-synthesis pathway in Escherichia coli: a theoretical and experimental approach. Biochem J 356, 433–444.

Chechik, G., Oh, E., Rando, O., Weissman, J., Regev, A., Koller, D., 2008. Activity motifs reveal principles of timing in transcriptional control of the yeast metabolic network. Nature Biotechnology 26, 1251–1259.

Chipman, K., Singh, A., 2009. Predicting genetic interactions with random walks on biological networks. BMC Bioinformatics 10, 17.

Chiu, H.-C., Marx, C.J., Segrè, D., 2012. Epistasis from functional dependence of fitness on underlying traits. Proc. R. Soc. B.

Chou, H.-H., Chiu, H.-C., Delaney, N.F., Segrè, D., Marx, C.J., 2011. Diminishing Returns Epistasis Among Beneficial Mutations Decelerates Adaptation. Science 332, 1190–1192.

Christie, K.R., Weng, S., Balakrishnan, R., Costanzo, M.C., Dolinski, K., Dwight, S.S., Engel, S.R., Feierbach, B., Fisk, D.G., Hirschman, J.E., Hong, E.L., Issel-Tarver, L., Nash, R., Sethuraman, A., Starr, B., Theesfeld, C.L., Andrada, R., Binkley, G., Dong, Q., Lane, C., Schroeder, M., Botstein, D., Cherry, J.M., 2004. Saccharomyces Genome Database (SGD) provides tools to identify and analyze sequences from Saccharomyces cerevisiae and related sequences from other organisms. Nucleic Acids Res. 32, D311–314.

Collins, S.R., Miller, K.M., Maas, N.L., Roguev, A., Fillingham, J., Chu, C.S., Schuldiner, M., Gebbia, M., Recht, J., Shales, M., Ding, H., Xu, H., Han, J., Ingvarsdottir, K., Cheng, B., Andrews, B., Boone, C., Berger, S.L., Hieter, P., Zhang, Z., Brown, G.W., Ingles, C.J., Emili, A., Allis, C.D., Toczyski, D.P., Weissman, J.S., Greenblatt, J.F., Krogan, N.J., 2007. Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. Nature 446, 806–810.

Collins, S.R., Schuldiner, M., Krogan, N.J., Weissman, J.S., 2006. A strategy for extracting and analyzing large-scale quantitative epistatic interaction data. Genome Biol 7, R63–R63.

Cordell, H.J., 2002. Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. Hum. Mol. Genet. 11, 2463–2468.

Cordell, H.J., 2009. Detecting gene–gene interactions that underlie human diseases. Nature Reviews Genetics 10, 392–404.

Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E.D., et al., 2010. The Genetic Landscape of a Cell. Science 327, 425–431.

Covert, M.W., Knight, E.M., Reed, J.L., Herrgard, M.J., Palsson, B.O., 2004. Integrating high-throughput and computational data elucidates bacterial networks. Nature 429, 92–96.

Covert, M.W., Schilling, C.H., Famili, I., Edwards, J.S., Goryanin, I.I., Selkov, E., Palsson, B.O., 2001. Metabolic modeling of microbial strains in silico. Trends Biochem. Sci. 26, 179–186.

Csardi, G., & Nepusz, T. 2006. The igraph software package for complex network research. *InterJournal, Complex Systems*, *1695*, 38.

Dandekar, T., Snel, B., Huynen, M., Bork, P., 1998. Conservation of gene order: a fingerprint of proteins that physically interact. Trends Biochem Sci 23, 324–8.

Decourty, L., Saveanu, C., Zemam, K., Hantraye, F., Frachon, E., Rousselle, J.-C., Fromont-Racine, M., Jacquier, A., 2008. Linking functionally related genes by sensitive and quantitative characterization of genetic interaction profiles. PNAS 105, 5821–5826.

Dekel, E., Alon, U., 2005. Optimality and evolutionary tuning of the expression level of a protein. Nature 436, 588–592.

Demerec, M., 1964. Clustering of Functionally Related Genes in Salmonella Typhimurium. Proc Natl Acad Sci U S A 51, 1057–60.

Dettman, J.R., Sirjusingh, C., Kohn, L.M., Anderson, J.B., 2007. Incipient speciation by divergent adaptation and antagonistic epistasis in yeast. Nature 447, 585–588.

Deutscher, D., Meilijson, I., Kupiec, M., Ruppin, E., 2006. Multiple knockout analysis of genetic robustness in the yeast metabolic network. Nature Genetics 38, 993–998.

Dixon, S.J., Costanzo, M., Baryshnikova, A., Andrews, B., Boone, C., 2009. Systematic Mapping of Genetic Interaction Networks. Annual Review of Genetics 43, 601–625.

Dixon, S.J., Fedyshyn, Y., Koh, J.L.Y., Prasad, T.S.K., Chahwan, C., Chua, G., Toufighi, K., Baryshnikova, A., Hayles, J., Hoe, K.-L., Kim, D.-U., Park, H.-O., Myers, C.L., Pandey, A., Durocher, D., Andrews, B.J., Boone, C., 2008. Significant conservation of synthetic lethal genetic interaction networks between distantly related eukaryotes. PNAS 105, 16653–16658.

Drummond, D.A., Bloom, J.D., Adami, C., Wilke, C.O., Arnold, F.H., 2005. Why highly expressed proteins evolve slowly. PNAS 102, 14338–14343.

Duarte, N.C., Herrgård, M.J., Palsson, B.Ø., 2004. Reconstruction and validation of Saccharomyces cerevisiae iND750, a fully compartmentalized genome-scale metabolic model. Genome Res. 14, 1298–1309.

Edwards, J.S., Ibarra, R.U., Palsson, B.O., 2001. In silico predictions of Escherichia coli metabolic capabilities are consistent with experimental data. Nature Biotechnology 19, 125–130.

Fani, R., Brilli, M., Lio, P., 2005. The origin and evolution of operons: the piecewise building of the proteobacterial histidine operon. J Mol Evol 60, 378–90.

Förster, J., Famili, I., Palsson, B.O., Nielsen, J., 2003. Large-scale evaluation of in silico gene deletions in Saccharomyces cerevisiae. OMICS 7, 193–202.

Garfinkel, D., Garfinkel, L., Pring, M., Green, S.B., Chance, B., 1970. Computer Applications to Biochemical Kinetics. Annual Review of Biochemistry 39, 473–498.

Giaever, G., Chu, A.M., Ni, L., Connelly, C., Riles, L., et al., 2002. Functional profiling of the Saccharomyces cerevisiae genome. Nature 418, 387–391.

Glansdorff, N., 1999. On the origin of operons and their possible role in evolution toward thermophily. J Mol Evol 49, 432–8.

Goldberg, D. S., & Roth, F. P., 2003. Assessing experimentally derived interactions in a small world. Proceedings of the National Academy of Sciences, 100(8), 4372-4376.

Gutteridge, A., Kanehisa, M., Goto, S., 2007. Regulation of metabolic networks by small molecule metabolites. BMC Bioinformatics 8, 88.

Hacker, J., Carniel, E., 2001. Ecological fitness, genomic islands and bacterial pathogenicity. A Darwinian view of the evolution of microbes. EMBO Rep 2, 376–81.

Harrison, R., Papp, B., Pál, C., Oliver, S.G., Delneri, D., 2007. Plasticity of genetic interactions in metabolic networks of yeast. Proceedings of the National Academy of Sciences 104, 2307–2312.

Hartman, J.L., Garvik, B., Hartwell, L., 2001. Principles for the Buffering of Genetic Variation. Science 291, 1001–1004.

He, X., Qian, W., Wang, Z., Li, Y., Zhang, J., 2010. Prevalent positive epistasis in Escherichia coli and Saccharomyces cerevisiae metabolic networks. Nature Genetics 42, 272–276.

Heinrich, R., Klipp, E., 1996. Control Analysis of Unbranched Enzymatic Chains in States of Maximal Activity. Journal of Theoretical Biology 182, 243–252.

Heinrich, R., Rapoport, T.A., 1974. A Linear Steady-State Treatment of Enzymatic Chains. European Journal of Biochemistry 42, 89–95.

Holme, P., Huss, M., Lee, S.H., 2011. Atmospheric Reaction Systems as Null-Models to Identify Structural Traces of Evolution in Metabolism. PLoS ONE 6, e19759.

Holtz, W.J., Keasling, J.D., 2010. Engineering Static and Dynamic Control of Synthetic Pathways. Cell 140, 19–23.

Hoops, S., Sahle, S., Gauges, R., Lee, C., Pahle, J., Simus, N., Singhal, M., Xu, L., Mendes, P., Kummer, U., 2006. COPASI—a COmplex PAthway SImulator. Bioinformatics 22, 3067–3074.

Horn, T., Sandmann, T., Fischer, B., Axelsson, E., Huber, W., Boutros, M., 2011. Mapping of signaling networks through synthetic genetic interaction analysis by RNAi. Nature Methods 8, 341–346.

Hurst, L.D., Pal, C., Lercher, M.J., 2004. The evolutionary dynamics of eukaryotic gene order. Nat Rev Genet 5, 299–310.

Huttenhower, C., Hibbs, M., Myers, C., Troyanskaya, O.G., 2006. A scalable method for integration and functional analysis of multiple microarray datasets. Bioinformatics 22, 2890–2897.

Huynen, M.A., Bork, P., 1998. Measuring genome evolution. Proc Natl Acad Sci U S A 95, 5849–56.

Ibarra, R. U., Edwards, J. S., & Palsson, B. O., 2002. Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. Nature, 420(6912), 186-189.

Itoh, T., Takemoto, K., Mori, H., Gojobori, T., 1999. Evolutionary instability of operon structures disclosed by sequence comparisons of complete microbial genomes. Mol Biol Evol 16, 332–346.

Jacob, F., Monod, J., 1961. Genetic regulatory mechanisms in the synthesis of proteins. J Mol Biol 3, 318–56.

Jamshidi, N., Palsson, B.Ø., 2008. Formulating genome-scale kinetic models in the post-genome era. Molecular Systems Biology 4.

Janga, S.C., Salgado, H., Martínez-Antonio, A., 2009. Transcriptional regulation shapes the organization of genes on bacterial chromosomes. Nucl. Acids Res. 37, 3680–3688.

Jeong, H., Mason, S.P., Barabási, A.L., Oltvai, Z.N., 2001. Lethality and centrality in protein networks. Nature 411, 41–42.

Jeong, H., Tombor, B., Albert, R., Oltvai, Z.N., Barabási, A.-L., 2000. The large-scale organization of metabolic networks. Nature 407, 651–654.

Joshi, A., Palsson, B.O., 1989. Metabolic dynamics in the human red cell. Part I--A comprehensive kinetic model. J. Theor. Biol. 141, 515–528.

Kacser, H., Burns, J., 1973. The control of flux. Symp. Soc. Exp. Biol. 27, 65–104.

Karp, P.D., Caspi, R., 2011. A Survey of Metabolic Databases Emphasizing the MetaCyc Family. Arch Toxicol 85, 1015–1033.

Keseler, I.M., Bonavides-Martínez, C., Collado-Vides, J., Gama-Castro, S., Gunsalus, R.P., Johnson, D.A., Krummenacker, M., Nolan, L.M., Paley, S., Paulsen, I.T., Peralta-Gil, M., Santos-Zavaleta, A., Shearer, A.G., Karp, P.D., 2009. EcoCyc: a comprehensive view of Escherichia coli biology. Nucleic Acids Res 37, D464–470.

Khan, A.I., Dinh, D.M., Schneider, D., Lenski, R.E., Cooper, T.F., 2011. Negative Epistasis Between Beneficial Mutations in an Evolving Bacterial Population. Science 332, 1193–1196.

Klipp, E., Heinrich, R., Holzhutter, H.G., 2002. Prediction of temporal gene expression. Metabolic opimization by re-distribution of enzyme activities. Eur J Biochem 269, 5406–13.

Koch, E., Costanzo, M., Bellay, J., Deshpande, R., Chatfield-Reed, K., Chua, G., D'Urso, G., Andrews, B., Boone, C., Myers, C., 2012. Conserved rules govern genetic interaction degree across species. Genome Biology 13, R57.

Koh, J.L.Y., Ding, H., Costanzo, M., Baryshnikova, A., Toufighi, K., Bader, G.D., Myers, C.L., Andrews, B.J., Boone, C., 2009. DRYGIN: a database of quantitative genetic interaction networks in yeast. Nucleic Acids Research 38, D502–D507.

Kolesov, G., Wunderlich, Z., Laikova, O.N., Gelfand, M.S., Mirny, L.A., 2007. How gene order is influenced by the biophysics of transcription regulation. PNAS 104, 13948–13953.

Kolsto, A.B., 1997. Dynamic bacterial genome organization. Mol Microbiol 24, 241–8.

Kovács, K., Hurst, L.D., Papp, B., 2009. Stochasticity in Protein Levels Drives Colinearity of Gene Order in Metabolic Operons of Escherichia coli. PLoS Biol 7, e1000115.

Kuepfer, L., Sauer, U., Blank, L.M., 2005. Metabolic functions of duplicate genes in Saccharomyces cerevisiae. Genome Res. 15, 1421–1430.

Lathe, W.C., Snel, B., Bork, P., 2000. Gene context conservation of a higher order than operons. Trends Biochem Sci 25, 474–9.

Lawrence, J.G., 2003. Gene organization: selection, selfishness, and serendipity. Annu Rev Microbiol 57, 419–40.

Lawrence, J.G., Roth, J.R., 1996. Selfish operons: horizontal transfer may drive the evolution of gene clusters. Genetics 143, 1843–60.

Lee, J.W., Na, D., Park, J.M., Lee, J., Choi, S., Lee, S.Y., 2012. Systems metabolic engineering of microorganisms for natural and non-natural chemicals. Nature Chemical Biology 8, 536–546.

Lehner, B., 2007. Modelling genotype–phenotype relationships and human disease with genetic interaction networks. J Exp Biol 210, 1559–1566.

Lehner, B., Crombie, C., Tischler, J., Fortunato, A., Fraser, A.G., 2006. Systematic mapping of genetic interactions in Caenorhabditis elegans identifies common modifiers of diverse signaling pathways. Nature Genetics 38, 896–903.

Lewontin, R.C., 1966. On the Measurement of Relative Variability.

Liaw, A., Wiener, M., 2002. Classification and Regression by randomForest. Resampling Methods in R: The boot Package 18.

Lim, H.N., Lee, Y., Hussein, R., 2011. Fundamental relationship between operon organization and gene expression. Proceedings of the National Academy of Sciences.

Ling, X., He, X., Xin, D., 2009. Detecting gene clusters under evolutionary constraint in a large number of genomes. Bioinformatics 25, 571–577.

Lu, T.K., Khalil, A.S., Collins, J.J., 2009. Next-generation synthetic gene networks. Nature Biotechnology 27, 1139–1150.

Lynch, M., 2007. The frailty of adaptive hypotheses for the origins of organismal complexity. Proceedings of the National Academy of Sciences 104, 8597–8604.

Mahadevan, R., Palsson, B.O., 2005. Properties of Metabolic Networks: Structure versus Function. Biophys J 88, L07–L09.

Mani, R., St.Onge, R.P., Hartman, J.L., Giaever, G., Roth, F.P., 2008. Defining genetic interaction. PNAS 105, 3461–3466.

Marashi, S.-A., Bockmayr, A., 2011. Flux coupling analysis of metabolic networks is sensitive to missing reactions. Biosystems 103, 57–66.

Maxwell, C., Moreno, V., Solé, X., Gómez, L., Hernández, P., Urruticoechea, A., Pujana, M., 2008. Genetic interactions: the missing links for a better understanding of cancer susceptibility, progression and treatment. Molecular Cancer 7, 4.

Medema, M.H., Raaphorst, R. van, Takano, E., Breitling, R., 2012. Computational tools for the synthetic design of biochemical pathways. Nature Reviews Microbiology 10, 191–202.

Mewes, H.W., 2004. MIPS: analysis and annotation of proteins from whole genomes. Nucleic Acids Research 32, 41D–44.

Mo, M.L., Palsson, B.O., Herrgård, M.J., 2009. Connecting extracellular metabolomic measurements to intracellular flux states in yeast. BMC Syst Biol 3, 37.

Moore, J.H., Williams, S.M., 2005. Traversing the conceptual divide between biological and statistical epistasis: systems biology and a more modern synthesis. BioEssays 27, 637–646.

Moreno-Hagelsieb, G., Collado-Vides, J., 2002. A powerful non-homology method for the prediction of operons in prokaryotes. Bioinformatics 18, S329–S336.

Moreno-Hagelsieb, G., Trevino, V., Perez-Rueda, E., Smith, T.F., Collado-Vides, J., 2001. Transcription unit conservation in the three domains of life: a perspective from Escherichia coli. Trends Genet 17, 175–7.

Mori, H., 2004. From the sequence to cell modeling: comprehensive functional genomics in Escherichia coli. J. Biochem. Mol. Biol. 37, 83–92.

Muro, E.M., Mah, N., Moreno-Hagelsieb, G., Andrade-Navarro, M.A., 2011. The pseudogenes of Mycobacterium leprae reveal the functional relevance of gene order within operons. Nucleic Acids Research 39, 1732 –1738.

Mushegian, A.R., Koonin, E.V., 1996. Gene order is not conserved in bacterial evolution. Trends Genet 12, 289–90.

Musso, G., Costanzo, M., Huangfu, M., Smith, A.M., Paw, J., Luis, B.-J.S., Boone, C., Giaever, G., Nislow, C., Emili, A., Zhang, Z., 2008. The extensive and condition-dependent nature of epistasis among whole-genome duplicates in yeast. Genome Res. 18, 1092–1099.

Nishizaki, T., Tsuge, K., Itaya, M., Doi, N., Yanagawa, H., 2007. Metabolic engineering of carotenoid biosynthesis in Escherichia coli by ordered gene assembly in Bacillus subtilis. Appl. Environ. Microbiol 73, 1355–1361.

Notebaart, R.A., Teusink, B., Siezen, R.J., Papp, B., 2008. Co-Regulation of Metabolic Genes Is Better Explained by Flux Coupling Than by Network Distance. PLoS Comput Biol 4, e26.

Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., Kanehisa, M., 1999. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 27, 29–34.

Okuda, S., Kawashima, S., Kobayashi, K., Ogasawara, N., Kanehisa, M., Goto, S., 2007. Characterization of relationships between transcriptional units and operon structures in Bacillus subtilis and Escherichia coli. BMC Genomics 8, 48.

Omelchenko, M.V., Makarova, K.S., Wolf, Y.I., Rogozin, I.B., Koonin, E.V., 2003. Evolution of mosaic operons by horizontal gene transfer and gene displacement in situ. Genome Biol 4, R55.

Onge, R.P.S., Mani, R., Oh, J., Proctor, M., Fung, E., Davis, R.W., Nislow, C., Roth, F.P., Giaever, G., 2007. Systematic pathway analysis using high-resolution fitness profiling of combinatorial gene deletions. Nature Genetics 39, 199–206.

Osbourn, A.E., Field, B., 2009. Operons. Cellular and Molecular Life Sciences 66, 3755–3775.

Ouzounis, C.A., Karp, P.D., 2000. Global Properties of the Metabolic Map of Escherichia coli. Genome Res. 10, 568–576.

Overbeek, R., Fonstein, M., D'Souza, M., Pusch, G.D., Maltsev, N., 1999. The use of gene clusters to infer functional coupling. Proc Natl Acad Sci U S A 96, 2896–901.

Oyarzún, D.A., Stan, G.-B.V., 2012. Synthetic gene circuits for metabolic control: design trade-offs and constraints. J. R. Soc. Interface.

Pál, C., Papp, B., Hurst, L.D., 2001. Highly Expressed Genes in Yeast Evolve Slowly. Genetics 158, 927–931.

Pal, C., Papp, B., Lercher, M.J., 2005. Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. Nat Genet 37, 1372–1375.

Paladugu, S., Zhao, S., Ray, A., Raval, A., 2008. Mining protein networks for synthetic genetic interactions. BMC Bioinformatics 9, 426.

Pan, X., Yuan, D.S., Xiang, D., Wang, X., Sookhai-Mahadeo, S., Bader, J.S., Hieter, P., Spencer, F., Boeke, J.D., 2004. A Robust Toolkit for Functional Profiling of the Yeast Genome. Molecular Cell 16, 487–496.

Papin, J.A., Price, N.D., Wiback, S.J., Fell, D.A., Palsson, B.O., 2003. Metabolic pathways in the post-genome era. Trends in Biochemical Sciences 28, 250–258.

Papin, J.A., Reed, J.L., Palsson, B.O., 2004. Hierarchical thinking in network biology: the unbiased modularization of biochemical networks. Trends in Biochemical Sciences 29, 641–647.

Papp, B., Pal, C., Hurst, L.D., 2003. Dosage sensitivity and the evolution of gene families in yeast. Nature 424, 194–7.

Papp, B., Teusink, B., Notebaart, R.A., 2009. A critical view of metabolic network adaptations. HFSP J 3, 24–35.

Parter, M., Kashtan, N., Alon, U., 2007. Environmental variability and modularity of bacterial metabolic networks. BMC Evolutionary Biology 7, 169.

Pfeiffer, T., Soyer, O.S., Bonhoeffer, S., 2005. The Evolution of Connectivity in Metabolic Networks. PLoS Biol 3, e228.

Poelwijk, F.J., Kiviet, D.J., Weinreich, D.M., Tans, S.J., 2007. Empirical fitness landscapes reveal accessible evolutionary paths. Nature 445, 383–386.

Price, M.N., Arkin, A.P., Alm, E.J., 2006. The Life-Cycle of Operons. PLoS Genet 2, e96.

Price, N.D., Reed, J.L., Palsson, B.Ø., 2004. Genome-scale models of microbial cells: evaluating the consequences of constraints. Nature Reviews Microbiology 2, 886–897.

Pu, S., Wong, J., Turner, B., Cho, E., Wodak, S.J., 2009. Up-to-date catalogues of yeast protein complexes. Nucleic Acids Res. 37, 825–831.

Qi, Y., Suhail, Y., Lin, Y., Boeke, J.D., Bader, J.S., 2008. Finding friends and enemies in an enemies-only network: A graph diffusion kernel for predicting novel genetic interactions and co-complex membership from yeast genetic interactions. Genome Res 18, 1991–2004.

Quinn, G.P., Keough, M.J., 2002. Experimental Design and Data Analysis for Biologists. Cambridge University Press.

R Development Core Team, 2009. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Raj, A., Oudenaarden, A. van, 2008. Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. Cell 135, 216–226.

Ravasz, E., Somera, A.L., Mongru, D.A., Oltvai, Z.N., Barabási, A.-L., 2002. Hierarchical Organization of Modularity in Metabolic Networks. Science 297, 1551–1555.

Reams, A.B., Neidle, E.L., 2004. Selection for gene clustering by tandem duplication. Annu Rev Microbiol 58, 119–42.

Rocha, E.P., 2006. Inference and analysis of the relative stability of bacterial chromosomes. Mol Biol Evol 23, 513–22.

Rocha, E.P.C., 2008. The Organization of the Bacterial Genome. Annual Review of Genetics 42, 211–233.

Rogozin, I.B., Makarova, K.S., Murvai, J., Czabarka, E., Wolf, Y.I., Tatusov, R.L., Szekely, L.A., Koonin, E.V., 2002. Connected gene neighborhoods in prokaryotic genomes. Nucleic Acids Res 30, 2212–23.

Roguev, A., Bandyopadhyay, S., Zofall, M., Zhang, K., Fischer, T., Collins, S.R., Qu, H., Shales, M., Park, H.-O., Hayles, J., Hoe, K.-L., Kim, D.-U., Ideker, T., Grewal, S.I., Weissman, J.S., Krogan, N.J., 2008. Conservation and Rewiring of Functional Modules Revealed by an Epistasis Map in Fission Yeast. Science 322, 405–410.

Rubinstein, N.D., Zeevi, D., Oren, Y., Segal, G., Pupko, T., 2011. The Operonic Location of Auto-transcriptional Repressors Is Highly Conserved in Bacteria. Mol Biol Evol.

Ryan, C., Greene, D., Cagney, G., Cunningham, P., 2010. Missing value imputation for epistatic MAPs. BMC Bioinformatics 11, 197.

Ryan, C.J., Roguev, A., Patrick, K., Xu, J., Jahari, H., Tong, Z., Beltrao, P., Shales, M., Qu, H., Collins, S.R., Kliegman, J.I., Jiang, L., Kuo, D., Tosti, E., Kim, H.-S., Edelmann, W., Keogh, M.-C., Greene, D., Tang, C., Cunningham, P., Shokat, K.M., Cagney, G., Svensson, J.P., Guthrie, C., Espenshade, P.J., Ideker, T., Krogan, N.J., 2012. Hierarchical Modularity and the Evolution of Genetic Interactomes across Species. Molecular Cell 46, 691–704.

Salgado, H., Moreno-Hagelsieb, G., Smith, T.F., Collado-Vides, J., 2000. Operons in Escherichia coli: Genomic analyses and predictions. PNAS 97, 6652–6657.

Sasson, V., Shachrai, I., Bren, A., Dekel, E., Alon, U., 2012. Mode of regulation and the insulation of bacterial gene expression. Mol. Cell 46, 399–407.

Savageau, M.A., 1977. Design of molecular control mechanisms and the demand for gene expression. Proc Natl Acad Sci U S A 74, 5647–51.

Schomburg, I., Chang, A., Ebeling, C., Gremse, M., Heldt, C., Huhn, G., Schomburg, D., 2004. BRENDA, the enzyme database: updates and major new developments. Nucleic Acids Res. 32, D431–433.

Schuldiner, M., Collins, S.R., Thompson, N.J., Denic, V., Bhamidipati, A., Punna, T., Ihmels, J., Andrews, B., Boone, C., Greenblatt, J.F., Weissman, J.S., Krogan, N.J., 2005. Exploration of the Function and Organization of the Yeast Early Secretory Pathway through an Epistatic Miniarray Profile. Cell 123, 507–519.

Segre, D., DeLuna, A., Church, G.M., Kishony, R., 2005. Modular epistasis in yeast metabolism. Nat Genet 37, 77–83.

Shinar, G., Dekel, E., Tlusty, T., Alon, U., 2006. Rules for biological regulation based on error minimization. Proc. Natl. Acad. Sci. U.S.A. 103, 3999–4004.

Sierro, N., Makita, Y., de Hoon, M., Nakai, K., 2008. DBTBS: a database of transcriptional regulation in Bacillus subtilis containing upstream intergenic conservation information. Nucleic Acids Res 36, D93–D96.

Sing, T., Sander, O., Beerenwinkel, N., Lengauer, T., 2005. ROCR: visualizing classifier performance in R. Bioinformatics 21, 3940–3941.

Smallbone, K., Simeonidis, E., Swainston, N., Mendes, P., 2010. Towards a genome-scale kinetic model of cellular metabolism. BMC Systems Biology 4, 6.

Stark, C., Breitkreutz, B.-J., Reguly, T., Boucher, L., Breitkreutz, A., Tyers, M., 2006. BioGRID: a general repository for interaction datasets. Nucleic Acids Res 34, D535–539.

Stelling, J., 2004. Mathematical models in microbial systems biology. Current Opinion in Microbiology 7, 513–518.

Steuer, R., Junker, B.H., 2008. Computational Models of Metabolism: Stability and Regulation in Metabolic Networks, in: Rice, S.A. (Ed.), Advances in Chemical Physics. John Wiley & Sons, Inc., pp. 105–251.

Swain, P.S., 2004. Efficient Attenuation of Stochasticity in Gene Expression Through Post-transcriptional Control. Journal of Molecular Biology 344, 965–976.

Szappanos, B., Kovács, K., Szamecz, B., Honti, F., Costanzo, M., Baryshnikova, A., Gelius-Dietrich, G., Lercher, M.J., Jelasity, M., Myers, C.L., Andrews, B.J., Boone, C., Oliver, S.G., Pál, C., Papp, B., 2011. An integrated approach to characterize genetic interaction networks in yeast metabolism. Nature Genetics 43, 656–662.

Szathmáry, E., 1993. Do deleterious mutations act synergistically? Metabolic control theory provides a partial answer. Genetics 133, 127–132.

Tamames, J., Casari, G., Ouzounis, C., Valencia, A., 1997. Conserved clusters of functionally related genes in two bacterial genomes. J Mol Evol 44, 66–73.

Tamames, J., Gonzalez-Moreno, M., Mingorance, J., Valencia, A., Vicente, M., 2001. Bringing gene order into bacterial shape. Trends Genet 17, 124–6.

Tănase-Nicola, S., ten Wolde, P.R., 2008. Regulatory Control and the Costs and Benefits of Biochemical Noise. PLoS Comput Biol 4, e1000125.

Taniguchi, Y., Choi, P.J., Li, G.-W., Chen, H., Babu, M., Hearn, J., Emili, A., Xie, X.S., 2010. Quantifying E. coli Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells. Science 329, 533–538.

Tischler, J., Lehner, B., Fraser, A.G., 2008. Evolutionary plasticity of genetic interaction networks. Nature Genetics 40, 390–391.

Tong, A.H.Y., Boone, C., 2006. Synthetic genetic array analysis in Saccharomyces cerevisiae. Methods Mol. Biol. 313, 171–192.

Tong, A.H.Y., Evangelista, M., Parsons, A.B., Xu, H., Bader, G.D., Pagé, N., Robinson, M., Raghibizadeh, S., Hogue, C.W.V., Bussey, H., Andrews, B., Tyers, M., Boone, C., 2001. Systematic Genetic Analysis with Ordered Arrays of Yeast Deletion Mutants. Science 294, 2364–2368.

Tong, A.H.Y., Lesage, G., Bader, G.D., Ding, H., Xu, H., Xin, X., Young, J., Berriz, G.F., Brost, R.L., Chang, M., Chen, Y., Cheng, X., Chua, G., Friesen, H., Goldberg, D.S., Haynes, J., Humphries, C., He, G., Hussein, S., Ke, L., Krogan, N., Li, Z., Levinson, J.N., Lu, H., Ménard, P., Munyana, C., Parsons, A.B., Ryan, O., Tonikian, R., Roberts, T., Sdicu, A.-M., Shapiro, J., Sheikh, B., Suter, B., Wong, S.L., Zhang, L.V., Zhu, H., Burd, C.G., Munro, S., Sander, C., Rine, J., Greenblatt, J., Peter, M., Bretscher, A., Bell, G., Roth, F.P., Brown, G.W., Andrews, B., Bussey, H., Boone, C., 2004. Global mapping of the yeast genetic interaction network. Science 303, 808–813.

Tucker, C.L., Fields, S., 2003. Lethal combinations. Nature Genetics 35, 204–205.

Ulitsky, I., Krogan, N.J., Shamir, R., 2009. Towards accurate imputation of quantitative genetic interactions. Genome Biol 10, R140.

Ullmann, A., Joseph, E., Danchin, A., 1979. Cyclic AMP as a modulator of polarity in polycistronic transcriptional units. Proc Natl Acad Sci U S A 76, 3194–3197.

van Wageningen, S., Kemmeren, P., Lijnzaad, P., Margaritis, T., Benschop, J.J., de Castro, I.J., van Leenen, D., Groot Koerkamp, M.J.A., Ko, C.W., Miles, A.J., Brabers, N., Brok, M.O., Lenstra, T.L., Fiedler, D., Fokkens, L., Aldecoa, R., Apweiler, E., Taliadouros, V., Sameith, K., van de Pasch, L.A.L., van Hooff, S.R., Bakker, L.V., Krogan, N.J., Snel, B., Holstege, F.C.P., 2010. Functional overlap and regulatory links shape genetic interactions between signaling pathways. Cell 143, 991–1004.

VanderSluis, B., Bellay, J., Musso, G., Costanzo, M., Papp, B., Vizeacoumar, F.J., Baryshnikova, A., Andrews, B., Boone, C., Myers, C.L., 2010. Genetic interactions reveal the evolutionary trajectories of duplicate genes. Molecular Systems Biology 6.

Visser, J.A.G.M. de, Elena, S.F., 2007. The evolution of sex: empirical insights into the roles of epistasis and drift. Nature Reviews Genetics 8, 139–149.

Wagner, A., 2000. Robustness against mutations in genetic networks of yeast. Nature Genetics 24, 355–361.

Warnes, G. R., Bolker, B., Lumley, T., & Johnson, R. C. (2006). gmodels: Various R programming tools for model fittingWarren, P.B., ten Wolde, P.R., 2004. Statistical Analysis of the Spatial Distribution of Operons in the Transcriptional Regulation Network of Escherichia coli. Journal of Molecular Biology 342, 1379–1390.

Watanabe, H., Mori, H., Itoh, T., Gojobori, T., 1997. Genome plasticity as a paradigm of eubacteria evolution. J Mol Evol 44 Suppl 1, S57–64.

Wessely, F., Bartl, M., Guthke, R., Li, P., Schuster, S., Kaleta, C., 2011. Optimal regulatory strategies for metabolic pathways in Escherichia coli depending on protein costs. Molecular Systems Biology 7.

Westerhoff, H.V., Palsson, B.O., 2004. The evolution of molecular biology into systems biology. Nature Biotechnology 22, 1249–1252.

Wilcox, R.R., 2005. Introduction to Robust Estimation and Hypothesis Testing. Elsevier academic press.

Winzeler, E.A., Shoemaker, D.D., Astromoff, A., Liang, H., Anderson, K., et al., 1999. Functional Characterization of the S. cerevisiae Genome by Gene Deletion and Parallel Analysis. Science 285, 901–906.

Wolf, Y.I., Rogozin, I.B., Kondrashov, A.S., Koonin, E.V., 2001. Genome alignment, evolution of prokaryotic genome organization, and prediction of gene function using genomic context. Genome Res 11, 356–72.

Wong, S.L., Zhang, L.V., Tong, A.H.Y., Li, Z., Goldberg, D.S., King, O.D., Lesage, G., Vidal, M., Andrews, B., Bussey, H., Boone, C., Roth, F.P., 2004. Combining biological networks to predict genetic interactions. Proceedings of the National Academy of Sciences of the United States of America 101, 15682–15687.

Wright, S., 1932. The roles of mutation, inbreeding, crossbreeding, and selection in evolution. Proceedings of the Sixth International Congress on Genetics.

Xie, G., Keyhani, N.O., Bonner, C.A., Jensen, R.A., 2003. Ancient origin of the tryptophan operon and the dynamics of evolutionary change. Microbiol Mol Biol Rev 67, 303–42, table of contents.

Yang, Q., Sze, S.-H., 2008. Large-scale analysis of gene clustering in bacteria. Genome Res. 18, 949–956.

Yang, W.S., Stockwell, B.R., 2008. Synthetic Lethal Screening Identifies Compounds Activating Iron-Dependent, Nonapoptotic Cell Death in Oncogenic-RAS-Harboring Cancer Cells. Chemistry & Biology 15, 234–245.

Ye, P., Peyser, B.D., Pan, X., Boeke, J.D., Spencer, F.A., Bader, J.S., 2005. Gene function prediction from congruent synthetic lethal interactions in yeast. Molecular Systems Biology 1, E1–E12.

Yu, H., Braun, P., Yildirim, M.A., Lemmens, I., Venkatesan, K., Sahalie, J., Hirozane-Kishikawa, T., Gebreab, F., Li, N., Simonis, N., Hao, T., Rual, J.-F., Dricot, A., Vazquez, A., Murray, R.R., Simon, C., Tardivo, L., Tam, S., Svrzikapa, N., Fan, C., de Smet, A.-S., Motyl, A., Hudson, M.E., Park, J., Xin, X., Cusick, M.E., Moore, T., Boone, C., Snyder, M., Roth, F.P., Barabási, A.-L., Tavernier, J., Hill, D.E., Vidal, M., 2008. High-quality binary protein interaction map of the yeast interactome network. Science 322, 104–110.

Zaslaver, A., Mayo, A.E., Rosenberg, R., Bashkin, P., Sberro, H., Tsalyuk, M., Surette, M.G., Alon, U., 2004. Just-in-time transcription program in metabolic pathways. Nature Genetics 36, 486–491.

Zhang, H., Yin, Y., Olman, V., Xu, Y., 2012. Genomic Arrangement of Regulons in Bacterial Genomes. PLoS One 7.

Zuk, O., Hechter, E., Sunyaev, S.R., Lander, E.S., 2012. The mystery of missing heritability: Genetic interactions create phantom heritability. PNAS 109, 1193–1198.

Összefoglalás

Az utóbbi években elérhetővé vált genomikai és nagy léptékű fenotípusos adatok analízise és matematikai modellbe integrálása révén elsőként vált lehetővé a nukleotidszekvenciától a látható fenotípusos jellegekig vezető út szisztematikus, nagy léptékű feltérképezése. Mivel az anyagcsere talán a legrégebb óta kutatott és legrészletesebben feltérképezett sejtes alrendszer, kézenfekvő választás lehet a fenti cél eléréséhez. Disszertációm két munkát tárgyal (Szappanos et al., 2011; Kovács et al., 2009), melyek mindegyike az anyagcsere szerkezetének, az itt ható génpárok relatív viszonyainak genetikai és evolúciós következményeit vizsgálja. Az első kutatási témámban az anyagcserehálózat, mint lehetséges eszköz jelent meg, ami segíthet a gének közötti kölcsönhatások jobb megértésében, a másodikban pedig adaptív hipotéziseket vizsgáltunk meg az anyagcsere működésére vonatkozóan, illetve az anyagcserére ható szelekció lehetséges hatását a genom szerveződésére. Disszertációm kérdésfelvetései a genom anatómiájának vizsgálatával egyrészt a genomikához, másrészt az anyagcserehálózatok rendszerszintű vizsgálatával a rendszerbiológia területéhez kapcsolódnak.

Genetikai interakciók modularitása és prediktálhatósága a sörélesztő anyagcserehálózatában

Munkánk során elsőként vizsgálhattuk az anyagcsere nagyléptékű genetikai interakciós térképét (Szappanos et al., 2011). Két gén genetikai interakcióban (episztázis) áll egymással, ha az egyes gének mutációinak hatása nem független egymástól. A genetikai interakciók fontos szerepet játszanak többek között a gének funkcionális kapcsolatainak feltérképezésében, a sokgénes öröklődésű betegségekben, illetve az evolúció lehetséges útvonalainak meghatározásában (Phillips, 2008). A génkölcsönhatások központi jelentősége ellenére a jelenség mechanisztikus háttere és a kölcsönhatások gének közötti megoszlása még kevéssé ismert. Kutatásainkban a *Saccharomyces cerevisiae* genetikai interakcióinak az anyagcserehálózat funkcionális moduljainak való viszonyát vizsgáltuk, valamint előrejelezhetőségét funkcionális genomikai és anyagcsehálózati információk valamint genomskálájú anyagcseremodell segítségével.

Vizsgálataink alapja a kísérletes kollaborátorunk (Boone labor, Torontó) által rendelkezésünkre bocsátott genetikai interakció adatsor volt, amelynek segítségével kb. 185 000, a modellünkben szereplő metabolikus génpár kvantitatív genetikai interakcióját elemezhettük. Ezek között a génpárok között el tudtuk különíteni az egyszeres génkiütések

alapján vártnál magasabb és alacsonyabb rátermettségű kettős mutánsokat, vagyis a pozitív és negatív genetikai interakciókat (multiplikatív modell szerint). A rátermettséget a haploid élesztőtörzsek kolóniamérete alapján becsültük.

A kísérletes eredményeket egy korábbi számítógépes modell alapján tett előrejelzésekkel vetettük össze. Ezek szerint i) a genetikai interakciók feldúsulnak az azonos funkciójú csoportban ii) e csoportok között a genetikai interakciók gyakran kizárólag negatívak vagy pozitívak (monokromatikusság) (Segre et al., 2005). A kísérletes eredmények részben alátámasztották a fenti mintázatok létét, de egyben a modellel tett előrejelzések általánosíthatóságának határaira is rámutattak. A hagyományosan definiált funkcionális modulokon belüli feldúsulás negatív és pozitív interakció esetében is szignifikáns, de kis mértékű volt (1,4 és 2,2-szeres). Hasonló eredményt kapunk a funkcionális kapcsolat biokémiailag pontosan értelmezett definíciójával az ún. fluxus kapcsoltsággal (Papin et al., 2004; Price et al., 2004). A kapcsolt fluxus azt jelenti, hogy az egyik reakció aktivitása minden alkalommal a másik reakció aktivitásával is jár, ami lehet egyirányú (irányított kapcsoltság) vagy kölcsönös (teljes kapcsoltság). Intuitív módon fluxus kapcsoltság esetén pozitív genetikai interakciót várnánk, hiszen ilyenkor definíció szerint az egyik gén kiütésével lenullázott fluxus a másik génhez tartozó reakcióaktivitás megszűnésével jár, ha az nincs is kiütve. Azonban a hagyományos annotációs csoportokhoz hasonló módon a negatív és pozitív genetikai interakciók feldúsulása itt is kismértékű (1,4 és 2,3-szoros). Az interakciók döntő többsége nem kapcsolt génpárok között fordul elő, míg a fluxus kapcsolt párok mindössze néhány százaléka van genetikai interakcióban. Vagyis míg a negatív és pozitív genetikai interakciók egyaránt feldúsulnak a tradicionális funkcionális modulokban és a biokémiailag definiált kapcsolatokban, a legtöbb genetikai interakció mégis különböző funkciókat köt össze. A modell második predikcióját tesztelve hasonló eredményt kaptunk: a monokromatikusság mértékét szignifikánsan nagyobbnak találtuk a randomizációval kapott nulleloszlás alapján vártnál, de ez mindössze 24-34 %-kal több monokromatikus modulpárt jelentett.

Ezután azt vizsgáltuk, mennyiben tudjuk prediktálni a genetikai interakciókat az anyagcsere génjeire vonatkozó egyéb ismereteink segítségével. Kétféle módszer előrejelzéseit vizsgáltuk: funkcionális genomikai, és az anyagcserehálózatra vonatkozó ismereteinken alapuló statisztikus modellezést és az anyagcsere nagyléptékű biológiai modelljét (FBA) mely képes egyes és kettes génkiütött törzsek növekedési rátájának becslésére. A statisztikus modellezéshez egyrészt genomszintű génpár jellemzőket használtunk korábbi munkákat követve (pl. koexpresszió, közös transzkripciós faktor) (Wong et al., 2004; Ulitsky et al.,

2009), másrészt anyagcserehálózati jellemzőket (pl. kódolt reakciók legrövidebb távolsága). Ezután a fenti jellemzők alapján klasszikus statisztikai (logisztikus regresszió) és egy újabb, döntési fák együttesén alapuló (ensemble) adatbányászati módszert (random forest (Breiman, 2001)) alkalmazva osztályoztuk genetikai interakció adatainkat. Az FBA modell esetében a prediktált genetikai interakciók empirikusan alátámasztott aránya akár 50%, illetve 11% lehet negatív illetve pozitív genetikai interakciók esetében, alátámasztva a legerősebbnek előrejelzett genetikai interakciók fiziológiás jelentőségét. Ugyanakkor azonos küszöbértékeknél a modell az empirikus genetikai interakció adatoknak csak igen alacsony arányát jelzi előre (2,8%, illetve 12,9%; negatív illetve pozitív genetikai interakció). A genomikai és anyagcserehálózati adatokon alapuló statisztikai modellezés, elsősorban a random forest módszer, az FBA-nál legtöbb esetben jobb predikciót eredményezett, de ez az előrejelzések 10%-os empirikus találati aránya felett még mindig csak a kísérletesen negatív genetikai interakciók 30% -át, illetve a pozitív genetikai interakciók 25%-ának előrejelzését jelenti. A becslés jósága akkor sem változik jelentősen, ha az FBA modell által prediktált adatokat (rátermettség és genetikai interakció értékek) hozzáadjuk a statisztikai modellhez, ugyanakkor negatív genetikai interakció esetében növeli a maximális precision értékét. Ez arra utal, hogy az anyagcseremodellben található olyan komplementer információ, ami a funkcionális genomikai és hálózati tulajdonságokból közvetlenül nem nyerhető ki. Összefoglalva, a genetikai interakciók többségét sem a biokémiai modellel, sem a funkcionális genomikai adatsorok és anyagcserehálózati adatok adatbányászati integrálásával nem tudjuk megbízható pontossággal megjósolni.

Az operonon belüli génsorrend evolúciója

Jól ismert, hogy a bakteriális génsorrend nem véletlenszerű és az operonok gyakran azonos anyagcsereutak vagy fehérjekomplexek tagjait tartalmazzák. Kutatásunkban azt vizsgáltuk, hogy vajon van-e az operonon belül valamilyen rendezettség a gének sorrendjében és ha igen, annak mi az oka. A kérdés megválaszolásához az *Escherichia coli* operonjait vizsgáltuk a rendelkezésre álló nagyszámú anyagcsereútvonalra vonatkozó és operonszerkezeti adat miatt (Keseler et al., 2009).

Eredményünk szerint a gének operonbeli sorrendje hajlamos az általuk kódolt, azonos metabolikus útvonalban részt vevő enzimek útvonalbeli sorrendjét tükrözni (kolinearitás). Megfigyelések szerint egyazon operonon belül a gének expressziós időpontja időben késleltetett az operon transzkripciós kezdőpontjától távolodva (Alpers & Tomkins, 1965,

1966) ezért ahogy matematikai modellünkben kimutattuk, a megfelelő sorrend a metabolikus útvonal bekapcsolásakor a végtermék gyorsabb megjelenését eredményezi.

Ez az elméleti előny csupán átmeneti, steady-state állapotban az egyes enzimek mennyisége független a génsorrendtől, ezért összetettebb magyarázatot kerestünk a kolinearitásra:

1. hipotézis: A kolinearitás előnye elsősorban az anyagcsereutak gyors aktiválását teszi lehetővé környezeti változás esetén. Amennyiben így van, azt várnánk, hogy elsősorban a különböző környezetekben különbözőféleképpen expresszálódó operonok mutatnak kolinearitást. Bioinformatikai elemzéseink szerint azonban a kolinearitás mértéke nem függ a különböző környezetek között mutatott expressziós változatosságtól.

2. hipotézis: A kolinearitás előnyös lehet, ha a gének expressziós szintje monoton csökkenést mutat az 5' végtől a 3' vég irányába (Nishizaki et al., 2007). E hipotézisnek ellentmond, hogy az *E.coli* operonjai esetében nem találtunk összefüggést a csökkenő expressziós trend és a kolinearitás mértéke között.

3. hipotézis: A „sztochaszikus leállás" hipotézis szerint alacsony génexpresszió esetén az útvonal kevés példányszámú enzimei véletlenszerűen elveszhetnek a sejt osztódása ill. a fehérjék lebomlása miatt. Matematikai modellünk segítségével bemutattuk, hogy így a kolinearitás ezekben a génekben folyamatos előnyt jelenthet, ugyanis az útvonal gyors újraindítására gyakran van szükség (t.i. állandó környezeti körülmények között is). Hipotézisünket alátámasztja, hogy valóban csak alacsonyan expresszálódó operonok esetén találunk szignifikáns kolinearitást, ahol az útvonal sztochasztikus leállása valószínűsíthető.

Összefoglalva, elsőként mutattuk ki szisztematikusan az operonon belüli génsorrend nem véletlenszerű mintázatát: az alacsonyan expresszálódó metabolikus operonokban a gének a kódolt enzimek reakciósorrendjét tükrözik (Kovács et al., 2009). Több alternatív adaptív hipotézist matematikai modellel vizsgáltunk, majd azok predikcióit empirikus *E. coli* adatokon teszteltük. Konklúziónk szerint a kolinearitás oka az alacsony expressziójú operonok esetében jelentkező sztochasztikus útvonalleállás idejének minimalizálása lehet.

Summary

With the recent availability of large-scale genomic and phenotypic datasets it has become possible, for the first time, to study the mapping from genotype to visible phenotypic traits in a systematic way and on an unprecedented scale. Bioinformatics analysis and integration of genome-scale datasets into large-scale mathematical models emerged as important methods to accomplish this goal. Metabolism is arguably the best characterized cellular subsystem which renders it as an excellent candidate to examine the link between genotype and phenotype. My thesis consists of two separate studies, both of which examine how the structure of the metabolic network influences the relations of the involved gene pairs. In the first part, the metabolic network is used as a tool to better understand interactions between mutations. In the second part, we investigated how natural selection acting on the performance of metabolic pathways might shape genome structure. My thesis is connected to the field of genomics by examining genome anatomy, and to systems biology by the system-level investigation of metabolic networks.

Modularity and predictability of genetic interactions in the *Saccharomyces cerevisiae* metabolic network

Our work is the first systematic, large-scale analysis of genetic interactions in a metabolic network (Szappanos et al., 2011). Genetic interactions, the non-independence of mutation effects, underlie various biological phenomena and illuminate gene functions. Despite efforts to globally map epistasis in model organisms, it remains poorly understood how genetic interactions arise from the operation of biomolecular networks. Our work analyzed the genetic interactions of the *Saccharomyces cerevisiae* to better understand how it is related to the functional modularity of the metabolic network and estimate its predictability based on genomic and metabolic network data.

Our work is based on the first large scale empirical dataset of genetic interaction among genes encoding metabolic enzymes, including ~185 000 double gene deletions with quantitative genetic interaction data (data produced in collaborator Charles Boone's lab). Among these gene pairs we defined double deletions that result in higher or lower fitness than expected (based on a multiplicative model) meaning positive or negative genetic interactions. Fitness was estimated based on colony size of haploid yeast strains.

We empirically tested earlier predictions about the distribution of interactions within and between metabolic functional modules. Specifically, a prior computational study based on FBA suggested that i) genetic interactions are enriched within metabolic annotation groups, and ii) interactions between different functional groups tend to be either exclusively negative or exclusively positive, a property termed 'monochromaticity' (Segre et al. 2005).

Using our large-scale genetic interaction map we found partial support for the above theoretical expectations. We report a modest but significant enrichment of both negative (1.6-fold) and positive (2.5-fold) genetic interactions within traditionally defined functional modules. However, the majority of genetic interactions occur between genes assigned to different metabolic functions (93% of negative and 90% of positive).

As an alternative to functional groups defined based on classical biochemical pathways, flux coupling provides a biochemically sound, unbiased definition of functional relatedness and has strong physiological and evolutionary relevance (Papin et al., 2004; Price et al., 2004). We used computationally identified flux-coupled gene pairs, that is, pairs of reactions where the activity of one reaction implies the activity of the other, either reciprocally or in one direction. In agreement with results obtained using annotation groups, although we find that both negative (1.4-fold) and positive (2.3-fold) interactions are significantly enriched in flux-coupled pairs, the overwhelming majority (> 97%) of both forms of interactions occur between uncoupled genes. In conclusion, both definitions of functional relatedness reveal that most genetic interactions connect across distinct functional modules.

Next, we asked whether interactions between different functional groups tend to be either exclusively negative or positive. In agreement with theoretical predictions, we found a statistically significant excess of monochromaticity among pairs of functional groups in the real data compared to randomized interaction maps. Nevertheless, monochromaticity in our genetic interaction map is modest: only ~24−34% more monochromatic pairs were found than expected by chance.

Next, we asked how well we can predict genetic interactions based on our knowledge of metabolic genes. We assessed the predictive power of two computational approaches: a genome-scale biochemical model of the metabolic network (FBA) which computes the growth of single and double deletant strains and a statistical/data mining method. In the second approach we compiled a dataset of gene-pair characteristics (e.g. coexpression), following earlier studies (Wong et al., 2004; Ulitsky et al., 2009) and metabolic network features (e.g. shortest path of reactions) but omitting any information on genetic interactions.

We used a classical statistical method (logistic regression) and a new data-mining method (random forest (Breiman, 2001)) to classify genetic interactions based on these features.
Using the biochemical model we can predict negative and positive interactions up to 50% and 11% rates of true predicted interactions (precision), respectively. Although this confirms that the highest predicted interaction scores have high physiological relevance, we find that only a minority of empirical interactions are captured by the model (2.8% and 12.9% for negative and positive interactions, respectively).
The statistical approach using genomic and metabolic network data gives better predictions. Although an increased fraction of *in vivo* interactions could be retrieved, ~70% of negative and ~75% of positive interactions were still predicted with very low (<10%) precision. Notably, incorporating fitness and genetic interaction scores derived from the biochemical model into statistical models boosts the precision of negative interaction predictions, indicating that biochemical modeling provides unique information that is not captured by purely statistical data integration. We conclude that the majority of genetic interactions are not well understood either in terms of biochemical processes or statistical associations.

Colinearity of gene order in *E. coli* metabolic operons

It is well established that gene order in prokaryotic genomes is not random. This is most evident when looking at operons, these often encoding enzymes involved in the same metabolic pathway or proteins from the same complex. However, it is almost completely unexplored whether gene order within operons is governed by chance or could have any functional significance. In particular, we investigated whether gene order within operons reflects the functional order of the encoded enzymes in a metabolic pathway (colinearity). To find out whether there are any sign of colinearity in real operons, we focused on *E. coli*, where high-quality and high-coverage data are available on both biochemical pathways and operon structures (Keseler et al., 2009). Our results showed that there is a significant trend for colinearity: approximately 60% of the intra-operonic gene pairs showing it, compared to 50% expected if gene order was random.
There is no known mutational bias which can result in colinearity thus we looked for adaptive scenarios. We argued that colinearity might have a fitness advantage (i.e. increased growth rate) by increasing pathway productivity. To gain insight into the potential interplays between gene oder and the flux of a metabolic pathway, we built general mathematical models of operon expression coupled to a linear metabolic pathway with four enzymes. At first sight, colinearity is unexpected as gene order should not affect the steady-state pathway

productivity. Indeed, this is confirmed by our mathematical model. We considered three extensions of the steady-state model that could potentially account for colinearity.

Hypothesis 1: colinearity in polar operons can increase steady-state pathway flux, where polarity refers to a decreasing mRNA abundance profile along the operon. The hypothesis is based on the theoretical finding that decreasing enzyme concentrations along the path can increase the flux along the pathway when the total enzyme concentration is fixed (Heinrich & Klipp, 1996). Thus, in a polar operon colinear arrangement can increase steady-state flux. This theoretical prediction is corroborated by our simulations and it predicts that we should observe colinearity in polar operons only. However, in contrast to this prediction, we failed to find an enrichment of colinearity in polar operons in *E.coli*.

Hypothesis 2: faster metabolic processing immediately after up-regulation upon environmental change. According to experimental measurements there is a time delay between the expression of two consecutive genes in an operon (Alpers & Tomkins, 1965, 1966). Thus right after activating the operon, the end-product can appear faster if the gene order is colinear. This verbal argument is confirmed by our model. This hypothesis predicts that operons showing high expression variation across conditions should more often display colinearity compared to constitutively expressed operons. However, using publicly available gene expression data we failed to find support for this hypothesis.

Hypothesis 3: metabolic stalling owing to stochastic protein loss. Small numbers of molecules are frequently involved in the process of gene expression and could lead to significant stochasticity in protein abundance (Elowitz et al., 2002). Whereas enzymes encoded in a highly expressed operon are likely to be always present in the cell whenever the operon is induced, stochasticity might play an important role in weakly expressed operons as enzymes could either decay or be diluted by cell division between two expression episodes (Cai et al. 2006), hence recurrently stalling metabolism. Colinearity could minimize the effect of such stochastic enzyme losses by speeding up the reinitiation of stalled metabolic transformations, in a similar manner as it provides a transient advantage after up-regulation of an inactive pathway. This argument is also supported by our model. This hypothesis specifically predicts that colinearity should be restricted to lowly expressed operons. Indeed, we found that only genes with low mRNA abundance show colinearity, hence supporting the „stochastic stalling" hypothesis.

To sum up, our work is the first reporting a non-random pattern of operonic gene order: in lowly expressed metabolic operons gene order reflects the functional order of the encoded enzymes (Kovács et al., 2009). Empirical tests of different adaptive scenarios for colinearity

in *E. coli* supports the hypothesis that the advantage of colinearity is to minimize metabolic stalling owing to stochastic protein loss.

Függelék

Az operon és anyagcsereút kapcsolt modelljének egyenletei

A modell egy négy génes operonból és egy lineáris anyagcsereútból áll, ami 5 metabolitot és 4 enzimet tartalmaz, melyeket az operon génjei kódolnak. Az RNS-polimeráz promóterhez (D) való reverzibilis kötődésével jön létre a C állapot f_0 asszociációs és b_0 disszociációs rátákkal modellezzük.

$$\frac{dD}{dt} = C \times (b_0 + k_0) - D \times f_0$$

$$\frac{dC}{dt} = D \times f_0 - C \times (b_0 + f_0)$$

A nyílt iniciációs komplex és az transzkcripció iniciációja egyirányú folyamatként szerepel k_0 rátával. Csak az mRNS leader régiója (M) szerepel a modellben, melyet a transzkripciót végző RNS polimeráz (T) v_0 rátával ír át. Az mRNS mf_0 rátával bomlik le (mC^1) és d rátával hígul. A riboszómák a degradoszómákkal versengenek a leader mRNS-hez való reverzibilis kötődésért (mf_1 asszociációs and $mb1$ disszociációs rátával). A transzláció az mC2 állapotból veszi kezdetét k_1 rátával ami után az M mRNS szakasz szabaddá válik a további riboszóma vagy degradoszóma kötéshez.

$$\frac{dT}{dt} = C \times k_0 - T \times v_0$$

$$\frac{dmC^1}{dt} = M \times mf_0$$

$$\frac{dM}{dt} = T \times v_0 + mC^2 \times (mb_1 + k_1) - M \times (mf_0 + d + mf_1)$$

$$\frac{dmC^2}{dt} = M \times mf_0 - mC^2 \times (mb_1 + k_1)$$

A négy enzim transzlációja az mT állapotban v_1 rátával zajlik, lebomlása és hígulása pedig α rátával ($\alpha = D+kdegr$). A „read-through" operon modellben csak az első génnek van riboszóma kötő helye így a transzlációt végző riboszóma, mT2 az E1 enzim megtermelése után írja át a következő enzimet (mT3 állapotban).

$$\frac{dmT1}{dt} = mC^2 \times k_1$$
$$\frac{dmT2}{dt} = mT1 \times v_1$$
$$\frac{dmT3}{dt} = mT2 \times v_1$$
$$\frac{dmT4}{dt} = mT3 \times v_1$$
$$\frac{dmT5}{dt} = mT4 \times v_1$$
$$\frac{dE1}{dt} = mT2 \times v_1 - E1 \times \alpha$$
$$\frac{dE2}{dt} = mT3 \times v_1 - E2 \times \alpha$$
$$\frac{dE3}{dt} = mT4 \times v_1 - E3 \times \alpha$$
$$\frac{dE4}{dt} = mT5 \times v_1 - E4 \times \alpha$$

A négy enzim standard Michaelis-Menten kinetikát mutat és enzimkinetikai paraméterei megegyeznek. Az első három termék metabolitkoncentrációinak időbeli változását az alábbi egyenlet írja le (i = 1,2,3):

$$\frac{dS_i}{dt} = k_{cat} \cdot E_i \cdot \frac{S_{i-1}}{S_{i-1} + K_m} - k_{cat} \cdot E_{i+1} \cdot \frac{S_i}{S_i + K_m} - D \cdot S_i$$

ahol k_{cat} a katalitikus konstans, D dilúciós ráta (sejt növekedési rátája), K_m a Michaelis konstans .

A végtermék (S_4) termelése (az S_4 összesített termelt mennyisége érdekes számunkra, ezért a végtermék dilúciója nem szerepel):

$$\frac{dS_4}{dt} = k_{cat} \cdot E_4 \cdot \frac{S_3}{S_3 + K_m}$$

1. táblázat A genetikai interakciók monokromatikussága funkcionális annotációs csoportok esetén

Több funkcionális csoporthoz tartozó géneket kizártuk az analízisből. A monokromitás mértékét az MC értékkel mértük (lásd *Módszerek 2.2.3*), és egy funkcionális csoportpárt akkor tekintettünk monokromatikusnak, ha $|MC_{ij}| > 0.5$. A tapasztalati monokromatikusság szignifikanciáját randomizációs teszttel határoztuk meg (lásd *Módszerek 2.2.3*). Mivel azok az annotációs csoportpárok, melyek között csak egy interakció van mindig monokromatikusak lennének, csak az egymás között legalább 2 vagy legalább 3 GI-t mutató csoport párokat vizsgáltuk.

	Funkcionális csoportpárok legalább 2 GI-val	**Funkcionális csoport párok legalább 3 GI-val**
Összes vizsgált funkcionális csoportpár	451	339
Monokromatikus funkcionális csoportpárok az adatsorban	177	111
Monokromatikus csoportpárok randomizált GI hálózatok alapján várt eloszlásának átlaga és szórása	142,87 (8,70)	82,75 (7,15)
A monokromatikus csoportpárok relatív többlete a valódi GI hálózatban a randomizált hálózatokhoz képest	23,89%	34,13%
A tapasztalati monokromitás szignifikancia szintje	$p < 10^{-4}$	$p < 10^{-4}$
Pozitív/összes GI háttér értéke (*bpr*)	0,341	0,337

2. táblázat Teljesen kapcsolt génpárok közötti monokromitás

	Teljesen kapcsolt csoportpárok legalább 2 GI-val	Teljesen kapcsolt csoportpárok legalább 3 GI-val
Összes vizsgált funkcionális csoportpár	144	80
Monokromatikus funkcionális csoportpárok az adatsorban	36	13
Monokromatikus csoportpárok randomizált GI hálózatok alapján várt eloszlásának átlaga és szórása	30,18 (3,08)	9,88 (2,20)
A monokromatikus csoportpárok relatív többlete a valódi GI hálózatban a randomizált hálózatokhoz képest	19%	32%
A tapasztalati monokromitás szignifikancia szintje	p = 0,043	p = 0,11
Összes vizsgált funkcionális csoportpár	0,25	0,29

3. táblázat Paraméterek és konstansok a metabolizmus és operon expresszió modelljében

Folyamat / Konstans	**Paraméterek**
Katalitikus konstans	k_{cat} = 50 s^{-1}, kerekített érték az *E. coli* aszpartát kináz I enzimen mért érték alapján (Chassagnole et al., 2001)
Michaelis konstans	K_m = 1 mM, kerekített érték az *E. coli* aszpartát kináz I enzimen mért érték alapján (Chassagnole et al. 2001)
Dilúciós ráta	D = 0,00019254 s^{-1} (60 perces sejtciklussal számolva)
Fehérje degradációs ráta	k_{degr} = 6.42 * 10^{-5} s^{-1} (Swain, 2004)
Fehérje dilúciós és degradációs ráta	α = 0.00025674 s^{-1} ($\alpha = D + k_{degr}$)
E. coli sejttérfogat	V = 7 * 10^{-16} l (forrás : redpoll.pharmacy.ualberta.ca/CCDB/cgi-bin/STAT_NEW.cgi)
RNS polimeráz DNS kötés (f_0) és disszociációs ráta (b_0)	f_0 = 0.42 s^{-1} (Swain 2004), lásd *6. ábra*, b_0 = 0.1 magasan expresszált operonoknál és b_0 = 1000 alacsonyan expresszált operonoknál.
Transzkripciós iniciációs ráta	k_0 = 0.1 s^{-1} (Swain 2004), lásd *6. ábra*
Riboszóma kötő hely létrejötte (v_0) és degradációja (mf_0) az mRNS-en	v_0 = 0.03 s^{-1}, mf_0 = 0.114 s^{-1} (Swain 2004), lásd *6. ábra*
Riboszóma kötés (mf_1) és disszociáció (mb_1) rátája	mf_1 = 4 s^{-1}, mb_1 = 0.4 s^{-1} (Swain 2004), lásd *6. ábra*
Transzláció rátája	k_1 = 0.3 s^{-1} (Swain 2004), v_1 = 0.017 s^{-1} (az egymás utáni géntermékék megjelenése közti 60 mp eléréséhez finomhangolt), lásd *6. ábra*

4. táblázat Anyagcsereútvonal steady-state fluxusa különböző génsorrendek esetén.

Az operon modell két determinisztikus szimulációjának összehasonlítása: egy tökéletesen kolineáris génsorrend esetén (ABCD) egy pedig anti-kolineáris elrendezéssel (DCBA). Az utolsó enzimen (E_4) átmenő fluxusokat hasonlítottuk össze egy olyan időpontban, miután az útvonal mindegyik köztes termékének koncentrációja állandó (8 tizedesjegy pontossággal)

Génsorrend	**Fluxus E_4 enzimen keresztül (mmol * s^{-1})**
ABCD	$1.64033624*10^{-16}$
DCBA	$1.64033624*10^{-16}$

5. táblázat Az útvonal steady-state fluxusa különböző operonális génsorrenddel poláritás esetén

Az operon polaritását feltételezve összehasonlítottuk a modell két determinisztikus szimulációjának eredményét tökéletesen kolineáris és antikolineáris génsorrend esetén. A polaritás a riboszóma-kötött mRNS intermedierek (mT_2, mT_3, mT_4) degradációjával lett bevezetve. A degradációs paraméter úgy lett beállítva, hogy egyharmad enzimszint csökkenést eredményezzen az egymás melletti géneket összehasonlítva (a transzlációs értéknek a fele). Mivel a degradáció csökkenti az össz enzim szintet, a modell paramétereket úgy változtattuk meg, hogy az össz enzim szint változatlan maradjon. Az utolsó (E_4) enzimen átmenő fluxusokat hasonlítottuk össze a kétféle génsorrend esetén, egy olyan időpontban, miután az összes átmeneti termék koncentrációja állandó (8 tizedesjegy pontossággal). A fenti műveletet magas és alacsony expressziós szintekre is megismételtük. A szimulációk azt mutatják, hogy a kolinearitásnak nyilvánvaló előnye van polaritás jelenlétében és az előny nagyobb, ha az expressziós szint magas (50% vs. 30.5%).

Génsorrend	**Fluxus E_4 enzimen keresztül (mmol * s^{-1})**	
	alacsony expresszió	**magas expresszió**
ABCD	$1.5087192*10^{-20}$	$1.61728822*10^{-16}$
DCBA	$1.0479932*10^{-20}$	$0.80871126*10^{-16}$

6. táblázat Determinisztikus szimuláció robosztussága a szubsztrát koncentráció és Km érték változásával szemben

Megismételtük a modell determinisztikus szimulációit különböző szubsztrátkoncentrációknál (S_0) és Michaelis-Menten konstansnál.

Szubsztrát koncentráció	Michaelis-Menten konstans	Kolinearitás relatív előnye	
		1 generáció után	50 generáció után
0.01 mM	1	8.64%	0.0995%
0.1 mM	1	8.63%	0.0996%
1 mM	1	8.49%	0.0990%
10 mM	1	5.12%	0.0812%
100 mM	1	2.45%	0.0490%
1 mM	0.01	1.11%	0.0517%
1 mM	0.1	7.48%	0.0926%
1 mM	10	8.64%	0.0970%
1 mM	100	10.38%	0.0878%

7. táblázat A sztochasztikus szimuláció robosztussága a szubsztrát koncentráció változásával szemben

Megismételtük a sztochasztikus szimulációkat különböző szubsztrát koncentrációkkal (S_0) és azt tláltuk, hogy a kolineritás relatív előnye 50 sejtgeneráció után konzisztensen 30 – 100-szor nagyobb alacsony expressziós szinten, mint magas expressziós szinten

Szubsztrát koncentráció	**A kolinearitás relatív előnye**	
	alacsony expresszió	**magas expresszió**
0.01 mM	3.48%	0.1%
1 mM	4.65%	0.1%
100 mM	5.45%	0.05%