

Génkölcsönhatások és evolúciós újítások vizsgálata anyagcsere- hálózati megközelítéssel

Doktori értekezés



Szappanos Balázs

Témavezető: Dr. Papp Balázs, Ph.D.

Biológia Doktori Iskola

MTA Szegedi Biológiai Kutatóközpont Biokémiai Intézet

Szeged

2016

Tartalomjegyzék

1. BEVEZETÉS.....	3
1.2. Az anyagcsere modellezése	4
1.3. A kényszer-alapú anyagcsere-modellezés alkalmazása.....	8
2. GENETIKAI INTERAKCIÓK VIZSGÁLATA ÉS MODELLEZÉSE AZ ANYAGCSERÉBEN.....	11
2.1. Bevezetés	11
2.1.1. Genetikai interakció meghatározása	11
2.1.2. A genetikai interakciók jelentősége	12
2.1.3. Genetikai interakciók modellezése az anyagcserében	13
2.2. Módszerek.....	15
2.2.1. Az élesztő anyagcsere kísérletesen meghatározott genetikai interakciós térképe	15
2.2.2. Az anyagcseremodell fitneszének kiértékelése figyelembe véve a környezeti és/vagy genetikai perturbációkat	16
2.2.3. Felhasznált anyagcseremodellek.....	17
2.2.4. Funkcionális pleiotrópia számolása az anyagcseremodell génjeire	18
2.2.5. Az anyagcseremodell automatizált feljavítása genetikai interakciós adatok felhasználásával	18
2.2.6. Az automatizált modelljavítási eljárás ellenőrzése keresztvalidálással	27
2.2.7. Az eredmények kiértékelése	28
2.2.8. Aszpartát oxidáz és kinolinát szintáz homológ gének keresése élesztőben	30
2.3. Eredmények.....	32
2.3.1. A gének genetikai interakciós fokszáma szorosan összefügg a gén fontosságával és pleiotrópiájával ..	32
2.3.2. Az anyagcseremodell genetikai interakciókat előrejelző ereje	37
2.3.3. Az anyagcseremodell predikciós ereje automatizált módszerrel feljavítható	41
2.3.4. A NAD bioszintézis útvonalak kijavítása az anyagcseremodellben	42
2.4. Diskusszió	45
3. REJTETT ENZIMAKTIVITÁSOK HÁLÓZATA ÉS EVOLÚCIÓS JELENTŐSÉGE	48
3.1. Bevezetés	48
3.2. Módszerek.....	51
3.2.1. Látens enzimaktivitásokat tartalmazó anyagcseremodell rekonstrukciója	51
3.2.2. Látens reakciók beköthetőségének vizsgálata.....	52
3.2.3. Környezetek meghatározása, ahol a látens reakciók fitneszhatását tesztelni lehet	52
3.2.4. Innovációt nyújtó látens reakciók azonosítása.....	53
3.2.5. Látens aktivitások genomléptékű feltárása fehérje túltermeltetési kísérletekkel	54
3.3. Eredmények.....	55
3.3.1. A látens reakciók beköthetőek a natív <i>E. coli</i> anyagcsere-hálózatba	55
3.3.2. A látens reakciók jelenléte fitneszelőnyt biztosít bizonyos környezetekben	55
3.3.3. A látens reakciók egyedi hatásának vizsgálata	57
3.3.4. Az <i>in silico</i> előrejelzések jó egyezést mutatnak a kísérletesen meghatározott metabolikus újtásokkal ..	59

3.4. Diskusszió	63
4. ÖSSZEGZÉS ÉS KITEKINTÉS	66
5. KÖSZÖNETNYILVÁNÍTÁS	70
6. IRODALOMJEGYZÉK.....	71
7. ÖSSZEFOGLALÁS	78
8. SUMMARY	81
9. FÜGGELÉK	84

1. Bevezetés

A biológiai rendszerek és jelenségek megértésének kulcsa, hogy minél több információt gyűjtsünk róluk és ezeket az információkat feldolgozzuk. Az elmúlt néhány évtizedben a kísérletes technikák nagymértékű fejlődése azonban az általuk szolgáltatott nyers adatmennyiség exponenciális növekedéséhez vezetett. Az adatok feldolgozásához új módszereket kellett kifejleszteni, a számítógépes adatfeldolgozás egyre inkább előtérbe került. A nagy mennyiségű adat azonban azt is lehetővé tette, hogy az összetett biológiai rendszereknek ne csak az egyes összetevőit, például egyes géneket és fehérjéket vizsgáljunk, hanem ezeket az összetevőket együttesen és a köztük fellépő kölcsönhatások figyelembevételével elemezzük. Ez a holisztikus megközelítés vezetett a rendszerbiológia létrejöttéhez. A rendszerbiológia célja leírni az élőlény működését a sejtet felépítő molekulák és kölcsönhatásaik megismerésével. Ehhez a matematikai modellezés eszköztárát használja.

Az anyagcsere az egyik legjobban feltárt alrendszere a sejtnak, és ezáltal kiemelten alkalmas arra, hogy rendszerbiológiai problémákat vizsgáljunk rajta. Nemcsak az anyagcserében részt vevő vegyületekről, enzimekről, génekről tudunk sokat, de a köztük lévő kölcsönhatásokról is. Ezeket az ismereteket felhasználhatjuk az anyagcsere, mint hálózat rendszerszintű vizsgálatára. Az anyagcseréről meglévő ismereteink összegyűjtése lehetővé tette olyan matematikai modellek építését is, amelyek előrejelzéseket tudnak tenni az anyagcserét érő zavarások, például génkiütések fitneszre gyakorolt hatásáról. Ezek a modellek, ahogy később látni fogjuk, képesek mechanisztikus magyarázatot adni az anyagcserével kapcsolatos jelenségekre azáltal, hogy az anyagcsere építőkövei (metabolitok, enzimek, stb.) közötti kölcsönhatásokat a biológiához hűen reprezentálják.

A rendszerszintű megközelítés az evolúcióbiológiában is új lehetőségekhez vezet. Az evolúcióbiológia kutatásában ugyanis kulcsszerepe van annak, hogy megértsük, hogyan határozza meg az élőlények genotípusa a fenotípust. Ezt a genotípus-fenotípus leképezést az anyagcseremodellek segítségével akár több száz génre kiterjedően, szisztematikusan tudjuk vizsgálni (Papp et al., 2011).

Doktori munkám során az volt a célom, hogy az anyagcsere-hálózatok felhasználásával az evolúciós genetika és az evolúciós innovációk néhány alapvető problémáját vizsgáljam. Egyrészt megvizsgáltam, hogy az anyagcseregének kiütésének hatásai között van-e kölcsönhatás, ha két gént együttesen ütünk ki, valamint hogy ezek a genetikai kölcsönhatások

milyen rendszerszintű mintázatokat mutatnak. A genetikai kölcsönhatások meghatározó szerepet töltenek be az evolúciós folyamatokban, a legtöbb fenotípusos jelleget ugyanis egynél több gén határozza meg. Így a genetikai kölcsönhatások tanulmányozása az evolúcióbiológia alapvető kérdéseinek megválaszolásán túl a rákkutatásban és más, genetikai háttérű betegségek kutatásában is kiemelkedő szerephez jut.

Doktori disszertációm második részében pedig az enzimek járulékos aktivitásainak jelentőségét részletezem új anyagcsereutak evolúciójában. Az élőlény új biokémiai útvonalak kialakulásával tud alkalmazkodni egy új környezethez, vagy toxikus anyagok, például antibiotikumok jelenlétéhez. Az enzimek járulékos, úgynevezett látens aktivitásai pedig az új útvonalakat felépítő biokémiai funkciók kialakulásának csírái lehetnek, ezáltal vizsgálatuk elősegítheti az adaptációs folyamatok jobb megértését.

1.2. Az anyagcsere modellezése

Az anyagcsere több okból is ideális alany biológiai kérdések rendszerszintű tanulmányozására. Az élő szervezetek életében központi szerepet tölt be, így vizsgálatával az élőlény működésének egészéről szerezhetünk információkat. Emellett az anyagcsere széleskörűen kutatott, nagy mennyiségű információ áll rendelkezésre róla.

Az anyagcserében részt vevő biokémiai útvonalak leírásán túl azok matematikai modellezésére is hamar felmerült az igény. Megjelentek az egyes útvonalakat leképező kinetikai modellek, melyeket számítógéppel értékelték ki (Garfinkel et al., 1970).

Idővel egyre több információ gyűlt össze az anyagcsere-útvonalakról, ezeket az adatokat pedig olyan adatbázisokban rendszerezték, mint a KEGG vagy a MetaCyc (Kanehisa and Goto, 1999; Caspi et al., 2012). A mind több adatnak köszönhetően lehetővé vált egyre összetettebb anyagcseremodellek felépítése. A teljes genomszekvenciák megjelenése pedig elősegítette az anyagcserében szerepet játszó gének azonosítását. Ez a folyamat az elmúlt évtizedekben eljutott oda, hogy a genom méretével összevethető nagyságú modelleket is fel lehet építeni (Covert et al., 2001; Palsson, 2009). A kísérletes oldalon a nagy áteresztőképességű technikák megjelenése adott újabb lökést a területnek, segítségükkel az anyagcsere szisztematikusan is vizsgálhatóvá vált, valamint szerepet kapnak az anyagcseremodellek felépítésében és előrejelzéseiknek az ellenőrzésében is.

Az anyagcsere-hálózat számítógépes modellezésére több eltérő módszert is kidolgoztak. Fontos jellemzője ezeknek a módszereknek, hogy segítségükkel mekkora rendszert lehet leképezni. Leginkább a modell építéséhez szükséges biológiai információk hozzáférhetősége korlátozza nagyobb modellek építését. Evolúciós és rendszerbiológiai kérdések megválaszolásához nincs feltétlen szükség genomléptékű modellekre, nagyobb modelleken azonban az egész rendszerre vonatkozó mintázatok is megfigyelhetők és az egyes adatpontoknak önmagukban kisebb a jelentőségük, így egy hibás adatpont kevésbé torzítja az eredményeket. Ilyen megfontolásokból a munkámhoz igyekeztem minél nagyobb léptékű modellt felhasználni.

Az anyagcseremodelleknek három kategóriáját különböztetjük meg aszerint, hogy mennyi információt tartalmaznak: kinetikai, topológiai és kényszer-alapú modellekről beszélhetünk.

A legegyszerűbb felépítése a topológiai modelleknek van. Már a topológiából is levonhatóak azonban a hálózat egészére vonatkozó fontos összefüggések, például az, hogy a hálózat elemei modulokba rendeződnek és ezek a modulok hierarchikus szerveződést mutatnak (Ravasz et al., 2002). A topológiai megközelítés hátránya, hogy az anyagcserét alkotó biokémiai reakciókról az alkotóelemek kapcsolódásán túl nem tartalmaz információt, hiányoznak a sztöchiometriai és kinetikai paraméterek. Ezáltal a topológiai modellek csak korlátozottan alkalmasak az anyagcsere-folyamatok mechanikus magyarázatára, illetve azok evolúciójának vizsgálatára (Stelling, 2004).

A három modellezési technika közül a kinetikai a legrészletesebb. A kinetikai modellekben minden egyes metabolit koncentrációjának időbeli változása le van írva a megfelelő kinetikai paraméterekkel a többi metabolit és az enzimek koncentrációjának függvényében. Ezekkel a modellekkel az anyagcsere egyes elemei közötti dinamikai kapcsolatok is modellezhetők. A kinetikai megközelítés hátránya azonban, hogy a modellezéshez szükséges mérési adatok mennyisége limitált (Costa et al., 2011). Ezen a téren történtek jelentős előrelépések, a kinetikai modellek azonban jelenleg nem alkalmasak genomszálajú modellek építésére (Smallbone et al., 2010; Stanford et al., 2013).

A kényszer-alapú modellek a hálózat topológiáján túl sztöchiometriai paramétereket valamint biológiai és fizikai-kémiai kényszereket is tartalmaznak. Ezek a modellek a topológiai modelleknél részletesebbek, de információtartalmuk alacsonyabb, mint a kinetikai modelleknek. A módszer előnye, hogy a modellépítéshez szükséges adatok – biokémiai reakciók sztöchiometriája, reverzibilitása, a tápanyagok (nutriensek) listája, stb. –

hozzáférhetőek, így genomskálájú modellek is könnyen rekonstruálhatóak. Ugyan a kinetikai modellekkel ellentétben a rendszer dinamikáját kis időintervallumon nem lehet követni (bár vannak módszerek ennek a hiányosságnak a kiküszöbölésére (Mahadevan et al., 2002)), erre leggyakrabban nincs is szükség (Ruppin et al., 2010).

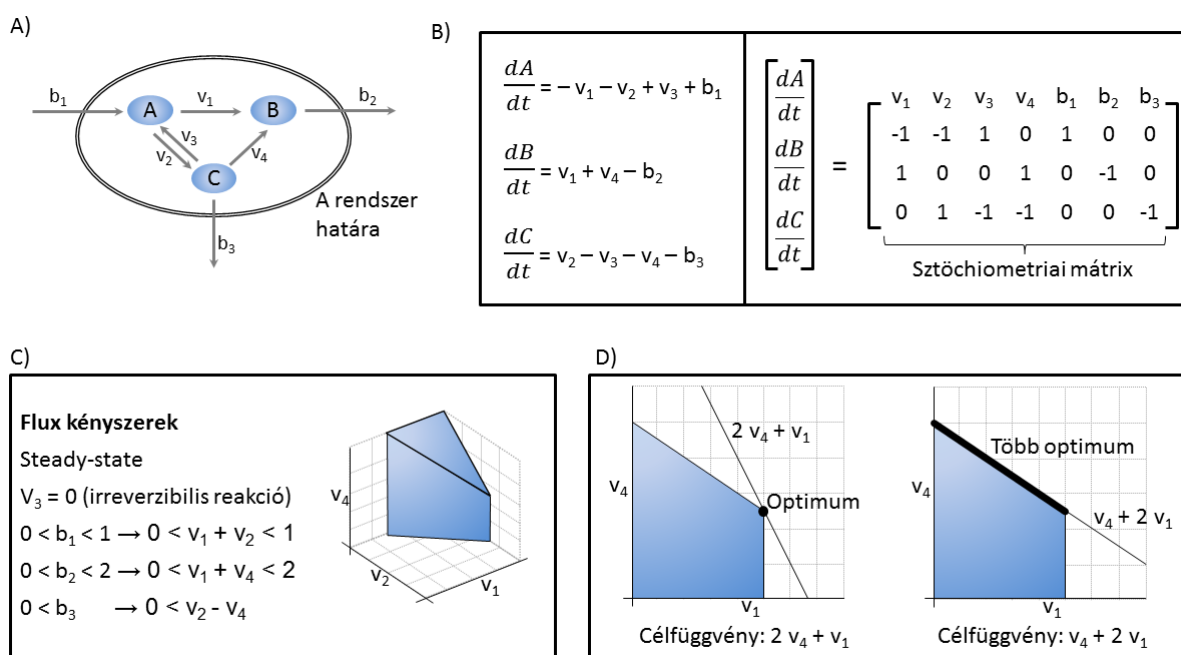
Napjainkban a kényszer-alapú modellezési technika a legelterjedtebb a genomskálájú modellek építésére és nagyszámú matematikai módszert fejlesztettek ki a modellek kiértékelésére (Lewis et al., 2012). Nevét a biológiai, fizikokémiai és környezeti kényszerekről kapta, amelyek a modell lehetséges megoldásainak terét szűkítik. A legfontosabb kényszer a metabolitok és a biokémiai reakciók közötti összefüggést leíró sztöchiometria, amit a modellben a sztöchiometriai mátrix ír le (1/A,B. ábra). Ez egy olyan mátrix, amelynek minden sora egy metabolitnak, minden oszlopa pedig egy reakciónak felel meg és a mátrix egy cellájának értéke az adott metabolit és reakció kapcsolatát leíró sztöchiometriai állandó. Abban az esetben, ha a metabolit szubsztrátja a reakciónak, a sztöchiometriai állandó negatív előjelű, míg ha terméke, akkor pozitív előjelű. Amennyiben egy metabolit nem szerepel egy adott reakcióban, a sztöchiometriai együttható értéke nulla (Orth et al., 2010; Kauffman et al., 2003).

Egy, az anyagcserében részt vevő metabolit koncentrációjának időbeli változását felírhatjuk a metabolitot termelő és felhasználó reakciók rátájának, más néven fluxusának felhasználásával (1/B. ábra). A kényszer-alapú modellezés során általában feltételezzük, hogy a rendszer steady-state állapotban van, tehát a metabolitok koncentrációja időben állandó. Ezáltal a sztöchiometrikus mátrixot felfoghatjuk úgy is, mint egy lineáris egyenletrendszert, ahol a reakcióráta a változók. Ez az egyenletrendszer azonban határozatlan, azaz több változót (reakció) tartalmaz, mint egyenletet (metabolit). Emiatt nincs egyértelmű megoldása, hanem több (gyakorlatilag végtelen) egyedi megoldása lehetséges. Biológiai ismereteinket felhasználhatjuk, hogy a megoldások terét leszűkítsük, és bizonyos feltételekkel megoldjuk az egyenletrendszert. Erre a legelterjedtebb módszer a fluxus balansz analízis (flux balance analysis, FBA). Az FBA során az eddig említett kényszereket (steady-state, sztöchiometria) kiegészítik a reakciók reverzibilitásával, a környezetből felvehető tápanyagok listájával, stb. (Price et al., 2004). Ezáltal a megoldási tér tovább szűkül, amit ebben a formájában fluxus tölcsérnek is hívnak (1/C. ábra).

Ha feltételezzük, hogy az anyagcsere-hálózat egy bizonyos feladat ellátására adaptálódott (pl. biomassza termelés), akkor optimalizációs eljárással meghatározhatók a reakciók fluxusai

(Kauffman et al., 2003). Ehhez szükség van egy célfüggvényre, ami lehet bizonyos metabolitok optimális termelése, maximális ATP termelés vagy a belső fluxusok összegének minimalizálása (Feist and Palsson, 2010; Schuetz et al., 2007). Egysejtű mikroorganizmusok esetén feltételezhetjük, hogy a növekedésük maximalizálására törekednek, amit az FBA keretein belül a növekedéshez szükséges metabolitok (biomassza összetevők) optimális arányban történő megtermelésével írhatunk le (Schuetz et al., 2007; Segrè et al., 2002).

A megfelelő célfüggvény kiválasztása után a reakciók fluxusainak meghatározása optimalizációs eljárások felhasználásával lehetővé válik (1/D. ábra). Az optimalizálásra használt eljárás függ a célfüggvénytől. Leggyakrabban lineáris programozást használnak, ami alkalmazható, ha a célfüggvény metabolitok termelésének vagy reakciók fluxusának maximalizálása (tehát a biomassza termelés maximalizálása esetén is) (Kauffman et al., 2003). Az optimalizáció eredménye a maximális biomassza termelés rátája, illetve a reakciók fluxus értékei. Fontos megjegyezni, hogy az egyes reakciókra kiszámolt fluxusok értéke többnyire nem egyedi, hanem bizonyos határok között változhat. Ennek oka, hogy az egyenletrendszer még mindig határozatlan.



1. ábra: kényszer-alapú modellezés (Orth et al., 2010. nyomán). A) Példa hálózat 3 metabolittal (A, B, C) és 7 reakcióval (b_1 , b_2 , b_3 , v_1 , v_2 , v_3 , v_4). A „v” jelű reakciók a metabolitok egymásba alakulásáért felelősek, míg a „b” jelűek a metabolitok rendszerbe való bejutásáért és onnan történő kijutásáért – ez megfelel a tápanyagok felvételének, illetve az

anyagcsere-végtermékek környezetbe való kijutásának. B) A hálózatban található metabolitok koncentrációjának időbeli változása a reakcióráták (fluxusok) függvényében, valamint a példahálózat sztöchiometriai mátrixa. C) Fluxus kényszerek és következményeik a megoldási térre. A példahálózatra alkalmazva a megadott kényszereket a megoldási tér leszűkül, ezt hívjuk fluxus tölcsernek. D) A rendszer optimalizálása különböző célfüggvények mellett. Az első esetben egyetlen optimális megoldás létezik, míg a másodikban több optimuma is van a rendszernek egy él mentén.

1.3. A kényszer-alapú anyagcsere-modellezés alkalmazása

Minden modellezési technika esetén felmerül, hogy mennyire pontos és milyen keretek között lehet használni. Számos tanulmány bizonyította a kényszer-alapú modellek használhatóságát. Az esszenciális géneket akár 90 százalékos pontossággal is képesek előrejelezni (Orth et al., 2011; Heavner et al., 2013). Az anyagcsere-modellezés során leggyakrabban használt két modellfaj, az *Escherichia coli* baktérium és a *Saccharomyces cerevisiae* sörélesztő esetében kimutatták, hogy a modellek előrejelzései az életképességre a vad törzs és az egyszeres génkiütések esetében a standard környezetben túl más környezetekben is megbízhatóak (Orth and Palsson, 2012; Snitkin et al., 2008).

Az FBA során optimalizációval számoljuk ki a rendszer állapotát, emiatt bármilyen perturbáció (mutáció, környezeti változás) esetében nem a közvetlen a perturbáció utáni állapotot kapjuk vissza, hanem az új kényszerek mellett elérhető ideális állapotot. A valóságban az organizmusnak azonban adaptálódnia kell a megváltozott körülményekhez. Biológiai értelemben tehát az FBA eredménye egy adaptált állapotot tükröz. Egy tanulmányban kimutatták, hogy az *E. coli* baktérium növekedési rátája glicerolon mint szénforráson növesztve 700 generáció alatt a kezdeti szuboptimális értékről az anyagcseremodell által előrejelzett optimális értékre nőtt (Ibarra et al., 2002). Az anyagcseremodellekkel tehát az adaptív evolúció eredményeként beállt anyagcsere-változások is vizsgálhatóak.

Noha a perturbáció utáni közvetlen állapotnál az FBA jobban tükrözi az adaptált állapotot, a génkiütések hatását kellő pontossággal jelzi előre (Orth et al., 2011; Heavner et al., 2013). Kidolgoztak ugyanakkor olyan módszereket, amelyek jobban közelítik génkiütés hatására megváltozott anyagcsere állapotát. Ezek a módszerek a reakciók fluxus értékeit pontosabban írják le mutáns törzsekben, azonban a géndelégiók növekedésre gyakorolt hatásának

előrejelzésében már nincs jelentős különbség az FBA-hoz képest (Segrè et al., 2002; Shlomi et al., 2005).

A kényszer-alapú anyagcseremodelleket széleskörűen alkalmazzák a biológia számos területén (McCloskey et al., 2013; Oberhardt et al., 2009). Az egyik iparilag is fontos terület az anyagcsere-mérnökség. Az anyagcsere-mérnökség (metabolic engineering) során az élő szervezeteket arra optimalizálják, hogy bizonyos vegyületeket minél hatékonyabban termeljenek (Yang et al., 1998). A genom skálájú anyagcseremodellekkel lehetővé vált az anyagcseregének végrehajtott módosításoknak az egész anyagcserére gyakorolt hatásának meghatározása, hatékonyabbá téve az optimális mérnöki stratégia kidolgozását (Park and Lee, 2008; Toya and Shimizu, 2013). Egy szemléletes példa az anyagcseremodellek használatára az anyagcsere-mérnökség területén (Lee et al., 2005) munkája, melyben a borostyánkősav túltermelésének lehetőségét vizsgálták *E. coli* baktériumban. Az *E. coli* anyagcseremodellt felhasználva génkiütések kombinált hatását határozzák meg a baktérium növekedésére és a borostyánkősav (szukcinát) termelésre. A modell sikeresen előrejelezte, mely géneket kell együttesen kiütni az optimális borostyánkősav termeléshez. Eredményeiket kísérletesen igazolták.

A kényszer-alapú modellezés nemcsak mikroorganizmusok esetében alkalmazható. Az anyagcsere-hálózat célfüggvényének meghatározása komplexebb, többsejtű élőlények esetében nehezebb feladat, erre azonban léteznek megoldások (Lee et al., 2012). Így ma már léteznek modellek növényi és emlős sejtekre, beleértve az embert is (Dal'molin et al., 2014; Orman et al., 2011; Thiele et al., 2013a). Az emberi anyagcsere és különböző betegségek, anyagcsere-rendellenességek vizsgálatára is sikerrel alkalmazzák az anyagcseremodelleket, mint például a rák anyagcseréje (Lewis and Abdel-Haleem, 2013), elhízás, cukorbetegség (Väremo et al., 2013), antibiotikum célpontok meghatározása (Chavali et al., 2012). A kényszer-alapú modellezési kereten belül lehetséges több faj anyagcseréje közötti kölcsönhatás vizsgálata is, ezáltal pedig a gazda-patogén kapcsolat vagy akár az emlős tápcsatorna mikrobiomjának modellezése (Thiele et al., 2013b; Levy and Borenstein, 2013).

Az evolúcióbiológia területén is hasznos eszköz a kényszer-alapú modellezés. Az evolúció kimenetének, illetve az evolúció során végbemenő folyamatoknak az előrejelzése komoly nehézségekbe ütközik. Egyfelől az evolúcióban az előre meghatározott események mellett véletlenszerű események is jelen vannak, másfelől az evolúció előrejelzéséhez ismerni kell az evolúcióban szerepet játszó biológiai egységeket (például géneket) és a köztük fellépő

kölcsönhatásokat. Ebben nyújtanak segítséget a rendszerbiológia mechanisztikus modelljei (Papp et al., 2011). A genetikai kölcsönhatások fontos szerepet töltenek be az evolúciós folyamatokban, emiatt a genetikai kölcsönhatásokért felelős mechanizmusoknak és a kölcsönhatások gének közötti eloszlásának megismerése az evolúcióbiológia egyik fontos célja. Azáltal, hogy az anyagcseremodellek képesek előrejelezni a mutációk fenotípusos hatását, képesek modellezni a mutációk között esetlegesen fellépő kölcsönhatásokat is, illetve az egyes kölcsönhatásoknak és a kölcsönhatások által mutatott mintázatoknak a háttérben meghúzódó mechanizmusokra is fényt deríthetnek. Ugyancsak fontos kérdés az evolúcióbiológiában, hogy hogyan tudjuk előrejelezni az adaptív változásokat és azok genetikai háttérét. Az anyagcseremodellek erre is lehetőséget nyújtanak, segítségükkel ugyanis rendszerszinten modellezhetjük új biokémiai reakciók megjelenésének hatását a fenotípusra. Mindezt ráadásul a gének szintjén is vizsgálhatjuk, ugyanis az anyagcseremodellekben az enzimek és gének közötti hozzárendelések is jelen vannak.

2. Genetikai interakciók vizsgálata és modellezése az anyagcserében

2.1. Bevezetés

Genetikai kölcsönhatásról vagy interakcióról akkor beszélünk, ha a két mutáció együttes hatása eltér attól, amit a mutációk önálló hatása alapján várnánk. A genetikai kölcsönhatás lehet negatív vagy pozitív, attól függően, hogy a kapott kettős mutáció hatása erősebb vagy gyengébb a vártnál.

Munkám során a gének közötti genetikai interakciókat vizsgáltam a gének teljes kiütésén keresztül. A továbbiakban ennek megfelelően tárgyalom a genetikai kölcsönhatást.

2.1.1. Genetikai interakció meghatározása

Negatív génkölcsönhatást mutat két gén, ha együttes kiütésük hatása a fitneszre¹ a vadhoz képest nagyobb, mint az egyszeres mutációk összegzett hatása. Ha például két gén által kódolt enzimek azonos reakciót katalizálnak (izoenzimek), akkor külön-külön kiütve őket csak csekély fitneszhatást tapasztalunk, mert a fennmaradó gén terméke kompenzálja a kiütés hatását. Együttesen kiütve viszont enzim hiányában a reakció nem játszódhat le, ami a reakció fontosságától függően akár letális is lehet az élőlényre.

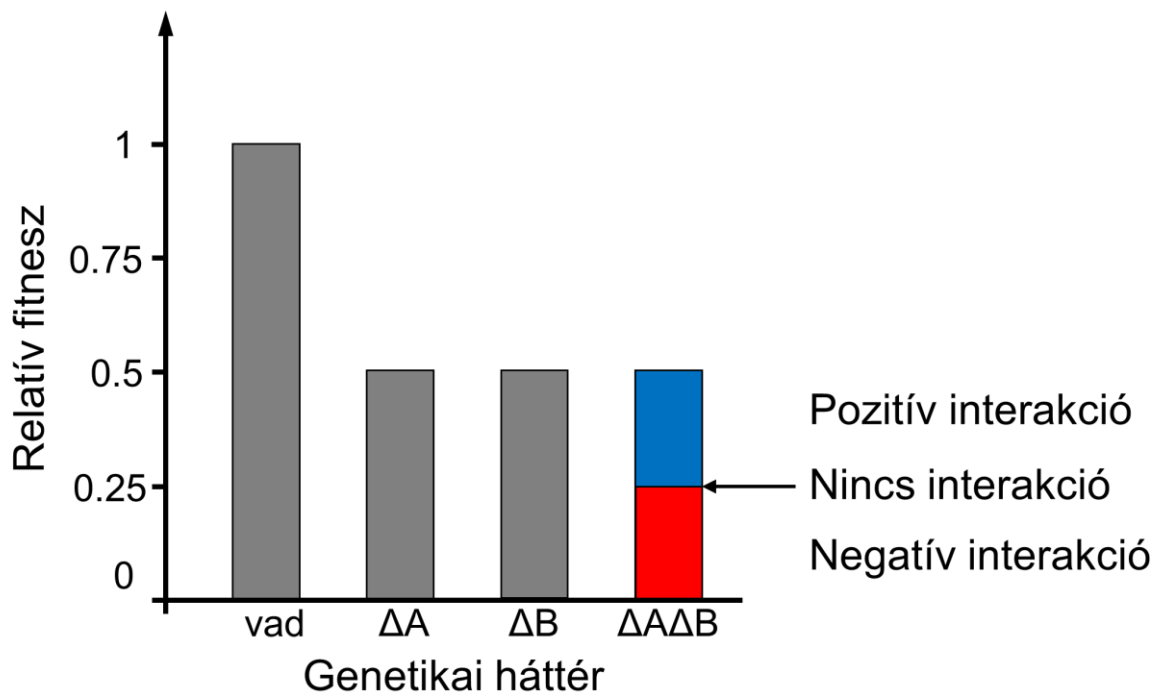
A gének közötti kölcsönhatás másik formája, amikor a kettős kiütés hatása kisebb, mint azt az egyszeres mutánsok fitnesze alapján várnánk. Ezt hívjuk pozitív génkölcsönhatásnak. Példaként egy lineáris útvonal génjeit vehetjük: már egyetlen gén kiütésével is kikapcsolható a teljes útvonal, így egy második gén kiütése nem, vagy csak kevésbé befolyásolja a fitneszt.

¹ Fitnesz alatt a dolgozatban a mikroorganizmusok növekedését jellemző mérőszámokat értjük, leggyakrabban a növekedési rátát.

Arra, hogy a gének egyszeres kiütéséből hogyan számoljuk a gének együttes kiütésének hatását, többféle megközelítés is létezik, leggyakrabban azonban az úgynevezett multiplikatív episztázis modellt alkalmazzák (Dixon et al., 2009). Eszerint a kettős mutáns relatív fitnessze az egyszeres mutánsok relatív fitnessének szorzata, amennyiben nem várunk interakciót a gének között (relatív fitness alatt az adott genotípus vadhoz viszonyított fitnessét értjük). A negatív illetve pozitív interakciót az ettől az értéktől való negatív vagy pozitív eltérésként definiáljuk (2. ábra). A genetikai interakció mértéke a multiplikatív modell szerint az alábbi képlettel számolható:

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

ahol f_{12} a kettős mutáns relatív fitnessze, f_1 és f_2 pedig az egyszeres mutánsok relatív fitnessze.



2. ábra: genetikai interakció a multiplikatív modell alapján. A vad típus relatív fitnessze 1, míg mindkét egyszeres mutánsé 0.5. A kettős mutáns relatív fitnessének 0.25-t várunk, amennyiben nincs interakció. Ha a tényleges érték ennél nagyobb, akkor pozitív interakció, ha kisebb, akkor negatív interakció áll fenn.

2.1.2. A genetikai interakciók jelentősége

A genetikai interakciók elméleti és gyakorlati jelentősége egyaránt kiemelkedő. A legtöbb fenotípusos bélyeg nem egyetlen gén hatása, hanem több gén együttes hatásának eredménye.

Ennek megfelelően a genetikai interakcióknak fontos szerepe van a genetikai háttérű betegségek kialakulásában (Lehner, 2007; Maxwell et al., 2008; Cordell, 2009). Ugyancsak fontos szerepet kapnak a genetikai interakciók ismeretlen funkciójú gének leírásában, egy új gén funkciója ugyanis megismerhető már ismert génekkel mutatott kölcsönhatásai alapján (Boone et al., 2007). A genetikai interakciók formálják az adaptív tájképet is, ezáltal befolyásolják az adaptáció menetét (Phillips, 2008; Poelwijk et al., 2007). Negatív genetikai interakciót mutat két gén, ha funkciójuk átfed, a funkcionális redundancia pedig védeltséget jelent véletlen mutációk ellen, emiatt gyakran megjelenik kulcsfontosságú sejtfolymatok esetében, mint például a sejtosztódás (Goldstein, 1993). Ilyen esetekben a funkcióért felelős egyik gén elvesztése még nem halálos az élőlény számára, csak ha minden olyan gén sérül, mely az adott funkciót el tudja látni. Az evolúció során ez egy lehetőséget is biztosít az élőlény számára: egy redundáns funkciót kódoló gén könnyebben megváltozhat, hiszen az eredeti funkcióját egy másik gén ellátja. Ennek egyik legegyszerűbb esete a génduplikáció (Ohno, 1970).

2.1.3. Genetikai interakciók modellezése az anyagcserében

Ahogy korábban láttuk, az anyagcseremodellek képesek megjósolni az egyszeres génkiütések fitneszre gyakorolt hatását akár genomskálán is. Ennek hatására felmerült, hogy a genetikai interakciók modellezésére is alkalmasak lehetnek. Korábban több tanulmány is felhasználta a genetikai interakciók modellezésére az anyagcseremodelleket (Segrè et al., 2005; Deutscher et al., 2006; Harrison et al., 2007; He et al., 2010). Az első ilyen jellegű munkát a *Saccharomyces cerevisiae* élesztő anyagcseremodellén végezték és kimutatták, hogy a genetikai interakciók hálózata modulokba szervezhető, azaz bizonyos funkciós csoportokon belül feldúsulnak az interakciók. Emellett azt is megfigyelték, hogy a modulokon belül és a modulpárok között túlnyomórészt egyféle interakció típus jelenik meg, például a pentóz foszfát útvonal enzimeit jellemzően negatív genetikai interakcióban állnak egymással és a glikolízis modullal, míg az ATP szintáz modullal pozitív genetikai interakciót mutatnak (Segrè et al., 2005). Eredményeik rávilágítottak, hogy a genetikai interakciók rendszerszintű vizsgálatával az anyagcsere-hálózat egészen új tulajdonságaira deríthetünk fényt. Ezek a korábbi munkák azonban teljesen vagy nagyrészt számítógépes szimuláción alapultak, az eredmények nagyskálájú kísérletes ellenőrzésére nem volt lehetőség. A korábban elérhető kísérletes adatok elsősorban esettanulmányok voltak, melyek egy-egy gén között fennálló kölcsönhatásra összpontosítottak, arról azonban szinte egyáltalán nem volt információ, hogy mely gének között nincs kölcsönhatás. Azóta megjelent egy új kísérletes módszer, az SGA

(synthetic genetic array), amely lehetővé teszi a genetikai interakciók automatizált feltérképezését sörélesztőben genomi szinten (Tong and Boone, 2006; Baryshnikova et al., 2010). Az SGA módszer segítségével a genetikai interakciók rendszerszintű vizsgálata immár a kísérletes oldalról is lehetővé vált. Egy átfogó munkában Costanzo és munkatársai 5.4 millió génpárra mérték le a genetikai interakció irányát és mértékét, ezzel lefedve a *Saccharomyces cerevisiae* élesztő génjeinek 75%-át (Costanzo et al., 2010).

Kutatásainknak három célja is volt. Egyrészt szisztematikusan tesztelni, hogy milyen hatékonysággal jelzi előre a biokémiai ismereteinket összegző anyagcseremodell a genetikai interakciós hálózat általános, nagyléptékű tulajdonságait és a konkrét génpárok közötti genetikai kölcsönhatásokat. Habár a genetikai interakciós hálózatok általános tulajdonságait kísérletes munkákkal már elkezdtek felderíteni (Costanzo et al., 2010; Tong et al., 2004), a megfigyelt mintázatok mögötti molekuláris mechanizmusokat sok esetben nem sikerült feltárni. Kiderült például, hogy azok az egyszeres génkiütések, amelyeknek jelentősebb a fitneszhatása, jellemzően több genetikai interakciót mutatnak, ezt a jelenséget azonban mindezidáig nem sikerült megmagyarázni (Costanzo et al., 2010). Az anyagcsere-hálózat számítógépes modellezése viszont egyedülálló lehetőséget kínál a genetikai interakciók mögött álló mechanizmusok feltárására.

További célunk volt egy automatizált számítógépes eljárást kifejleszteni, amely segítségével a modell hibás előrejelzései javíthatók és ezáltal a genetikai interakciós adatsor alapján új biológiai hipotéziseket javasolhatunk. Harmadik célunk pedig az ily módon automatikusan kinyert hipotézisek ellenőrzése volt.

2.2. Módszerek

2.2.1. Az élesztő anyagcsere kísérletesen meghatározott genetikai interakciós térképe

A nagyskalájú genetikai interakciós adatokat Charles Boone kanadai kutató csoportja (Toronto, Kanada) mérte le, majd bocsátotta rendelkezésünkre. Az SGA kísérletes módszer segítségével lemérték, hogy mi a hatása a *Saccharomyces cerevisiae* sörélesztő fitnessére, ha kiütik az élőlény anyagcseréjében részt vevő enzimeit kódoló géneket egyenként és párosával. Az SGA (synthetic genetic array) módszer két, egyszeres génkiütéses könyvtárat használ fel (Costanzo et al., 2010). Mindkét élesztő könyvtárban a génkiütéseket PCR-alapú génkiütéses módszerrel végezték, melynek során a kódoló szekvenciát (és csak azt) lecserélik egy antibiotikum rezisztencia kazettára (a két könyvtárban eltérő kazettát használnak). A génhez tartozó promoter és terminátor szakaszok változatlanul megmaradnak, tehát a génkiütés során nem történik frameshift (Baudin et al., 1993). A kettős mutánst a két egyszeres génkiütéses könyvtár keresztezésével hozták létre. Fontos megjegyezni, hogy a két élesztő könyvtárban szereplő gének listái között van eltérés, ugyanis bizonyos génkiütéseket technikai okokból nem sikerült mindkét könyvtárba bevinni. Emellett bizonyos génpárok szintén kiestek a mutánsok létrehozásához használt SGA technológiával való inkompatibilitás miatt és minőség-ellenőrzési okok miatt.

A kísérlet során a fitnesszt a szilárd agaron képzett telepek méretével jellemezték (Baryshnikova et al., 2010). A gének, amelyekkel a kísérletek során dolgoztak, az általunk a szimulációkhoz használt *S. cerevisiae* anyagcsere-hálózat alapján lettek kiválasztva (Mo et al., 2009). A kanadai kutatók egy korábbi munkájukban a *S. cerevisiae* élesztőre már mérték genetikai interakciókat nagy skálán, az az adatsor azonban nem koncentrált az anyagcserében szerepet játszó génekre, az anyagcseregének egy része hiányzott belőle (Costanzo et al., 2010). Az általunk használt adatsor az új, anyagcsere-specifikus méréseknek és a korábbi munkának az anyagcserére vonatkozó méréseinek egyesítésével állt elő és 652 nem esszenciális anyagcsere gént valamint 176 821 génpárt tartalmaz, amelyből 2668 mutat negatív, 1415 pedig pozitív genetikai interakciót.

A gének közötti genetikai interakció meghatározásához a kanadai kutatók korábbi munkája során optimalizált kritériumokat használtuk fel: két gén között genetikai interakció áll fenn, ha $|\epsilon| > 0.08$ és $p < 0.05$, ahol ϵ a multiplikatív episztázis értéke, a p értéket pedig a kísérletes replikátumok és az egyszeres mutánsok hibaeloszlásai alapján számolták (Baryshnikova et al., 2010; Costanzo et al., 2010). Ezt a határértéket úgy húzták meg, hogy az ideális legyen a

genetikai kölcsönhatások felhasználásához. Azért volt szükség egy határérték meghatározására, mert a kísérletesen meghatározott genetikai kölcsönhatások előjelében jobban megbízunk, mint a kvantitatív értékükben. Emiatt általában más funkcionális vizsgálatokban sem veszik figyelembe az interakció erejét azon túl, hogy a küszöb felett van vagy sem (Costanzo et al., 2010). Ezt az elvet követve a szimulációk során a kísérletes genetikai kölcsönhatásoknak mi is csak az irányát vettük figyelembe, a nagyságát nem.

Emellett egy megbízhatóbb adatsort is összeállítottunk szigorúbb feltételeket alkalmazva, hogy kiszűrjük a hamisan megállapított interakciókat. Ehhez csak olyan génpárokat vettünk figyelembe, amelyek le voltak mérve mind a korábbi, nagyskálájú adatsorban, mind az új, anyagcseregénekre koncentráló adatsorban. Két gén között akkor állapítottunk meg interakciót, ha ε iránya mindkét mérésnél azonos, $p < 0.05$ mindkét mérésre és legalább az egyik mérés esetén igaz, hogy $|\varepsilon| > 0.08$. Ha két gén közül egyikre sem igaz, hogy $|\varepsilon| > 0.08$ és $p < 0.05$, akkor a gének nincsenek interakcióban. Azok a génpárok, amelyek egyik fent említett csoportba sem tartoztak, ki lettek zárva az adatsorból. Ez a szigorúbb adatsor 122 875 génpárból áll, ebből 529 mutat negatív és 194 mutat pozitív genetikai interakciót.

2.2.2. Az anyagcseremodell fitneszének kiértékelése figyelembe véve a környezeti és/vagy genetikai perturbációkat

Az élesztősejt fitneszének szimulációjára a korábban bemutatott FBA módszert alkalmaztuk. Az FBA módszer előfeltétele, hogy a rendszer steady-state állapotban legyen, ez azonban nem mond ellent a kísérleteknek. A kísérletes fitneszt ugyanis több generáció követése határozza meg, tehát a kísérlet olyan időtartamban zajlott, ahol fennáll a steady-state állapot. Emiatt azt várjuk, hogy a legtöbb általunk megfigyelt fenotípus leírható a steady-state feltételezésével, tehát az anyagcseremodell jó közelítése a kísérleteknek.

A kísérletben alkalmazott táptalajt a számítógépes modellben is reprezentáljuk úgy, hogy a tápanyagok felvételéért felelős reakciókon átfolyó fluxus értékét korlátozzuk a kívánt mértékben (lásd lentebb). Génkiütés esetében pedig a gén által kódolt fehérjéhez kizárólagosan köthető reakciókon átfolyó fluxus értékét kényszerítjük nullára (Raman and Chandra, 2009). Az FBA szimulációkat az R programnyelvben végeztük (R Core Team, 2013) a Sybil programcsomaggal (Gelius-Dietrich et al., 2013), amely különösen alkalmas nagy mennyiségű szimuláció futtatására.

2.2.3. Felhasznált anyagcseremodellek

A szimulációkhoz a *S. cerevisiae* frissített anyagcseremodelljét használtuk, a modell neve: iMM904 (Mo et al., 2009). Ez a modell 904 anyagcseregént tartalmaz, továbbá 1412 biokémiai reakciót. A modell 8 sejtkompartmentet tartalmaz. Az eredményeink ellenőrzésére a szimulációkat megismételtük egy régebbi modellen is (iLL672, (Kuepfer et al., 2005)). Az iLL672 modell 672 gént és 906 biokémiai reakciót, valamint csupán 2 kompartmentet tartalmaz (citoplazma és mitokondrium), ezért kevesebb előfeltevéssel él a kompartmentek közötti transzportfolyamatokat illetően.

In silico környezetek

Az FBA szimulációkhoz használt *in silico* környezetet úgy állítottuk össze, hogy minél hűebben tükrözze a genetikai interakciós kísérletek során alkalmazott médiumot. A kísérletek során alkalmazott környezet: synthetic complete médium, amiből hiányzott a hisztidin, arginin és lizin (SD/MSD –His/Arg/Lys, (Tong and Boone, 2006)). A különböző tápanyagok felvételének maximális sebességét (Snitkin et al., 2008) alapján határoztuk meg. A teljes *in silico* környezet megtalálható a *Függelék: 1. táblázatban*.

A génkiütések szimulációjához a (Snitkin et al., 2008) munkában leírt *in silico* környezeteket használtuk.

In silico élesztő törzsek

A kísérletek során használt élesztő törzsekben gyakran ki vannak ütve bizonyos marker gének, amelyeknek szerepe van az anyagcserében. Mivel ezek a kiütések hatással lehetnek a különböző környezeteken való növekedésre, igyekeztünk minél hűbben leképezni a szimulációkra is. A Boone csoport a genetikai interakciós kísérletekhez olyan törzset használt, amelyben a *can1*, *lyp1*, *ura3*, *leu2* és *met17* gének ki voltak ütve (Costanzo et al., 2010). A modellben az ezeknek megfelelő reakciókat kiütöttük, feltéve, hogy más gén nem kódolta az adott reakciót (OMPDC, IPMD, AHSERL, AHSERL2 reakciók).

A modell feljavítása során biztosak akartunk lenni abban, hogy ha valamit változtatunk azért, hogy a genetikai interakciókat pontosabban jelezzük előre, akkor ne rontsuk le a modell más tulajdonságait. A modellek minőségét leggyakrabban azon méri le, hogy egyszeres génkiütések fitneszhatását milyen hatékonysággal jelzik előre különböző környezetekben. Emiatt a modellezés során felhasználtunk egy olyan adatsort is, amelyben 465 gén kiütésének

hatását mérték le 16 környezetben (Snitkin et al., 2008). Ebben az adatsorban azonban egy másik *Saccharomyces cerevisiae* törzset használtak, amelyben a *his3*, *leu2* és *ura3* gének voltak kiütve (Snitkin et al., 2008). Ennek megfelelően, amikor ezzel az adatsorral vetettük össze a szimulációs eredményeket, akkor a fent felsorolt génekhez rendelt IGPDH, IPMD és OMPDC reakciókat töröltük a modellből.

2.2.4. Funkcionális pleiotrópia számolása az anyagcseremodell génjeire

Egy adott gén pleiotrópiája alatt a dolgozatomban olyan mérőszámot értek, mely megadja, hogy a gén hány különböző biológiai folyamatban vesz részt. Korábban már kimutatták, hogy a genetikai interakciós hálózatban csomópontként szereplő gének jellemzően erősen pleiotrópák is (Costanzo et al., 2010), emiatt a pleiotrópia vizsgálata rávilágíthat a genetikai kölcsönhatások anyagcsere-hálózaton belüli mintázatainak hátterére.

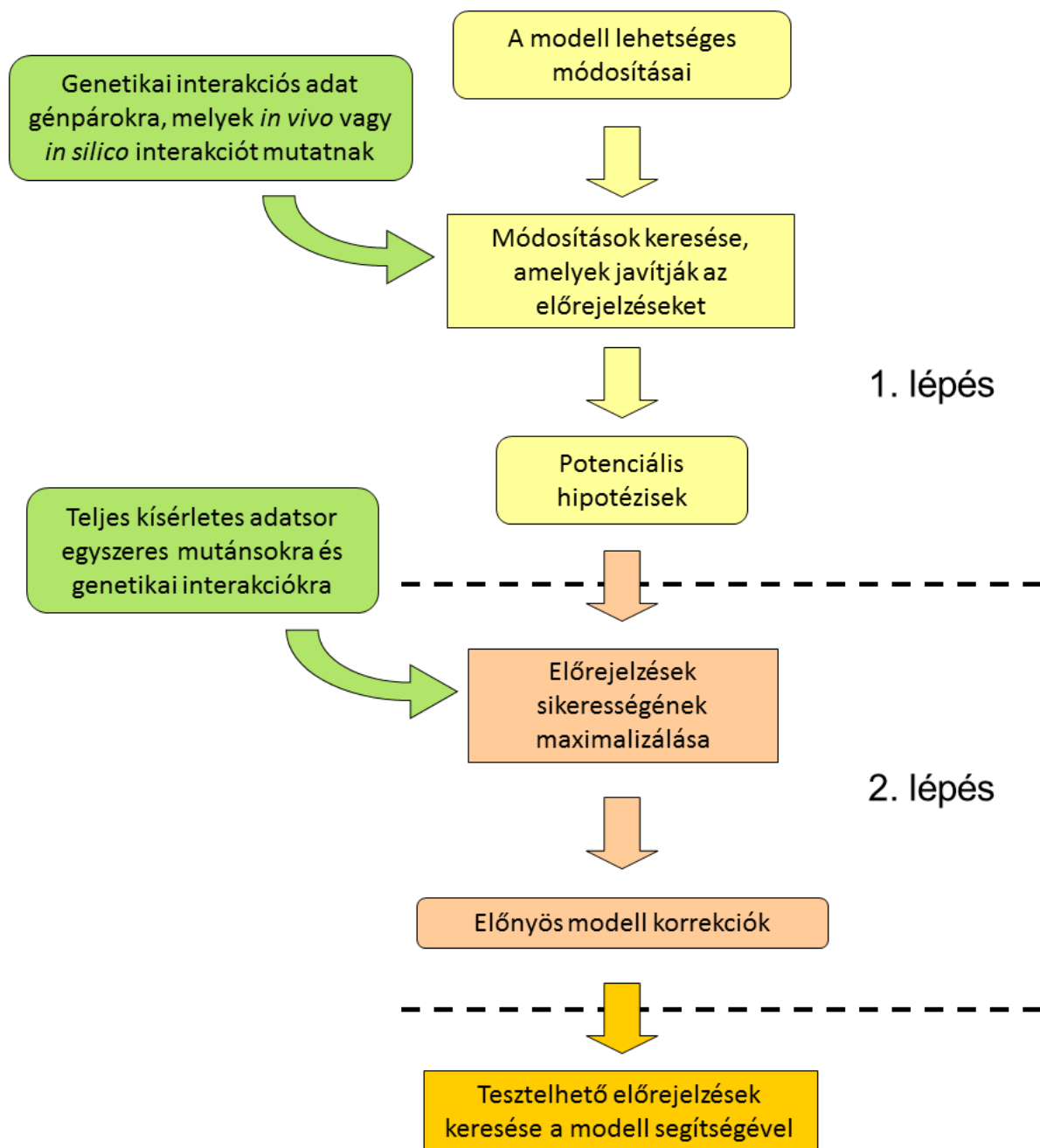
Az anyagcseregének pleiotrópiájának kiszámolásához az anyagcseremodellt használtuk fel. A modell 54 olyan metabolitot (biomassza komponens) határoz meg, melyeknek megtermelése létfontosságú az *in silico* élőlény életben maradásához és növekedéséhez (Függelék: 2. táblázat).

Első lépésben minden biomassza komponensre kiszámoltuk a modell által elérhető maximális termelés mértékét oly módon, hogy bevezettünk egy, a komponens a sejtől kiválasztó mesterséges reakciót, és meghatároztuk az ezen a reakción átmehető maximális fluxus értéket. Ezután minden egyes génre kiszámoltuk, hogy a kiütése csökkenti-e egy adott biomassza komponens maximális termelését. Végül minden génre megszámoltuk, hogy a kiütése hány biomassza komponens maximális termelését csökkenti. Ez a szám adja meg az adott gén pleiotrópiáját, tehát azt, hogy az adott gén hány eltérő folyamatot befolyásol az anyagcsere-hálózaton belül.

2.2.5. Az anyagcseremodell automatizált feljavítása genetikai interakciós adatok felhasználásával

Az anyagcseremodell feljavítására egy gépi tanuláson alapuló módszert fejlesztettünk ki, amely automatikusan generál hipotéziseket, amelyek javítják a negatív genetikai interakciók előrejelzését. Egy korábbi módszerhez képest, amely a kísérletes és számítógépes adatok közötti eltérést minden mutánsra külön-külön vizsgálta (Kumar and Maranas, 2009), a mi módszerünk az összehasonlítást globálisan végzi, tehát minden elérhető információt egyszerre vesz figyelembe. A módszer alapja egy kétszintű genetikai algoritmus, mely javítja a

kísérletes és prediktált adatok közötti egyezést (3. ábra). A genetikai algoritmusok a természetes szelekció inspirálta heurisztikus optimalizációt alkalmaznak, ezáltal olyan problémák megoldását teszik lehetővé, ahol az összes lehetséges állapot kiszámolása a paraméterek nagy száma miatt nem lehetséges. Az optimalizáció során egy populáció evolúciója zajlik, ahol a populáció egyedei mind potenciális megoldások (különböző modellek meghatározott változtatásokkal az eredeti modellhez képest), és a paramétertér egy-egy pontját reprezentálják (Goldberg, 1989). Hasonló algoritmusokat korábban sikeresen alkalmaztak az anyagcsere-mérnökség területén, hogy egy kívánt metabolit termelését maximalizáló génkiütéseket azonosítsanak (Rocha et al., 2008). A mi algoritmusunk egy olyan populációból indul ki, amelynek egyedei véletlenszerű változtatásokat hordozó („mutáns”) modellek. Az algoritmus minden modellre kiszámolja a modell kísérletes adatokkal való egyezését (ezt nevezik az egyed fitnessének), és a legpontosabb modellekből kialakítja bizonyos szabályok szerint (lásd alább) a következő generációt. A genetikai interakciók nagy száma miatt minden interakciót kiszámolni meglehetősen időigényes, így egy két lépcsős eljárást fejlesztettünk ki.



3. ábra: A kétlépéses modelljavító eljárás sémája. Az optimalizáció két lépésből áll. Az első lépésben csak azokat a génpárokat használtuk fel, amelyek kísérletesen vagy *in silico* negatív genetikai interakciót mutatnak. Ezáltal sikerült felgyorsítani ezt a lépést, ami lehetővé tette, hogy nagyobb populációmérettel és több generációval futtassuk a genetikai algoritmust. Az optimalizáció során esetlegesen megjelenő új, hibás interakciókat az algoritmus második lépése szűrte ki, melyben már az összes lehetséges génpáron teszteltük a genetikai algoritmus által generált modelleket. A második lépés során a genetikai algoritmus paraméterterét az első lépés eredményeként kapott módosítások képezték.

Az optimalizáció első lépésében azok közül a génpárok közül, amelyek mind a genetikai kölcsönhatásokat leíró kísérletes adatsorban, mind a modellben szerepelnek, csak azokat használtuk fel, amelyek vagy kísérletesen, vagy *in silico*, vagy mindkettőben negatív genetikai kölcsönhatást mutatnak. A genetikai interakciók ritkák (a génpárok 0.5%-a mutat negatív genetikai interakciót), így a legtöbb génpár nem mutat interakciót és kihagyásuk jelentősen felgyorsítja a paramétertér feltérképezését. Ennek azonban az az ára, hogy a modellen végrehajtott módosítások a nem tesztelt génpárokból új *in silico* genetikai kölcsönhatásokat idézhetnek elő, és ezeket a fals pozitív találatokat az optimalizáció első lépése nem veszi figyelembe. Ez két okból nem jelent problémát. Egyrészt a modell által kihagyott valódi genetikai interakciók komolyabb problémát jelentenek, mint a modell által hibásan jelzett interakciók (lásd Eredmények). Másrészt az optimalizáció során esetlegesen megjelenő új, hibás interakciókat az algoritmus második lépése szűri ki.

Az optimalizáció második lépésében az első lépés eredményeként kapott legjobb predikciós pontosságú modellekből indultunk ki, és egy újabb genetikai algoritmust futtattunk, ekkor azonban már a populációkban minden modell predikciós pontosságát az összes lehetséges génpáron lemértük. A teljes génpárlista felhasználása ebben a lépésben biztosítja, hogy a végső modellekben nem lesznek olyan módosítások, melyek aránytalanul sok fals pozitív negatív genetikai kölcsönhatást okoznak.

Az anyagcseremodellek predikciós pontosságának számolása

Az algoritmus mindkét lépésében hasonlóan történt a modellek értékelése. Először kiszámoltuk az *in silico* genetikai interakció értékét (ϵ) FBA segítségével, majd ezeket az értékeket binárisra alakítottuk (van/nincs interakció) és összehasonlítottuk a kísérletes adatokkal. Korábban már említettem, hogy a kísérletes adatok esetében a genetikai kölcsönhatások előjele megbízhatóbb, mint a mértékük, és ez a szimulációkra is igaz. Emiatt a prediktált kölcsönhatásokra is kiválasztottunk egy küszöbértéket, ahol a legpontosabb az előrejelzés. Az adatok binárisra transzformálásánál az erős negatív genetikai interakciókra összpontosítottunk: negatív volt az interakció, ha $\epsilon < -0.5$, és nem volt interakció, ha ez a feltétel nem teljesült. Ezen a küszöbértéken az eredeti modell nagy pontossággal jelezte előre a kísérletes genetikai interakciókat. Előzetes tesztelések alapján pedig azt találtuk, hogy a modell előrejelzéseinek pontossága robusztus a határérték változtatására, tehát más határértékek mellett is hasonló eredményeket kapnánk.

A bináris *in silico* genetikai interakciók kísérletes értékekkel való összevetéséhez a nagyobb megbízhatóságú, szigorúbb kísérletes adatsort használtuk fel. Az optimalizálás első lépésében a modelleket csak azokon a génpárokon értékeltük ki, amelyek kísérletesen vagy *in silico* genetikai interakciót mutatnak az eredeti modell szerint. Ennek megfelelően az optimalizáció ezen szakasza azokat a módosításokat részesítette előnyben, amelyek növelték a helyesen prediktált interakciókat (valódi pozitív, VP, lásd még 1. táblázat) vagy csökkentették a modell által hibásan interakciónak jelzett eseteket (fals pozitív, FP). A modellek kísérletes adatokhoz való illeszkedését az alábbi képlet alapján számoltuk:

$$Fitness\ (első\ lépés) = VP - FP$$

Az optimalizáció második lépésében minden elérhető mérési adatot felhasználtunk, ez összesen 67 517 a modellel átfedő génpárt jelent. A predikciók pontosságát Matthews korrelációs koefficienssel (Matthews Correlation Coefficient, MCC) számoltuk ki (Baldi et al., 2000). A MCC előnye, hogy egyforma súllyal veszi figyelembe az igazságmátrix² mind a négy kategóriáját, ami a genetikai interakciók ritkasága miatt különösen fontos, ugyanis a valódi negatív interakciók száma nagyságrendekkel meghaladja a többi kategóriát. A MCC kiszámolása a következő egyenlettel történik:

$$MCC = \frac{VP \times VN - FP \times FN}{\sqrt{(VP + FP)(VP + FN)(VN + FP)(VN + FN)}}$$

Az élesztő anyagcseremodell automatizált feljavítása során fontos szempont volt, hogy a módosítások hatására a modell azon képessége, hogy az egyszeres génkiütéseket megbízhatóan előrejelzi, ne vesszen el. Emiatt az optimalizáció második lépésében nem csak a genetikai interakciók predikcióját próbáltuk javítani, hanem az is szempont volt, hogy az egyszeres génkiütések predikciója ne romoljon. Az esszenciális gének előrejelzésének minősítéséhez egy korábbi tanulmány kísérletes adatsorából a YPD médiumon mért génkiütések eredményeit használtuk, a szimulációkhoz pedig a kéziratban leírt *in silico* környezetet alkalmaztuk (Segrè et al., 2002). Így 566 egyszeres génkiütés predikcióját tudtuk tesztelni. Az egyszeres génkiütések esetében is binárisra alakítottuk az FBA eredményét oly módon, hogy ha a vadra normált relatív FBA érték 0.6 alatt volt, akkor úgy tekintettük, hogy a

² Az igazságmátrix olyan 2x2 elemű mátrix, amelybe a vizsgált mintákat (itt génpárokat) sorolják be aszerint, hogy kísérletesen és a modellezés során mit mutatnak. Az 1. táblázat példájával élve a sorok megadják, hogy a modellezés során láttunk interakciót vagy sem, míg az oszlopok azt mutatják, hogy a kísérletes adatsorban láttunk-e interakciót az adott génpárra.

génkiütés letális volt. A modell predikciós sikerét a genetikai interakciókhoz hasonlóan a génkiütések esetében is Matthews korreláció számolásával végeztük.

Egy modell végső predikciós erejét a genetikai interakciókra és az egyszeres génkiütésekre számolt MCC értékekből számoltuk a következő formulával:

$$\text{Fitness (második lépés)} = (MCC_{\text{interakciók}} + MCC_{\text{génkiütések}}) / 2 - \pi,$$

ahol π egy büntetés, ami arányos a modellen végrehajtott módosítások számával (lásd alább). Ahogy az látható, ez a formula egyenlő arányban veszi figyelembe a kölcsönhatások és az egyedi génkiütésekre kapott előrejelzések pontosságát. Azért döntöttünk így, mert előzetes számításaink azt mutatták, hogy a két mérőszám eltérő súlyozása nem befolyásolja érdemben a modell feljavítását. Ennek oka, hogy az eredeti anyagcseremodell már optimalizálva van az egyedi génkiütések minél pontosabb előrejelzésére, emiatt ezen a területen nem vártunk jelentős javulást. $MCC_{\text{génkiütések}}$ tehát elsősorban akkor érvényesül, ha egy változtatás rontaná az egyszeres génkiütések előrejelzését.

		Kísérletesen megfigyelt interakció	
		Van	Nincs
Modell által jelzett interakció	Van	Valódi pozitív (VP)	Fals pozitív (FP)
	Nincs	Fals negatív (FN)	Valódi negatív (VN)

1. táblázat: a kísérletes megfigyelések és a predikció összevetését bemutató igazságmátrix. Minden egyes génpár, amelyre genetikai interakciót számoltunk, besorolható a táblázatban szereplő négy kategória egyikébe.

Az eljárás során a modelleken végrehajtott lehetséges módosítások

Kétféle változtatást engedélyeztünk a modellen, hasonlóan egy korábbi munkához (Kumar and Maranas, 2009). Az első a biomassa-összetevők listájának módosítása. Ezek azok a metabolitok, amelyeknek megfelelő arányban történő maximális termelése az anyagcseremodell célja az FBA során. Bizonyos metabolitok jelenléte a biomassa-alkotóelemek között megkérdőjelezhetetlen (aminosavak, nukleotidok), másoké azonban kétséges (pl. egyes vitaminok, tartalék tápanyagok). A sörélesztő anyagcseréjének

szimulációjára több modellt is kidolgoztak, melyekben a biomassa összetétele eltér. Ennek több oka van. Egyrészt a biomassa összetételére vonatkozó adatok kísérletes mérésekből származnak (Förster et al., 2003; Mo et al., 2009) és a mérések között lehetnek eltérések, valamint a biomassa összetétel is változhat a környezet függvényében. Továbbá a biomassa modellben található reprezentációja az esszenciális metabolitokat hivatott tartalmazni, és elképzelhető, hogy egy metabolit, amely a mérések során jelen van a biomasszában, nem esszenciális a növekedéshez. Végezetül a biomassa modellben található összetétele csupán egy közelítés, összeállításánál pedig fontos szempont, hogy a modell előrejelzései minél pontosabbak legyenek, még akkor is, ha ehhez bizonyos metabolitokat, amelyek a kísérletek alapján a biomassa részei, el kell hagyni belőle. Emiatt a modell biomassa összetételének optimalizálása szerves részét képezi az anyagcsere-hálózatok rekonstrukciójának (Kuepfer et al., 2005).

Azokat a metabolitokat, amelyek nem voltak jelen mindegyik élesztő modell biomassa-összetevői között, módosítható biomassa-komponenseknek vettük és jelenlétüket vagy hiányukat az optimalizációs algoritmus szabadon változtathatta. A módosítható biomassa-komponensek listájának összeállításához négy publikált modellt használtunk fel: iFF708 (Förster et al., 2003), iND750 (Duarte et al., 2004), iLL672 (Kuepfer et al., 2005), iMM904 (Mo et al., 2009). A 21 módosítható biomassa-komponens listáját a *Függelék: 3. táblázat* tartalmazza.

A második megengedett módosítás a modellen a reakciók reverzibilitásának változtatása. A reverzibilitási fok növekedését nem engedjük, csak újabb kényszerek modellbe vezetését (tehát csak reverzibilis \rightarrow irreverzibilis és irreverzibilis \rightarrow inaktív változásokat engedélyeztünk). Sem a reverzibilitási fok növekedését, sem új reakciók hozzáadását nem engedélyeztük, ugyanis az eredeti modell azért nem találja meg a kísérletes negatív genetikai interakciókat, mert a kettős génkiütés fitnessét túlbecsli. Ezt a problémát pedig nem lehet új reakciók hozzáadásával vagy irreverzibilis reakció reverzibilissé alakításával orvosolni, melyek csak új potenciális alternatív útvonalat jelentenének. Az *in silico* esszenciálisnak talált reakciókat és azokat a reakciókat, melyek kísérletesen esszenciálisnak meghatározott génekhez vannak kapcsolva az anyagcseremodellben, nem módosítottuk (Giaever et al., 2002; Segrè et al., 2002). Emellett nem módosítottuk az úgynevezett blokkolt reakciókat sem, amelyeknek a fluxus értéke semmilyen körülmények között nem lehet nullától eltérő érték (például azért, mert nincs olyan reakció, amely megtermelné a blokkolt reakció szubsztrátját (Burgard et al., 2004)). Hogy tovább csökkentsük az optimalizáció számításigényét, a teljesen

kapcsolt reakciók csoportjait egyetlen reakcióként képeztük le a paramétertérben. Teljesen kapcsoltnak akkor vehetünk két reakciót, ha a fluxusuk tökéletesen korrelál minden körülmények között (például az első reakció kizárólagosan termel egy metabolitot, amelyet egyedül a második reakció tud feldolgozni (Burgard et al., 2004)). Összességében 454 módosítható reakciót gyűjtöttünk össze. A reverzibilis reakciók esetében azonban a reakció két lehetséges irányát külön-külön is lehet módosítani, emiatt ezek a reakciók az irányoknak megfelelően két bináris paraméterként lettek leképezve, míg az irreverzibilis reakciók egy bináris paraméterként. Összesen 615 bináris paraméter írta le a reakciók reverzibilitásának és inaktiválásának változását.

A genetikai algoritmus részletei

A genetikai algoritmus egy populáció evolúcióját végzi, ahol minden egyed egy bináris paraméterekből álló vektor és minden paraméter egy bizonyos módosítást jelent az eredeti anyagcseremodellhez képest. A paraméter állapota 1, ha a módosítást alkalmazzuk a modellre és 0, ha nem. A kiindulási populációt véletlenszerűen generáltuk úgy, hogy a paraméterek többségét nullára állítottuk, és csak egy kis részüket állítottuk egyre. Az egyre állított paraméterek számát minden modell esetében egy véletlenül kiválasztott szám adta meg, amit egy $\lambda = 1$ várható értékű Poisson eloszlásból húztunk. Az így kapott számot növeltük eggyel, biztosítva, hogy legalább 1 változtatás legyen a kiindulási populáció minden egyedében. Ezután véletlenszerűen kiválasztottunk a kapott számnak megfelelő paramétert és egyre állítottuk az értéküket, míg a többi paraméterét nullára.

A populáció minden egyedére kiszámoltuk a fitneszt a módosított modell predikcióinak kísérletes adatokkal való egyezése alapján a fentebb leírt módon. A következő generáció egyedeit a jelenlegi generáció egyedeinek fitnesze alapján határoztuk meg a következő operátorok alkalmazásával: elitizmus, keresztezés, mutáció, beillesztés, kiütés. Első lépésként a legmagasabb fitneszű egyedeket változatlan formában átvittük az új generációba (elitizmus). Ezután létrehoztunk egy halmazt, melynek mérete feleakkora volt, mint a populáció mérete, és annak a valószínűsége, hogy egy egyed belekerült, arányos volt az egyed fitneszével (ennek a halmaznak az angol neve mating pool). Az egyedek halmazba kerülésének valószínűségét egy lineáris rang eljárás (linear ranking procedure) alapján határoztuk meg (Bäck and Hoffmeister, 1991). Ezután a halmazból véletlenszerűen kiválasztottunk egy vagy két egyedet („szülők”) és az operátorok egyikével létrehoztunk egy új egyedet az új

generációba. Minden új egyedet egyetlen operátor használatával hoztunk létre. Az egyes operátorok részleteikben:

- *Keresztezés*: két szülőt választottunk és minden paraméter esetében az értéket az egyik szülőtől örökölte 50%-50% eséllyel.
- *Mutáció*: egy szülőt választottunk, annak pedig két paraméterét, egyik 0, a másik 1 állapotban, és a két paraméter értékét megcseréltük, így az egyesek és nullák száma változatlan maradt.
- *Beillesztés és kiütés*: véletlenszerűen kiválasztott számú nullás (vagy egyes a kiütés esetében) állapotú paramétert az ellenkező állapotra állítottunk. Azt, hogy hány paramétert váltunk át, egy, a $\lambda = 1$ paraméterű exponenciális eloszlásból húzott véletlenszerű szám határozta meg, mely azonban legalább 1 kellett, hogy legyen.

Az optimalizáció első lépésében az evolúció 1000 generáción át tartott és a populáció 100 egyedből állt. Az előzetes szimulációk alapján azt vártuk, hogy az 1000 generáció elegendő ahhoz, hogy a genetikai algoritmus egy optimumba konvergáljon (azaz a legjobb egyed fitnessze már ne javuljon tovább). Az utódok megoszlása aszerint, hogy mely operátorral hoztuk őket létre: 5% elitizmus, 45% keresztezés, 40% mutáció, 5% beillesztés és 5% kiütés.

Az optimalizáció első lépésének eredményeként összegyűjtöttünk minden olyan modellt, ami az optimalizáció során létrejött, és amelynek fitnessze elérte a maximális fitnesszt. Ezekből a modellekből kiválasztottuk a leggyakoribb módosításokat (azokat, amelyek az átlagnál gyakrabban fordultak elő a modellekben). Ezek a módosítások képezték a lehetséges paraméterek terét az optimalizáció második lépésében (lásd 3. ábra). Ezáltal a paraméterter a második lépésre jelentősen lecsökkent, jellemzően kevesebb, mint 50 bináris paraméterre, szemben az első lépés 615 paraméterével. Ennek következtében a második lépésben már lehetővé vált, hogy a teljes genetikai interakciós adatsoron tovább optimalizáljuk a modellt az első lépés eredményeként kapott ~50 paraméterrel.

Az optimalizáció második körében a populáció 50 egyedből állt és az evolúció 20 generáción keresztül zajlott, az egyedek fitnesszét pedig minden kísérletes adat figyelembe vételével számoltuk. Hiába volt a paraméterter sokkal kisebb, az optimalizáció második lépése még így is sokkal időigényesebb volt, mint az első lépés, a nagy mennyiségű génpár miatt, amelyet ki kellett értékelni. Az előzetes szimulációk azonban azt mutatták, hogy 20 generáció alatt konvergál az algoritmus egy optimumba.

A második lépés kezdő populációjának minden egyedét véletlenszerűen hoztuk létre úgy, hogy minden paramétert 50% eséllyel kapcsolunk be. A neutrális módosítások kiszűrése érdekében egy büntetést vezettünk be (π), amelynek értéke:

$$\pi = 10^{-7} \times \frac{\text{paraméterek száma az egyes állapotban}}{\text{paraméterek száma}}$$

A büntetés értékét úgy választottuk meg, hogy az semmilyen pozitív hatást nem tud elnyomni, amit egy új módosítás okozna. Ha viszont két egyed fitnessze egyforma lenne, akkor a büntetés hatására az lesz az előnyösebb, amelyik kevesebb módosítást tartalmaz. Emellett a genetikai algoritmus operátorainak arányát is megváltoztattuk az első lépéshez képest a következők szerint: 5% elitizmus, 32.5% keresztezés, 27.5% mutáció, 5% beillesztés és 30% kiütés, tehát a kiütés arányának növelésével a módosítások számának csökkentését preferáltuk.

Az optimalizáció második lépésének eredményét tovább egyszerűsítettük azáltal, hogy a még benn maradt neutrális módosításokat kivettük. A teljes kétlépéses optimalizációt 8 alkalommal futtattuk le, ezzel mértük fel, hogy a kapott optimális modell egyedi, vagy több megoldás is van, illetve hogy szélesítsük az eredményül kapott hipotézisek körét. Az optimalizáció ugyanis felfogható a modell egyfajta evolúciójának is, és két optimalizáció eredménye jelentősen eltérhet, egyrészt a véletlen módosításoknak köszönhetően, másrészt a módosítások között esetlegesen fellépő kölcsönhatások miatt.

2.2.6. Az automatizált modelljavítási eljárás ellenőrzése keresztvalidálással

Az anyagcseremodell automatikus feljavítására kifejlesztett módszerünket keresztvalidálással ellenőriztük a következőképpen:

1. A kísérletes genetikai interakciós adatsort és az egyszeres génkiütéses adatsort is véletlenszerűen két, egyforma részre osztottuk (egyik a tanuló, másik a teszt adatsor). A felosztást oly módon végeztük, hogy mindkét csoportba egyforma számú genetikai interakció (illetve esszenciális gén) kerüljön.
2. Lefuttattuk a kétlépéses genetikai algoritmust a tanuló adatsoron.
3. Kiértékeljük az optimalizált modell(ek)e)t a teszt adatsoron az optimalizációnál leírt módszerrel. Ha több, egyformán jól teljesítő modellt kaptunk eredményül, akkor mindet kiértékeljük a teszt adatsoron és az átlagos teljesítményüket vettük. Összehasonlításként az eredeti, módosítatlan modellt is kiértékeljük a teszt adatsoron.
4. Felcseréltük a tanuló és teszt adatsorokat és megismételtük a 2-3 lépéseket.

5. Az 1-4 lépéseket ötször megismételtük.

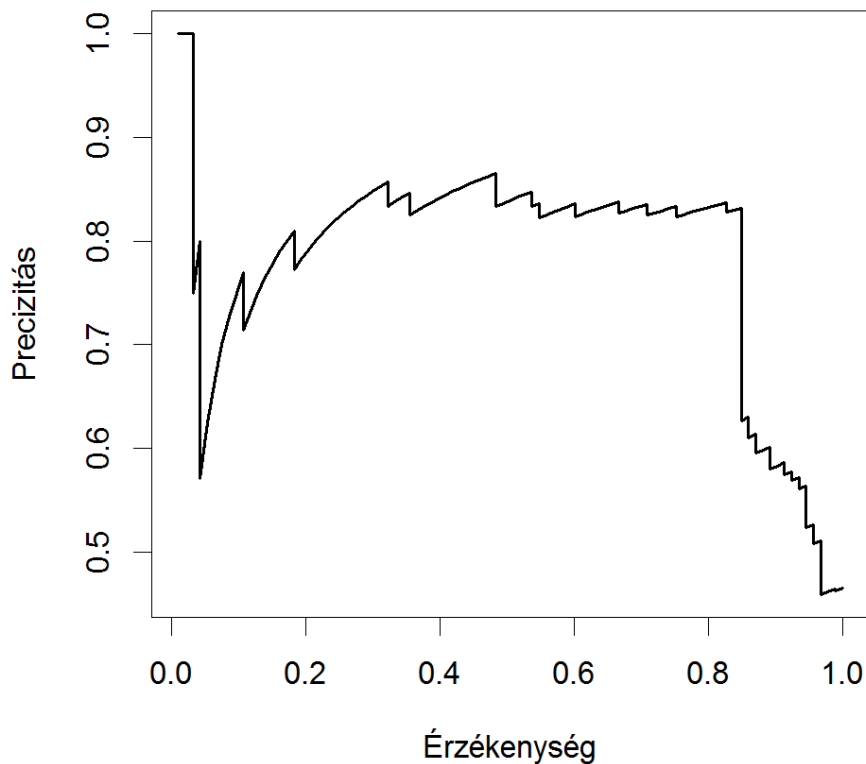
Ezzel az eljárással 10 független becslését adtuk meg az automatikus modelljavító eljárásunknak.

2.2.7. Az eredmények kiértékelése

Precizitás-érzékenység görbe

A különféle számítógépes előrejelzések minőségének jellemzésére egy szemléletes módszer az úgynevezett precizitás-érzékenység görbe készítése (angolul precision-recall, a görbe elnevezésére ugyan nincs bevett magyarítás, de mind a precision, mind a recall szavakra van, így a dolgozatban a görbe magyarított nevét használom). A precizitás (precision) megadja, hogy a szimuláció előrejelzéseinek mekkora része valós eset (ahol a valóság jellemzően egy kísérlet eredménye). Az igazságmátrixból kiindulva a precizitást leírhatjuk a következő egyenlettel: $\text{precizitás} = \text{VP} / (\text{VP} + \text{FP})$. Az érzékenység (recall) pedig a valódi esetek azon része, amelyeket a szimuláció előrejelez. Az érzékenység képlete: $\text{érzékenység} = \text{VP} / (\text{VP} + \text{FN})$. Ha példaként az esszenciális gének előrejelzését nézzük, akkor a precizitás megadja, hogy a szimulációnk által talált esszenciális géneknek mekkora hányada valóban esszenciális, míg az érzékenység megmondja, hogy a valóban esszenciális gének közül mennyit találtunk meg.

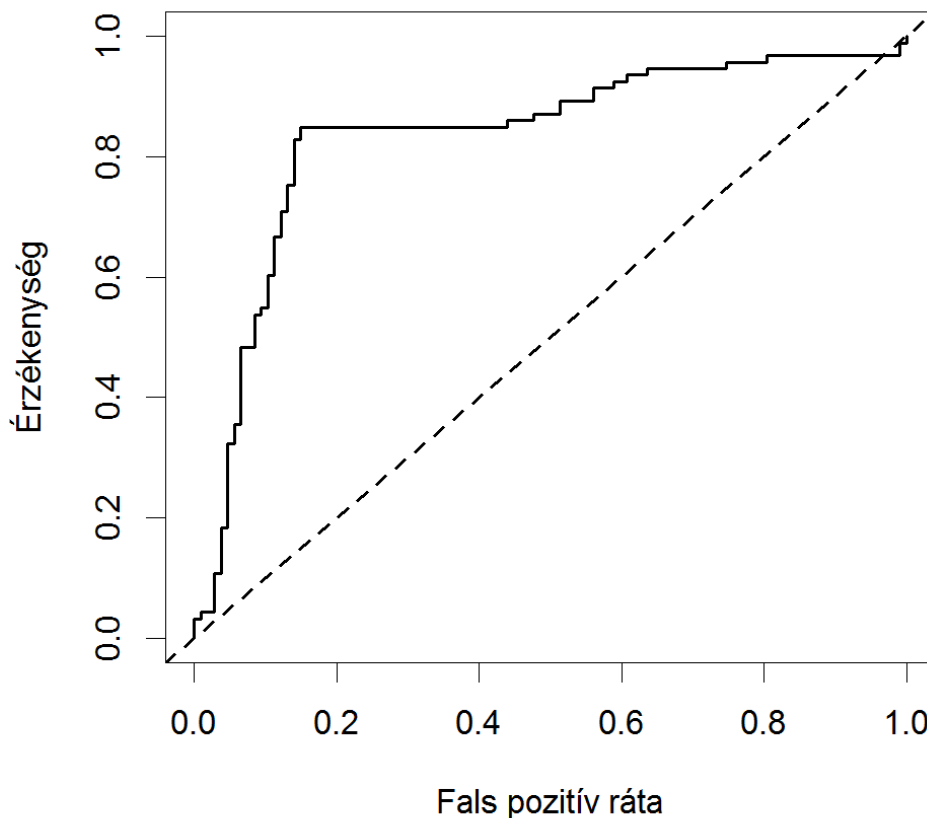
A precizitás-érzékenység görbét úgy vesszük fel, hogy a szimuláció eredményét különböző küszöbértékek mellett binárisra alakítjuk (minden esetre megadjuk, hogy a modell szerint találat vagy sem), majd kiszámoljuk a precizitás és érzékenység értéket minden küszöbértékre és ábrázoljuk (4. ábra). Munkám során a precizitás-érzékenység görbék készítéséhez az ROCR programcsomagot alkalmaztam (Sing et al., 2005).



4. ábra: precizitás-érzékenység görbe szemléltetés. Az ábrán az $ROCR()$ függvény tesztadatai láthatók.

ROC görbe

A vevő működési karakterisztika (Receiver Operating Characteristic, ROC) görbe egy másik grafikus módja a predikciós erő kifejezésének. Az ROC görbe az érzékenység és a fals pozitív ráta közötti kompromisszum megítélésére használható fel (5. ábra). Az érzékenység fogalmát lásd fentebb a precizitás-érzékenység görbe leírásánál, a fals pozitív ráta pedig megadja, hogy azoknak az eseteknek, amelyek a valóságban nem találatok, mekkora részét mondja a szimuláció találatnak. Képlettel felírva: $\text{fals pozitív ráta (FPR)} = \text{FP} / (\text{FP} + \text{VN})$. A precizitás-érzékenység görbéhez hasonlóan itt is különböző küszöbértékek mellett binárisítjuk a szimuláció eredményét. Ha a szimuláció tökéletesen megfelel a valóságnak, akkor az ROC görbén egyetlen pont van a bal felső sarokban (érzékenység = 1, $\text{FPR} = 0$). Annál jobb egy előrejelzés, minél jobban megközelíti az ROC görbéje ezt a pontot. Az átlóban helyezkednek el az előrejelzései egy olyan predikciós eljárásnak, amely véletlenszerűen dönti el, hogy egy génpár tagjai kölcsönhatnak-e egymással.



5. ábra: ROC görbe szemléltetés. Az ábrán az $ROC()$ függvény tesztadatai láthatók. A szaggatott vonal a véletlenszerű előrejelzéseknek felel meg.

2.2.8. Aszpartát oxidáz és kinolinát szintáz homológ gének keresése élesztőben

Escherichia coli baktériumban az aszpartát oxidáz (EC szám: 1.4.3.16) és kinolinát szintáz (EC 2.5.1.72) enzimek alkotják a *de novo* NAD bioszintézis útvonalat (Flachmann et al., 1988). Az enzimek által kódolt reakciók megtalálhatóak az IMM904 élesztő anyagcseremodellben is (Mo et al., 2009), azonban nincs gén hozzájuk rendelve. Annak érdekében, hogy kiderítsük, valóban hiányoznak-e ezek az enzimek az élesztőből, iteratív PSI-BLAST keresést futtattunk (Altschul et al., 1997). A PSI-BLAST algoritmus pozíció-specifikus pontozómátrixot (position-specific score matrix, PSSM) épít fel egy megelőző BLAST keresésben legjobb eredményt elért szekvenciák többszörös illesztéséből, majd a következő illesztéshez a PSSM-et használva iteratíván keres további szekvenciákat, minden lépésben frissítve a PSSM-et. Ezáltal még a szekvenciák közötti gyenge kapcsolatokat is megtalálja. A *nadA* (kinolinát szintáz) és *nadB* (L-aszpartát oxidáz) *E. coli* géneket használtuk fel kiindulási szekvenciának a BLASTP (BLAST Protein) kereséshez. Az E-érték küszöbérték

10 volt (ez az alapértelmezett), az iterációk számát pedig 20-ra állítottuk. Három eltérő módszerrel kerestünk élesztő homológ géneket:

- 1) csak *S. cerevisiae* fehérjeszekvenciák között kerestünk,
- 2) a keresést az összes elérhető organizmus fehérjeszekvenciáin végeztük, majd a végső listából kiválogattuk az élesztő fehérjét,
- 3) a keresést az összes elérhető organizmus fehérjeszekvenciáin indítottuk, majd az első iteráció után elmentettük a PSSM-ot. Ezután egy PSI-BLAST futást indítottunk a lementett PSSM-szal mint kiindulási szekvencia profillal és kizárólag az élesztő szekvenciák között kerestünk.

2.3. Eredmények

A genetikai interakciós hálózat tulajdonságait átfogó kísérletes munkák során leírták ugyan (Costanzo et al., 2010; Tong et al., 2004), de ezek a munkák nem láttak bele a kapott mintázatokért felelős mechanizmusokba. Ezt a hiányosságot mi a genetikai interakciók genomléptékű elméleti modellezésével kívántuk feloldani. Emellett a kísérletes adatokat felhasználva első alkalommal teszteltük, hogy a modell milyen mértékben képes előrejelezni a genetikai interakciókat. A modellt egy automatizált eljárással feljavítottuk és egy javasolt változtatást kísérletesen is ellenőriztünk.

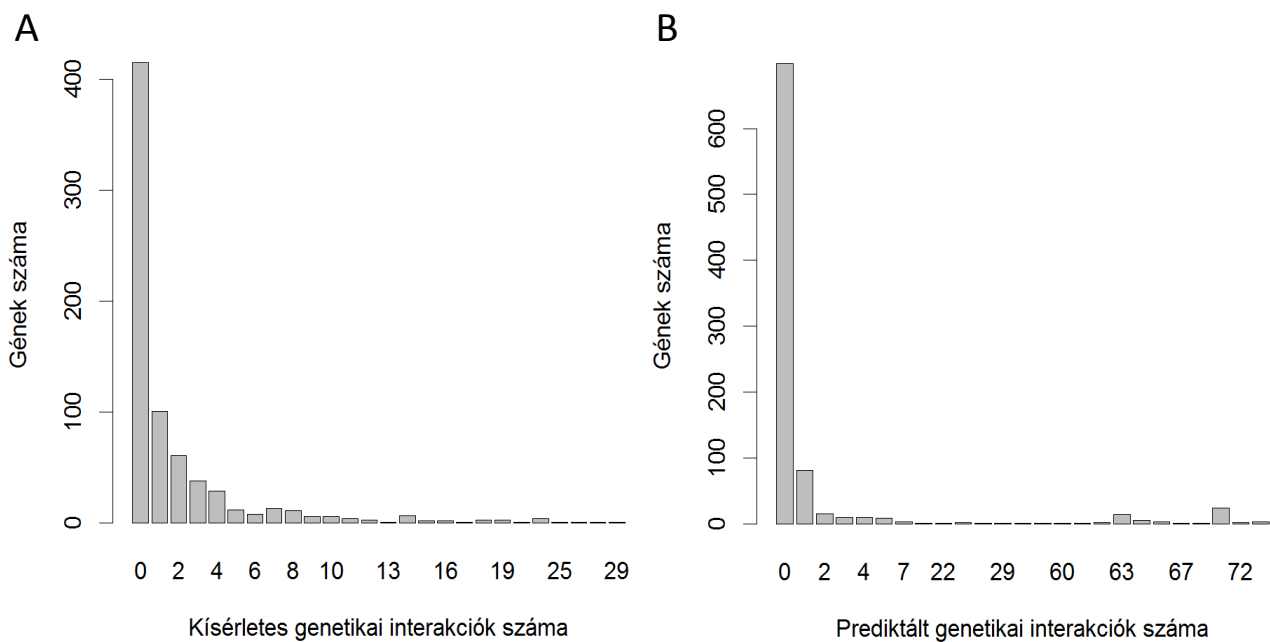
2.3.1. A gének genetikai interakciós fokszáma szorosan összefügg a gén fontosságával és pleiotrópiájával

A genetikai interakciós hálózat jellemző tulajdonsága, hogy a gének nagy részének kevés kapcsolata van más génekkel, míg a gének egy kis része felelős a kapcsolatok többségéért (6/A ábra). Emellett Costanzo és munkatársai kimutatták, hogy egy gén kiütésének hatása az élőlény fitneszére összefüggésben áll a gén által mutatott genetikai interakciók számával (Costanzo et al., 2010). Minél több interakcióban vesz részt egy gén, annál károsabb a kiütése az élőlény számára. Ez mind a pozitív, mind a negatív interakciók esetében igaz. Erre a jelenségre eddig nem sikerült magyarázatot találni. Egy lehetséges magyarázat, hogy egy olyan génkiütés, amely nagymértékben káros az élőlényre nézve, képes elfedni több kevésbé káros génkiütés hatását is, ezáltal pedig a gének között pozitív interakciót látunk (ún. masking kölcsönhatásokat). Ezt a logikát követve azonban a negatív interakciók hiányát várnánk a káros génkiütések esetében, ugyanis a fitnesz már csak kis mértékben tud tovább romlani egy újabb génkiütés hatására. Ez a kép azonban nem felel meg a valóságnak, ugyanis a negatív genetikai kölcsönhatások is feldúsulnak az olyan gének körül, melyek kiütése jelentős hátrányt jelent.

Az anyagcseremodell vizsgálata során először arra voltunk kíváncsiak, hogy a modell képes-e visszaadni a genetikai interakciós hálózat fentebb leírt tulajdonságait. Ennek vizsgálatához első lépésben kiszámoltuk az egyszeres génkiütések hatását és a genetikai interakciók számát az anyagcseremodell génjeire az FBA eljárás segítségével.

A genetikai interakciók számának eloszlása a modell esetében nagyon hasonlít ahhoz, amit a kísérletes adatoknál láttunk. A gének alig ~12 százaléka felelős az interakciók ~85 százalékáért (6/B ábra). Az ábrán azonban az is látszik, hogy az előrejelzett kölcsönhatások

száma bizonyos génekre sokkal nagyobb, mint a kísérletes megfigyelt. A különbség abból adódik, hogy a kísérletes adatokkal összehasonlítva az anyagcseremodell lényegesen több gyenge pozitív genetikai kölcsönhatást mutat. A kísérletes módszerek nem elég érzékenyek, hogy ilyen gyenge pozitív kölcsönhatásokat kimutassanak, emiatt az még nyitott kérdés, hogy vajon ezek a kölcsönhatások tényleg jelen vannak a gének között, vagy ez a jelenség a modell sajátossága.



6. ábra: a kísérletesen detektált (A) és a modell által előrejelzett (B) genetikai interakciók számának eloszlása. Az ábrán a negatív és pozitív interakciók összege van ábrázolva.

A modell emellett erős negatív korrelációt mutat az egyes gének kiütésének fitneszhatása és a gének kölcsönhatásainak száma között, mind a negatív, mind a pozitív genetikai interakciók esetében (Spearman-féle rangkorreláció, $\rho = -0.89$ a pozitív és $\rho = -0.66$ a negatív génkölcsönhatásokra). Ez a mintázat akkor is megmarad, ha kizárjuk az elemzésből azokat a géneket, melyek kiütése semmilyen hatással nincs a modell fitneszére (Spearman $\rho = -0.59$, p érték $<10^{-3}$ a pozitív kölcsönhatásokra és $\rho = -0.47$, p érték = 0.005 a negatív interakciókra, 7. ábra). Ebből következik, hogy a mintázatért nem ezek a „csendes” gének a felelősek.

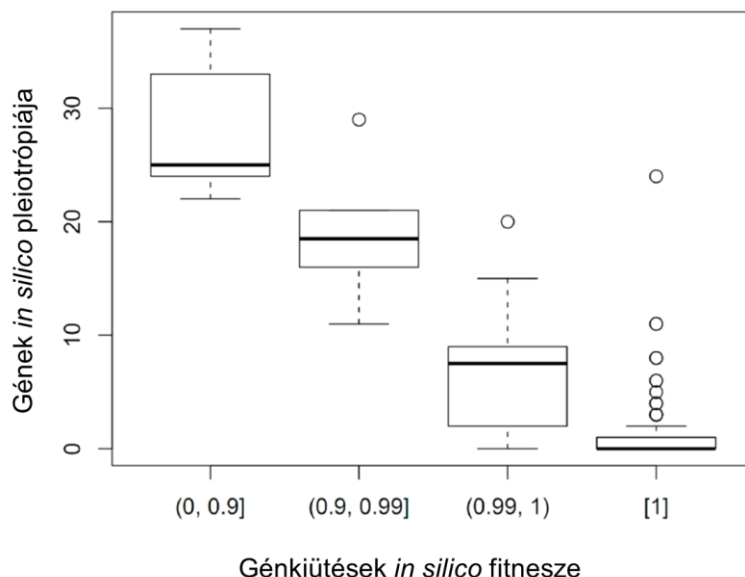


7. ábra: a genetikai kölcsönhatások száma és a génkiütések fitnesszhatása közötti összefüggés az anyagcseremodell alapján. Az egyes értékek rang transzformált alakban szerepelnek.

Az anyagcseremodellek egyik legnagyobb előnye a rendszerbiológiai vizsgálatok során, hogy képesek a különböző jelenségekre mechanisztikus magyarázatot adni. A modell e tulajdonságát mi arra akartuk felhasználni, hogy magyarázatot találjunk a gének genetikai interakciós fokszáma és a kiütés fitnessze közötti összefüggésre. Egy lehetséges magyarázat arra, hogy egy bizonyos génnek sok genetikai kölcsönhatása van, az lehet, hogy a gén nagyszámú biológiai folyamatban vesz részt (erősen pleiotróp), emiatt a kiütésének hatását sok más gén kiütése befolyásolhatja pozitív vagy negatív irányba. Egy korábbi munkában már valóban kimutatták, hogy a genetikai interakciós hálózatban csomópontként szereplő gének gyakran több biológiai funkcióban is részt vesznek (Costanzo et al., 2010). Ha a sokfunkciós gének kiütésének (átlagosan) nagy a fitnesszhatása, akkor negatív korrelációt várunk a génkiütéses mutánsok fitnessze és a genetikai interakciós foksám között. A gének pleiotrópiája összetett tulajdonság, amelyet nehéz empirikusan vizsgálni, azonban az FBA eljárás lehetővé teszi a számítógépes vizsgálatát. Ehhez meghatároztuk, hogy egy adott gén kiütése hány, a sejt növekedéséhez szükséges metabolit (biomassza összetevők) megtermelését korlátozza (Shlomi et al., 2007a).

Az elemzés kimutatta, hogy a hipotézisnek megfelelően erős negatív korreláció van egy gén által érintett bioszintetikus folyamatok száma (érintett biomassza összetevők) és a gén kiütésének hatása között (Spearman $\rho = -0.83$, p érték $<10^{-9}$; a korábbi elemzéseknek megfelelően csak azokat a géneket vizsgáltuk, amelyek kiütésének hatása van a fitnesszra; 8.

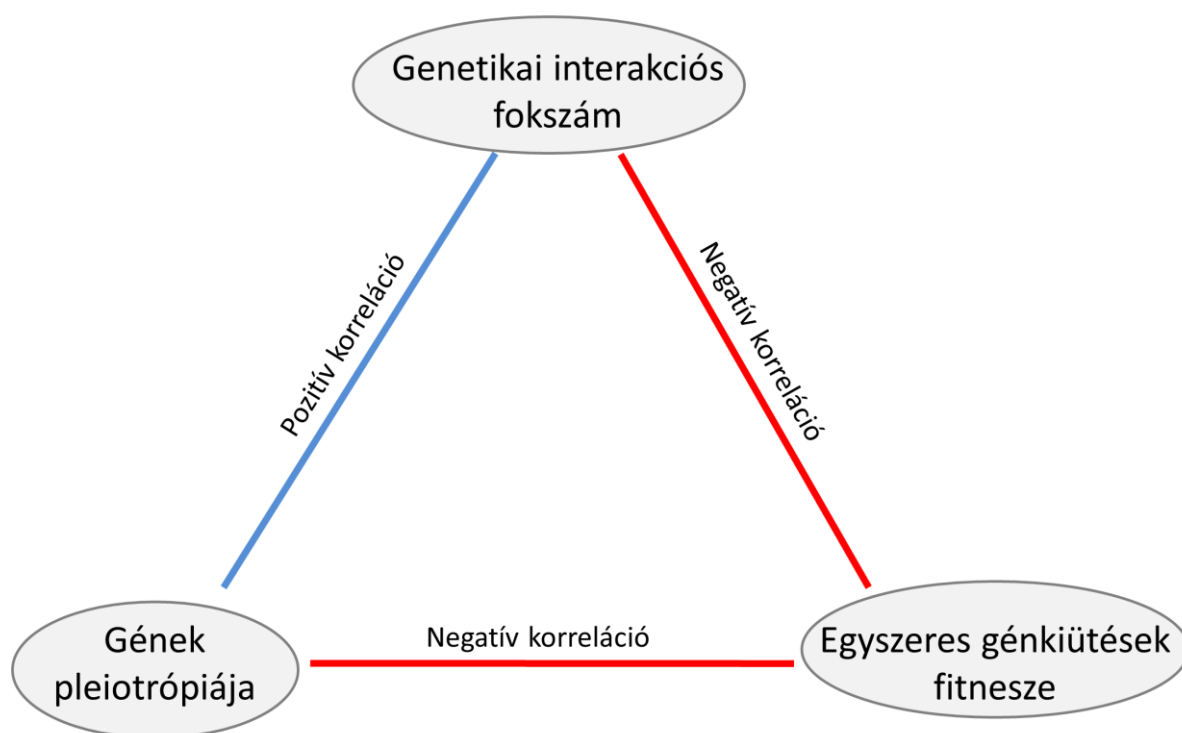
ábra). Ezen kívül a pleiotrópia és a genetikai kölcsönhatások száma is korrelációban van. Ez a mintázat mind a kísérletes (Spearman $\rho = 0.55$ a negatív és $\rho = 0.24$ a pozitív interakciókra, p érték $<10^{-8}$ mindkét esetre), mind a modell segítségével szimulált (Spearman $\rho = 0.62$ a negatív és $\rho = 0.25$ a pozitív interakciókra, p érték $<10^{-8}$ mindkét esetre) genetikai interakciók esetében fennáll.



8. ábra: a gének kiütésének *in silico* fitnesszhatása összefüggésben van az anyagcseremodell alapján számolt pleiotrópiával (egy gén pleiotrópiája egyenlő azon biomassza komponensek számával, amelyek bioszintézisét a gén kiütése gátolja).

A pleiotrópia és az *in silico* egyszeres mutáns fitness között látott korreláció arra sarkallt minket, hogy együttesen is megvizsgáljuk ezen géntulajdonságok hatását a genetikai kölcsönhatásokra. Ahogy korábban láttuk, a pleiotrópia és a genetikai interakciós foksám negatívan korrelál az egyszeres génkiütések fitnessével, míg egymás között pozitív korrelációt mutatnak (lásd 9. ábra). A további vizsgálatokhoz parciális korrelációs elemzéseket végeztünk (a parciális korreláció két változó korrelációja, miután bizonyos további változók hatását többszörös lineáris regresszióval kiküszöböltük (Reiczigel et al., 2007)). Parciális Spearman rang korrelációs együtthatókat számoltunk két változó között, miközben kontrolláltunk a harmadik változó hatására. A vizsgált változók egyike mindig valamilyen genetikai kölcsönhatás (negatív vagy pozitív), a másik pedig vagy a pleiotrópia (ekkor az egyszeres mutáns fitnessre kontrolláltunk), vagy az egyszeres mutáns fitness (ekkor

a pleiotrópiára kontrolláltunk). A számításokhoz a `pcor.test` függvényt alkalmaztuk, amelyet az R programnyelvben implementáltak (Kim, 2012; R Core Team, 2013).



9. ábra: összefüggés a genetikai interakciós foksám, a gének pleiotrópiája és a génkiütések fitnesze között. A kék vonal pozitív, míg a piros vonal negatív korrelációt jelez.

Az elemzésből ezúttal is kihagytuk azokat a géneket, melyek kiütésének nincs fitneszhatása, ezeknek a géneknek ugyanis a pleiotrópiája is nulla (nincsenek hatással egyetlen biomassza komponensre sem) és jellemzően nem mutatnak genetikai interakciót más génekkel, ennélfogva hamis korrelációt okozhatnak. Továbbá gének bizonyos csoportjai mindig azonos fenotípust mutatnak a szimulációk során, emiatt ezekre a génekre nem lehet független adatpontként tekinteni a statisztikai elemzéseknél (például közös fehérje komplexben lévő gének, illetve olyan gének, amelyek által kódolt reakciók rátája kapcsolt a hálózatban (pl. egy lineáris útvonal, ahol egy köztes metabolitot csak egy reakció termel és csak egy reakció fogyaszt)). Hogy elkerüljük az ebből fakadó torzulást az eredményekben, minden korrelált géncsoportot egyetlen, véletlenszerűen kiválasztott gén képviselt az elemzésekben. A szűrések eredményeképpen 34 gén maradt.

A parciális korrelációs elemzések feltárták, hogy bár a modellben a pozitív interakciós foksámot a génkiütések hatása határozza meg (ez összhangban áll azzal a hipotézissel, hogy

a nagy káros mutációk elfedhetik sok, kevésbé káros mutáció hatását), a negatív interakciós fokszámot a pleiotrópia okozza (2. táblázat).

	Negatív interakciós fokszám		Pozitív interakciós fokszám	
	Spearman parciális ρ	P-érték	Spearman parciális ρ	P-érték
Korreláció az interakciós fokszám és a pleiotrópia között kontrollálva az egyszeres mutáns fitnessre	0.447	0.005	-0.112	0.531
Korreláció az interakciós fokszám és az egyszeres mutáns fitness között kontrollálva a pleiotrópiára	0.089	0.620	-0.451	0.005

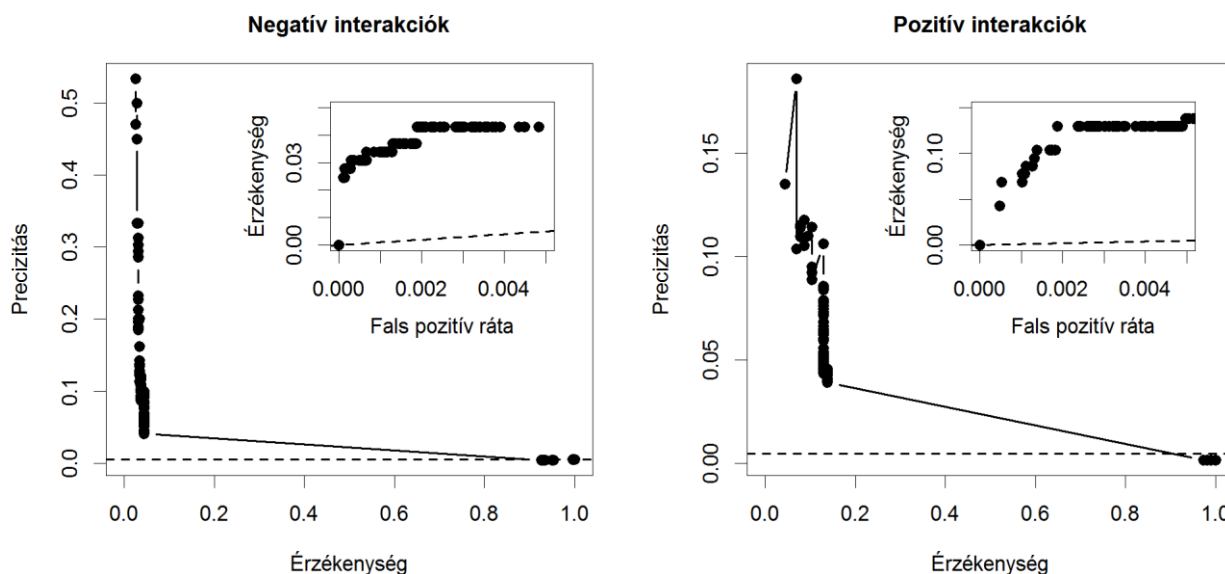
2. táblázat: parciális korrelációs elemzés a modell által előrejelzett genetikai interakciós fokszám, az egyszeres mutáns fitness és a pleiotrópia között.

2.3.2. Az anyagcseremodell genetikai interakciókat előrejelző ereje

A Boone csoporttól származó átfogó kísérletes genetikai interakciós adatsor páratlan lehetőséget nyújt arra, hogy megvizsgáljuk, milyen mértékben képes az anyagcseremodell az FBA segítségével előrejelezni a genetikai kölcsönhatásokat. Az adatsor teljes, tehát nemcsak azt tudjuk, hogy mely génpárok között van kölcsönhatás, de azt is, hogy melyek között nincs.

A szimuláció során kizártuk az elemzésből az esszenciális géneket és az olyan géneket, amelyek kizárólag blokkolt reakciót kódolnak. Az úgynevezett blokkolt reakciókon az FBA során nem mehet át fluxus, különben a modell nem képes növekedésre (ennek oka lehet, hogy az adott reakció csak más környezetben való növekedés során kap szerepet, de az is, hogy a reakció az anyagcserének egy kevésbé feltérképezett részén helyezkedik el, ahonnan még hiányoznak további reakciók (Burgard et al., 2004)). Így a modell és a kísérletes adatsor között összesen 67 517 nem-esszenciális génpár egyezik meg, melyek között 325 negatív és 116 pozitív kísérletes genetikai kölcsönhatás található. Az FBA segítségével kiszámoltuk, hogy mely génpárok mutatnak genetikai kölcsönhatást, és azt találtuk, hogy közöttük erős feldúsulást mutatnak a kísérletes interakciók. A negatív interakciók 100-szoros feldúsulást mutatnak a véletlenszerűen várthoz képest (50%-os precizitás mellett), míg a pozitív interakciók 60-szorost (a precizitás 11%, 10. ábra). A precizitás értékét erre az esetre úgy

értelmezhetjük, hogy például a negatív interakciók esetében a szimuláció által kölcsönhatásban lévőnek mondott génpárok fele a kísérletes adatok szerint is kölcsönhatásban van. Mivel a genetikai interakciók ritkák, ezek rendkívül magas predikciós sikernek számítanak (ezért kapunk 100x feldúsulást).



10. ábra: A kísérletesen megfigyelt és számítógéppel számolt genetikai interakciók összevetése precizitás-érzékenység és ROC görbékkel (utóbbi beágyazva) különböző *in silico* interakciós küszöbértékek mellett. A szaggatott vonal jelöli a véletlen előrejelzést figyelembe véve a genetikai interakciók gyakoriságát a kísérletes adatok alapján. A precizitás értékek más skálán helyezkednek el a két ábrán. Az ROC ábrák részlegesek, csak a görbék bal alsó sarkát mutatják, ahol az adatpontok elhelyezkednek.

Ezek az eredmények alátámasztják, hogy az *in silico* előrejelzett genetikai interakciók biológiailag relevánsak (Harrison et al., 2007), azonban a kísérletesen látott kölcsönhatásoknak csak egy kis hányadát találjuk meg a szimulációval: az érzékenység 2.8% a negatív és 12.9% a pozitív interakciókra. Ezen az eredményen az sem változtatott, ha az FBA helyett egy alternatív algoritmust használtunk (Segrè et al., 2002), más képlettel számoltuk a genetikai kölcsönhatást (Segrè et al., 2005), vagy egy kevésbé kompartmentalizált élesztő modellt használtunk (Kuepfer et al., 2005). Érdemes megjegyezni, hogy kísérletesen genetikai kölcsönhatást mutató génpároknak csak elenyésző hányada mutat ellentétes irányú kölcsönhatást a szimulációk során, tehát az alacsony érzékenység értékek

oka nem a genetikai interakciók hibás besorolása, hanem az, hogy a szimuláció az interakciók nagy részét nem találja meg.

Felmerülhet a kérdés, hogy vajon mennyire megbízhatóak a kísérletes adatok? Az SGA egy nagyskálájú módszer lévén szükségszerűen kevésbé pontos, mint azok a vizsgálatok, amelyek csupán néhány gén közötti kölcsönhatás mérésére összpontosítanak. A korábban, főleg esettanulmányok során azonosított negatív interakciókkal összehasonlítva valóban azt látjuk, hogy a kísérletes adatsor érzékenysége 35%, tehát az SGA adatsor a más forrásból származó negatív genetikai interakciók 35%-át visszaadja. Az erős negatív és pozitív interakciók azonban nagymértékben feldúsultak a korábban azonosított kölcsönhatásokban (140-szeres feldúsulás a negatív és 80-szoros feldúsulás a pozitív interakciók esetében (Costanzo et al., 2010)).

A genetikai kölcsönhatások előrejelzésére léteznek statisztikai módszerek is, melyek számos különböző biológiai információt használnak fel az interakciók előrejelzésére (Wong et al., 2004; Ulitsky et al., 2009). Az egyik munkában genomikai és proteomikai adatokat integráltak egy statisztikai modellbe és 80%-os érzékenységgel sikerült előrejelezniük a negatív genetikai interakciókat az általuk használt kísérletes adatsorhoz viszonyítva. Azt találták, hogy meglepő módon a szekvencia homológia csak az interakcióban lévő génpárok 3%-ára volt jellemző, és a negatív genetikai interakciót leginkább az jelezte előre, ha két gén ugyanazt a fenotípust mutatta kiütés esetén, ha fizikai interakcióban voltak, ha azonos funkciót töltöttek be, vagy ha hasonló volt a génpár körüli lokális hálózati topológia (Wong et al., 2004). Ezek a módszerek azonban a meglévő genetikai interakciós adatokat is felhasználják az előrejelzéshez, így csak olyan organizmusok esetében használhatóak, ahol kellő mennyiségű genetikai interakciós adat áll rendelkezésre, szemben az anyagcseremodellel (Wong et al., 2004; Ulitsky et al., 2009). Csoportunk ezért megvizsgálta, hogy a genetikai interakciós adatok kizárásával, génpárok funkcionális genetikai és más jellegű tulajdonságai alapján milyen mértékben lehet előrejelezni a génkölcsönhatásokat statisztikai módszerekkel. Ezeket a vizsgálatokat Kovács Károly munkatársam végezte (Szappanos et al., 2011). Statisztikai módszerünk az FBA-nál jobb precizitás és érzékenység értékekkel rendelkezett, de még mindig csak a kísérletesen negatív kölcsönhatások 30 százalékát, míg a pozitív interakciók 25 százalékát találta meg. Ebből levonhatjuk a következtetést, hogy a genetikai interakciók többségét sem a kényszer-alapú modellezés, sem a statisztikai módszerek nem képesek visszaadni. Ugyanakkor az FBA segítségével számolt mutáns fitness és genetikai interakciós adatok statisztikai modellbe építése javítja a negatív

genetikai interakciók előrejelzését, tehát a kényszer-alapú modellek tartalmaznak olyan információt, amelyhez nem lehet pusztán statisztikai módszerekkel hozzájutni (Szappanos et al., 2011).

Több oka is lehet, hogy a kényszer-alapú modellezés az interakciók nagy részét nem találja meg. Az egyik lehetséges magyarázat, hogy az egyszeres génkiütések hatásának az előrejelzése nem tökéletes (Mo et al., 2009; Snitkin et al., 2008), így előfordulhat, hogy két nem-esszenciális gén kölcsönhatását azért nem látjuk, mert az egyik gént esszenciálisnak mutatja a modell. A kísérletes adatokat összevetve a modell előrejelzéseivel azt látjuk, hogy a negatív interakciók ~24%-ában, míg a pozitív interakciók ~22%-ában az egyik gén valóban esszenciális *in silico*. A hibásan esszenciálisnak mutatott gének kizárásával ugyan nő az érzékenység, de az interakciók többségét még mindig nem jelzi előre a modell.

Egy másik magyarázat az lehet, hogy a gyenge *in vivo* interakciókat az anyagcseremodell kevésbé tudja visszaadni. Az érzékenység valóban sokat javul, ha csak az erős kísérletes kölcsönhatások előrejelzését nézzük (~17% ha $\epsilon \leq -0.5$ és 25% ha $\epsilon \geq 0.15$), azonban még az erős interakciók többsége is rejtve marad a modell elől.

Az FBA az anyagcsere optimális működését feltételezi, és nem vesz figyelembe regulációs mechanizmusokat a szimuláció során. Így azonban olyan gének, amelyek *in vivo* gátlás alatt állnak, a modellben kompenzálni tudják más gének kiütését. Ennek hatására a modell alábecsülheti a mutációk hatását. Ez mind az egyszeres, mind a kettős génkiütéseket érinti, és ezt igazolja az is, hogy a modell csak a kísérletes negatív interakciók 3.7%-ára mutat erős negatív kölcsönhatást. Ennek a jelenségnek a vizsgálatához egy transzkriptomikai adatsort használtunk fel (Nagalakshmi et al., 2008), hogy meghatározzuk azokat az anyagcsere géneket, melyeknek az expressziója gátolt a vizsgált környezetben, majd a génekhez tartozó reakciók aktivitását nullára kényszerítettük (Åkesson et al., 2004). Ez a lépés azonban gyakorlatilag nem változtatta az érzékenységet a szimulációnak, ami arra utal, hogy a jövőben szükség lesz ennél kifinomultabb módszerekre. Egy lehetséges megoldás a regulációs FBA módszer alkalmazása, melynek során a regulációs hatások, mint logikai kapcsolók be vannak építve az FBA eljárásba (Shlomi et al., 2007b). A modell ugyancsak nem veszi figyelembe, hogy egyes közti-termékek toxikusak lehetnek a sejt számára. Ha egy szubsztrát feldúsulása toxikus a sejt számára, akkor a szubsztrátot fogyasztó enzimek kiütése ezen a toxicitáson keresztül is csökkenti az élőlény fitneszt, a modellben azonban a metabolitok koncentrációja

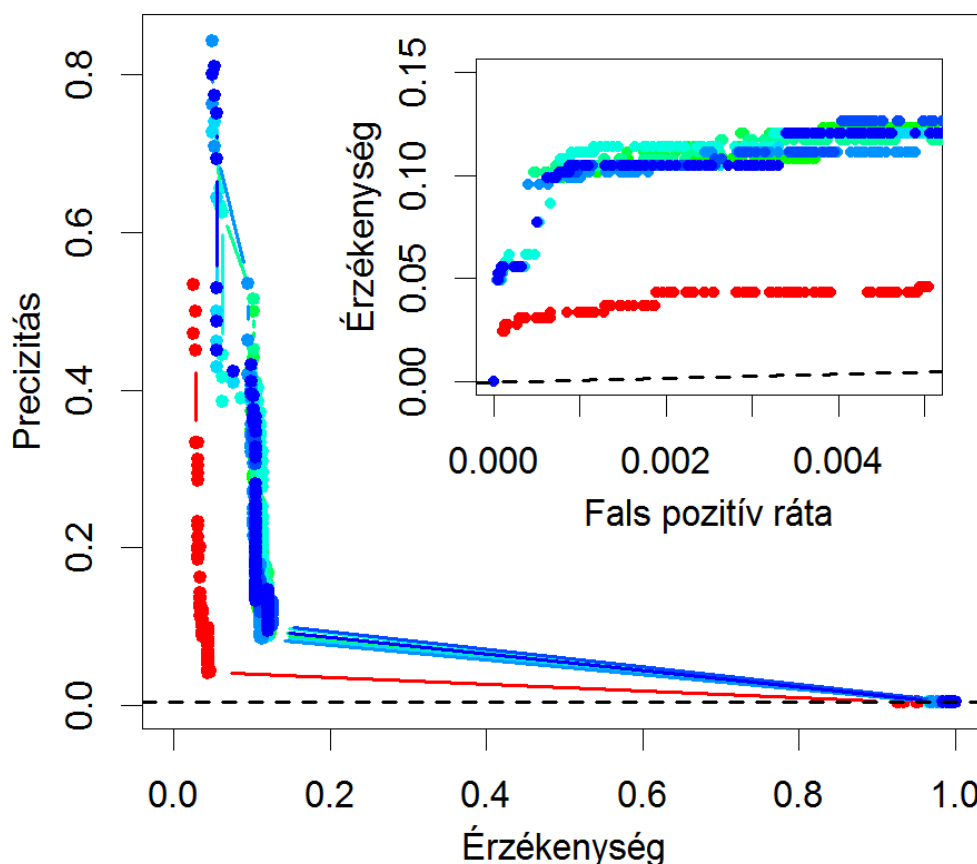
nincsen definiálva (a steady-state miatt a koncentrációk állandóak), így a többlettermelést sem tudja szimulálni.

Ahogy az anyagcseréről alkotott ismereteink, úgy az anyagcsere rekonstrukciója is szükségszerűen nem teljes. A hiányos tudás és az esetlegesen hibás annotáció szintén felelősek lehetnek a fals előrejelzések egy részéért.

2.3.3. Az anyagcseremodell predikciós ereje automatizált módszerrel feljavítható

Ahogy korábban láttuk, az anyagcseremodell a genetikai kölcsönhatások egy jelentős részét nem képes előrejelezni. Annak érdekében, hogy javítsunk a modell szimulációs erején, kifejlesztettünk egy gépi tanuláson alapuló eljárást. Módszerünk automatikusan képes hipotéziseket gyártani, hogy megmagyarázza a gének között fellépő negatív genetikai interakciókat. A modellen különböző módosításokat hajt végre, majd a módosítások hatását globálisan teszteli.

8 alkalommal futtattuk le javító algoritmusunkat az élesztő anyagcseremodellen. Automatizált módszerünk több változtatást is javasolt a modellen (Függelék: 4. táblázat), amelyek jelentősen javították az előrejelzések sikerességét: az érzékenység 100-267 %-kal, míg a precizitás 44-59 %-kal nőtt (11. ábra). Bár módszerünk a pozitív genetikai kölcsönhatásokat nem veszi figyelembe, a javasolt változtatásokat leellenőriztük, és nem befolyásolják az előrejelzés sikerességét a pozitív interakciókra. A javasolt módosítások között volt több is, amelyek mind a 8 futás alkalmával feljöttek. Ezek egy része a NAD bioszintézis útvonalat érintette, amit a következő fejezetben bővebben kifejtünk, de emellett feljött még egy változtatás, amely kiveszi a glikogént a növekedéshez létfontosságú metabolitok köréből. Ez a módosítás csökkenti a modellt alkotó kényszerek listáját (nem szükséges glikogént termelni a túléléshez), emiatt nem új negatív genetikai interakciókat vezet be, hanem két, hibásan jelzett interakciót szüntet meg.



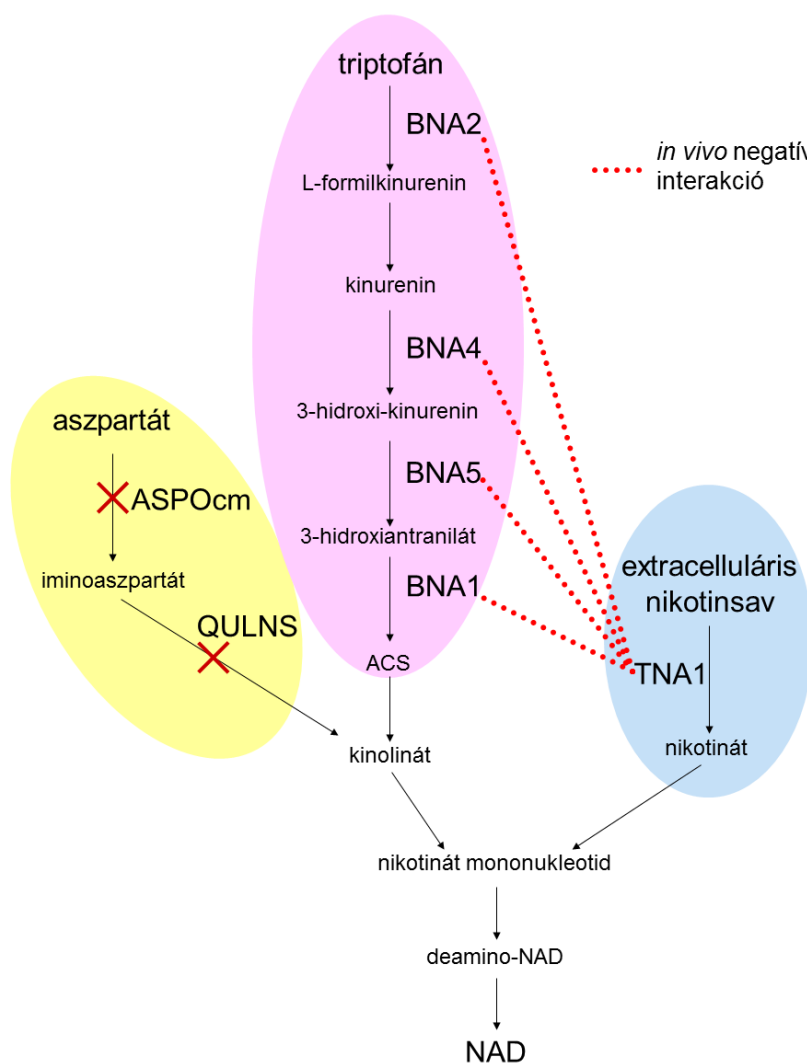
11. ábra: 8 független modell-optimalizáció eredménye. Precizitás-érzékenység görbék és ROC görbék (beágyazva) mutatják a különbséget az eredeti (piros) és a feljavított (kék és zöld, összesen 8 görbe) modellek között. A szaggatott vonalak a véletlen előrejelzést mutatják.

Keresztvalidálás segítségével megvizsgáltuk, hogy az általunk kifejlesztett módszer hogyan teljesít a modell-optimalizációhoz nem használt adatokon. 10-szeres keresztvalidálást végeztünk, tehát 10 alkalommal osztottuk két részre az adatokat tanuló és tesztelő adatszetekre. A teszt szetten leértük a modell MCC értékét javítás előtt és után, majd Wilcoxon-féle előjeles rangpróbával összehasonlítottuk őket, és azt találtuk, hogy módszerünk szignifikánsan javítja a modell negatív kölcsönhatás előrejelző képességét olyan adatokon is, amelyeket nem használtunk fel a modell javításához (p érték <0.002). A keresztvalidálás szerint a negatív interakciók előrejelzésének érzékenysége átlagosan $\sim 87\%$ -kal nőtt.

2.3.4. A NAD bioszintézis útvonalak kijavítása az anyagcseremodellben

A legjelentősebb javítást az algoritmusunk a NAD bioszintézis módosításával érte el, ahol egy vagy két reakció megváltoztatásával négy új, kísérletesen is látott negatív genetikai

kölcsönhatás jelent meg. A szimulációhoz használt élesztő anyagcseremodell (Mo et al., 2009) három NAD bioszintézis útvonalat tartalmaz: I) kinurenin útvonal, ami triptofánból indul ki II) nikotinsav útvonal, ami az extracelluláris nikotinsavat alakítja át, és III) aszpartát útvonal, ami aszpartátból indul ki. A kísérletes adatok azt mutatják, hogy a kinurenin és nikotinsav útvonal génjei kompenzálni tudják egymást, ezt azonban a modell nem mutatja, mert az aszpartát útvonal kompenzálja a kettős mutánsok fitneszhatását. Az algoritmus az aszpartát útvonal törlését javasolta, mi pedig utánajártunk, hogy ez a hipotézis megállja-e a helyét (12. ábra).



12. ábra: NAD bioszintézis útvonalak az élesztő anyagcseremodellben. A kinurenin útvonal génjei negatív kölcsönhatást mutatnak *in vivo* a nikotinsav transzporter génjével, ezt azonban nem látjuk *in silico*. Az automatikus modelljavító eljárás az aszpartát útvonal reakcióinak törlését javasolta (ASPOcm, aszpartát oxidáz; QULNS, kinolinát szintáz).

Az aszpartát útvonal két reakcióból áll, és *Escherichia coli*-ban az aszpartát oxidáz (NadB) és kinolinát szintáz (NadA) enzimek kódolják. Az *E. coli* enzimekből kiindulva homológ fehérjéket kerestünk PSI-BLAST algoritmussal (Altschul et al., 1997) *Saccharomyces cerevisiae*-ben. A NadA fehérjének nem találtuk homológját élesztőben. A NadB fehérje feltekeredési topológiája hasonló más oxidoreduktáz fehérjékhez (szukcinát dehidrogenáz és fumarát reduktáz (Mattevi et al., 1999)), ennek megfelelően a szekvencia keresés legerősebb élesztő találatai szukcinát dehidrogenáz és fumarát reduktáz fehérjék voltak (pl. *SDH1*, *OSM1*, *FRD1*, stb.). Ezeken kívül találtunk két gént (*CIR2* és *HER1*) nagyon alacsony szekvencia hasonlósággal, melyeknek nem ismert a funkciója. Az eredményekből arra következtetünk, hogy bár homológia keresésünk ugyan talált az aszpartát oxidázzal rokon fehérjéket élesztőben, azonban aszpartát oxidáz aktivitású fehérje ezek között nem volt, így nincs bizonyíték arra, hogy jelen lenne L-aszpartát oxidáz gén az élesztő genomában.

A következő lépésben olyan specifikus előrejelzéseit kerestük a modellnek, amelyeket kísérletesen tesztelhetünk és igazolhatják, hogy az aszpartát útvonal valóban hibásan szerepel az anyagcseremodellben. Azt találtuk, hogy csak a javított modell jelzi esszenciálisnak a kinurenin útvonal génjeit (*BNA1*, *BNA2*, *BNA4* és *BNA5*) ha hiányzik a nikotinsav a környezetből. Ezeket az előrejelzéseket kísérletesen teszteltük (a labormunkát Szamecz Béla munkatársam végezte (Szappanos et al., 2011)) és kimutattuk, hogy az említett 4 gén valamelyikének kiütésével valóban nikotinsav-függővé tehető az élesztő. Minden eredmény arra utal tehát, hogy az aszpartátból kiinduló NAD bioszintézis útvonal élesztőben nincs jelen (lásd még (Panozzo et al., 2002)).

2.4. Diszkusszió

A kényszer-alapú anyagcsere-hálózat rekonstrukció lehetőséget ad a genetikai kölcsönhatások rendszerszintű vizsgálatára és a kölcsönhatások mögött álló mechanizmusok feltárására. Mindeddig azonban nem volt olyan nagy-skálájú kísérletes adatsor, amelyen ellenőrizni tudtuk volna a modell által szolgáltatott előrejelzéseket (Harrison et al., 2007). Charles Boone és csoportja azonban a közelmúltban kidolgozta az SGA kísérletes módszert, amellyel vizsgálhatóvá vált a nem-esszenciális génpárok között fellépő teljes interakciós tér a sörélesztőben (Costanzo et al., 2010). Az általuk biztosított kísérletes adatok felhasználásával megvizsgáltuk, milyen mértékben képes az élesztő genomléptékű anyagcseremodellje előrejelezni a genetikai interakciókat, illetve a genetikai kölcsönhatások hálózatában található mintázatokat, valamint a modell segítségével megkíséreltük megmagyarázni a kapott eredményeket.

Egy, a kísérleti adatokon látott érdekes mintázat a genetikai interakciós foksám (azaz egy gén hány genetikai interakcióban vesz részt) és az adott gén kiütésének fitneszhatása közötti összefüggés (Costanzo et al., 2010). Ezt a mintázatot az anyagcseremodell is mutatja. Emellett azonban a modell magyarázatot is tudott adni a jelenségre: segítségével kimutattuk, hogy mind a negatív genetikai interakciós foksám, mind a génkiütések fitneszhatása összefügg a génnek pleiotrópiájával (azzal, hogy egy génnek hány biológiai funkcióban van szerepe). Ez alapján azt feltételezzük, hogy a több biológiai funkcióban részt vevő (pleiotrópikus) génnek kiütésének hatását nagyszámú más gén mutációja befolyásolja, ahol az egyes gének a pleiotrópikus gén más-más funkcióját kompenzálják.

Az anyagcseremodell nem csak a nagyobb léptékű mintázatokat képes előrejelezni a genetikai kölcsönhatások hálózatában, hanem magukat a génpárok között fellépő egyes interakciókat is. A modell által előrejelzett genetikai kölcsönhatások az esetek felében valódi találatok, ahogy azt a kísérletes adatokkal való összevetésünk alátámasztotta. Figyelembe véve, hogy mennyire ritkák a genetikai kölcsönhatások (a génpároknak csak 0.5%-a mutat negatív interakciót), ez az érték kimagaslóan jónak számít. Ugyanakkor a valódi kölcsönhatásoknak a modell csak kis hányadát találja meg. Ennek több oka is lehet. Egyrészt maga az anyagcsere-hálózat rekonstrukciója sem teljes, illetve az annotációban is előfordulhatnak hibák. Másrészt a modell nem veszi figyelembe a reguláció különböző szintjeit (génexpresszió, metabolikus reguláció), holott ezeknek a folyamatoknak fontos szerepük van az anyagcsere működésében (Heinemann and Sauer, 2010). Emellett előfordulhatnak olyan genetikai kölcsönhatások is,

amelyeket azért nem lát az FBA módszer, mert steady-state állapotot feltételez, például káros intermediér molekulák akkumulációja. Ezeket az eseteket a nagyobb léptékű kinetikai modellek fogják feloldani, ha majd elérhetővé válnak a jövőben. Emellett léteznek már olyan kísérletes módszerek is, melyekkel nagyobb időbeli és fenotípusbeli felbontással vizsgálható egy kezelés vagy génkiütés hatása, például a dinamikus proteomika, mely több fehérje szintjét tudja egyszerre követni időben (Geva-Zatorsky et al., 2010).

A genetikai kölcsönhatások kísérletes adatsorát felhasználhatjuk arra, hogy az élesztő anyagcseremodellt feljavítsuk. Erre a célra kifejlesztettünk egy automatizált gépi tanulási eljárást. Módszerünk különböző változtatásokat javasolt a modellen, melyek hatására jelentősen javult a modell előrejelzéseinek sikere. Az egyik javaslat szerint a glikogén termelése nem létfontosságú az élesztő számára, és ez a javaslat kijavít két, hibásan jelzett genetikai interakciót. Ez a javaslat összhangban áll azzal, hogy a glikogén, mint tartalék szénhidrát, csak éhezés és stresszhatás alatt válik fontossá az élesztő számára (François and Parrou, 2001). Ezen felül a NAD bioszintézisben javasolt módosítások is megalapozottnak bizonyultak.

Fontos megjegyezni, hogy a modellt feljavító algoritmusunk több olyan esetet is talált, amikor a javasolt módosítások között kölcsönhatás van, azaz két változtatás együtt nagyobb javulást eredményez az előrejelzésekben, mint külön-külön tennék. Egy ilyen példa a dolgozatomban mellékletként 4. táblázatában szereplő egyik módosítás, ami alternatív AMP bioszintézis útvonalak lezárását eredményezi és 3 biokémiai reakció együttes kikapcsolására van szükség hozzá, ezek: ADPT, PUNP1, PUNP5. Ezek közül a reakciók közül egyiknek a módosítása sem javít önállóan a genetikai kölcsönhatások előrejelzésén. Ezeknek a módosítások között fellépő kölcsönhatásoknak a megléte indokolja a genetikai algoritmus használatát.

Napjainkban a tudományos jellegű információ mennyisége robbanásszerűen növekszik, ebben a környezetben pedig egyre nagyobb szükség van az adatok számítógépes, automatizált feldolgozására és az automatikus hipotézis-generálásra. Ezek a módszerek lehetővé teszik, hogy a kutatók minél kevesebb manuális munkával minél több összefüggést tudjanak felismerni a felhasznált adatokban, majd a kapott hipotéziseket kísérletes úton teszteljék. Eredményeink azt mutatják, hogy a gépi tanulási eljárások nagy áteresztőképességű kísérletes adatokkal egybekötve képesek hipotézisek automatikus előállítására és a hipotézisek tesztelésére is, ez pedig legalább részben automatizált biológiai felfedezésekhez vezet (Covert et al., 2004; King et al., 2009).

3. Rejtett enzimaktivitások hálózata és evolúciós jelentősége

3.1. Bevezetés

Az enzimek biológiai katalizátorok, amelyek bizonyos biokémiai reakciók sebességének felgyorsítására alakultak ki az evolúció során. Páratlan katalitikus hatékonyságuk mellett hagyományosan ki szokták emelni a szubsztrátkötés specifikusságát is. Mindazonáltal már régóta ismert, hogy az enzimek képesek olyan biokémiai reakciókat is felgyorsítani melléktermékként, amelyeknek a katalizálására nem specializálódtak (Khersonsky and Tawfik, 2010). Ezeket a szelekciós nyomás alatt nem álló enzimaktivitásokat rejtett vagy látens aktivitásoknak hívjuk, szemben az enzimek natív vagy főaktivitásával. Jellemző rájuk, hogy ezek a rejtett aktivitások nagyságrendekkel gyengébbek a natív aktivitásoknál, így közvetlen élettani hatásuk csekély, mutációk révén azonban felerősödhetnek (Khersonsky and Tawfik, 2010).

Fontos megjegyezni, hogy a rejtett aktivitások megléte nem csak az enzimekre, hanem szinte minden más fehérjére is jellemző (Copley, 2012; Atkins, 2015). A látens aktivitásoknak kezdetben nem tulajdonítottak nagy jelentőséget, kivételes esetnek gondolták azokat a fehérjéket, amelyek rejtett aktivitással bírtak. Az elmúlt évtizedekben azonban egyre több munka foglalkozott ezeknek a rejtett aktivitásoknak a jelenlétével és lehetséges biológiai szerepével és az eredmények alapján megállapíthatjuk, hogy a rejtett aktivitások széles körben elterjedtek a fehérjék között (Khersonsky and Tawfik, 2010). Ennek több oka is van.

Egyrészt a nagymértékű szubsztrát-specifikusság megnöveli a szubsztrátkötés energetikai költségét, ezáltal nagyobb aktivációs energiához és alacsonyabb katalitikus sebességi állandóhoz (k_{cat}) vezet, gyakorlatilag korlátozza a fehérje hatékonyságát (Fersht, 1999). Még a leginkább specifikus enzimek esetében is kimutatható bizonyos mértékű elfogadása alternatív szubsztrátoknak. Ilyen esetekben gyakran javító mechanizmusok biztosítják a hibás működés kiküszöbölését (Fersht, 1999).

Másrészt a fehérjék specifikusságát az evolúció formálja, így ha egy rejtett aktivitás káros, akkor az ellen szelekció hat. Olyan vegyületekkel szemben azonban, amelyekkel az organizmus nem találkozik, nem alakulhat ki szelekciós nyomás. Jó példa erre a tirocidin szintetáz 1 (TycA), a tirocidin antibiotikum bioszintézisében szerepet játszó enzim. A TycA adenilációs doménje erősen szelektív L-fenilalaninra, viszont más természetes aminosavakat

nem fogad el szubsztrátnak. A mesterséges D-fenilalanint viszont a természetes szubsztráttal összemérhető hatékonysággal alakítja át (Khersonsky and Tawfik, 2010; Villiers and Hollfelder, 2009).

Harmadrészt a rejtett aktivitások jelenléte előnyös is lehet az élőlény számára, a látens enzimaktivitások ugyanis evolúciós utak kiindulópontjai lehetnek és új biokémiai funkcióhoz vezethetnek. Bár a látens aktivitások gyengék, génduplikáció után a látens aktivitás felerősödésével az új funkció kifejeződik, azaz fiziológiai szerepet kaphat. Ezáltal új biokémiai útvonalak megjelenéséhez vezethetnek a látens reakciók (Jensen, 1976). Ennek fontos szerepe van az antibiotikumokkal szembeni rezisztencia kialakulásában, különféle stresszhatások kivédésében (Soo et al., 2011), valamint új tápanyagokhoz való alkalmazkodásban (Jensen, 1976). Ugyancsak ezt használják ki az irányított enzimevolúciós kísérletekben, ahol a látens aktivitásokat mutációkkal felerősítik, ezáltal új, például iparilag fontos funkcióhoz juttatják az enzimet (Griffiths and Tawfik, 2003; Martínez and Schwaneberg, 2013).

Annak megértése, hogy hogyan alakulnak ki új útvonalak az anyagcsere-hálózatban, az evolúcióbiológia és a rendszerbiológia egyik kulcskérdése. Az uralkodó elmélet szerint új útvonalak az anyagcserében már meglévő enzimekből épülnek fel, melyek a hálózat különböző pontjairól származnak (Jensen, 1976; Schmidt et al., 2003). Ebben az evolúciós folyamatban kulcsszerepet játszanak a látens aktivitások, lehetővé téve, hogy az enzimek olyan funkciót is végrehajtsanak (még ha nagyon alacsony aktivitással is), amelyekre nem adaptálódtak (Khersonsky and Tawfik, 2010). Több bizonyíték is alátámasztja ezt az elméletet (Rison et al., 2002; Schmidt et al., 2003). Szép példája az elméletnek Jürgens és munkatársai munkája (Jürgens et al., 2000), akik a hisztidin és a triptofán bioszintézis útvonalak két, hasonló struktúrájú enzimét vizsgálták. Sikerült kimutatniuk, hogy a hisztidin bioszintézisben szerepet játszó HisA enzim egyetlen pontmutációval szert tett a triptofán bioszintézisben közreműködő TrpF enzim katalitikus aktivitására.

A bizonyítékok ellenére csak nagyon keveset tudunk a látens enzimaktivitások, mint evolúciós nyersanyag tényleges evolúciós potenciáljáról. Vajon a látens aktivitások milyen gyakran eredményeznek evolúciós újítást (új fenotípust)? Vajon tipikusan a natív hálózattól elszigetelten helyezkednek el, vagy a többségük beköthető abba és ezáltal elvben bekapcsolódhat a sejt anyagcsereforgalmába? És milyen fenotípusos előnnyel járhat a látens aktivitások felerősödése? Ezeknek a kérdéseknek a megválaszolásához átfogó enzimaktivitási

adatokra és rendszerszintű (hálózatos) megközelítésre van szükség, ami lehetővé teszi a genotípus – fenotípus leképezést. Jelenleg ugyan nincs elérhető szisztematikus, torzítatlan adatsor a látens aktivitásokról, azonban *Escherichia coli* baktériumra nagy mennyiségű adat áll rendelkezésre az irodalomban (Scheer et al., 2011). Ennélfogva összegyűjtöttünk egy átfogó listát az *Escherichia coli* enzimek kísérletesen meghatározott látens aktivitásaiból, majd ezeket a biokémiai reakciókat beépítettük egy genomléptékű anyagcsere-hálózati modellbe (Feist et al., 2007). Integrált anyagcsere-hálózatos megközelítésünkkel arra a kérdésre kerestük a választ, hogy a látens aktivitások jelenléte a natív hálózatban milyen fenotípusos előnyökkel jár.

3.2. Módszerek

3.2.1. Látens enzimaktivitásokat tartalmazó anyagcseremodell rekonstrukciója

A munka első lépéseként a már kísérletesen leírt látens reakciókat gyűjtöttük össze, majd ezeket a reakciókat integráltuk egy már létező *Escherichia coli* anyagcseremodellbe, hogy a fenotípusra kifejtett hatásukat vizsgálhassuk. A látens reakciókat Richard Notebaart munkatársunk gyűjtötte ki, míg a reakciók modellbe építését Richard és én közösen végeztük el (Notebaart et al., 2014).

A látens reakciók elsődleges forrása a BRENDA adatbázis volt (Scheer et al., 2011). Az adatbázisban az enzimek EC-szám (Enzyme Commission number (Webb and NC-IUBMB, 1992)) alapján vannak rendszerezve és minden EC-számhoz biokémiai reakciók vannak rendelve irodalmi források alapján. Az adatbázis azt is tartalmazza, hogy az adott reakciót milyen organizmusból vagy organizmusokból mutatták ki, illetve hogy az adott reakció natív reakciója-e az enzimnek. Ezt kihasználva kigyűjtöttük az *Escherichia coli* baktériumban leírt natív és látens reakciókat. A BRENDA adatbázis mellett az elsődleges irodalomból is gyűjtöttünk további látens reakciókat, a látens reakciók teljes listáját lásd a *Függelék: 5. táblázatban*.

A kigyűjtött látens reakciókat több lépésben ellenőriztük. Először meggyőződünk róla, hogy a látens reakció valóban hozzá van rendelve az adott enzimhez a forrásként megjelölt közleményben. Ezután kigyűjtöttük a szubsztrát és termék molekulák összegképletét különböző adatbázisokból (KEGG (Kanehisa and Goto, 1999), PubChem (Kim et al., 2016), ChEBI (Hastings et al., 2013), és ChemIDPlus (Tomasulo, 2002)). Amennyiben a szerzett információk konzisztensek voltak, megvizsgáltuk, hogy a reakció sztöchiometriailag korrekt, vagy sem, azaz a szubsztrátok atom-összetétele megegyezik-e a termékek atom-összetételével. Eltérés esetén megpróbáltuk korrigálni a sztöchiometriát a sztöchiometriai együtthatók változtatásával vagy vízmolekula hozzáadásával. Ha nem sikerült korrigálni a reakciót vagy a metabolitok valamelyike nem volt egyértelmű, a reakciót kivettük a látens reakciók listájából.

A látens reakciókat egy korábban publikált, kézzel ellenőrzött genomléptékű *Escherichia coli* anyagcseremodellhez adtuk hozzá (iAF1260 modell (Feist et al., 2007)). Ez a modell 1260 anyagcseregént tartalmaz, valamint 2077 biokémiai reakciót. A modellt 262 látens reakcióval és 277 olyan új metabolittal egészítettük ki, amelyek csak a látens reakciókban szerepelnek

(lásd *Függelék: 5-6. táblázat*). A látens reakciókat a hozzájuk tartozó EC-szám alapján rendeltük enzimekhez. A modell kiegészült továbbá 65 új transzporter reakcióval, melyek a BRENDA és BIGG adatbázisokból (Scheer et al., 2011; Schellenberger et al., 2010), valamint az irodalomból származtak és különféle tápanyagok felvételét tették lehetővé. Ezeket az új transzporter reakciókat is felhasználtuk, amikor a látens reakciók különböző környezetekben kifejtett hatását elemeztük.

3.2.2. Látens reakciók beköthetőségének vizsgálata

A látens reakciókat három kategóriába soroltuk aszerint, hogy hogyan kapcsolódnak a natív hálózathoz: teljesen bekötött, részlegesen bekötött és nem bekötött reakciókra. Amennyiben egy látens reakcióban kizárólag olyan metabolitok szerepelnek, amelyek a natív modellben is jelen vannak, a reakció teljesen beköthető a hálózatba. Részlegesen beköthető reakcióról beszélünk, ha vagy a szubsztrátok, vagy a termékek között jelen vannak olyan új metabolitok, amelyek nem fordulnak elő a natív modellben, és nem beköthető egy reakció, ha mind a szubsztrát, mind a termék oldalon tartalmaz új metabolitokat.

3.2.3. Környezetek meghatározása, ahol a látens reakciók fitneszhatását tesztelni lehet

A látens reakciókban rejlő evolúciós potenciál felmérésének érdekében megvizsgáltuk, hogy a standard körülményektől eltérő környezetekben biztosítanak-e valamilyen fitneszelőnyt, ha hozzáadjuk őket az anyagcseremodellhez, azaz növelik-e a baktérium biomassza-termelését. Ehhez létrehoztunk egy átfogó *in silico* környezetlistát.

Első lépésben meghatároztuk a standard glukóz minimál környezetet. Az *E. coli* anyagcsere-modellhez lepublikált környezetet vettük alapul (Feist et al., 2007), meggátoltuk azonban a CO₂, proton és kobalamin felvételét a médiumból a következő okokból. A CO₂ felvételének gátlására azért volt szükség, hogy elkerüljük a CO₂-on mint egyedüli szénforráson való növekedést a modellben. Bizonyos körülmények között ugyanis néhány reakció hozzáadása a modellhez lehetővé teszi a CO₂-on növekedést. A látens reakciók között azonban nem volt ilyen eset, így ez a korlátozás felesleges óvatosságnak bizonyult. Ugyan a CO₂ jelenléte befolyásolhatja a pH-t, azonban a modell a pH pontos értékét nem veszi figyelembe, így a pH élőlényre kifejtett hatását sem képes előrejelezni. Mindazonáltal a CO₂ gyorsan kidiffundál a sejtből, a médiumban pedig nincs jelen CO₂ nagy koncentrációban, így a nettó CO₂ felvétel az *E. coli* számára nem releváns (Merlin et al., 2003). Emellett a proton tetszőleges felvétele a környezetből ingyen energiához juttathatná a modellt, amely így szénforrás felvétele nélkül is

képes lenne ADP-ből ATP-t előállítani. A kobalamint pedig a modell nem használja fel az anyagcsere-folyamatokban, így a felvétele sem történik meg *in silico*.

Második lépésben összeállítottunk egy listát a glukóz minimál környezettől eltérő *in silico* környezetekből. A listában minden környezet csak egy tápanyagban különbözött a glukóz minimál környezettől: a standard szén-, nitrogén-, foszfor- vagy kénforrás egyikét (glukóz, ammónium, foszfát, szulfát) lecseréltük egy másik, a modell által felvehető tápanyagra. Emellett a listában minden környezet szerepel oxigén hiányában és jelenlétében is. A teljes lista 2754 környezetet számlál, amelyben tehát minden, a modell által felvehető tápanyag szerepel, mint potenciális szén-, nitrogén-, foszfor- és kénforrás.

3.2.4. Innovációt nyújtó látens reakciók azonosítása

Az anyagcseremodellek fitnessét különböző környezetekben FBA szimulációval határoztuk meg, amelyhez az R programnyelvet és a Sybil programcsomagot használtuk (R Core Team, 2013; Gelius-Dietrich et al., 2013). A szimuláció során egy vagy több látens reakciót adtunk a natív modellhez, majd az FBA segítségével meghatároztuk, hogy a natív modellhez képest nőtt-e a modell fitnessze. A látens reakcióknak akkor tulajdonítottunk fitnesszhatást, ha a biomassza-termelés a natív modellhez képest legalább 5 százalékkal nagyobb. Ennek a küszöbértéknek az a szerepe, hogy kiszűrje az irrelevánsan kicsi különbségeket. Az FBA az optimalizálás természetéből fakadóan új reakciók hozzáadására nem tud alacsonyabb biomassza-termeléssel válaszolni, így a látens reakciókkal kiegészített modellnek várhatóan a legtöbb környezetben némileg nagyobb lesz a fitnessze.

Először azt vizsgáltuk meg, hogy mi a látens reakciókban rejlő maximális potenciál, azaz mi történik, ha az összes látens reakciót szabadon felhasználhatja az anyagcseremodell a növekedéshez. A natív *E. coli* modellhez hozzáadtuk az összes látens reakciót, majd az így kapott modell növekedését megvizsgáltuk minden *in silico* környezetben, és meghatároztuk, hogy mely környezetekben nő jobban a natív modellnél.

A következő lépésben az egyes látens reakciók egyedi hatását akartuk tesztelni. Ehhez egy, az előzőhöz hasonló vizsgálatot végeztünk, azonban a natív modellt egyszerre csak egyetlen látens reakcióval egészítettük ki, majd a kapott új modell növekedését meghatároztuk minden környezetben, ahol a korábbi elemzés azt mutatta, hogy a látens reakciók előnyt biztosítanak. Ezáltal sikerült olyan eseteket meghatároznunk, ahol egy adott környezetben egy adott látens

reakció növekedési előnyt jelent, egyes esetekben pedig akár esszenciális is lehet a növekedéshez.

3.2.5. Látens aktivitások genomléptékű feltárása fehérje túltermeltetési kísérletekkel

A látens reakciók különböző környezetekben tanúsított hatásait kísérletesen is teszteltük. Ezeket a kísérleteket Kintses Bálint kollégám végezte, így itt csak vázlatosan tárgyalom őket (Notebaart et al., 2014). A kísérletek alapja, hogy egy adott enzim látens aktivitásait felerősíthetjük, ha megnöveljük az enzim koncentrációját a sejtben, tehát túltermeltetjük (Patrick et al., 2007). Kísérleteinkben az ASKA klónkönyvtárat használtuk fel, egy olyan *E. coli* törzskollekciót, amelynek minden tagjában más *E. coli* gén van nagy kópiaszámú plazmidra klónozva egy indukálható promoter mögé (Kitagawa et al., 2005).

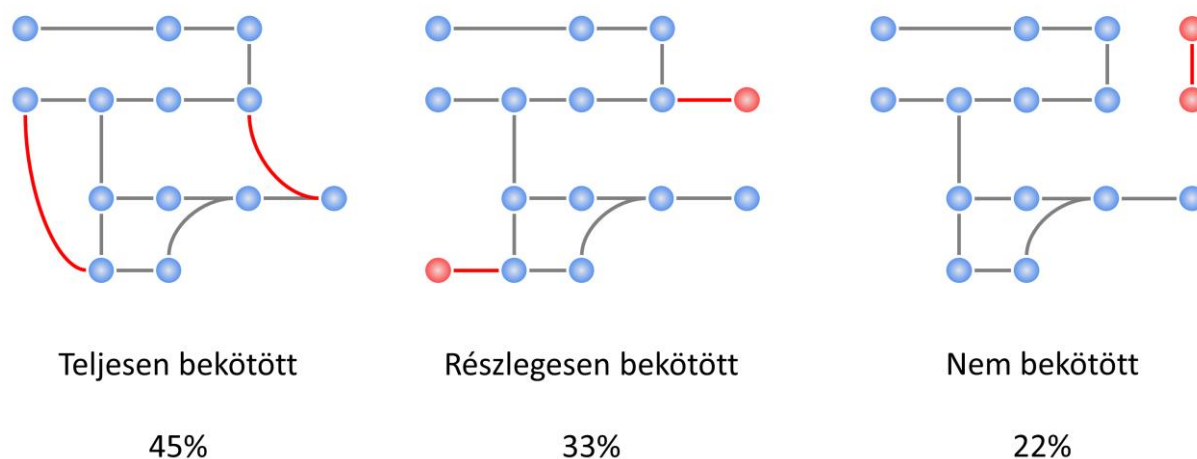
Az általunk használt módszer lényege, hogy egy egységnyi térfogatban levő folyékony tápoldatban (ami egy adott környezetet képvisel) növesztünk egy baktériumpopulációt, melynek minden sejtjében más és más gén van túltermelve és a sejtek között versengés folyik a tápanyagokért (Soo et al., 2011). Miután az előnyös fehérje túltermelést hordozó törzsek feldúsultak a környezetben, szekvenálással azonosítottuk ezeket. A gyakorlatban ez úgy nézett ki, hogy egy 96 lyukú mikrotiter lemez minden kútjába más tápoldatot tettünk, és minden kútban lezajlott a versengés a baktériumpopulációban a különböző genotípusok között. A módszer előnye, hogy párhuzamos, így sok környezetben lehet tesztelni gyakorlatilag az összes *E. coli* gén túltermelésének hatását. A mikrotiter lemezekben egy kútba 4078 különböző klónt tettünk, melyek mindegyikében más gén volt túltermelve. Ez a teljes ASKA fehérje túltermeléses könyvtár *E. coli*-ban. Minden környezetbe ugyanebből a 4078 klónból álló populációból vittünk át megközelítőleg 6 millió sejtet.

A környezetek teszteléséhez felhasználtuk a Biolog cég által összeállított PM1 és PM2A lemezeket, melyek együttesen 190, különböző szénforrást tartalmazó minimál médiumot foglalnak magukba. Ezeket a Biolog lemezeket mikroorganizmusok fenotípusának elemzésére dolgozták ki (Bochner et al., 2001). Ezt a listát 4 további szénforrással is kiegészítettük, melyekben az *in silico* előrejelzések szerint bizonyos látens reakciók szükségesek a növekedéshez, de hiányoztak a PM1 és PM2A lemezekről (etilén-glikol, L-gliceraldehid, D-lixóz és fenilgalaktozid). Összesen tehát 194 környezetet teszteltünk, melyekben csak a szénforrás különbözött. Minden környezetben csak egy szénforrás volt, a szénforrásokat nem vizsgáltuk kombinációban. Minden egyes környezetből két replikátum mérést csináltunk.

3.3. Eredmények

3.3.1. A látens reakciók beköthetőek a natív *E. coli* anyagcsere-hálózatba

Vajon a látens reakciók elkülönülnek a natív anyagcsere-hálózattól, vagy ugyanazokat a metabolitokat alakítják át, mint a natív reakciók? Új biokémiai reakciók hozzáadásával keresztkötések és zsákutcaák jöhetnek létre az anyagcsere-hálózatban, illetve az új reakciók elszigeteltek is lehetnek a hálózat többi részétől. Megvizsgáltuk a látens reakciókat, és azt találtuk, hogy a látens reakciók leggyakrabban keresztkötéseket alkotnak: 45% teljesen beköthető a natív *E. coli* anyagcsere-hálózatba, 33% alakít ki zsákutcat, és csupán 22%, mely izoláltan helyezkedik el a natív hálózathoz képest (lásd 13. ábra). Azáltal, hogy a látens reakciók többsége beköthető a natív hálózatba, képesek új biokémiai útvonalakat kialakítani a natív reakciókkal közreműködve.



13. ábra: a látens reakciók beköthetősége. Ezen a szemléltető ábrán a kék körök a natív modellben is előforduló metabolitokat jelölik, míg a piros körök olyan metabolitokat, melyek kizárólag látens reakciókban szerepelnek. A szürke vonalak natív reakciókat, míg a piros vonalak látens reakciókat reprezentálnak.

3.3.2. A látens reakciók jelenléte fitnesselőnyt biztosít bizonyos környezetekben

Miután kiderült, hogy a látens reakciók többsége illeszkedik a natív anyagcsere-hálózatba, feltettük a kérdést, hogy vajon képesek-e növelni a modell fitnessét, azaz a megszerzésük evolúciós előnyt jelent-e bizonyos körülmények között. Ennek a kérdésnek a megválaszolására kiszámoltuk a látens reakciók fitness-hozzájárulását nagyszámú *in silico* környezetben.

A 2754 tesztelt környezetből a natív modell 645 esetben nőtt, míg az összes látens reakcióval kibővített modell 664 esetben. 19 környezetben tehát esszenciális a látens reakciók jelenléte, azaz valódi evolúciós újítást eredményeznek (3. táblázat). Ez a 19 reakció 2.9%-kal növelte meg az anyagcseremodell által felhasználható nutriensek számát. Ez a szám elsőre kevésnek tűnhet, azonban ezt úgy érték el, hogy a modell reakcióinak számát a teljesen beköthető látens reakciók hozzáadásával alig 10.8%-kal terjesztették ki (1257-ről 1393-ra növelve azon reakciók számát, amelyeken flux mehet át). Ez az eredmény megerősíti a látens anyagcsere lehetséges szerepét, mint evolúciós nyersanyag. Emellett további 31 környezetben legalább 5 %-os növekedési előnyt jelentettek a látens reakciók.

Szénforrás	Nitrogénforrás	Foszforforrás	Kénforrás	Aerob	Látens reakció
Fenilgalaktozid	Ammónium	Foszfát	Szulfát	Igen	u0019
L-glicerinaldehyd	Ammónium	Foszfát	Szulfát	Igen	u0130, u0229
D-arabinóz	Ammónium	Foszfát	Szulfát	Igen	u0271
L-szorbóz	Ammónium	Foszfát	Szulfát	Igen	u0260
Etilén-glikol	Ammónium	Foszfát	Szulfát	Igen	u0228
D-lixóz	Ammónium	Foszfát	Szulfát	Igen	u0194
D-tartarát	Ammónium	Foszfát	Szulfát	Igen	u0095
5-amino-4-oxopentanoát	Ammónium	Foszfát	Szulfát	Igen	u0020
D-arabitol	Ammónium	Foszfát	Szulfát	Igen	u0312
D-2-dezoxiribóz	Ammónium	Foszfát	Szulfát	Igen	u0103, u0270
L-arginin	Ammónium	Foszfát	Szulfát	Nem	u0192
Fenilgalaktozid	Ammónium	Foszfát	Szulfát	Nem	u0019
L-glicerinaldehyd	Ammónium	Foszfát	Szulfát	Nem	u0130, u0229
D-arabinóz	Ammónium	Foszfát	Szulfát	Nem	u0271
L-szorbóz	Ammónium	Foszfát	Szulfát	Nem	u0260
D-lixóz	Ammónium	Foszfát	Szulfát	Nem	u0194
D-tartarát	Ammónium	Foszfát	Szulfát	Nem	u0095
D-arabitol	Ammónium	Foszfát	Szulfát	Nem	u0312
D-2-dezoxiribóz	Ammónium	Foszfát	Szulfát	Nem	u0103, u0270

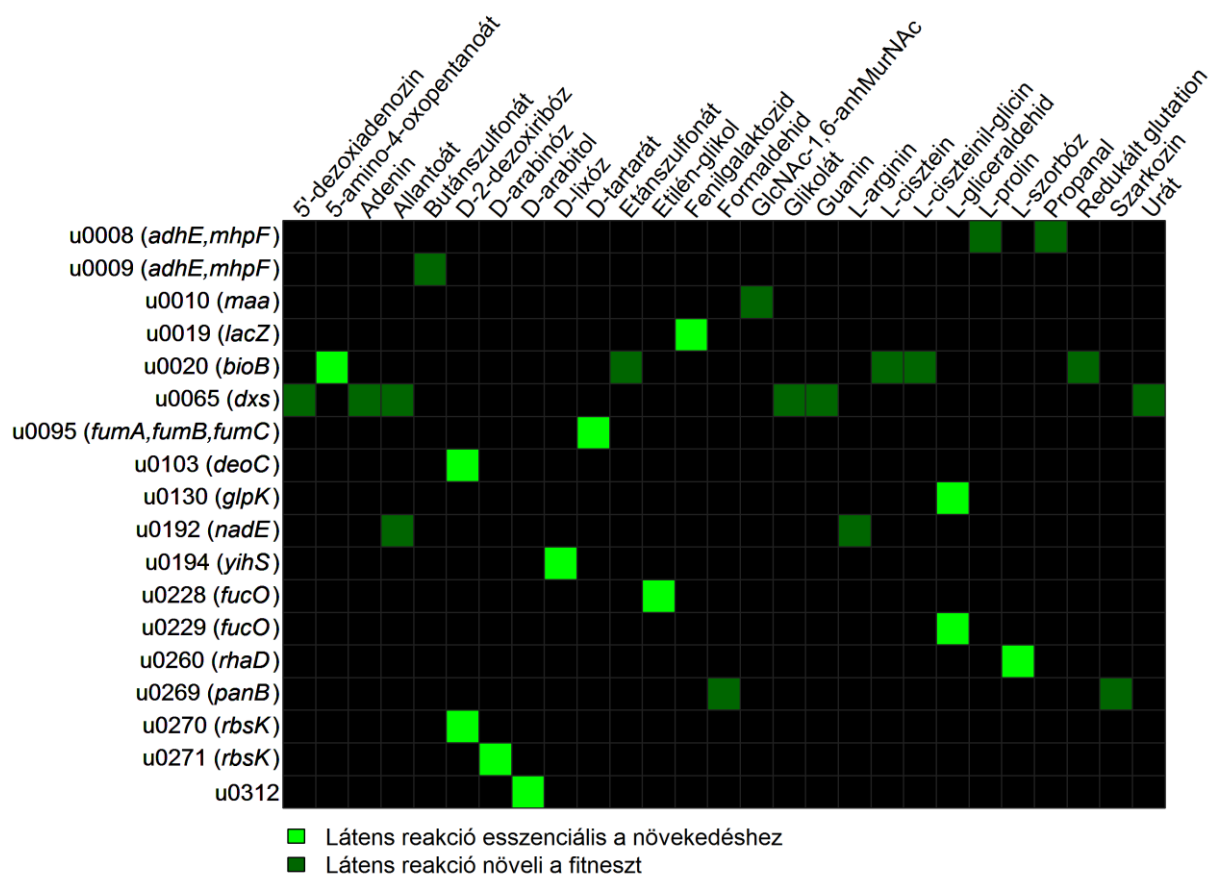
3. táblázat: Környezetek, ahol a látens reakciók jelenléte esszenciális. A „Látens reakció” oszlopban azok a reakciók vannak felsorolva, amelyek biztosítják a növekedést az adott szénforráson. Abban az esetben, ha több látens reakció van felsorolva ebben az oszlopban, a növekedéshez elég, ha az egyik jelen van.

Azok a környezetek, ahol a látens reakciók új funkciót kölcsönöznek a natív modellnek (tehát esszenciálisak), kivétel nélkül alternatív szénforrást tartalmaznak. Emellett megfigyelhető,

hogy az aerob és anaerob szimulációk eredménye között jelentős az átfedés: a látens reakciók hatása kevésbé függ az oxigén jelenlététől.

3.3.3. A látens reakciók egyedi hatásának vizsgálata

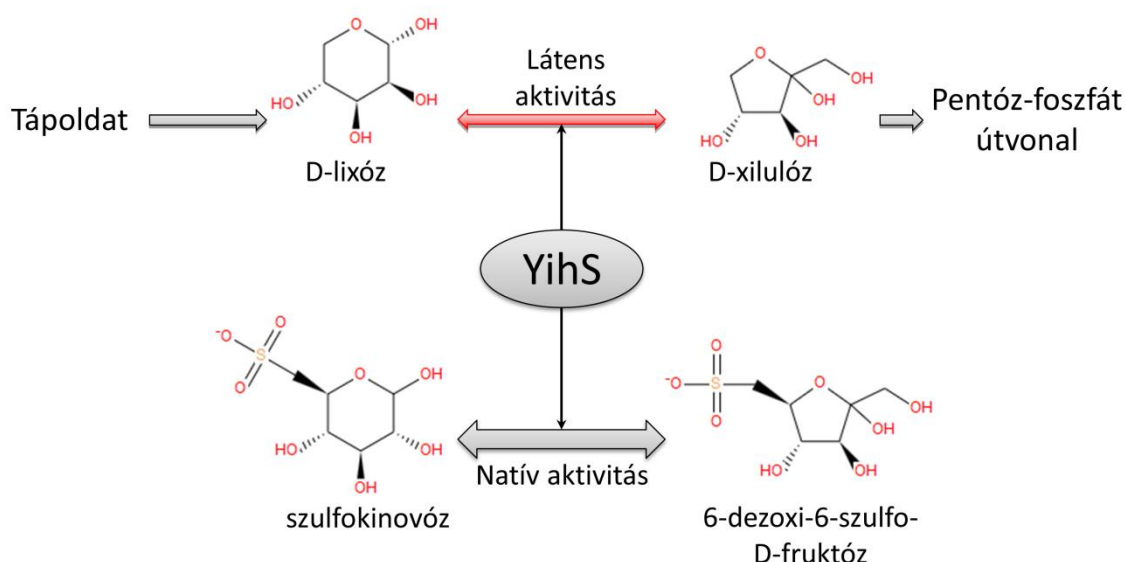
A következő lépésben megvizsgáltuk, hogy egyetlen látens reakció hozzáadása a natív modellhez milyen előnnyel jár a korábban vizsgált környezetekben. Kiderült, hogy minden olyan környezetben, ahol a látens reakciók hiányában a natív modell nem tud nőni, elegendő egyetlen látens reakció a növekedéshez. Ugyanakkor ritka, hogy ugyanaz a látens reakció több környezetben is fitnesselőnyt nyújtson, tehát a látens reakciók hatása környezetfüggő (a 14. ábra mutatja a különböző aerob szénforrásokon fitnesselőnyt adó látens reakciókat). Ezzel látszólag ellentmond, hogy a látens reakciók hatása kevésbé függ az oxigén jelenlététől, ahogy korábban láttuk. Az oxigén jelenlététől valóban kevésbé függ a szimulációk során a látens reakciók hatása, a szénforrásokra azonban erősen specifikus, hogy melyik látens reakció teszi lehetővé a felhasználását. Ez arra utal, hogy egy új környezetben az adaptációt nagyobb mértékben befolyásolja a szénforrás kémiai jellege, mint az oxigén jelenléte. Fontos azonban megjegyezni, hogy a modell nem veszi figyelembe, hogy aerob és anaerob körülmények között eltér a gének regulációs állapota, emiatt alulbecsli az oxigén jelenlétének hatását.



14. ábra: Aerob szénforrások és az adott környezetben fitnesselőnyt nyújtó látens reakciók. A reakciók neve után a reakcióhoz köthető gén neve szerepel zárójelben. A fekete cellák esetében a látens reakció nem biztosít semmilyen előnyt a növekedéshez, a sötétzöld cellák az 5%-nál nagyobb előnyt jelzik, míg a világoszöld cellákhoz tartozó látens reakció esszenciális az adott környezetben.

A látens reakciók innovatív szerepének egyik példája a D-lixóz esete. Eredményeink alapján a natív anyagcseremodell képtelen felhasználni ezt a szénforrást. Ezzel egybevág, hogy az elsődleges irodalom alapján a vad típusú *Escherichia coli* baktérium kísérletesen sem képes növekedni, ha D-lixóz az egyetlen szénforrás (Stevens and Wu, 1976). Az *in silico* szimulációk alapján azonban elegendő egyetlen látens reakció hozzáadása a modellhez, hogy a D-lixóz felhasználható legyen, mégpedig úgy, hogy a látens reakció a D-lixózt D-xilulózzá alakítja, a D-xilulóz pedig natív reakciókon keresztül beépül a központi anyagcserébe (Stevens and Wu, 1976). A látens reakció az *yihS* gén által kódolt enzimhez kötődik, melynek a natív funkciója a szulfokinovóz izomerizációja, lehetővé téve a szulfokinovóz felhasználását szénforrásként (15. ábra). A 15. ábrán megfigyelhető, hogy a natív és a látens aktivitás

kémiaailag rokon molekulákon történik meg. Enzim-túltermeléses kísérletekkel sikerült is igazolnunk, hogy az *yihS* gén terméke valóban képes végrehajtani a D-lixóz átalakítását és így biztosítani a növekedést ezen a szénforráson (lásd a következő fejezetet).



15. ábra: az *yihS* gén által kódolt enzim natív és látens aktivitása. Az enzim natív szubsztrátja a szulfokinovóz, míg a hozzá hasonló szerkezetű D-lixózt mint látens szubsztrátot csak alacsony hatékonysággal alakítja át az enzim.

3.3.4. Az *in silico* előrejelzések jó egyezést mutatnak a kísérletesen meghatározott metabolikus újításokkal

Szerettük volna az eredményeinket laborkísérletekkel is alátámasztani, ráadásul egy, az *in silico* szimulációkhoz hasonló rendszerszintű vizsgálattal. Erre a célra végeztük el az ASKA fehérje túltermeltetési kísérleteket, melyek erénye, hogy gyakorlatilag minden *Escherichia coli* gén túltermelésének hatását felfedik, ha az növeli a fitneszt egy adott környezetben, ráadásul számos környezeten végezhető el párhuzamosan. Ezen túl a kapott eredmények teljesen függetlenek a jelenlegi biokémiai tudásunktól, hiszen egyszerű genetikai kísérleten

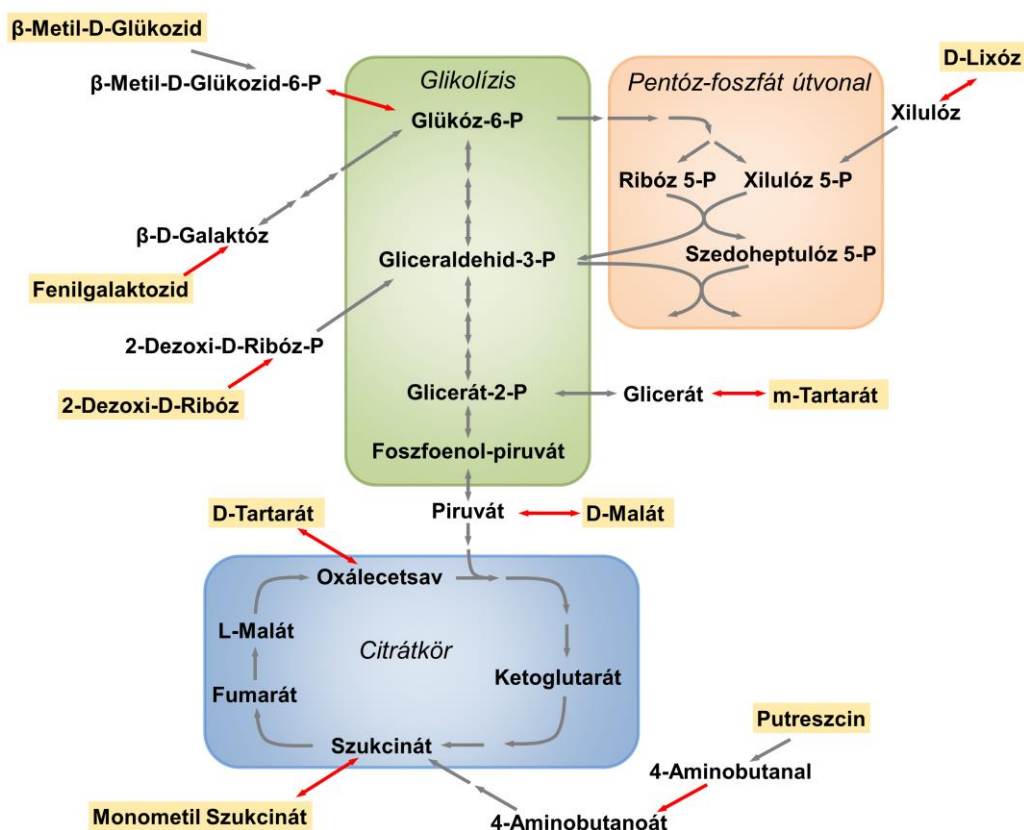
alapulnak. Emiatt ez a módszer kitűnő módja az *in silico* eredmények kísérletes ellenőrzésének.

Az ASKA túltermeltetési kísérletekkel 17 gént sikerült azonosítani, melyek túltermelése növelte a fitnesszt legalább egy specifikus szénforráson. Összesen 17 szénforráson sikerült fitnesszelőnyt kimutatni, ez a környezetek 9%-a (Notebaart et al., 2014). Fontos megjegyezni, hogy a feljött gének többsége enzimet kódol, és a gének 59%-ánál (10 a 17-ből) feltételezhető *E. coli*-ból vagy más baktériumból származó enzimológiai adatok vagy kémiai hasonlóság alapján, hogy látens aktivitás áll a fitnesszelőny hátterében (lásd 4. táblázat). Genomléptékű kísérletekkel is sikerült tehát igazolni, hogy a látens aktivitások új környezeti adaptációkat hoznak létre, ráadásul ehhez gyakran elég egyetlen lépés, avagy mutáció is.

A kísérleteket az *in silico* szimulációkkal összevetve elmondható, hogy a modellezés segítségével sikerült előrejelezni a kísérletesen feljött környezetek 44%-át (4. táblázat). A 16. ábra bemutatja, hogy a 4. táblázatban felsorolt látens aktivitások hogyan kötik be a kérdéses szénforrásokat a központi anyagcserébe. Megfigyelhető, hogy a látens aktivitások különböző pontokon tudnak bekapcsolódni a natív anyagcserébe, ezáltal változatos új anyagcsere-útvonalak nyílnak meg az organizmus számára.

Szénforrás	Enzim	Kísérletes fenotípus	<i>In silico</i> egyezés
D-lixóz	Szulfokinovóz izomeráz (<i>yihS</i>)	++	●
D-2-dezoxiribóz	Ribokináz (<i>rbsK</i>)	++	●
D-tartarát	Fumaráz A (<i>fumA</i>), Fumaráz B (<i>fumB</i>)	++	●
Fenilgalaktozid	β -galaktozidáz (<i>lacZ</i>)	+	●
m-tartarát	D-malát dehidrogenáz (<i>dmlA</i>)	++	○
Putreszcín	Acetaldehid dehidrogenáz (<i>mhpF</i>)	++	○
D-malát	3-izopropilmalát dehidrogenáz (<i>leuB</i>)	+	○
β -metil-D-glükozid	6-foszfo- β -glükozidáz (<i>bglB</i>)	+	○
Monometil szukcinát	Karboxilészter hidroláz (<i>ybfF</i>)	+	○

4. táblázat: azon enzimek listája, melyeknek kísérletes túltermelése lehetővé teszi (++) vagy javítja (+) egy adott szénforrás felhasználását. A modell által is előrejelzett környezet-enzim párokat (●) jelöli, míg azokat, amelyeket *in silico* nem észleltünk, (○).



16. ábra: a központi anyagcsere vázlatos rajza. Piros nyíllal vannak feltüntetve azok a látens aktivitások, amelyek kísérletesen növekedési előnyt jelentenek bizonyos szénforrások (sárgával kiemelve) mellett.

Kiértékeljük a kísérletes tesztek és az *in silico* előrejelzések közötti átfedést is. Csak olyan géneket vettünk figyelembe az összehasonlításban, melyek szerepelnek az *E. coli* anyagcserehálózatban valamint az ASKA könyvtárban is, ez összesen 112 gént jelent. Hasonlóan jártunk el a környezetekkel is, csak azokat a szénforrásokat néztük, melyek mind a kísérletekben, mind a modellezés során szerepeltek (105 környezet). Ezáltal 112×105 gén-szénforrás párt értékeltünk ki. A statisztikai összehasonlítást Fisher-egzakt teszttel végeztük, az eredmény alapján az átfedés jelentős (p érték $<10^{-13}$, lásd 5. táblázat). Az 5. táblázat alapján a kísérletes és *in silico* vizsgálatok során feljött környezet – gén párok közötti átfedés várható értéke 0.01, a valóságban tapasztalt 5 eset tehát több mint 2 nagyságrenddel felülmúlja a véletlenszerűen vártat.

Kísérletesen meghatározott növekedési előny		
<i>In silico</i> előrejelzett növekedési előny	Igen	Nem
	Igen	Nem
Igen	5	11
Nem	1	11743

5. táblázat: a kísérletesen kimutatott és *in silico* előrejelzett növekedési előnyök kontingencia táblázata. A táblázatban a környezet-enzim párok száma szerepel, tehát a bal felső cellában szereplő 5-ös azt jelenti, hogy 5 olyan eset van, ahol mind *in silico*, mind kísérletesen egy adott enzimhez köthető látens reakció előnyt biztosít egy adott környezetben. A 4. táblázathoz képest azért csak 5 környezet-enzim pár szerepel itt, mert a 4. táblázatban felsorolt szénforrások között van három, amelyeknek a felvételét a modell nem tudja szimulálni, mert a modellből hiányzik az adott metabolit felvételét biztosító transzport reakció és a metabolit sem szerepel a modellben. Ezek: *m*-tartarát, β -metil-D-glükózid, Monometil szukcinát. Ezekben a környezeteken tehát a modell nem tudja előrejelezni a látens reakciók hatását.

3.4. Diszkusszió

Az enzimek szubsztrátspecifitása korlátozott, és az ennek eredményeképpen jelenlévő látens aktivitások egyfajta evolúciós nyersanyagot szolgáltatnak új biokémiai útvonalak és anyagcsere-funkciók kialakulásához (Khersonsky and Tawfik, 2010). Ennek vizsgálatára azonban mindezidáig csak esettanulmányok álltak rendelkezésre. Munkánk az első olyan elemzés, amelyben a látens reakciók evolúciós szerepe az egész anyagcsere-hálózat szintjén van felmérve.

Különböző forrásokból összegyűjtve látens aktivitásokat sikerült a natív *Escherichia coli* anyagcseremodellt 12.6%-kal megnövelni. Az általunk összegyűjtött látens reakciók többsége beköthető a natív hálózatba, ezáltal új biokémiai útvonalakat tudnak kialakítani a natív reakciókkal együttműködve. Ezek a látens reakciók ~3%-kal kiterjesztették azoknak a környezeteknek a listáját, amelyeken az anyagcseremodell képes növekedni, tehát megállapítható, hogy a rejtett enzimaktivitásoknak egy tekintélyes hányada evolúciós újítást eredményez. Ehhez minden esetben elegendő volt egyetlen látens reakcióval kibővíteni a natív modellt, mintegy meghosszabbítani a meglévő útvonalakat egy újabb lépéssel, azaz nem volt szükség bonyolult, több génes útvonalak kialakulására. A látens reakciók mutációkkal történő felerősítése egy egyszerű módja az adaptációnak, mely során nincs szükség külső genetikai anyag beépítésére. Feltételezhető, hogy a horizontális géntranszfer prokariótákhoz képest kisebb szerepe miatt eukariótáknál ennél is nagyobb jelentősége van a látens aktivitásoknak az új környezetekhez való adaptációban (Pál et al., 2005).

A látens anyagcserének további jelentősége is lehet az új tápanyagok hasznosítása mellett. A látens reakciók átkötéseket biztosítanak az anyagcsere-hálózaton belül, így elősegíthetik a hálózat gyors reakcióját enzim- vagy metabolitkoncentrációban bekövetkezett zavarok esetén (Wagner and Fell, 2001). Továbbá új útvonalak kialakításán keresztül a látens aktivitások új másodlagos metabolitok termeléséhez is vezethetnek (Weng et al., 2012).

A látens aktivitások károsak is lehetnek az élőlény számára, például azáltal, hogy toxikus mellékterméket termelnek, vagy kevésbé optimális útvonalakra terelik az anyagáramlást az anyagcsere-hálózatban. Ezeket a jelenségeket Pál Ferenc és Györkei Ádám kollégáim vizsgálták, ezért itt csak röviden tárgyalom az eredményeiket. Azt találták, hogy a hálózat szintjén a látens aktivitások által termelt metabolitok nem toxikusabbak, mint a natív reakciók által termelték, valamint azok az útvonalak, amelyekben látens reakciók is szerepelnek, hatékonyságban nem maradnak el a csak natív reakciókból felépülő útvonalaktól. Ezek

azonban általános mintázatok, egyedi látens reakciók esetében előfordulhat, hogy csökkentik az élőlény fitneszt. Erre jó példa az a kísérletes munka, amelyben egy új piridoxál-foszfát bioszintézis útvonalat hoztak létre *E. coli* baktériumban bizonyos enzimek túltermelésével. Az új útvonal intermedier metabolitjai között vannak olyanok, amelyek nem natív metabolitok az *E. coli* sejtekben, és kiderült, hogy ezek a vegyületek gátolják bizonyos aminosavak bioszintézisét (Kim and Copley, 2012).

Az evolúciós rendszerbiológia központi kérdése, hogy hogyan lehet előrejelezni a mutációk fenotípusos hatását és az evolúció lehetséges útvonalait (Heckmann et al., 2013; Papp et al., 2011). Bár történtek jelentős előrelépések a génkiütések fitneszhatásának és genetikai interakcióinak genomléptékű anyagcsere-hálózatban történő előrejelzésében (Papp et al., 2011), az adaptív evolúciós újítások genetikai hátterét jelenleg kevésbé értjük. Eredményeink azt mutatják, hogy a rejtett enzimaktivitások segítségével előrejelezhetőek új fenotípusok és azok genetikai háttere az anyagcsere szintjén.

Fontos megjegyezni, hogy mind az általunk gyűjtött, mind a kísérletek során felfedezett rejtett enzimaktivitások alulbecslik a látens biokémiai reakciók tényleges számát és evolúciós potenciálját. Mind a kísérletekben, mind a szimulációk során olyan eseteket vizsgáltunk, ahol az organizmusnak egy új tápanyag feldolgozásához kellett adaptálnia. Más módon is előnyös lehet azonban egy látens reakció: kialakíthat egy hatékonyabb bioszintetikus útvonalat egy, az élőlény számára kulcsfontosságú metabolithoz vagy előmozdíthatja egy káros anyag, például egy antibiotikum lebontását (Soo et al., 2011). Az is belátható, hogy az irodalom alapján ismert látens reakciók listája nem teljes. Maga az *E. coli* anyagcsere-hálózat rekonstrukciója sem teljes, így bizonyos látens reakciók előnyös hatása a szükséges natív reakciók hiányában nem kimutatható.

Kísérletes megközelítésünk szintén nem ad teljes képet az *E. coli* organizmusban található rejtett enzimaktivitásokról. Egyrészt a kísérletek során csak azok a biokémiai újítások jelentek meg, melyekhez egyetlen látens reakció felerősítése elegendő volt. Azokat az eseteket, ahol egy többlépéses útvonal kialakulása szükséges az előnyös fenotípus eléréséhez, nem látjuk. Másrészt a túltermeléses kísérletek során rejtve maradhatnak azok a látens enzimaktivitások, melyek csupán kismértékben biztosítanak fitneszelőnyt. Harmadrészt az enzim túltermelésén kívül más módokon is növekedhet egy látens aktivitás katalitikus hatékonysága mutációk hatására. Ugyancsak előfordulhat, hogy egy látens aktivitás felerősítése nem tudja kifejteni

pozitív hatását, mert az ehhez szükséges natív reakciók valamelyike nem hozzáférhető gátló regulációs hatások miatt.

Annak ellenére, hogy a tudásunk a rejtett enzimaktivitásokról közel sem teljes, az *in silico* és kísérletesen azonosított genotípus-fenotípus párok közötti nagymértékű átfedés arra utal, hogy a rejtett aktivitások általunk összeállított hálózata már így is jelentős részét lefedi annak az evolúciós nyersanyagának, mely az új környezetekhez való rövidtávú adaptáció során az élőlény rendelkezésére áll. Jó példa erre a D-lixóz mint szénforrás esete. A D-lixóz ritkán fordul elő a természetben (Ahmed, 2001) és a vad típusú *E. coli* nem képes lebontani (ezt kísérleteink is igazolták). A modellezés során kiderült, hogy egyetlen biokémiai reakció elegendő a D-lixóz lebontásához és szénforrásként való felhasználásához, ez a reakció pedig a mannóz izomeráz enzim (Yihs) mellékreakciója. A kísérleteink igazolták, hogy az Yihs, és csakis az Yihs túltermelése teszi lehetővé a D-lixózon való növekedést. Ugyanakkor a kísérletek során feltárt, de a modell által előre nem jelzett új fenotípusok olyan rejtett enzimaktivitásokhoz vezetnek, melyek jelenleg nincsenek leírva *E. coli*-ban. Ez alapján elmondhatjuk, hogy még nagyszámú olyan látens aktivitás van, mely felfedezésre vár.

A látens enzimaktivitások ismeretén alapuló új fenotípusok előrejelzésének fontos gyakorlati alkalmazásai is vannak. A rejtett enzimaktivitások szisztematikus tesztelése számítógépes modellezéssel egybekötve új bioszintetikus útvonalakat tárhat fel az iparban is fontos vegyületek számára, akár új, gazdaságilag előnyösebb környezetekben is (Carbonell and Faulon, 2010). Emellett ez a megközelítés segíthet megérteni az enzimek másodlagos aktivitásainak és funkcionyeréses mutációinak a szerepét a tumorok evolúciójában (Dang et al., 2009). A tumorok megfékezése szempontjából a fehérjék látens aktivitásai károsak lehetnek, mert elősegíthetik a tumorsejtek alkalmazkodását akár a gazdaszervezethez, akár bizonyos tumorellenes kezelésekhez. A látens aktivitások feltárásával jobban megismerhetjük a tumorsejtek ezen lehetséges fejlődési irányait is, ami elősegítheti hatékonyabb tumorellenes kezelések kifejlesztését.

4. Összegzés és kitekintés

Doktori munkám során a rendszerbiológia eszköztárát használtam fel arra, hogy az evolúcióbiológia kérdéseit vizsgáljam. Elsősorban az anyagcsere kényszer-alapú modellezését, mint módszert használtam, melynek segítségével többek között különböző genetikai módosítások fenotípusos hatása vizsgálható genomléptékben. Az anyagcsere az egyik legjobban ismert és leginkább kutatott sejtes alrendszer, így ideális modell az evolúciós folyamatok vizsgálatára.

A genetikai kölcsönhatások fontos szerepet töltenek be az evolúciós folyamatokban, a fenotípust ugyanis általában nem egyetlen gén határozza meg, hanem gének sokasága. A 2. fejezetben bemutatott munkánkban arra a kérdésre kerestük a választ, hogy a kényszer-alapú modellek képesek-e előrejelezni a genetikai kölcsönhatásokat. Azt találtuk, hogy bár az anyagcseremodellek érzékenysége a genetikai kölcsönhatások előrejelzésében alacsony, a modell előrejelzései még így is jelentősen jobbak, mint azt véletlenszerűen várnánk (10. ábra). Ennek köszönhetően a modell képes reprodukálni a genetikai kölcsönhatások által mutatott mintázatokat, illetve mechanisztikus magyarázatot is tud adni rájuk (6-9 ábra).

Látván a modell viszonylagos alacsony érzékenységét a genetikai kölcsönhatásokra, felmerült bennünk, hogy lehetséges-e ezen javítani. Kifejlesztettünk egy eljárást, amely genomléptékű kísérletes adatokat felhasználva automatikusan, ismétlődő lépésekben feljavítja a modellt. Ez az eljárás képes volt új hipotéziseket generálni, azaz lehetséges módosításokat javasolni a modellen. Ezek a hipotézisek kísérletesen tesztelhetők, és egy, a NAD bioszintézis útvonalra tett javaslatot sikeresen igazoltunk is. Ez és az ehhez hasonló módszerek lehetővé teszik, hogy iteratív módon növeljük ismereteinket, melynek során legalább részben automatikusan történik meg az új hipotézisek felállítása és tesztelése, ezzel felgyorsítva a kutatói munkát (King et al., 2009).

Munkánk során sikerült feljavítani az anyagcseremodell genetikai kölcsönhatásokat előrejelző képességét, ezzel közel duplájára növelni az előrejelzések érzékenységét. Ugyanakkor a genetikai kölcsönhatások többségét még így sem látja a modell. Vannak azonban olyan módszerek, amelyekkel javítani lehetne az anyagcseremodellek ezen képességén. Egyfelől integrálni kellene egy regulációs szintet a modellekbe, így azok figyelembe tudnák venni, hogy egy adott gén átíródik-e az adott körülmények között. Erre ad lehetőséget többek között az rFBA, SR-FBA, OptORF és a PROM módszer (Covert et al., 2004; Shlomi et al., 2007b;

Kim and Reed, 2010; Chandrasekaran and Price, 2010). Ezeknek a módszereknek a használatához szükség van arra, hogy ismerjük a sejtben jelen lévő regulációs kölcsönhatásokat, ezek az ismeretek azonban még a jól ismert modellélőlények esetében is hiányosak.

Egy másik megközelítés transzkriptomikai mérések eredményét beépíteni a modellbe, azaz egy adott reakció fluxát az adott enzimet kódoló gén expressziójának függvényévé tenni (Shlomi et al., 2008). Ennek a módszernek előnye, hogy nem kell ismerni a regulációs kölcsönhatásokat, elegendő lemérni a kérdéses sejtek génexpressziós profilját az adott környezetben. Problémát jelent azonban, hogy a gének expressziója és a biokémiai reakciókon átmenő flux közti összefüggés nem lineáris és sok tényezőtől függ, emiatt nem egyértelmű, hogy hogyan tesszük a fluxokat függővé az expressziós adatoktól. Emellett egy munkában szisztematikusan tesztelték a kényszer-alapú modellek fluxát génexpressziós adatok alapján kényszerítő módszereket, és azt találták, hogy ezek a módszerek általában nem jelzik előre a fluxokat jobban, mint a módosíthatatlan FBA (Machado and Herrgård, 2014).

A fent említett technikákon túl is nagyszámú módszer áll rendelkezésre, melyekkel növelni lehet a modellbe épített biológiai információ mennyiségét, ami pedig segíthet a genetikai kölcsönhatások előrejelzésében (Lewis et al., 2012). Minden újabb hozzáadott réteggel nő azonban a lehetséges hibák száma a modellben, illetve bizonyos adatoknál (például génexpresszió) a modell elemeihez való hozzárendelés nem egyértelmű. Ahogy azonban egyre több genomléptékű kísérletes adatsor áll rendelkezésre, ezek a módszerek is tovább fejlődnek, lehetővé téve az evolúcióbiológia jelenségeinek részletesebb vizsgálatát az anyagcseremodellekkel.

Dolgozatom 3. fejezetében leírt munkánkban az evolúciós innovációk előrejelezhetőségét vizsgáltuk. Abból a feltételezésből indultunk ki, hogy az enzimek rejtett aktivitásai evolúciós nyersanyagként szolgálhatnak és adaptáció során ezek a kezdetben még jelentéktelen funkciók felerősödve új fenotípust eredményezhetnek. Az irodalomban ismert látens aktivitásokat összegyűjtöttük és beépítettük az *Escherichia coli* anyagcsere-hálózatába, majd szimulációkkal kimutattuk, hogy a látens aktivitások segítségével új környezetekhez tud adaptálódni az organizmus. Genomléptékű kísérletes tesztekkel ezt megerősítettük, ráadásul a kísérletes és *in silico* eredmények között jelentős átfedést tapasztaltunk. Sikert tehát kimutatnunk, hogy az evolúciós újítások előrejelezhetőek egy rendszerbiológiai modell felhasználásával.

Ahogy dolgozatomban 2. fejezetében bemutattam, a genetikai kölcsönhatások terén az anyagcseremodell előrejelzései alacsony érzékenységgel bírnak. Emiatt felmerülhet, hogy vajon hogyan képes a modell mégis előrejelezni az evolúciós újításokat. Fontos megjegyezni azonban, hogy az általunk vizsgált *in silico* evolúciós újítások mind egy lépésesek voltak, azaz minden új környezetben elegendő volt egyetlen látens reakció ahhoz, hogy az adott környezetben a modell növekedni tudjon. Emellett a nagyskálájú labor kísérletekkel sem teszteltük több fehérje egyidejű túltermelésének hatását. Az ilyen, egy lépéses genetikai beavatkozások kimenetelét az anyagcseremodell nagy hatásfokkal képesek előrejelezni, a génkiütések esetében akár 90 százalékos pontossággal (Heavner et al., 2013; Orth et al., 2011). Annak eldöntésére, hogy komplexebb evolúciós újítások (például új, többlépéses útvonalak) kialakulását milyen pontosan képesek előrejelezni a kényszer alapú anyagcseremodell, további modellezésre és nagyskálájú kísérletekre lesz szükség a jövőben.

A látens aktivitásoknak, mint evolúciós nyersanyagoknak fontos szerepük van az adaptáció során. Az általunk létrehozott látens aktivitás adatbázis azonban bár genomléptékű, nem szisztematikus, hanem esettanulmányok sokaságán alapszik. Az intenzívebben tanulmányozott enzimekre és metabolitokra több információ érhető el, így vélhetőleg több látens aktivitás köthető hozzájuk a mi gyűjteményünkben. Jelenleg nincs elérhető szisztematikus felmérése a látens aktivitásoknak, csupán néhány, egy adott enzimszaládra korlátozódó tanulmány létezik (Huang et al., 2015; Kuznetsova et al., 2006). Emiatt a látens reakciók anyagcsere-hálózaton belüli eloszlására vonatkozó kérdéseket jelenleg nem tudjuk megválaszolni. Márpedig a látens reakciók eloszlásának ismerete fontos kérdés, ugyanis az, hogy a látens reakciók milyen útvonalakban dúsulnak fel, befolyásolhatja az élőlény adaptációs képességeit.

A metabolomika egy lehetséges megoldást kínál erre a problémára. A metabolitok koncentrációjának mérésével, illetve időbeli változásuk nyomon követésével megállapítható, hogy egy adott enzim milyen metabolitokat fogad el szubsztrátjaként. A modern metabolomikai módszerekkel pedig akár több száz metabolit egyidejű mérése is lehetővé vált (Fuhrer et al., 2011). Emellett ma már elérhető több olyan törzskönyvtár is, amelyek az enzimek termelését és kinyerését könnyítik meg, ezáltal nagyszámú enzim is gyorsan előállítható és vizsgálható *in vitro* (Kitagawa et al., 2005). Ezek a módszerek a közeljövőben lehetővé tehetik, hogy egy adott organizmus biokémiai reakcióit és azok genetikai hátterét

szisztematikusan feltérképezzük, tekintettel arra, hogy egy adott reakció natív vagy látens aktivitása egy adott enzimnek.

5. Köszönetnyilvánítás

Mindenekelőtt szeretnék köszönetet mondani témavezetőmnek, Papp Balázsnak, amiért felvett a csoportjába PhD hallgatónak, valamint köszönöm Neki a munkám során tőle kapott szakmai támogatást és iránymutatást. Köszönöm Neki a jelen dolgozat alapos átnézését is.

Köszönetet szeretnék mondani Pál Csabának, akinek szakmai tanácsai és segítsége nélkülözhetetlen volt a munkám során.

A 2. fejezetben bemutatott munka egy nagyobb projekt részeként került publikálásra (Szappanos et al., 2011). A projektben való közreműködésért szeretnék köszönetet mondani:

Kovács Károlynak a közös munkáért,
Szamecz Bélának az élesztő nikotinsav-függőségének kísérletes teszteléséért,
Martin Lerchernek és Gabriel Gelius-Dietrichnek a Sybil programcsomag fejlesztéséért,
Charles Boone laboratóriumának (Toronto, Kanada) a kísérletes adatokért és
Jelasity Márknak a szakmai segítségért.

A 3. fejezetben bemutatott munkában (Notebaart et al., 2014) való közreműködésért köszönetet szeretnék mondani:

Richard Notebaartnak a látens reakciók összegyűjtéséért és modellbe építéséért, valamint a közös munkáért,
Kintses Bálintnak az ASKA kollekcióval végzett kísérletes munkájáért,
Pál Ferencnek és Györkei Ádámnak az általuk végzett bioinformatikai elemzésekért és
Bogos Balázsnak, Lázár Viktóriának, Spohn Rékának és Csörgő Bálintnak a kísérletes munkában való részvételükért.

Köszönetet mondok az Evolúciós Rendszerbiológiai Műhely minden tagjának, akikkel volt szerencsém együtt dolgozni a doktori éveim alatt.

6. Irodalomjegyzék

1. Ahmed, Z. (2001). Production of natural and rare pentoses using microorganisms and their enzymes. *Electron. J. Biotechnol.* 4, 1–16.
2. Åkesson, M., Förster, J., and Nielsen, J. (2004). Integration of gene expression data into genome-scale metabolic models. *Metab. Eng.* 6, 285–293.
3. Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402.
4. Atkins, W.M. (2015). Biological messiness vs. biological genius: Mechanistic aspects and roles of protein promiscuity. *J. Steroid Biochem. Mol. Biol.* 151, 3–11.
5. Bäck, T., and Hoffmeister, F. (1991). Extended Selection Mechanisms in Genetic Algorithms. (Morgan Kaufmann), pp. 92–99.
6. Baldi, P., Brunak, S., Chauvin, Y., Andersen, C.A.F., and Nielsen, H. (2000). Assessing the accuracy of prediction algorithms for classification: an overview. *Bioinformatics* 16, 412–424.
7. Baryshnikova, A., Costanzo, M., Kim, Y., Ding, H., Koh, J., Toufighi, K., Youn, J.-Y., Ou, J., San Luis, B.-J., Bandyopadhyay, S., et al. (2010). Quantitative analysis of fitness and genetic interactions in yeast on a genome scale. *Nat. Methods* 7, 1017–1024.
8. Baudin, A., Ozier-Kalogeropoulos, O., Denouel, A., Lacroute, F., and Cullin, C. (1993). A simple and efficient method for direct gene deletion in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* 21, 3329–3330.
9. Bochner, B.R., Gadzinski, P., and Panomitros, E. (2001). Phenotype microarrays for high-throughput phenotypic testing and assay of gene function. *Genome Res.* 11, 1246–1255.
10. Boone, C., Bussey, H., and Andrews, B.J. (2007). Exploring genetic interactions and networks with yeast. *Nat. Rev. Genet.* 8, 437–449.
11. Burgard, A.P., Nikolaev, E.V., Schilling, C.H., and Maranas, C.D. (2004). Flux coupling analysis of genome-scale metabolic network reconstructions. *Genome Res.* 14, 301–312.
12. Carbonell, P., and Faulon, J.-L. (2010). Molecular signatures-based prediction of enzyme promiscuity. *Bioinforma. Oxf. Engl.* 26, 2012–2019.
13. Caspi, R., Altman, T., Dreher, K., Fulcher, C.A., Subhraveti, P., Keseler, I.M., Kothari, A., Krummenacker, M., Latendresse, M., Mueller, L.A., et al. (2012). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.* 40, D742–D753.
14. Chandrasekaran, S., and Price, N.D. (2010). Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in *Escherichia coli* and *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. U. S. A.* 107, 17845–17850.
15. Chavali, A.K., D’Auria, K.M., Hewlett, E.L., Pearson, R.D., and Papin, J.A. (2012). A metabolic network approach for the identification and prioritization of antimicrobial drug targets. *Trends Microbiol.* 20, 113–123.
16. Copley, S.D. (2012). Moonlighting is mainstream: paradigm adjustment required. *BioEssays News Rev. Mol. Cell. Dev. Biol.* 34, 578–588.
17. Cordell, H.J. (2009). Detecting gene–gene interactions that underlie human diseases. *Nat. Rev. Genet.* 10, 392–404.
18. Costa, R.S., Machado, D., Rocha, I., and Ferreira, E.C. (2011). Critical perspective on the consequences of the limited availability of kinetic data in metabolic dynamic modelling. *IET Syst. Biol.* 5, 157–163.

19. Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E.D., Sevier, C.S., Ding, H., Koh, J.L.Y., Toufighi, K., Mostafavi, S., et al. (2010). The genetic landscape of a cell. *Science* 327, 425–431.
20. Covert, M.W., Schilling, C.H., Famili, I., Edwards, J.S., Goryanin, I.I., Selkov, E., and Palsson, B.O. (2001). Metabolic modeling of microbial strains in silico. *Trends Biochem. Sci.* 26, 179–186.
21. Covert, M.W., Knight, E.M., Reed, J.L., Herrgard, M.J., and Palsson, B.O. (2004). Integrating high-throughput and computational data elucidates bacterial networks. *Nature* 429, 92–96.
22. Dal'molin, C.G.O., Quek, L.-E., Palfreyman, R.W., and Nielsen, L.K. (2014). Plant genome-scale modeling and implementation. *Methods Mol. Biol. Clifton NJ* 1090, 317–332.
23. Dang, L., White, D.W., Gross, S., Bennett, B.D., Bittinger, M.A., Driggers, E.M., Fantin, V.R., Jang, H.G., Jin, S., Keenan, M.C., et al. (2009). Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* 465, 966.
24. Deutscher, D., Meilijson, I., Kupiec, M., and Rupp, E. (2006). Multiple knockout analysis of genetic robustness in the yeast metabolic network. *Nat. Genet.* 38, 993–998.
25. Dixon, S.J., Costanzo, M., Baryshnikova, A., Andrews, B., and Boone, C. (2009). Systematic Mapping of Genetic Interaction Networks. *Annu. Rev. Genet.* 43, 601–625.
26. Duarte, N.C., Herrgård, M.J., and Palsson, B.Ø. (2004). Reconstruction and Validation of *Saccharomyces cerevisiae* iND750, a Fully Compartmentalized Genome-Scale Metabolic Model. *Genome Res.* 14, 1298–1309.
27. Feist, A.M., and Palsson, B.O. (2010). The biomass objective function. *Curr. Opin. Microbiol.* 13, 344–349.
28. Feist, A.M., Henry, C.S., Reed, J.L., Krummenacker, M., Joyce, A.R., Karp, P.D., Broadbelt, L.J., Hatzimanikatis, V., and Palsson, B.Ø. (2007). A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol. Syst. Biol.* 3, 121.
29. Fersht, A. (1999). *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding* (W. H. Freeman).
30. Flachmann, R., Kunz, N., Seifert, J., Gütlisch, M., Wientjes, F.J., Läufer, A., and Gassen, H.G. (1988). Molecular biology of pyridine nucleotide biosynthesis in *Escherichia coli*. Cloning and characterization of quinolinate synthesis genes *nadA* and *nadB*. *Eur. J. Biochem. FEBS* 175, 221–228.
31. Förster, J., Famili, I., Fu, P., Palsson, B.Ø., and Nielsen, J. (2003). Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res.* 13, 244–253.
32. François, J., and Parrou, J.L. (2001). Reserve carbohydrates metabolism in the yeast *Saccharomyces cerevisiae*. *FEMS Microbiol. Rev.* 25, 125–145.
33. Fuhrer, T., Heer, D., Begemann, B., and Zamboni, N. (2011). High-throughput, accurate mass metabolome profiling of cellular extracts by flow injection-time-of-flight mass spectrometry. *Anal. Chem.* 83, 7074–7080.
34. Garfinkel, D., Garfinkel, L., Pring, M., Green, S.B., and Chance, B. (1970). Computer Applications to Biochemical Kinetics. *Annu. Rev. Biochem.* 39, 473–498.
35. Gelius-Dietrich, G., Desouki, A.A., Fritzemeier, C.J., and Lercher, M.J. (2013). Sybil—efficient constraint-based modelling in R. *BMC Syst. Biol.* 7, 125.
36. Geva-Zatorsky, N., Dekel, E., Cohen, A.A., Danon, T., Cohen, L., and Alon, U. (2010). Protein Dynamics in Drug Combinations: a Linear Superposition of Individual-Drug Responses. *Cell* 140, 643–651.

37. Giaever, G., Chu, A.M., Ni, L., Connelly, C., Riles, L., Véronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., André, B., et al. (2002). Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* *418*, 387–391.
38. Goldberg, D.E. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning* (Addison-Wesley).
39. Goldstein, L.S. (1993). Functional redundancy in mitotic force generation. *J. Cell Biol.* *120*, 1–3.
40. Griffiths, A.D., and Tawfik, D.S. (2003). Directed evolution of an extremely fast phosphotriesterase by in vitro compartmentalization. *EMBO J.* *22*, 24–35.
41. Harrison, R., Papp, B., Pál, C., Oliver, S.G., and Delneri, D. (2007). Plasticity of genetic interactions in metabolic networks of yeast. *Proc. Natl. Acad. Sci.* *104*, 2307–2312.
42. Hastings, J., de Matos, P., Dekker, A., Ennis, M., Harsha, B., Kale, N., Muthukrishnan, V., Owen, G., Turner, S., Williams, M., et al. (2013). The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for 2013. *Nucleic Acids Res.* *41*, D456–463.
43. He, X., Qian, W., Wang, Z., Li, Y., and Zhang, J. (2010). Prevalent positive epistasis in *Escherichia coli* and *Saccharomyces cerevisiae* metabolic networks. *Nat. Genet.* *42*, 272–276.
44. Heavner, B.D., Smallbone, K., Price, N.D., and Walker, L.P. (2013). Version 6 of the consensus yeast metabolic network refines biochemical coverage and improves model performance. *Database* *2013*, bat059–bat059.
45. Heckmann, D., Schulze, S., Denton, A., Gowik, U., Westhoff, P., Weber, A.P.M., and Lercher, M.J. (2013). Predicting C4 photosynthesis evolution: modular, individually adaptive steps on a Mount Fuji fitness landscape. *Cell* *153*, 1579–1588.
46. Heinemann, M., and Sauer, U. (2010). Systems biology of microbial metabolism. *Curr. Opin. Microbiol.* *13*, 337–343.
47. Huang, H., Pandya, C., Liu, C., Al-Obaidi, N.F., Wang, M., Zheng, L., Toews Keating, S., Aono, M., Love, J.D., Evans, B., et al. (2015). Panoramic view of a superfamily of phosphatases through substrate profiling. *Proc. Natl. Acad. Sci. U. S. A.* *112*, E1974–1983.
48. Ibarra, R.U., Edwards, J.S., and Palsson, B.O. (2002). *Escherichia coli* K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature* *420*, 186–189.
49. Jensen, R.A. (1976). Enzyme recruitment in evolution of new function. *Annu. Rev. Microbiol.* *30*, 409–425.
50. Jürgens, C., Strom, A., Wegener, D., Hettwer, S., Wilmanns, M., and Sterner, R. (2000). Directed evolution of a ($\beta\alpha$)8-barrel enzyme to catalyze related reactions in two different metabolic pathways. *Proc. Natl. Acad. Sci. U. S. A.* *97*, 9925–9930.
51. Kanehisa, M., and Goto, S. (1999). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* *28*, 27–30.
52. Kauffman, K.J., Prakash, P., and Edwards, J.S. (2003). Advances in flux balance analysis. *Curr. Opin. Biotechnol.* *14*, 491–496.
53. Khersonsky, O., and Tawfik, D.S. (2010). Enzyme promiscuity: a mechanistic and evolutionary perspective. *Annu. Rev. Biochem.* *79*, 471–505.
54. Kim, S. (2012). ppcor: Partial and Semi-partial (Part) correlation.
55. Kim, J., and Copley, S.D. (2012). Inhibitory cross-talk upon introduction of a new metabolic pathway into an existing metabolic network. *Proc. Natl. Acad. Sci.* *109*, E2856–E2864.
56. Kim, J., and Reed, J.L. (2010). OptORF: Optimal metabolic and regulatory perturbations for metabolic engineering of microbial strains. *BMC Syst. Biol.* *4*, 53.

57. Kim, S., Thiessen, P.A., Bolton, E.E., Chen, J., Fu, G., Gindulyte, A., Han, L., He, J., He, S., Shoemaker, B.A., et al. (2016). PubChem Substance and Compound databases. *Nucleic Acids Res.* *44*, D1202-1213.
58. King, R.D., Rowland, J., Oliver, S.G., Young, M., Aubrey, W., Byrne, E., Liakata, M., Markham, M., Pir, P., Soldatova, L.N., et al. (2009). The Automation of Science. *Science* *324*, 85–89.
59. Kitagawa, M., Ara, T., Arifuzzaman, M., Ioka-Nakamichi, T., Inamoto, E., Toyonaga, H., and Mori, H. (2005). Complete set of ORF clones of Escherichia coli ASKA library (a complete set of E. coli K-12 ORF archive): unique resources for biological research. *DNA Res. Int. J. Rapid Publ. Rep. Genes Genomes* *12*, 291–299.
60. Kuepfer, L., Sauer, U., and Blank, L.M. (2005). Metabolic functions of duplicate genes in *Saccharomyces cerevisiae*. *Genome Res.* *15*, 1421–1430.
61. Kumar, V.S., and Maranas, C.D. (2009). GrowMatch: An Automated Method for Reconciling In Silico/In Vivo Growth Predictions. *PLoS Comput Biol* *5*, e1000308.
62. Kuznetsova, E., Proudfoot, M., Gonzalez, C.F., Brown, G., Omelchenko, M.V., Borozan, I., Carmel, L., Wolf, Y.I., Mori, H., Savchenko, A.V., et al. (2006). Genome-wide analysis of substrate specificities of the Escherichia coli haloacid dehalogenase-like phosphatase family. *J. Biol. Chem.* *281*, 36149–36161.
63. Lee, D., Smallbone, K., Dunn, W.B., Murabito, E., Winder, C.L., Kell, D.B., Mendes, P., and Swainston, N. (2012). Improving metabolic flux predictions using absolute gene expression data. *BMC Syst. Biol.* *6*, 73.
64. Lee, S.J., Lee, D.-Y., Kim, T.Y., Kim, B.H., Lee, J., and Lee, S.Y. (2005). Metabolic Engineering of Escherichia coli for Enhanced Production of Succinic Acid, Based on Genome Comparison and In Silico Gene Knockout Simulation. *Appl. Environ. Microbiol.* *71*, 7880–7887.
65. Lehner, B. (2007). Modelling genotype–phenotype relationships and human disease with genetic interaction networks. *J. Exp. Biol.* *210*, 1559–1566.
66. Levy, R., and Borenstein, E. (2013). Metabolic modeling of species interaction in the human microbiome elucidates community-level assembly rules. *Proc. Natl. Acad. Sci. U. S. A.* *110*, 12804–12809.
67. Lewis, N.E., and Abdel-Haleem, A.M. (2013). The evolution of genome-scale models of cancer metabolism. *Front. Physiol.* *4*, 237.
68. Lewis, N.E., Nagarajan, H., and Palsson, B.O. (2012). Constraining the metabolic genotype–phenotype relationship using a phylogeny of in silico methods. *Nat. Rev. Microbiol.* *10*, 291–305.
69. Machado, D., and Herrgård, M. (2014). Systematic evaluation of methods for integration of transcriptomic data into constraint-based models of metabolism. *PLoS Comput. Biol.* *10*, e1003580.
70. Mahadevan, R., Edwards, J.S., and Doyle III, F.J. (2002). Dynamic Flux Balance Analysis of Diauxic Growth in Escherichia coli. *Biophys. J.* *83*, 1331–1340.
71. Martínez, R., and Schwaneberg, U. (2013). A roadmap to directed enzyme evolution and screening systems for biotechnological applications. *Biol. Res.* *46*, 395–405.
72. Mattevi, A., Tedeschi, G., Bacchella, L., Coda, A., Negri, A., and Ronchi, S. (1999). Structure of L-aspartate oxidase: implications for the succinate dehydrogenase/fumarate reductase oxidoreductase family. *Structure* *7*, 745–756.
73. Maxwell, C.A., Moreno, V., Sole, X., Gomez, L., Hernandez, P., Urruticoechea, A., and Pujana, M.A. (2008). Genetic interactions: the missing links for a better understanding of cancer susceptibility, progression and treatment. *Mol. Cancer* *7*, 4.

74. McCloskey, D., Palsson, B.Ø., and Feist, A.M. (2013). Basic and applied uses of genome-scale metabolic network reconstructions of *Escherichia coli*. *Mol. Syst. Biol.* 9, 661.
75. Merlin, C., Masters, M., McAteer, S., and Coulson, A. (2003). Why is carbonic anhydrase essential to *Escherichia coli*? *J. Bacteriol.* 185, 6415–6424.
76. Mo, M.L., Palsson, B.Ø., and Herrgård, M.J. (2009). Connecting extracellular metabolomic measurements to intracellular flux states in yeast. *BMC Syst. Biol.* 3, 37.
77. Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., and Snyder, M. (2008). The Transcriptional Landscape of the Yeast Genome Defined by RNA Sequencing. *Science* 320, 1344–1349.
78. Notebaart, R.A., Szappanos, B., Kintsjes, B., Pál, F., Györkei, Á., Bogos, B., Lázár, V., Spohn, R., Csörgő, B., Wagner, A., et al. (2014). Network-level architecture and the evolutionary potential of underground metabolism. *Proc. Natl. Acad. Sci.* 111, 11762–11767.
79. Oberhardt, M.A., Palsson, B.Ø., and Papin, J.A. (2009). Applications of genome-scale metabolic reconstructions. *Mol. Syst. Biol.* 5, 320.
80. Ohno, S. (1970). *Evolution by Gene Duplication* (Heidelberg, Germany: Springer Berlin Heidelberg).
81. Orman, M.A., Berthiaume, F., Androulakis, I.P., and Ierapetritou, M.G. (2011). Advanced Stoichiometric Analysis of Metabolic Networks of Mammalian Systems. *Crit. Rev. Biomed. Eng.* 39, 511–534.
82. Orth, J.D., and Palsson, B. (2012). Gap-filling analysis of the iJO1366 *Escherichia coli* metabolic network reconstruction for discovery of metabolic functions. *BMC Syst. Biol.* 6, 30.
83. Orth, J.D., Thiele, I., and Palsson, B.Ø. (2010). What is flux balance analysis? *Nat. Biotechnol.* 28, 245–248.
84. Orth, J.D., Conrad, T.M., Na, J., Lerman, J.A., Nam, H., Feist, A.M., and Palsson, B.Ø. (2011). A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism—2011. *Mol. Syst. Biol.* 7, 535.
85. Pál, C., Papp, B., and Lercher, M.J. (2005). Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. *Nat. Genet.* 37, 1372–1375.
86. Palsson, B. (2009). METABOLIC SYSTEMS BIOLOGY. *FEBS Lett.* 583, 3900–3904.
87. Panozzo, C., Nawara, M., Suski, C., Kucharczyka, R., Skoneczny, M., Bécam, A.-M., Rytka, J., and Herbert, C.J. (2002). Aerobic and anaerobic NAD⁺ metabolism in *Saccharomyces cerevisiae*. *FEBS Lett.* 517, 97–102.
88. Papp, B., Notebaart, R.A., and Pál, C. (2011). Systems-biology approaches for predicting genomic evolution. *Nat. Rev. Genet.* 12, 591–602.
89. Park, J.H., and Lee, S.Y. (2008). Towards systems metabolic engineering of microorganisms for amino acid production. *Curr. Opin. Biotechnol.* 19, 454–460.
90. Patrick, W.M., Quandt, E.M., Swartzlander, D.B., and Matsumura, I. (2007). Multicopy suppression underpins metabolic evolvability. *Mol. Biol. Evol.* 24, 2716–2722.
91. Phillips, P.C. (2008). Epistasis — the essential role of gene interactions in the structure and evolution of genetic systems. *Nat. Rev. Genet.* 9, 855–867.
92. Poelwijk, F.J., Kiviet, D.J., Weinreich, D.M., and Tans, S.J. (2007). Empirical fitness landscapes reveal accessible evolutionary paths. *Nature* 445, 383–386.
93. Price, N.D., Reed, J.L., and Palsson, B.Ø. (2004). Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nat. Rev. Microbiol.* 2, 886–897.
94. R Core Team (2013). *R: A Language and Environment for Statistical Computing* (Vienna, Austria: R Foundation for Statistical Computing).

95. Raman, K., and Chandra, N. (2009). Flux balance analysis of biological systems: applications and challenges. *Brief. Bioinform.* 10, 435–449.
96. Ravasz, E., Somera, A.L., Mongru, D.A., Oltvai, Z.N., and Barabási, A.-L. (2002). Hierarchical Organization of Modularity in Metabolic Networks. *Science* 297, 1551–1555.
97. Reiczigel, J., Harnos, A., and Solymosi, N. (2007). *Biostatistika nem statisztikusoknak* (Pars Kft.).
98. Rison, S.C.G., Teichmann, S.A., and Thornton, J.M. (2002). Homology, pathway distance and chromosomal localization of the small molecule metabolism enzymes in *Escherichia coli*. *J. Mol. Biol.* 318, 911–932.
99. Rocha, M., Maia, P., Mendes, R., Pinto, J.P., Ferreira, E.C., Nielsen, J., Patil, K.R., and Rocha, I. (2008). Natural computation meta-heuristics for the in silico optimization of microbial strains. *BMC Bioinformatics* 9, 499.
100. Ruppín, E., Papin, J.A., de Figueiredo, L.F., and Schuster, S. (2010). Metabolic reconstruction, constraint-based analysis and game theory to probe genome-scale metabolic networks. *Curr. Opin. Biotechnol.* 21, 502–510.
101. Scheer, M., Grote, A., Chang, A., Schomburg, I., Munaretto, C., Rother, M., Söhngen, C., Stelzer, M., Thiele, J., and Schomburg, D. (2011). BRENDA, the enzyme information system in 2011. *Nucleic Acids Res.* 39, D670–676.
102. Schellenberger, J., Park, J.O., Conrad, T.M., and Palsson, B.Ø. (2010). BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions. *BMC Bioinformatics* 11, 213.
103. Schmidt, S., Sunyaev, S., Bork, P., and Dandekar, T. (2003). Metabolites: a helping hand for pathway evolution? *Trends Biochem. Sci.* 28, 336–341.
104. Schuetz, R., Kuepfer, L., and Sauer, U. (2007). Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. *Mol. Syst. Biol.* 3.
105. Segrè, D., Vitkup, D., and Church, G.M. (2002). Analysis of optimality in natural and perturbed metabolic networks. *Proc. Natl. Acad. Sci.* 99, 15112–15117.
106. Segrè, D., DeLuna, A., Church, G.M., and Kishony, R. (2005). Modular epistasis in yeast metabolism. *Nat. Genet.* 37, 77–83.
107. Shlomi, T., Berkman, O., and Ruppín, E. (2005). Regulatory on/off minimization of metabolic flux changes after genetic perturbations. *Proc. Natl. Acad. Sci. U. S. A.* 102, 7695–7700.
108. Shlomi, T., Herrgård, M., Portnoy, V., Naim, E., Palsson, B.Ø., Sharan, R., and Ruppín, E. (2007a). Systematic condition-dependent annotation of metabolic genes. *Genome Res.* 17, 1626–1633.
109. Shlomi, T., Eisenberg, Y., Sharan, R., and Ruppín, E. (2007b). A genome-scale computational study of the interplay between transcriptional regulation and metabolism. *Mol. Syst. Biol.* 3, 101.
110. Shlomi, T., Cabili, M.N., Herrgård, M.J., Palsson, B.Ø., and Ruppín, E. (2008). Network-based prediction of human tissue-specific metabolism. *Nat. Biotechnol.* 26, 1003–1010.
111. Sing, T., Sander, O., Beerenwinkel, N., and Lengauer, T. (2005). ROCr: visualizing classifier performance in R. *Bioinformatics* 21, 7881.
112. Smallbone, K., Simeonidis, E., Swainston, N., and Mendes, P. (2010). Towards a genome-scale kinetic model of cellular metabolism. *BMC Syst. Biol.* 4, 6.
113. Snitkin, E.S., Dudley, A.M., Janse, D.M., Wong, K., Church, G.M., and Segrè, D. (2008). Model-driven analysis of experimentally determined growth phenotypes for 465 yeast gene deletion mutants under 16 different conditions. *Genome Biol.* 9, R140.

114. Soo, V.W.C., Hanson-Manful, P., and Patrick, W.M. (2011). Artificial gene amplification reveals an abundance of promiscuous resistance determinants in *Escherichia coli*. *Proc. Natl. Acad. Sci.* *108*, 1484–1489.
115. Stanford, N.J., Lubitz, T., Smallbone, K., Klipp, E., Mendes, P., and Liebermeister, W. (2013). Systematic construction of kinetic models from genome-scale metabolic networks. *PloS One* *8*, e79195.
116. Stelling, J. (2004). Mathematical models in microbial systems biology. *Curr. Opin. Microbiol.* *7*, 513–518.
117. Stevens, F.J., and Wu, T.T. (1976). Growth on D-lyxose of a mutant strain of *Escherichia coli* K12 using a novel isomerase and enzymes related to D-xylase metabolism. *J. Gen. Microbiol.* *97*, 257–265.
118. Szappanos, B., Kovács, K., Szamecz, B., Honti, F., Costanzo, M., Baryshnikova, A., Gelius-Dietrich, G., Lercher, M.J., Jelasity, M., Myers, C.L., et al. (2011). An integrated approach to characterize genetic interaction networks in yeast metabolism. *Nat. Genet.* *43*, 656–662.
119. Thiele, I., Swainston, N., Fleming, R.M.T., Hoppe, A., Sahoo, S., Aurich, M.K., Haraldsdottir, H., Mo, M.L., Rolfsson, O., Stobbe, M.D., et al. (2013a). A community-driven global reconstruction of human metabolism. *Nat. Biotechnol.* *31*, 419–425.
120. Thiele, I., Heinken, A., and Fleming, R.M. (2013b). A systems biology approach to studying the role of microbes in human health. *Curr. Opin. Biotechnol.* *24*, 4–12.
121. Tomasulo, P. (2002). ChemIDplus-super source for chemical and drug information. *Med. Ref. Serv. Q.* *21*, 53–59.
122. Tong, A.H.Y., and Boone, C. (2006). Synthetic genetic array analysis in *Saccharomyces cerevisiae*. *Methods Mol. Biol. Clifton NJ* *313*, 171–192.
123. 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., et al. (2004). Global mapping of the yeast genetic interaction network. *Science* *303*, 808–813.
124. Toya, Y., and Shimizu, H. (2013). Flux analysis and metabolomics for systematic metabolic engineering of microorganisms. *Biotechnol. Adv.* *31*, 818–826.
125. Ulitsky, I., Krogan, N.J., and Shamir, R. (2009). Towards accurate imputation of quantitative genetic interactions. *Genome Biol.* *10*, R140.
126. Väre, L., Nookaew, I., and Nielsen, J. (2013). Novel insights into obesity and diabetes through genome-scale metabolic modeling. *Front. Physiol.* *4*, 92.
127. Villiers, B.R.M., and Hollfelder, F. (2009). Mapping the Limits of Substrate Specificity of the Adenylation Domain of TycA. *ChemBioChem* *10*, 671–682.
128. Wagner, A., and Fell, D.A. (2001). The small world inside large metabolic networks. *Proc. Biol. Sci.* *268*, 1803–1810.
129. Webb, E.C., and NC-IUBMB (1992). *Enzyme Nomenclature 1992: Recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology and the Nomenclature and Classification of Enzymes* (Academic Press).
130. Weng, J.-K., Philippe, R.N., and Noel, J.P. (2012). The rise of chemodiversity in plants. *Science* *336*, 1667–1670.
131. 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., et al. (2004). Combining biological networks to predict genetic interactions. *Proc. Natl. Acad. Sci. U. S. A.* *101*, 15682–15687.
132. Yang, Y.-T., Bennett, G.N., and San, K.-Y. (1998). Genetic and metabolic engineering. *Electron. J. Biotechnol.* *1*, 134–141.

7. Összefoglalás

Az elmúlt néhány évtizedben a kísérletes technikák fejlődése az elérhető biológiai adatok mennyiségének exponenciális növekedéséhez vezetett. Az adatok feldolgozásához új módszereket kellett kifejleszteni, a számítógépes adatfeldolgozás egyre inkább előtérbe került. Emellett lehetővé vált a biológiai rendszerek vizsgálata holisztikus megközelítéssel, ami a rendszerbiológia kialakulásához vezetett. A rendszerbiológiai kutatások egyik kiemelt célpontja az anyagcsere, mely az egyik legjobban feltárt sejtrendszer. Az anyagcseréről elérhető nagy mennyiségű adat lehetővé tette matematikai modellek kidolgozását. Ezekkel a modellekkel a genotípus-fenotípus leképezést akár több száz génre kiterjedően, szisztematikusan tudjuk vizsgálni, ezáltal pedig evolúciobiológiai kérdésekre is válaszokat adhatunk velük. Doktori munkám során két evolúciobiológiai kérdéskört jártam körül az anyagcseremodellek felhasználásával. Egyrészt az anyagcseregének közötti genetikai interakciók hálózatos viselkedését és előrejelezhetőségét vizsgáltam, másrészt az enzimek járulékos aktivitásainak evolúciós jelentőségét kíséreltem meg felderíteni.

Genetikai interakciók vizsgálata és modellezése az anyagcserében

Genetikai kölcsönhatásról akkor beszélünk, ha két mutáció együttes hatása eltér attól, amit a mutációk önálló hatása alapján várnánk. A genetikai kölcsönhatás lehet negatív vagy pozitív, attól függően, hogy a kapott kettős mutáció hatása erősebb vagy gyengébb a vártnál. Munkánk során kiszámoltuk a genetikai kölcsönhatásokat az élesztő anyagcserehálózatban az úgynevezett fluxus balansz analízis (FBA) módszerrel. Az eredményt egy nagyskalájú kísérletes adatsorral vetettük össze, melyet kollaborátorunk (Charles Boone és csoportja) mért le. Ez az adatsor több, mint 176.000 anyagcseregénpárra tartalmazza a genetikai kölcsönhatás értékét.

Először azt vizsgáltuk meg, hogy az anyagcseremodell képes-e visszaadni kísérletesen megfigyelt mintázatokat a genetikai kölcsönhatások hálózatában és mechanikus magyarázatot adni rájuk. Kísérletesen megfigyelték, hogy minél több interakcióban vesz részt egy gén, annál károsabb a kiütése az élőlény számára. Ezt a mintázatot a modell is mutatja, emellett pedig azt is megfigyeltük *in silico*, hogy erős negatív korreláció van egy gén által érintett bioszintetikus folyamatok száma (pleiotrópia) és a gén kiütésének hatása között, valamint hogy a pleiotrópia és a genetikai kölcsönhatások száma pedig pozitív korrelációban van. Ez alapján feltételezhetjük, hogy a több biológiai funkcióban részt vevő (pleiotrópikus) gének

kiütésének hatását nagyszámú más gén mutációja befolyásolja, ahol az egyes gének a pleiotrópikus gén más-más funkcióját kompenzálják.

A következő lépésben arra kerestük a választ, hogy a modell milyen sikerességgel jelzi előre az egyes genetikai kölcsönhatásokat. Azt találtuk, hogy az *in silico* kölcsönhatások között erős feldúsulást mutatnak a kísérletes interakciók a véletlenszerűen várthoz képest, azonban a kísérletesen látott kölcsönhatásoknak csak egy kis hányadát találjuk meg a szimulációval. Annak érdekében, hogy javítsunk a modell szimulációs erején, kifejlesztettünk egy gépi tanuláson alapuló eljárást. Módszerünk különböző változásokat hajt végre a modellen automatikusan és a kísérletes adatsoron teszteli, hogy javult-e az modell előrejelző képessége. Sikerült nagymértékben feljavítani a modellt és az egyik javasolt változtatást kísérletesen is igazoltuk. Összefoglalva, bár az anyagcseremodell által előrejelzett genetikai kölcsönhatások biológiailag relevánsak, a kísérletesen azonosított kölcsönhatások többségét nem látja a modell. Ezen a hiányosságán azonban tudunk javítani.

Rejtett enzimaktivitások hálózata és evolúciós jelentősége

Annak megértése, hogy hogyan alakulnak ki új útvonalak az anyagcsere-hálózatban, az evolúcióbiológia és a rendszerbiológia egyik kulcskérdése. Az uralkodó elmélet szerint új útvonalak az anyagcserében már meglévő enzimekből épülnek fel, melyek a hálózat különböző pontjairól származnak. Régóta ismert, hogy az enzimek képesek olyan biokémiai reakciókat is felgyorsítani melléktermékként, amelyeknek a katalizálására nem specializálódtak. Ezeket a szelekciós nyomás alatt nem álló enzimaktivitásokat rejtett vagy látens aktivitásoknak hívjuk, szemben az enzimek natív vagy főaktivitásával. Ezek a látens aktivitások gyengék, fiziológiai hatásuk elhanyagolható, azonban mutációk révén felerősödhetnek, ezáltal pedig új enzimaktivitások és azokon keresztül új biokémiai útvonalak kialakulásához vezethetnek. Munkánk során a látens aktivitások fenotípusos hatását vizsgáltuk új környezetekhez való adaptációban.

Első lépésként összegyűjtöttük az irodalomban és adatbázisokban elérhető látens aktivitásokat, melyeket *E. coli*-ban írtak le. Ezeket a reakciókat a natív *E. coli* anyagcserébe illesztettük. Azt találtuk, hogy a látens reakciók többsége teljesen beilleszthető a natív hálózatba és csak egy kis részük az, amely a natív hálózattól izoláltan helyezkedik el. Ezáltal a látens aktivitások képesek új biokémiai útvonalakat kialakítani a natív reakciókkal közreműködve.

Ezután FBA módszerrel megvizsgáltuk nagyszámú *in silico* környezetben, hogy milyen fitneszelőnyt biztosítanak a látens aktivitások a natív modellhez képest. A látens reakciók a környezetek 2.9%-ában esszenciálisnak bizonyultak a növekedéshez. Minden ilyen környezetben elegendő volt egyetlen látens aktivitást hozzáadni a natív modellhez, hogy biztosítsuk a növekedést. Ugyanakkor ritka volt az olyan látens reakció, amely több környezetben is előnyt biztosított, tehát a látens reakciók hatása környezetfüggő.

A látens reakciók hatása kísérletesen felerősíthető, ha az adott enzimet túltermeltetjük az élőlénnel. Ez lehetőséget ad arra, hogy *in silico* eredményeinket kísérletesen is teszteljük. Kísérleteinkben 4269 *E. coli* gén túltermeltetésének hatását vizsgáltuk meg 194, különböző szénforrást tartalmazó környezetben (a kísérleteket Kintses Bálint kollégám végezte). Az eredmények azt mutatták, hogy a kísérletek során feljött környezetek jelentős részét sikerült a modellel előrejelezni, valamint a kísérletes és *in silico* vizsgálatok során feljött környezet – gén párok közötti átfedés lényegesen nagyobb a véletlenszerűen vártnál. Összességében tehát elmondhatjuk, hogy nemcsak képesek voltunk a modellel előrejelezni a látens aktivitások új környezethez való adaptációban betöltött szerepét, de az általunk összegyűjtött rejtett aktivitások listája jelentős részét lefedi annak az evolúciós nyersanyagnak, mely az új környezetekhez való rövidtávú adaptáció során az élőlény rendelkezésére áll.

8. Summary

New experimental techniques led to the exponential growth of available biological data recently. In order to process the great amount of data new computational methods had to be developed. Systems biology emerged as the holistic approach to investigate biological systems. Metabolism is one of the most well-known cellular subsystems and therefore a suitable model for exploring and answering the questions of systems biology. Large-scale mathematical models of metabolism can be reconstructed and used to predict the growth effect of different perturbations like gene deletion. These metabolic models are also capable of mapping genotype to phenotype and therefore to answer the questions of evolutionary biology. In my thesis I describe two works using metabolic modelling to investigate the network of genetic interactions and the evolutionary potential of underground enzyme activities.

Investigating genetic interactions by using metabolic modelling

Genetic interaction occurs when the effect of the simultaneous deletion of two genes is different than what we expect based on the effects of the single gene deletions. Genetic interaction can be negative when the effect of the double mutant is stronger than expected and positive when the effect is weaker. We applied flux balance analysis (FBA) method on the yeast metabolic network to calculate the genetic interactions between yeast enzymes. The results were compared with a large-scale experimental data set produced by our collaborators (Charles Boone lab, Toronto). This experimental dataset contains genetic interaction data for more than 176 000 enzyme pairs.

First, we tested the capability of the metabolic model to predict general patterns of the network of genetic interactions. Experimental genetic interaction datasets show that the more interactions a gene has the higher the growth effect of the deletion of the gene will be. We confirmed this pattern on the metabolic model as well. Besides, the metabolic model shows strong negative correlation between the numbers of biological functions a gene has (pleiotropy) and the single gene deletion fitness values. Gene pleiotropy also correlates with the degree of genetic interactions. Based on this we assume that the effect of deletion of genes with multiple biological functions (highly pleiotropic) is influenced by several other genes, each which compensate a different biological function of the highly pleiotropic gene.

Next, we compared the *in silico* genetic interactions with the experimental dataset to estimate the prediction accuracy of the model. We found that experimentally confirmed genetic interactions are strongly enriched among *in silico* interactions but the model can find only a small fraction of experimental interactions. In order to improve the model we developed an automated model refinement method based on a machine learning algorithm. Our method minimizes model mispredictions globally by creating and evolving a population of modified models. Several modifications were suggested that, together, considerably improved the fit of the model to our genetic interaction map. One of the hypotheses of our method was verified experimentally. Taken together, the metabolic model is capable to predict biologically relevant genetic interactions but it misses most of the experimentally verified interactions. This ability of the model can be improved by automated model refinement.

The evolutionary importance of underground metabolic reactions

Understanding the way new metabolic pathways emerge is a main goal of evolutionary biology and systems biology. The prevailing theory is that new pathways are patched together from already existing enzymes coming from different parts of the network. It is well-known that enzymes are capable of catalyzing, albeit at low rates, reactions other than those for which they have evolved. These so called underground reactions are prevalent and their activity can be increased through mutations therefore they have a key role in evolving new metabolic functions. During our work we investigated the phenotypic effect of underground activities in the adaptation to new conditions.

First, we systematically collected a list of well-known underground activities of *E. coli* enzymes and integrated them in a genome-wide *E. coli* metabolic model. 45% of the underground reactions could be fully integrated in the native network and only a small fraction of them is completely isolated from the native reactions. Therefore underground activities are capable of forming new biochemical pathways combined with native reactions.

Next, we applied flux balance analysis to calculate the fitness advantage of the underground reactions in 2754 *in silico* conditions. Underground reactions were essential in 2.9% of the conditions. In each of these conditions adding a single underground reaction was enough to enable growth. Besides, most underground reactions showed fitness advantage in only one condition.

The effect of underground activities can be enhanced experimentally by overexpressing the enzyme responsible for the underground activity. In order to validate our *in silico* results we did a systematic overexpression screen of 4269 *E. coli* genes in 194 different conditions (each containing a different carbon source). These experiments were done by my colleague, Bálint Kintsés. We found that our *in silico* model could predict 44% of the conditions emerged from the experimental screen. Moreover the overlap between the experimental and predicted condition – gene pairs was 2 orders of magnitude higher than randomly expected. These results suggest that our network of *in silico* underground reactions already covers a significant part of the evolutionary raw material available for short-term adaptation to novel environments.

9. Függelék

1. táblázat: a genetikai interakciók számítógépes szimulációjához használt *in silico* környezet összetétele

A tápanyagok mellé írt felvételi ráták mmol/óra/(gramm száraz anyag) mértékegységben értendőek (mmol/h/gDW).

Tápanyag	Felvételi ráta
4-aminobenzoát	0.000002
Adenin	3.01
L-alanin	0.36
L-aszparagin	0.36
L-aszpartát	0.36
Biotin	0.00000142
L-cisztein	0.36
Fe ²⁺	1000
D-glukóz	22.6
L-glutamin	0.36
L-glutamát	3.6
Glicin	0.36
L-izoleucin	0.36
Mio-inozitol	0.11
K ⁺	4.44
L-leucin	1.8
L-metionin	0.36
Na ⁺	0.75
Nikotinsav	0.000002
O ₂	6.3
L-fenilalanin	0.36
Foszfát	0.89
(R)-pantoténsav	0.0002
L-prolin	0.36
Riboflavin	0.00092
L-szerin	0.36
Szulfát	100
Tiamin	0.0032
L-treonin	0.36
L-triptofán	0.36
L-tirozin	0.36
Uracil	3.63
L-valin	0.36

2. táblázat: az élesztő anyagcseremodell biomassa komponensei

1,3-béta-D-Glükán	L-Leucin
L-Alanin	L-Lizin
AMP	Mannán
L-Arginin	L-Metionin
L-Aszparagin	NAD
L-Aszpartát	Foszfátidát, élesztő-specifikus
ATP	Foszfátidil-kolin, élesztő-specifikus
cAMP	Foszfátidil-etanolamin, élesztő-specifikus
Kitin (monomer)	L-Fenilalanin
CMP	Protohem
Koenzim A	L-Prolin
L-Cisztein	Foszfátidil-szerin, élesztő-specifikus
dAMP	Foszfátidil-1-D-mio-inozitol, élesztő-specifikus
dCMP	Ubikinon-6
dGMP	Riboflavin
dTMP	L-Szerin
Ergoszterol	Szulfát
FAD	5,6,7,8-Tetrahidrofolát
L-Glutamin	Tiamin-trifoszfát
L-Glutamát	L-Treonin
Glicin	Trehalóz
Glikogén	Triglicerid, élesztő-specifikus
GMP	L-Triptofán
Redukált glutation	L-Tirozin
H ₂ O	UMP
L-Hisztidin	L-Valin
L-Izoleucin	Zimoszterol

3. táblázat: módosítható biomassa-komponensek listája

Ergoszterol
Glikogén
Trehalóz
cAMP
Kitin
Koenzim A
FAD
Redukált glutation
Protohem
Mannóz-(inozitol-P)2-ceramid, ceramid-1 (24C), élesztő-specifikus
Mannóz -(inozitol-P)2-ceramid, ceramid-1 (26C), élesztő-specifikus
Mannóz -(inozitol-P)2-ceramid, ceramid-2 (24C), élesztő-specifikus
Mannóz -(inozitol-P)2-ceramid, ceramid-2 (26C), élesztő-specifikus
Mannóz -(inozitol-P)2-ceramid, ceramid-3 (24C), élesztő-specifikus
Mannóz -(inozitol-P)2-ceramid, ceramid-3 (26C), élesztő-specifikus
NAD
Riboflavin
5,6,7,8-Tetrahidrofolát
Tiamin trifoszfát
Ubiquinon-6
Szulfát

4. táblázat: a javító algoritmus által javasolt módosítások az anyagcseremodellen

Érintett reakciók vagy biomassa komponens ^a		ASPOcm, QULNS, NMNAT, RNMK	ADPT, PUNP1, PUNP5	ASPOcm, QULNS	Glikogén	ACALDtm	Ubikinon-6
Javasolt módosítás		Kiütés (ASPOcm vagy QULNS) és (NMNAT vagy RNMK)	ADPT kiütése, PUNP1 és PUNP5 irreverzibilisre állítása	Kiütés	Kizárás a biomasszából	Irreverzibilisre állítás	Kizárás a biomasszából
Javított előrejelzés ^b	Negatív kölcsönhatás	9 új, helyesen előrejelzett negatív kölcsönhatás megjelent (TNA1-BNA1, TNA1-BNA2, TNA1-BNA4, TNA1-BNA5, NPT1-BNA1, NPT1-BNA2, NPT1-BNA4, NPT1-BNA5, NPT1-BNA6) valamint 3 hibásan jelzett negatív kölcsönhatás eltűnt (NMA1-TNA1, NMA1-PNC1, NMA1-NRK1)	5 új, helyesen előrejelzett negatív kölcsönhatás megjelent (HPT1-ADE1, HPT1-ADE2, HPT1-ADE4, HPT1-ADE5,7, HPT1-ADE6)	4 új, helyesen előrejelzett negatív kölcsönhatás megjelent (TNA1-BNA4, TNA1-BNA1, TNA1-BNA2, TNA1-BNA5)	2 hibásan jelzett negatív kölcsönhatás eltűnt (GSY1-GSY2, GLG1-GLG2)	1 új, helyesen előrejelzett negatív kölcsönhatás megjelent (ALD6-PDB1)	1 hibásan jelzett negatív kölcsönhatás eltűnt (ARO3-ARO4)

Érintett reakciók vagy biomassza komponens ^a		ASPOcm, QULNS, NMNAT, RNMK	ADPT, PUNP1, PUNP5	ASPOcm, QULNS	Glikogén	ACALDtm	Ubikinon-6
	Egyszeres mutáns	1 új, helyesen előrejelzett esszenciális gén megjelent (QNS1)	-	-	-	-	5 új, helyesen előrejelzett esszenciális gén megjelent (COQ2, COQ3, COQ5, COQ6, PPA2)
Elrontott előrejelzés	Negatív kölcsönhatás	1 helyesen előrejelzett negatív kölcsönhatás eltűnt (NMA1–NMA2) és 2 hibásan jelzett negatív kölcsönhatás megjelent (TNA1–BNA3, NPT1–BNA3)	1 hibásan jelzett negatív kölcsönhatás megjelent (ADE6–AAH1)	1 hibásan jelzett negatív kölcsönhatás megjelent (TNA1–BNA3)	-	2 hibásan jelzett negatív kölcsönhatás megjelent (ALD6–ACH1, ALD6–PDX1)	-
	Egyszeres mutáns	1 hibásan jelzett esszenciális gén megjelent (NMA1)	-	-	-	-	-
Leírás/Magyarázat		Tévesen van jelen egy alternatív NAD bioszintézis útvonal a modellben; hibás gén-reakció hozzárendelés és reakció reverzibilitás	Alternatív AMP bioszintézis útvonalak ^e	Tévesen van jelen egy alternatív NAD bioszintézis útvonal a modellben	A glikogén nem esszenciális a növekedéshez	Alternatív acetát bioszintézis útvonal ^e	Ubikinon-6 csak nem-fermentálható szénforráson esszenciális
Gyakoriság ^f		8	1	8	8	8	4
MCC növekmény ^d	Negatív kölcsönhatás	0.069842	0.040612	0.032538	0.006114	0.003958	0.002937
	Egyszeres mutáns	0.000775	0	0	0	0	0.032093

^a Reakciók rövidítése: ASPOcm, aszpartát oxidáz; QULNS, kinolinát szintáz; NMNAT, nikotinamid-nukleotid adenililtranszferáz; RNMK, ribozil-nikotinamid kináz; ADPT, adenin foszforibozil-transzferáz; PUNP1, purinnukleozid foszforiláz (Adenozin); PUNP5, purinnukleozid foszforiláz (Inozin); ACALDtm, acetaldehid mitokondriális diffúzió.

^b Az érintett gének és génpárok zárójelben vannak felsorolva.

^c Megadja, hogy 8 független optimalizációból hány esetben jött fel a kérdéses módosítás.

^d Az MCC növekmény a módosított és az eredeti anyagcseremodell előrejelzéseire számolt MCC értékek különbsége, megadja, hogy a kérdéses módosításnak mekkora hatása van az előrejelzések sikerességére. Az eredeti modell MCC értéke 0.111 a genetikai kölcsönhatásokra és 0.499 az egyszeres génkiütésekre.

^e Alternatív mechanizmusok lehetségesek ezért nincs pontos magyarázat megadva.

5. táblázat: látens reakciók

A reakció egyenletekben a metabolitok a modellben használt azonosítóikkal szerepelnek. A „Forrás” oszlop értéke 1, ha a látens reakció a Brenda adatbázisból származik, és 2, ha más forrásból. Utóbbi esetben a forráscikk PubMed azonosítója is fel van tüntetve.

EC-szám	Reakció azonosító	Reakció egyenlet	Társított gének	Forrás: BRENDA (1) vagy irodalom (2)
1.1.5.2	u0001	gal[c] + q8[c] <=> q8h2[c] + C02669[c]	b0124 or b0837	1
1.1.5.2	u0002	q8[c] + xyl-D[c] <=> q8h2[c] + C02266[c]	b0124 or b083	1
2.7.1.51	u0004	atp[c] + fru[c] -> adp[c] + f1p[c] + h[c]	b2803	1
6.3.5.4	u0005	asp-L[c] + gln-L[c] + gtp[c] + h2o[c] -> asn-L[c] + glu-L[c] + gmp[c] + h[c] + ppi[c]	b0674	1
3.5.4.1	u0006	h2o[c] + C00791[c] + h[c] -> nh4[c] + C02565[c]	b0337	1
6.3.3.3	u0007	ctp[c] + co2[c] + dann[c] -> cdp[c] + dtb[c] + (3) h[c] + pi[c]	b0778	1
1.2.1.10	u0008	coa[c] + nad[c] + ppal[c] <=> ppcoa[c] + nadh[c] + h[c]	b0351	1
1.2.1.10	u0009	coa[c] + nad[c] + btal[c] <=> btcoa[c] + nadh[c] + h[c]	b0351	1
2.3.1.79	u0010	accoa[c] + gam[p] <=> acgam[c] + coa[c] + h[c]	b0459	1
2.7.1.130	u0011	utp[c] + lipidAds[c] -> udp[c] + h[c] + lipidA[c]	b0915	1
2.7.1.130	u0012	gtp[c] + lipidAds[c] -> gdp[c] + h[c] + lipidA[c]	b0915	1
2.7.1.130	u0013	ctp[c] + lipidAds[c] -> cdp[c] + h[c] + lipidA[c]	b0915	1
2.1.2.1	u0014	C02115[c] + thf[c] <=> ala-D[c] + h2o[c] + mlthf[c]	b2551	1
2.1.2.1	u0015	C00872[c] <=> gly[c] + co2[c]	b2551	1
2.6.1.29	u0016	ptrc[c] + pyr[c] -> 4abutn[c] + ala-L[c]	b3073	1
2.6.1.29	u0017	15dap[c] + akc[c] -> glu-L[c] + C12455[c] + h[c]	b3073	1
3.2.1.23	u0018	6dgc[c] + h2o[c] -> gal[c] + glc-D[c]	b0344	1
3.2.1.23	u0019	C02578[c] + h2o[c] -> gal[c] + phenol[c]	b0344	1
2.8.1.6	u0020	cys-L[c] + 2 h[c] -> ala-L[c] + h2s[c]	b0775	1
2.7.8.8	u0021	cdpdddecg[c] + glyc[c] -> cmp[c] + h[c] + pg120[c]	b2585	1
2.7.8.8	u0022	cdpdtdecg[c] + glyc[c] -> cmp[c] + h[c] + pg140[c]	b2585	1
2.7.8.8	u0023	cdpdtdec7eg[c] + glyc[c] -> cmp[c] + h[c] + pg141[c]	b2585	1
2.7.8.8	u0024	cdpdhdecg[c] + glyc[c] -> cmp[c] + h[c] + pg160[c]	b2585	1
2.7.8.8	u0025	cdpdhdec9eg[c] + glyc[c] -> cmp[c] + h[c] + pg161[c]	b2585	1
2.7.8.8	u0026	cdpdodecg[c] + glyc[c] -> cmp[c] + h[c] + pg180[c]	b2585	1
2.7.8.8	u0027	cdpdodec11eg[c] + glyc[c] -> cmp[c] + h[c] + pg181[c]	b2585	1
1.2.1.39/1.2.1.8	u0028	bzal[c] + h2o[c] + nad[c] -> bz[c] + 2 h[c] + nadh[c]	b1385 or b0312	1
1.2.1.39	u0029	hxa[c] + h2o[c] + nad[c] -> hxa[c] + 2 h[c] + nadh[c]	b1385	1
1.2.1.39	u0030	hpa[c] + h2o[c] + nad[c] -> hpa[c] + 2 h[c] + nadh[c]	b1385	1
1.2.1.39	u0031	34dhpac[c] + h2o[c] + nad[c] -> 34dhpha[c] + 2 h[c] + nadh[c]	b1385	1
1.2.1.39	u0032	4hoxpacd[c] + h2o[c] + nad[c] -> 4hphac[c] + 2 h[c] + nadh[c]	b1385	1
6.3.2.1	u0033	ala-B[c] + ctp[c] + pant-R[c] -> cmp[c] + h[c] + pnto-R[c] + ppi[c]	b0133	1
6.3.2.1	u0034	ala-B[c] + gtp[c] + pant-R[c] -> gmp[c] + h[c] + pnto-R[c] + ppi[c]	b0133	1
6.3.2.1	u0035	ala-B[c] + itp[c] + pant-R[c] -> imp[c] + h[c] + pnto-R[c] + ppi[c]	b0133	1
6.3.2.1	u0036	ala-B[c] + utp[c] + pant-R[c] -> ump[c] + h[c] +	b0133	1

EC-szám	Reakció azonosító	Reakció egyenlet	Társított gének	Forrás: BRENDA (1) vagy irodalom (2)
		pnto-R[c] + ppi[c]		
1.1.1.95	u0037	akg[c] + 3 h[c] + nadh[c] -> C02630[c] + nad[c]	b2913	1
1.2.1.8	u0038	h2o[c] + nad[c] + ppal[c] -> 2 h[c] + nadh[c] + ppa[c]	b0312	1
1.2.1.8	u0039	glyald[c] + h2o[c] + nadp[c] -> glyc-R[c] + 2 h[c] + nadph[c]	b0312	1
1.2.1.8	u0040	glyald[c] + h2o[c] + nad[c] -> glyc-R[c] + 2 h[c] + nadh[c]	b0312	1
1.2.1.8	u0041	bzal[c] + h2o[c] + nadp[c] -> bz[c] + 2 h[c] + nadph[c]	b0312	1
1.2.1.8	u0042	mthgxl[c] + h2o[c] + nad[c] -> pyr[c] + 2 h[c] + nadh[c]	b0312	1
1.2.1.8	u0043	mthgxl[c] + h2o[c] + nadp[c] -> pyr[c] + 2 h[c] + nadph[c]	b0312	1
1.2.1.8	u0044	fald[c] + h2o[c] + nadp[c] -> for[c] + 2 h[c] + nadph[c]	b0312	1
1.2.1.8	u0045	fald[c] + h2o[c] + nad[c] -> for[c] + 2 h[c] + nadh[c]	b0312	1
1.2.1.8	u0046	btal[c] + h2o[c] + nadp[c] -> but[c] + 2 h[c] + nadph[c]	b0312	1
1.2.1.8	u0047	ptal[c] + h2o[c] + nadp[c] -> pta[c] + h[c] + nadph[c]	b0312	1
1.2.1.8	u0048	ptal[c] + h2o[c] + nad[c] -> pta[c] + h[c] + nadh[c]	b0312	1
1.2.1.8	u0049	gcald[c] + h2o[c] + nadp[c] -> glyclt[c] + 2 h[c] + nadph[c]	b0312	1
1.2.1.8	u0050	h2o[c] + nadp[c] + pacald[c] -> 2 h[c] + nadph[c] + pac[c]	b0312	1
1.2.1.8	u0051	h2o[c] + nadp[c] + udcal[c] -> 2 h[c] + nadph[c] + udca[c]	b0312	1
1.2.1.8	u0052	h2o[c] + nad[c] + udcal[c] -> 2 h[c] + nadh[c] + udca[c]	b0312	1
3.5.1.1	u0053	asn-D[c] + h2o[c] -> asp-D[c] + nh4[c]	b1767 or b0828	1
2.8.3.9	u0059	but[c] + accoa[c] -> acac[c] + btcoa[c]	b2221 and b2222	1
2.4.1.25	u0061	2 malt[c] -> glc-D[c] + malttr[c]	b3416	1
2.4.1.25	u0062	3 malttr[c] -> malttrr[c] + maltpt[c] + h2o[c]	b3416	1
3.1.3.10	u0063	C03360[c] + h2o[c] -> C00870[c] + pi[c]	b1002	1
2.7.4.3	u0064	gtp[c] + cdp[c] <=> gdp[c] + ctp[c]	b0474	1
2.2.1.7	u0065	hpyr[c] + g3p[c] + h[c] -> xu5p-D[c] + co2[c]	b0420	1
2.2.1.7	u0066	pyr[c] + r5p[c] + h2o[c] -> dDs7p[c] + co2[c]	b0420	1
2.2.1.7	u0067	pyr[c] + C01796[c] + h[c] -> dfu[c] + co2[c]	b0420	1
2.2.1.7	u0068	pyr[c] + gcald[c] + h[c] -> deryt[c] + co2[c]	b0420	1
2.2.1.7	u0069	pyr[c] + e4p[c] + h[c] -> df6p[c] + co2[c]	b0420	1
4.1.3.3	u0070	kdo[c] <=> pyr[c] + C00216[c]	b3225	1
2.7.1.17	u0071	ctp[c] + dxyl[c] -> cdp[c] + dxyl5p[c] + h[c]	b3564	1
2.7.1.17	u0072	utp[c] + dxyl[c] -> udp[c] + dxyl5p[c] + h[c]	b3564	1
2.7.1.17	u0073	itp[c] + dxyl[c] -> idp[c] + dxyl5p[c] + h[c]	b3564	1
2.7.1.17	u0074	gtp[c] + dxyl[c] -> gdp[c] + dxyl5p[c] + h[c]	b3564	1
2.7.1.17	u0075	atp[c] + xylt[c] + h[c] -> adp[c] + C02895[c]	b3564	1
6.1.1.6	u0076	atp[c] + arg-L[c] + trnalys[c] -> amp[c] + argtrnalys[c] + ppi[c]	b2890 or b4129	1
6.1.1.6	u0077	atp[c] + thr-L[c] + trnalys[c] -> amp[c] + thrtrnalys[c] + ppi[c]	b2890 or b4129	1
6.1.1.6	u0078	atp[c] + met-L[c] + trnalys[c] -> amp[c] + mettrnalys[c] + ppi[c]	b2890 or b4129	1
6.1.1.6	u0079	atp[c] + leu-L[c] + trnalys[c] -> amp[c] + leutrnalys[c] + ppi[c]	b2890 or b4129	1
6.1.1.6	u0080	atp[c] + ala-L[c] + trnalys[c] -> amp[c] + alatrnalys[c] + ppi[c]	b2890 or b4129	1

EC-szám	Reakció azonosító	Reakció egyenlet	Társított gének	Forrás: BRENDA (1) vagy irodalom (2)
6.1.1.6	u0081	atp[c] + ser-L[c] + trnals[c] -> amp[c] + sertrnals[c] + ppi[c]	b2890 or b4129	1
6.1.1.6	u0082	atp[c] + cys-L[c] + trnals[c] -> amp[c] + cystrnals[c] + ppi[c]	b2890 or b4129	1
6.1.1.6	u0083	atp[c] + glu-L[c] + trnals[c] -> amp[c] + glutrnals[c] + ppi[c]	b2890 or b4129	1
6.1.1.6	u0084	atp[c] + orn[c] + trnals[c] -> amp[c] + orntrnals[c] + ppi[c]	b2890 or b4129	1
3.6.3.2	u0087	atp[c] + cobalt2[p] + h2o[c] -> adp[c] + cobalt2[c] + h[c] + pi[c]	b4242	1
2.8.3.8	u0088	pta[c] + accoa[c] + 4 h[c] -> C00888[c] + ac[c]	b2221 and b2222	1
2.8.3.9	u0089	ac[c] + b2coa[c] + h[c] -> C01771[c] + accoa[c]	b2221 and b2222	1
2.8.3.9	u0090	accoa[c] + 3hbut-R[c] -> 3hbycoa[c] + ac[c]	b2221 and b2222	1
2.8.3.9	u0091	aacoa[c] + for[c] + 4 h[c] -> formcoa[c] + acac[c]	b2221 and b2222	1
1.2.4.2	u0092	akg[c] + C00248[c] + h[c] <=> sdhlm[c] + co2[c]	b0116 and b0726 and b0727	1
4.2.3.5	u0094	6fESEp[c] -> 3 h[c] + 6flu[c] + pi[c]	b2329	1
4.2.1.2	u0095	tartr-D[c] <=> h2o[c] + oaa[c]	b1612 or b4122 or b1611	1
4.2.1.2	u0096	h2o[c] + C03248[c] <=> oaa[c] + 2 h[c]	b1612 or b4122 or b1611	1
1.3.1.12	u0097	Largn[c] + nad[c] -> tyr-L[c] + nadh[c] + co2[c] + h[c]	b2600	1
2.7.4.22	u0098	gtp[c] + ump[c] <=> gdp[c] + udp[c]	b0171	1
6.2.1.5	u0102	gtp[c] + coa[c] + succ[c] <=> gdp[c] + pi[c] + succoa[c]	b0728 and b0729	1
4.1.2.4	u0103	drib[c] -> glyald[c] + acald[c]	b4381	1
2.4.2.14	u0107	prpp[c] + nh4[c] -> ppi[c] + pram[c]	b2312	1
6.1.1.9	u0108	atp[c] + trnavall[c] + ile-L[c] -> amp[c] + ppi[c] + isoleutrnava[c]	b4258	1
2.3.1.1	u0112	accoa[c] + adip[c] -> coa[c] + acahedio[c] + 2 h[c]	b2818	1
2.3.1.1	u0113	apim[c] + accoa[c] -> acahepdio[c] + coa[c]	b2818	1
2.3.1.1	u0114	glu-L[c] + ppcoa[c] + h[c] -> coa[c] + propglu[c]	b2818	1
2.7.4.16	u0115	itp[c] + thmmp[c] -> idp[c] + thmpp[c]	b0417	1
2.5.1.16	u0116	ametam[c] + spmd[c] -> 5mta[c] + h[c] + C00750[c]	b0121	1
2.5.1.15	u0117	C11169[c] + 2ahhmd[c] -> disulthi[c] + ppi[c]	b3177	1
2.5.1.15	u0118	C07458[c] + 2ahhmd[c] -> disulan[c] + ppi[c]	b3177	1
2.7.4.16	u0119	adp[c] + thmmp[c] -> amp[c] + thmpp[c]	b0417	1
2.7.4.12	u0120	datp[c] + dgm[c] <=> dadp[c] + dgdp[c]	b3648	1
2.7.4.12	u0121	datp[c] + dtm[c] <=> dadp[c] + dtdp[c]	b3648 or b1098	1
2.7.4.8	u0122	datp[c] + gmp[c] <=> dadp[c] + gdp[c]	b3648	1
2.7.1.39	u0123	atp[c] + hom-D[c] -> adp[c] + h[c] + pdhom[c]	b0003	1
2.7.1.39	u0124	atp[c] + 4hynor[c] + h[c] -> adp[c] + 4plnor[c]	b0003	1
2.7.1.39	u0125	atp[c] + l2amhy[c] + h[c] -> adp[c] + l2amphy[c]	b0003	1
2.7.1.33	u0126	atp[c] + npan[c] -> npanp[c] + adp[c] + h[c]	b3974	1
2.3.1.180	u0127	btcoa[c] + h[c] + malACP[c] -> 3ohexACP[c] + co2[c] + coa[c]	b1091	1
2.7.1.30	u0128	atp[c] + glyald[c] -> adp[c] + g3p[c] + h[c]	b3926	1
2.7.1.30	u0129	atp[c] + dha[c] -> adp[c] + dhap[c] + h[c]	b3926	1
2.7.1.30	u0130	atp[c] + C02426[c] -> adp[c] + g3p[c] + h[c]	b3926	1
4.2.3.1	u0131	athr-L[c] -> 2obut[c] + nh4[c]	b0004	1
4.2.3.1	u0132	C02635[c] + h2o[c] -> pyr[c] + nh4[c] + hcl[c]	b0004	1
4.2.3.1	u0133	vigly[c] + h2o[c] -> thr-L[c]	b0004	1

EC-szám	Reakció azonosító	Reakció egyenlet	Társított gének	Forrás: BRENDA (1) vagy irodalom (2)
4.1.2.13	u0134	f1p[c] <=> dhap[c] + glyald[c]	b2097 or b2925 or b1773	1
2.7.7.47	u0136	atp[c] + C17573[c] -> ppi[c] + 3adeblu[c]	x	1
2.7.7.47	u0137	atp[c] + C07655[c] + h[c] -> 3adebios[c] + ppi[c]	x	1
2.7.7.47	u0138	atp[c] + C02078[c] + h[c] -> 9adespec[c] + ppi[c]	x	1
5.4.2.5	u0140	C00586[c] + g1p[c] -> glc-D[c] + C06369[c]	x	1
5.4.2.5	u0141	glc-D[c] + f1p[c] -> g6p[c] + fru[c]	x	1
5.4.2.5	u0142	fru[c] + g1p[c] -> glc-D[c] + f6p[c]	x	1
5.4.2.5	u0143	C00247[c] + g1p[c] -> glc-D[c] + sor6p[c]	x	1
5.4.2.5	u0144	man[c] + g1p[c] -> glc-D[c] + man6p[c]	x	1
5.4.2.5	u0145	gal[c] + g1p[c] -> glc-D[c] + C01113[c]	x	1
5.4.2.5	u0146	3metglu[c] + g1p[c] -> glc-D[c] + 3metpglu[c]	x	1
6.1.1.15	u0147	atp[c] + cys-L[c] + trnapro[c] -> amp[c] + ppi[c] + cysttrnapro[c]	b0194	1
6.1.1.15	u0148	atp[c] + ala-L[c] + trnapro[c] -> amp[c] + ppi[c] + alatnapro[c]	b0194	1
3.5.1.16	u0149	h2o[c] + aceala[c] -> ac[c] + ala-L[c] + h[c]	b3957	1
3.5.1.16	u0150	benom[c] + h2o[c] -> bz[c] + orn[c]	b3957	1
3.5.1.16	u0151	aceglu[c] + h2o[c] -> gln-L[c] + ac[c]	b3957	1
3.5.1.16	u0152	acelys[c] + h2o[c] -> lys-L[c] + ac[c]	b3957	1
3.5.1.16	u0153	acedser[c] + h2o[c] -> ser-D[c] + ac[c]	b3957	1
3.5.1.16	u0154	acelser[c] + h2o[c] -> ser-L[c] + ac[c]	b3957	1
3.5.1.16	u0155	C02710[c] + h2o[c] -> leu-L[c] + ac[c]	b3957	1
3.5.1.16	u0156	aceasn[c] + h2o[c] -> asn-L[c] + ac[c]	b3957	1
3.5.1.16	u0157	acecys[c] + h2o[c] -> cys-L[c] + ac[c]	b3957	1
3.5.1.16	u0158	acelmet[c] + h2o[c] -> met-L[c] + ac[c]	b3957	1
3.5.1.16	u0159	acedmet[c] + h2o[c] -> met-D[c] + ac[c]	b3957	1
4.1.2.13	u0160	C00447[c] <=> e4p[c] + dhap[c]	b2097 or b2925 or b1773	1
4.4.1.8	u0161	C00491[c] + h2o[c] -> C01962[c] + nh4[c] + pyr[c]	b1622 or b3008	1
1.3.1.26	u0162	23dhdp[c] + h[c] + 3acepyrh[c] -> 3acepyr[c] + thdp[c]	b0031	1
3.1.1.45	u0164	napace[c] + h2o[c] -> C11714[c] + ac[c]	x	1
2.5.1.39	u0166	frdp[c] + 4hbz[c] -> farhyben[c] + ppi[c]	b4040	1
2.5.1.39	u0167	grdp[c] + 4hbz[c] -> gerhyben[c] + ppi[c]	b4040	1
2.5.1.39	u0168	C00353[c] + 4hbz[c] -> gergerhyben[c] + ppi[c]	b4040	1
3.2.1.52	u0169	niacedglu[c] + h2o[c] -> acgam[c] + C00870[c] + h[c]	b1107	1
2.7.7.58	u0170	24dhb[c] + atp[c] + h[c] -> 24dhba[c] + ppi[c]	b0594	1
2.7.7.58	u0171	25dhb[c] + atp[c] + h[c] -> 25dhba[c] + ppi[c]	b0594	1
2.7.7.58	u0172	2dhb[c] + atp[c] + h[c] -> 2dhba[c] + ppi[c]	b0594	1
2.7.7.58	u0173	234dhb[c] + atp[c] + h[c] -> 234dhba[c] + ppi[c]	b0594	1
6.3.4.4	u0174	asp-L[c] + xtp[c] + imp[c] -> dcamp[c] + xdp[c] + 2 h[c] + pi[c]	b4177	1
6.3.4.4	u0175	asp-L[c] + itp[c] + imp[c] -> dcamp[c] + idp[c] + 2 h[c] + pi[c]	b4177	1
6.3.4.4	u0176	asp-L[c] + utp[c] + imp[c] -> dcamp[c] + udp[c] + 2 h[c] + pi[c]	b4177	1
2.7.1.11	u0177	s7p[c] + atp[c] -> C00447[c] + adp[c] + h[c]	b3916 or b1723	1
2.7.1.11	u0179	ctp[c] + f6p[c] -> cdp[c] + fdp[c] + h[c]	b3916 or b1723	1

EC-szám	Reakció azonosító	Reakció egyenlet	Társított gének	Forrás: BRENDA (1) vagy irodalom (2)
2.7.1.11	u0180	itp[c] + f6p[c] -> idp[c] + fdp[c] + h[c]	b3916 or b1723	1
2.7.1.11	u0181	gtp[c] + f6p[c] -> gdp[c] + fdp[c] + h[c]	b3916 or b1723	1
2.7.1.11	u0182	utp[c] + f6p[c] -> udp[c] + fdp[c] + h[c]	b3916 or b1723	1
2.7.1.16	u0183	atp[c] + abt[c] -> adp[c] + h[c] + abt5p[c]	b0063	1
2.7.1.16	u0184	atp[c] + rbt[c] -> adp[c] + h[c] + rbt5p[c]	b0063	1
1.8.1.4	u0185	nadh[c] + C00324[c] <=> nad[c] + C01913[c] + h[c]	b0116	1
1.8.1.4	u0186	nadh[c] + C00248[c] + h[c] <=> nad[c] + C00579[c]	b0116	1
1.8.1.7	u0187	nadph[c] + C03170[c] <=> nadp[c] + C02090[c] + h[c]	b3500	1
1.8.1.2	u0188	5 h[c] + 3 nadph[c] + no2[c] -> 2 h2o[c] + 3 nadp[c] + nh4[c]	b2763 and b2764	1
1.8.1.2	u0189	2 h[c] + nadph[c] + C00192[c] -> h2o[c] + nadp[c] + nh4[c]	b2763 and b2764	1
6.3.1.5	u0191	atp[c] + dnad[c] + gln-L[c] + h2o[c] -> amp[c] + h[c] + nad[c] + ppi[c] + glu-L[c]	b1740	1
6.3.1.5	u0192	atp[c] + dnad[c] + asp-L[c] + h2o[c] -> amp[c] + 3 h[c] + nad[c] + ppi[c] + oaa[c]	b1740	1
6.3.1.5	u0193	datp[c] + dnad[c] + nh4[c] -> damp[c] + h[c] + nad[c] + ppi[c]	b1740	1
5.3.1.7	u0194	C00476[c] <=> xyly-D[c]	b3880	1 and 2: PMID: 18328504
1.1.1.282	u0195	nadp[c] + C00296[c] <=> 3dhq[c] + 2 h[c] + nadph[c]	b1692	1
2.7.4.9	u0196	ctp[c] + dtmp[c] <=> cdp[c] + dtdp[c]	b1098	1
2.7.4.9	u0197	dctp[c] + dtmp[c] <=> dcdp[c] + dtdp[c]	b1098	1
2.7.4.9	u0198	dgtp[c] + dtmp[c] <=> dgdp[c] + dtdp[c]	b1098	1
2.7.4.9	u0199	dtmp[c] + dttp[c] <=> 2 dtdp[c]	b1098	1
2.7.4.9	u0200	gtp[c] + dtmp[c] <=> gdp[c] + dtdp[c]	b1098	1
2.7.4.9	u0201	itp[c] + dtmp[c] <=> idp[c] + dtdp[c]	b1098	1
2.7.4.9	u0202	utp[c] + dtmp[c] <=> udp[c] + dtdp[c]	b1098	1
1.5.99.8	u0203	C00102[c] + pro-L[c] + 2 h[c] -> 1pyr5c[c] + C05720[c]	b1014	1
3.1.2.12	u0204	2napace[c] + h2o[c] -> C11713[c] + ac[c] + h[c]	b2154 or b0355	1
3.1.2.12	u0205	4nitpro[c] + h2o[c] -> C00870[c] + ppa[c] + h[c]	b2154 or b0355	1
3.1.2.12	u0206	4nitbut[c] + h2o[c] -> C00870[c] + but[c]	b2154 or b0355	1
3.1.2.12	u0207	4nitace[c] + h2o[c] -> C00870[c] + ac[c] + h[c]	b2154 or b0355	1
2.5.1.48	u0208	cys-L[c] + C01077[c] -> cyst-L[c] + h[c] + ac[c]	b3939	1
2.5.1.48	u0209	suchms[c] + h2o[c] -> 2obut[c] + h[c] + nh4[c] + succ[c]	b3939	1
2.6.1.85	u0211	chor[c] + nh4[c] -> 4adcho[c] + h2o[c]	b3360 and b1812	1
2.7.1.62	u0212	C02306[c] + C00586[c] + h[c] <=> deglu1p[c] + nh4[c]	x	1
2.7.1.62	u0213	C02306[c] + man[c] <=> man1p[c] + h[c] + nh4[c]	x	1
2.7.1.62	u0214	C02306[c] + fru[c] <=> f1p[c] + h[c] + nh4[c]	x	1
2.7.1.62	u0215	C02306[c] + gam[c] <=> gam1p[c] + h[c] + nh4[c]	x	1
3.5.4.25	u0216	dgtp[c] + 4 h2o[c] -> 25drapp[c] + for[c] + 4 h[c] + ppi[c]	b1277	1
3.5.4.25	u0217	4 h2o[c] + 25aripypp[c] -> 25drapp[c] + for[c] + 4 h[c] + ppi[c]	b1277	1
6.3.1.1	u0218	asp-L[c] + atp[c] + C00192[c] -> amp[c] + basphy[c] + ppi[c]	b3744	1
6.3.5.2	u0219	atp[c] + nh4[c] + xmp[c] -> amp[c] + ppi[c] + gmp[c] + 2 h[c]	b2507	1
1.14.11.17	u0221	akg[c] + o2[c] + butsul[c] -> btal[c] + co2[c] + 2 h[c] + so3[c] + succ[c]	b0368	1

EC-szám	Reakció azonosító	Reakció egyenlet	Társított gének	Forrás: BRENDA (1) vagy irodalom (2)
1.14.11.17	u0222	akg[c] + o2[c] + pensul[c] -> ptal[c] + co2[c] + 2 h[c] + so3[c] + succ[c]	b0368	1
5.1.3.2	u0223	dt dp glu[c] + 2 h[c] <=> C02097[c]	b0759	1
5.1.3.2	u0224	gal[c] <=> glc-D[c]	b0759	1
5.1.3.2	u0225	fru[c] <=> tag[c]	b0759	1
1.1.1.77	u0227	h[c] + p pal[c] + nadh[c] <=> C01845[c] + nad[c]	b2799	1
1.1.1.77	u0228	h[c] + g cald[c] + nadh[c] <=> C01380[c] + nad[c]	b2799	1
1.1.1.77	u0229	C02426[c] + h[c] + nadh[c] <=> glyc[c] + nad[c]	b2799	1
2.7.1.5	u0230	atp[c] + rib-D[c] -> adp[c] + h[c] + ru5p-D[c]	b3904	1
2.7.1.5	u0231	atp[c] + xylu-L[c] -> adp[c] + h[c] + C06441[c]	b3904	1
2.7.1.5	u0232	atp[c] + C01719[c] -> adp[c] + h[c] + C02976[c]	b3904	1
2.7.1.5	u0233	atp[c] + C00764[c] -> adp[c] + h[c] + sor1p-D[c]	b3904	1
2.6.1.42	u0235	val-L[c] + 3mop[c] <=> 3mob[c] + ile-L[c]	b3770 or b4054	1
2.6.1.42	u0236	val-L[c] + 4mop[c] <=> 3mob[c] + leu-L[c]	b3770 or b4054	1
2.6.1.42	u0237	ile-L[c] + 2obut[c] <=> 3mop[c] + C02356[c]	b3770 or b4054	1
2.6.1.42	u0238	C01933[c] + akg[c] + h[c] <=> glu-L[c] + C00902[c]	b3770 or b4054	1
2.6.1.42	u0239	C01826[c] + akg[c] + h[c] <=> glu-L[c] + C06255[c]	b3770 or b4054	1
1.4.99.1	u0240	arg-L[c] + h2o[c] + C00102[c] + 3 h[c] <=> nh4[c] + C05720[c] + C03771[c]	b1189	1
1.4.99.1	u0241	gln-L[c] + C00102[c] + h2o[c] <=> caroxbut[c] + C05720[c] + nh4[c]	b1189	1
1.4.99.1	u0242	met-D[c] + C00102[c] + h2o[c] <=> msuloxbut[c] + C05720[c] + nh4[c]	b1189	1
1.4.99.1	u0243	h2o[c] + asn-D[c] + C00102[c] + 5 h[c] <=> nh4[c] + C05720[c] + oxosuc[c]	b1189	1
1.4.99.1	u0244	h2o[c] + C00102[c] + his-D[c] + 3 h[c] <=> nh4[c] + C05720[c] + imi2oxpro[c]	b1189	1
1.4.99.1	u0245	h2o[c] + C00102[c] + phe-D[c] + 4 h[c] <=> nh4[c] + phpyr[c] + C05720[c]	b1189	1
1.4.99.1	u0246	ala-D[c] + h2o[c] + C00102[c] + 3 h[c] <=> nh4[c] + pyr[c] + C05720[c]	b1189	1
1.4.99.1	u0247	h2o[c] + C00102[c] + val-D[c] + 4 h[c] <=> 3mob[c] + nh4[c] + C05720[c]	b1189	1
1.5.1.2	u0250	nadph[c] + delpicar[c] -> nadp[c] + C00408[c]	b0386	1
2.3.1.39	u0251	2aceeth[c] + malcoa[c] + nh4[c] <=> coa[c] + acesmalcys[c] + 2 h[c]	b1094 and b1092	1
2.3.1.39	u0252	acealacys[c] + malcoa[c] + h[c] <=> coa[c] + acealamalcys[c]	b1094 and b1092	1
4.1.2.14	u0253	glyald[c] + pyr[c] <=> 2ddgln[c]	b1850	1
4.1.2.17	u0254	6dpsip[c] <=> dhap[c] + lald-D[c]	b2800 or b2738	1
4.1.2.17	u0255	rib1p[c] <=> dhap[c] + g cald[c]	b2800 or b2738	1
4.1.2.17	u0256	tag1p-D[c] <=> dhap[c] + C02426[c]	b2800 or b2738	1
4.1.2.17	u0257	psi1p[c] <=> dhap[c] + glyald[c]	b2800 or b2738	1
4.1.2.19	u0259	C06441[c] <=> dhap[c] + glyoald[c]	b3902	1
4.1.2.19	u0260	C02888[c] <=> dhap[c] + glyald[c]	b3902	1
4.1.2.19	u0261	sor1p-D[c] <=> dhap[c] + glyald[c]	b3902	1
4.1.2.19	u0262	dxy1p-L[c] <=> dhap[c] + acald[c]	b3902	1
4.1.2.19	u0263	ery1p-D[c] <=> dhap[c] + fald[c]	b3902	1
4.1.2.19	u0264	6dsor1p-D[c] <=> dhap[c] + lald-D[c]	b3902	1
2.7.1.45	u0265	2ddgln[c] + ctp[c] -> 2ddg6p[c] + cdp[c] + h[c]	b3526	1
2.7.1.45	u0266	2ddgln[c] + utp[c] -> 2ddg6p[c] + udp[c] + h[c]	b3526	1
2.7.1.45	u0267	2ddgln[c] + gtp[c] -> 2ddg6p[c] + gdp[c] + h[c]	b3526	1

EC-szám	Reakció azonosító	Reakció egyenlet	Társított gének	Forrás: BRENDA (1) vagy irodalom (2)
2.7.1.45	u0268	2ddgln[c] + itp[c] -> 2ddg6p[c] + idp[c] + h[c]	b3526	1
2.1.2.11	u0269	fald[c] + thf[c] -> h2o[c] + mlthf[c]	b0134	1
2.7.1.15	u0270	drib[c] + atp[c] -> 2dr5p[c] + adp[c] + h[c]	b3752	1
2.7.1.15	u0271	C00216[c] + atp[c] -> ara5p[c] + adp[c] + h[c]	b3752	1
2.7.1.15	u0272	xyl-D[c] + atp[c] -> C06814[c] + adp[c] + h[c]	b3752	1
1.13.11.16	u0277	C06728[c] + o2[c] -> 2h6x24dn[c] + h[c]	b0348	1
1.13.11.16	u0278	C00090[c] + o2[c] -> 2h6xh24dn[c] + h[c]	b0348	1
1.13.11.16	u0279	C02923[c] + o2[c] -> 2h6xhp24dn[c] + h[c]	b0348	1
2.7.2.4	u0280	C01044[c] + atp[c] + h[c] <=> nforaspp[c] + adp[c]	b0002 or b3940 or b4024	1
2.7.2.4	u0281	C01042[c] + atp[c] + h[c] <=> nacaspp[c] + adp[c]	b0002 or b3940 or b4024	1
2.7.2.4	u0282	nclrasp[c] + atp[c] + h[c] <=> nclraspp[c] + adp[c]	b0002 or b3940 or b4024	1
1.1.1.2	u0290	alein[c] + nadph[c] + h[c] <=> pno[c] + nadp[c]	b3011	1
1.1.1.2	u0291	ppal[c] + nadph[c] + h[c] <=> C01845[c] + nadp[c]	b3011	1
1.1.1.2	u0292	acald[c] + nadph[c] + h[c] <=> etoh[c] + nadp[c]	b3011	1
1.1.1.6	u0293	2 h[c] + mthgxl[c] -> lald-L[c]	b3945	1
3.1.3.-	u0300	3pg[c] + h2o[c] -> glyc-R[c] + pi[c]	b3715 or b3385 or b3198 or b0822	2: PMID: 16990279
3.1.3.5	u0302	h2o[c] + udp[c] -> pi[c] + ump[c] + h[c]	b4374	2: PMID: 16990279
3.1.3.18	u0304	2pg[c] + h2o[c] -> glyc-R[c] + pi[c]	b3385	2: PMID: 16990279
3.1.3.23	u0306	ru5p-L[c] + h2o[c] -> rbl-L[c] + pi[c]	b0822 or b2293 or b3697	2: PMID: 16990279
3.1.3.-	u0309	C00660[c] + h2o[c] -> g6p[c] + pi[c]	b0766 or b0844 or b1897	2: PMID: 16990279
3.1.3.11	u0310	C00665[c] + h2o[c] -> f6p[c] + pi[c]	b0200	2: PMID: 16990279
x	u0312	C01904[c] + nad[c] <=> xylu-D[c] + h[c] + nadh[c]	x	2: PMID: 181526
x	u0313	3php[c] + h2o[c] -> hpyr[c] + pi[c]	b1813	2: PMID: 21119630
3.1.3.6	u0314	h2o[c] + dtdp[c] -> pi[c] + dtmp[c] + h[c]	b4374	2: PMID: 16990279
3.1.3.-	u0315	h2o[c] + suc6p[c] -> suc[c] + pi[c]	b3812	2: PMID: 16990279
2.6.1.16	u0316	nh4[c] + gmglni[c] -> gln-L[c] + C02126[c] + h[c]	b3729	1
6.3.5.4	u0317	gghy[c] + h2o[c] -> C00192[c] + glu-L[c] + h[c]	b0674	1

6. táblázat: az *E. coli* anyagcseremodellhez hozzáadott új metabolitok, melyek csak a látens reakciókban jelennek meg

A táblázat feltünteti a metabolitok nevét, modellen belül használt azonosítóját, illetve azonosítóját a KEGG, ChEBI és PubChem adatbázisokban (Kanehisa and Goto, 1999; Hastings et al., 2013; Kim et al., 2016).

Metabolit azonosító	Metabolit	KEGG	ChEBI	PubChem
C02669	D-Galactono-1,5-lactone	C02669	15945	439781
C02266	D-Xylono-1,5-lactone	C02266	15867	439692
C02565	N-Methylhydantoin	C02565	16354	69217
C02115	2-Methylserine	C02115	17799	94309
C00872	Aminomalonnate	C00872	17475	100714
C12455	5-Aminopentanal	C12455	x	443849

Metabolit azonosító	Metabolit	KEGG	ChEBI	PubChem
6dg	D-Gal-alpha-1->6D-Glucose (or allolactose)	x	36229	x
C02578	Phenylgalactoside	C02578	x	102336
phenol	phenol	C00146	15882	996
C01300	2-amino-4-hydroxy-6-hydroxymethyl-7,8-dihydropteridine	C01300	17083	4518
chexano	cyclohexanone	C00414	17854	3704
C01880	hexano-6-lactone	C01880	17915	4993
bzal	benzaldehyde	C00261	17169	3559
bz	benzoate	C00180	30746	3480
hxal	hexanal	x	x	6184
hpal	heptanal	C14390	x	17395390
hpa	heptanoate	x	32362	93052
34dhpha	3,4-Dihydroxyphenylacetate	C01161	41941	4388
4hphac	4-Hydroxyphenylacetate	C00642	18101	3915
C02630	2-hydroxyglutarate	C02630	17084	5610
udcal	undecanal	x	x	10325220
udca	undecanoate	x	32369	16019974
ptal	pentanal	x	x	10328700
pta	Pentanoate	C00803	31011	4061
asn-D	D-asparagine	C01905	28159	5014
asp-D	D-aspartate	C00402	17364	3692
C03360	4-Nitrophenyl phosphate	C03360	17440	6198
2am3ox4p	2-amino-3-oxo-4-phosphohydroxybutyrate	x	x	x
dDs7p	1-deoxy-D-sedoheptulose 7-phosphate	x	x	x
dfu	1-deoxy-D-fructose	x	x	x
deryt	1-deoxy-L-erythrulose	x	x	x
df6p	1-deoxy-D-fructose 6-phosphate	x	x	x
cm	Chloramphenicol	C00918	17698	4172
C03601	Chloramphenicol 3-acetate	C03601	16730	6392
xylt	Xylitol	C00379	17151	3669
C02895	Xylitol 5-phosphate	C02895	16772	5824
argtrnals	L-arginyl-tRNALys	x	x	x
thrtrnals	L-threonyl-tRNALys	x	x	x
mettrnals	L-methionyl-tRNALys	x	x	x
leutrals	L-leucyl-tRNALys	x	x	x
alatrals	L-alanyl-tRNALys	x	x	x
sertrals	L-seryl-tRNALys	x	x	x
cysrals	L-cysteinyl-tRNALys	x	x	x
glutrals	L-glutamyl-tRNALys	x	x	x
orntrals	ornithyl-tRNALys	x	x	x
Ltrpam	L-Tryptophanamide	C00977	16533	4226
C01771	Crotonic acid	C01771	17217	4904
3hbut-R	(R)-3-Hydroxybutanoate	C01089	10983	4324
3hbycoa	(S)-3-Hydroxybutyryl-CoA	C01144	15453	4375

Metabolit azonosító	Metabolit	KEGG	ChEBI	PubChem
formcoa	Formyl-CoA	C00798	15522	4056
sdhlam	S-Succinylidihydrolipoamide	C01169	17432	4396
6fESEP	(6R)-6-fluoro-5-enolpyruvylshikimate-3-phosphate	x	x	x
6flu	6-fluorochorismate	x	x	x
tartr-D	D-Tartrate	C02107	15672	5189
C03248	Acetylenedicarboxylate	C03248	30781	6117
ctbt	crotonobetaine	x	x	5462194
Largn	L-Arogenate	C00826	17530	4084
5a2opntn	5-amino-2-oxopentanoate	C01110	17572	4342
24diaglu	2,4-Diaminoglutarate	x	x	x
4am2oxb	4-Amino-2-oxobutanoate	x	x	x
drib	D-2-Deoxyribose	C01801	27806	10786
rnam	N-Ribosylnicotinamide	C03150	6038	15927
isoleutnaval	L-isoleucyl-tRNAVal	x	x	x
C15811	[enzyme]-cysteine	C15811	x	47205136
C15812	[enzyme]-S-sulfanylcysteine	C15812	x	47205137
2ahhmd	2-amino-4-hydroxy-6-hydroxymethyl-7,8-dihydropteridine diphosphate	C04807	15998	7369
C02432	L-selenocysteine	C02432	9093	7995
C01529	selenium	C01529	27568	4691
5enpyr3p	5-Enolpyruvylshikimate-3-phosphate	x	x	739104
C14397	1-chloro-2,4-dinitrobenzene	C14397	x	17395397
C11175	S-(2,4-dinitrophenyl)-glutathione	C11175	8927	13357
adip	L-2-aminoadipate	C00956	37023	4207
acahedio	2-acetylaminohexanedioate	x	x	x
apim	DL-2-aminopimelate	x	x	679919
acahepdio	2-acetylaminooheptanedioate	x	x	x
propglu	N-propionyl-L-glutamate	x	x	x
C00888	pentanoyl-CoA	C00888	15536	4144
C02123	3-propylmalate	C02123	30850	5202
C00750	spermine	C00750	15746	4012
C07315	Sulfamethoxazole	C07315	9332	9523
disul	dihydropterin-sulfamethoxazole	x	x	x
C07659	Sulfathalidine	C07659	x	9861
disulthi	dihydropterin-sulfathiazole	x	x	x
C07458	Sulfanilamide	C07458	45373	9661
disulan	dihydropterin-sulfanilamide	x	x	x
hom-D	D-homoserine	x	x	675779
pdhom	O-phospho-D-homoserine	x	x	728728
4hynor	4-hydroxynorvaline	x	x	642221
4plnor	4-phospho-L-norvaline	x	x	191351
l2amhy	L-2-amino-5-hydroxyvalerate	x	x	x
l2amphy	L-2-amino-5-phosphovalerate	x	x	191351
npan	N-heptylpantothenamide	x	x	25244595

Metabolit azonosító	Metabolit	KEGG	ChEBI	PubChem
npanp	N-heptylpantothenamide 4-phosphate	x	x	x
C02426	L-Glyceraldehyde	C02426	27975	5454
vigly	DL-vinylglycine	x	x	677340
hcl	HCl	C01327	17883	4538
C00579	Dihydrolipoamide	C00579	17694	3858
C01136	S-Acetyldihydrolipoamide	C01136	16807	4367
12dipalpac	1,2-dipalmitoyl phosphatidic acid	x	x	x
cdpdipalgly	CDP-dipalmitoylglycerol	x	x	x
C00591	N-Acylneuraminate	C00591	16498	3870
C01064	CMP-N-acylneuraminate	C01064	16788	4303
C00413	Streptomycin	C00413	17076	3703
C03462	3"-Adenylylstreptomycin	C03462	29076	6282
C17573	Bluensomycin	C17573	x	96023918
3adeblu	3"-adenylylbluensomycin	x	x	x
C07655	Streptobiosamine	C07655	x	9857
3adebios	3"-adenylylstreptobiosamine	x	x	x
C02078	Spectinomycin	x	x	5164
9adespec	9-adenylylspectinomycin	x	x	x
C01935	Maltodextrin	C01935	x	5040
C06369	2-Deoxy-D-glucose 6-phosphate	C06369	16043	8605
C00247	L-Sorbose	C00247	48649	3546
sor6p	L-Sorbose 6-phosphate	x	x	x
C01113	D-Galactose 6-phosphate	C01113	4141	4345
3metglu	3-O-Methylglucose	x	x	152131
3metpglu	3-O-Methyl-D-glucose 6-phosphate	x	x	x
cystnapro	L-cysteinyI-tRNAPro	x	x	x
alatnapro	L-alanyI-tRNAPro	x	x	x
aceala	N-Acetyl-DL-alanine	x	x	675039
benorn	N-Benzoyl-L-ornithine	x	x	x
aceglu	N-acetyl-L-glutamine	x	x	680315
acelys	N-Acetyl-L-lysine	x	x	608105
acedser	N-acetyl-D-serine	x	x	352294
acelser	N-acetyl-L-serine	x	x	207214
C02710	N-Acetyl-L-leucine	C02710	17786	5673
aceasn	N-acetyl-L-asparagine	x	x	677074
acecys	N-acetyl-L-cysteine	x	28939	627486
acelmet	N-acetyl-L-methionine	x	x	413656
acedmet	N-acetyl-D-methionine	x	x	680113
C01962	Thiocysteine	C01962	28839	5063
C00491	L-cystine	C00491	16283	3774
3acepyrh	reduced 3-acetylpyridine	x	x	x
3acepyr	oxidized 3-acetylpyridine	x	x	x
4nitace	4-Nitrophenyl acetate	x	x	156596
napace	alpha-Naphthyl acetate	x	x	156600

Metabolit azonosító	Metabolit	KEGG	ChEBI	PubChem
C11714	1-Naphthol	C11714	10319	13879
C04145	Solanesyl pyrophosphate	C04145	18144	6830
solhyben	3-solanesyl-4-hydroxybenzoate	x	x	x
farhyben	farnesyl-4-hydroxybenzoate	x	x	x
gerhyben	3-geranyl-4-hydroxybenzoate	x	x	25244694
C00353	Geranylgeranyl diphosphate	C00353	x	3646
gergerhyben	geranylgeranyl-4-hydroxybenzoate	x	x	x
acehexd	N-Acetyl-beta-D-hexosaminides	x	x	x
acehexm	N-acetyl-D-hexosamines	x	x	x
niacedglu	p-nitrophenyl-beta-2-acetamido-2-deoxy-D-glucopyranoside	x	x	x
C00870	4-Nitrophenol	C00870	16836	4127
C01860	Trichodiene	C01860	15861	4975
24dhb	2,4-dihydroxybenzoic acid	x	x	840430
24dhba	(2,4-dihydroxybenzoyl)adenylate	x	x	x
25dhb	2,5-dihydroxybenzoic acid	C00628	17189	3901
25dhba	(2,5-dihydroxybenzoyl)adenylate	x	x	x
2dhb	2-hydroxybenzoic acid	x	16914	338
2dhba	(2-hydroxybenzoyl)adenylate	x	x	x
234dhb	2,3,4-dihydroxybenzoic acid	x	36062	72
234dhba	(2,3,4-dihydroxybenzoyl)adenylate	x	x	x
rbl-D	D-ribulose	C00199	17363	3499
abt	L-arabitol	C00532	18403	3814
abt5p	L-arabitol 5-phosphate	x	x	x
rbt	D-ribitol	x	15963	6912
rbt5p	D-ribitol 5-phosphate	C01068	16246	4307
C00324	ferricyanide	C00324	5020	3618
C01913	ferrocyanide	C01913	5032	5021
C03170	oxidized trypanothione	C03170	35490	6053
C02090	reduced trypanothione	C02090	17842	5175
thiben	Thionitrobenzoic acid	x	x	700368
C00216	D-Arabinose	C00216	17108	3516
C00476	D-Lyxose	C00476	16789	3759
C00296	L-Quinate	C00296	29751	3590
C05720	Reduced 2,6-Dichlorophenolindophenol	C05720	x	8025
2napace	2-Naphthyl acetate	x	x	593056
C11713	2-naphthol	C11713	10432	13878
4nitpro	4-Nitrophenyl propionate	x	x	217004
4nitbut	4-Nitrophenyl butyrate	x	x	218054
C01077	O-Acetyl-L-homoserine	C01077	16288	4315
C02306	Phosphoramidate	C02306	15907	5360
C00586	2-Deoxy-D-glucose	C00586	15866	3865
deglu1p	2-deoxy-D-glucose 1-phosphate	x	x	x
25aripyrpp	2-amino-5-formylamino-6-ribosylamino-4(3H)-pyrimidinone triphosphate	x	x	x

Metabolit azonosító	Metabolit	KEGG	ChEBI	PubChem
C00192	Hydroxylamine	C00192	15429	3492
basphy	beta-Aspartylhydroxamic acid	x	x	x
C00248	Lipoamide	C00248	17460	3547
butsul	Butanesulfonic acid	x	x	589531
pensul	Pentanesulfonic acid	x	x	93575753
C02097	dTDP-galactose	C02097	15848	5181
C03194	(R)-1-Aminopropan-2-ol	C03194	15675	6071
C01845	Propanol	C01845	17824	4963
C01380	Ethylene glycol	C01380	30742	4575
C06441	L-Xylulose 1-phosphate	C06441	28566	8675
C01719	L-Fructose	C01719	28120	4856
C02976	Fructose 1-phosphate	C02976	x	5887
C00764	D-sorbose	C00764	17317	4026
C02888	L-Sorbose 1-phosphate	C02888	38342	5819
C01904	D-arabitol	C01904	18333	5013
C02356	2-Aminobutanoate	C02356	35619	5403
C01933	L-Norleucine	C01933	18347	5038
C00902	2-Oxohexanoate	C00902	17308	4157
C01826	L-Norvaline	C01826	18314	4949
C06255	2-Oxopentanoate	C06255	33033	8494
C00734	Chitosan	C00734	16261	3997
C00102	2,6-Dichloroindophenol	C00102	945	3402
C03771	5-guanidino-2-oxopentanoate	C03771	28116	6525
caroxbut	4-Carbamoyl-2-oxobutanoate	x	x	x
msuloxbut	4-methylsulfanyl-2-oxobutanoate	x	x	x
oxosuc	2-oxosuccinamic acid	C02362	16327	5407
imi2oxpro	3-(1H-imidazol-4-yl)-2-oxopropanoate	x	x	x
his-D	D-Histidine	C06419	27947	8654
phe-D	D-Phenylalanine	C02265	16998	5326
val-D	D-valine	C06417	27477	8652
C00408	Pipecolic acid	C00408	30913	3698
delpicar	DELTA1-piperidine-6-carboxylate	x	x	57391681
2aceeth	2-Acetamidoethanethiol	x	x	x
acesmalcys	N-acetyl-S-malonylcysteamine	x	x	x
acealacys	N-(N-acetyl-beta-alanyl)cysteamine	x	x	x
acealamalcys	N-(N-acetyl-beta-alanyl)-S-malonylcysteamine	x	x	x
6dpsip	6-Deoxy-D-psicose 1-phosphate	x	x	x
rib1p	D-Ribulose 1-phosphate	x	x	x
tag1p-D	L-Tagatose 1-phosphate	x	x	x
psi1p	D-Psicose 1-phosphate	x	x	x
glyoald	glycoaldehyde	x	x	151997
sor1p-D	D-Sorbose-1-phosphate	x	x	151033
dxy1p-L	5-Deoxy-L-xylulose 1-phosphate	x	x	770301

Metabolit azonosító	Metabolit	KEGG	ChEBI	PubChem
ery1p-D	D-Erythrulose 1-phosphate	x	x	x
6dsor1p-D	6-Deoxy-D-sorbose 1-phosphate	x	x	x
etpyr	ethylpyruvate	x	x	155354
ethypro	ethyl-2-hydroxypropanoate	x	x	150406
etace	ethylacetoacetate	x	x	152020
ethybut	ethyl-2-hydroxybutanoate	x	x	521365
metace	methylacetoacetate	x	x	150848
methybut	methyl-3-hydroxybutanoate	x	x	x
etthioace	ethylthioacetoacetate	x	x	93575773
3hybues	3-hydroxythiobutyric acid S-ethyl ester	x	x	x
pnol	Propenol	x	x	x
alein	acrolein	C01471	15368	4646
bnol	butanol	x	x	178520
2h6x24dn	2-hydroxy-6-oxoocta-2,4-dienoate	x	x	x
C06728	3-Ethylcatechol	C06728	16572	8950
2h6xh24dn	2-hydroxy-6-oxohexa-2,4-dienoate	x	x	x
C00090	catechol	C00090	18135	3390
C02923	3-Methylcatechol	C02923	18404	5846
2h6xhp24dn	2-hydroxy-6-oxohepta-2,4-dienoate	x	x	x
C01044	N-Formyl-L-aspartate	C01044	48429	4286
nforaspp	N-formyl-4-phospho-L-aspartate	x	x	x
C01042	N-Acetyl-L-aspartate	C01042	16953	4284
nacraspp	N-acetyl-4-phospho-L-aspartate	x	x	x
nclrasp	N-Chloroacetyl-L-aspartate	x	x	x
nclraspp	N-chloroacetyl-4-phospho-L-aspartate	x	x	x
2shycys	S-2-hydroxyethyl-L-cysteine	x	x	695887
bmcap	benzylmercaptan	x	x	x
sbcys	S-benzyl-L-cysteine	x	x	481453
pmcap	phenylmercaptan	x	x	151083
sphcys	S-phenyl-L-cysteine	x	x	99224350
npmerc	n-propylmercaptan	x	x	150947
sprcys	S-propyl-L-cysteine	x	x	679417
emcap	ethylmercaptan	x	x	x
setcys	S-ethyl-L-cysteine	x	x	85085944
smetcys	S-methyl-L-cysteine	x	x	213085
thgly	Thioglycerol	x	x	150343
C00447	Sedoheptulose 1,7-bisphosphate	C00447	17969	3735
C01796	D-Erythrose	C01796	27904	4923
2ddribp	2-deoxy-D-ribose-5-phosphate	x	57651	99319312
C00791	Creatinine	C00791	16737	4049
C00663	beta-D-Glucose 1-phosphate	C00663	16218	3932
Pb2	Pb2	C06696	27889	8921
C00660	D-Glucose 1,6-bisphosphate	C00660	17680	3929

Metabolit azonosító	Metabolit	KEGG	ChEBI	PubChem
C00665	D-Fructose 2,6-bisphosphate	C00665	28602	3934
C11169	Sulfathiazole	C11169	9337	13351
C07329	Isovaleraldehyde	C07329	16638	9537
ival	Isovaleraldehydepd	x	x	x
vale	Valeraldehyde	x	x	8063
valep	Valeraldehydep	x	x	x
gmgluni	L-gamma-glutamyl-p-nitroanilide	x	x	81732
C02126	4-Nitroaniline	C02126	17064	5205
gghy	glutamate-gamma-hydroxamic acid	x	x	159636
tag	D-tagatose	C00795	4249	4053
C02635	3-Chloro-L-alanine	C02635	17403	5615
C06814	D-Xylose-5-phosphate	C06814	37492	9033