

Tomasz BORUTA, Marcin BIZUKOJC, Stanisław LEDAKOWICZ

e-mail: tomaszboruta85@gmail.com

Katedra Inżynierii Bioprocessowej, Wydział Inżynierii Procesowej i Ochrony Środowiska, Politechnika Łódzka, Łódź

Modelowanie metabolizmu w erze inżynierii biologicznej

Inżynieria metaboliczna systemów

Od czasów wprowadzenia przez *Baileya* pojęcia inżynierii metabolicznej [*Bailey, 1991*] dokonano znacznego postępu w naukach biologicznych. Sekwencjonowanie genomów jest dziś działaniem o charakterze wręcz rutynowym dzięki opracowaniu szybkich i stosunkowo niedrogich metod [*Metzker, 2010*]. Mamy dostęp do olbrzymiej ilości danych generowanych z użyciem wysokowydajnych technologii związanych z analizą X-omów [*Zhang i in., 2010*]. Bardzo popularne stało się, typowe dla nauk inżynierskich, podejście holistyczne (całościowe) do zagadnień biologicznych. Obserwujemy szybki rozwój biologii systemów (*Systems Biology*), której zastosowania sięgają od inżynierii metabolicznej [*Kim i in., 2012*] do medycyny [*Nielsen, 2012*]. Katalizatorem zmian w tym kierunku stało się zaangażowanie ekspertów z dziedzin inżynierii chemicznej i informatyki, którzy zaczęli stosować metody obliczeniowe w celu poznania mechanizmów rządzących złożonymi systemami biologicznymi. Doprowadziło to również do dynamicznego rozwoju biologii syntetycznej (*Synthetic Biology*), która czerpie z wiedzy i umiejętności inżynierów w projektowaniu układów biologicznych [*Marchisio i Stelling, 2011*]. Stosuje się w niej metody znane z inżynierii elektrycznej celem tworzenia zintegrowanych obwodów genetycznych [*Nandagopal i Elowitz, 2011*] oraz programy zainspirowane inżynierią mechaniczną służące do komputerowo wspomaganego projektowania biologicznego (BioCAD) [*Chen i in., 2012*]. Zgodnie z oczekiwaniami nadeszła era szeroko rozumianej inżynierii biologicznej, w której dąży się do wprowadzania pewnych standardów umożliwiających bezproblemową komunikację. Powstają specjalistyczne komputerowe formaty danych, które wciąż ewoluują i są dostosowywane do nowych osiągnięć w modelowaniu komórek, np. *Synthetic Biology Open Language* (SBOL) [*Galdzicki i in., 2011*]. Powstał zbiór standardowych elementów używanych w inżynierii biologicznej do tworzenia nowych układów (*Registry of Standard Biological Parts*), które mogą być w wygodny dla użytkownika sposób użyte w komputerowym wspomaganiu projektowania [*Marchisio, 2012*]. Projektowanie inżynierskie ugruntowało swoją pozycję w biologii i z pewnością będzie się w najbliższych latach dynamicznie rozwijać.

Podejście holistyczne dominuje obecnie również we współczesnej inżynierii metabolizmu, której zadaniem jest wprowadzenie modyfikacji genetycznych w sposób głęboko przemyślany i oparty na analizie systemu biologicznego. Jednocześnie czerpie ona z osiągnięć inżynierii biologicznej, która otwiera drogę do biosyntezy niespotykanych w naturze związków chemicznych. W literaturze funkcjonują obecnie pojęcia *systems metabolic engineering* oraz *synthetic metabolic engineering* [*Becker i Wittmann, 2012*], które ilustrują kierunek rozwoju inżynierii metabolizmu.

Modelowanie metabolizmu oparte na optymalizacji i ograniczeniach

Modelowanie matematyczne odgrywa niezwykle istotną rolę w nowoczesnym podejściu do inżynierii metabolicznej. Pozwala na sformułowanie pewnych strategii eksperymentalnych, które później mogą być zweryfikowane w laboratorium. Wyzwanie polega na zbudowaniu modelu, który pozwoli przewidzieć odpowiedź organizmu na modyfikację genetyczną. Przy ogromnej złożoności systemów biologicznych i braku pełnego zrozumienia zachodzących w nich procesów, modelowanie procesów wewnątrzkomórkowych jest wciąż ogromnym wyzwaniem. Przyjmuje się jednak szereg uproszczeń, które pozwalają podjąć próbę kwantyfikacji metabolizmu, a więc wyznaczenia wartości strumieni reakcji.

Punktem wyjściowym do modelowania są rekonstrukcje metabolizmu na skalę genomu, które obejmują wszystkie reakcje metaboliczne obecne w danym organizmie [*Thiele i Palsson, 2010*]. Rekonstruowanie sieci odbywa się na podstawie sekwencji genomu, który następnie poddany jest elektronicznej anotacji. Anotacja obejmuje wyszukiwanie przewidywanych genów oraz powiązanie ich z konkretną funkcją na podstawie podobieństwa do znanych i scharakteryzowanych eksperymentalnie sekwencji. Obecność genu kodującego dany enzym można interpretować jako obecność funkcjonalnego enzymu i odpowiadającej mu reakcji metabolicznej. Rekonstrukcje na podstawie automatycznej anotacji są jednak niekompletne i niedostosowane do specyficznych cech danego organizmu. Konieczna jest więc czasochłonna procedura sprawdzania i wprowadzania poprawek do sieci metabolicznej mającej zapewnić zgodność z danymi literaturowymi. Kwestią niezwykle istotną jest zapewnienie poprawnej stechiometrii reakcji, użycie właściwych kofaktorów oraz uwzględnienie obserwowanej w warunkach fizjologicznych odwracalności reakcji lub jej braku.

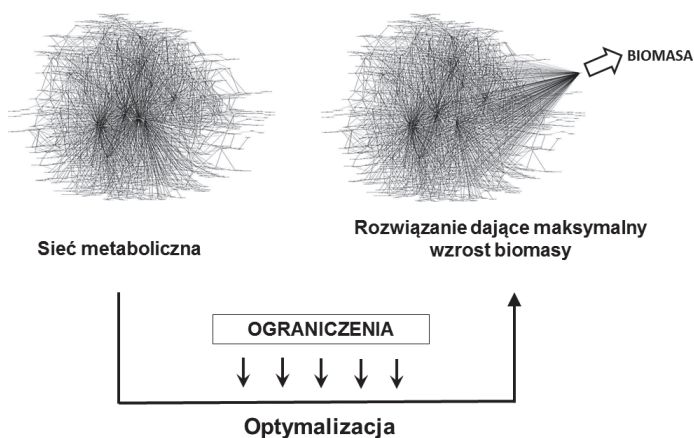
Mimo znacznych postępów w dziedzinie rekonstrukcji na skalę genomu, nie jest to wciąż czynność rutynowa. Brak jest uniwersalnych narzędzi do tworzenia wysokiej jakości sieci metabolicznych, dostosowanych do konkretnego przypadku. W przypadku organizmów o największym znaczeniu dla nauki i przemysłu, modele są w sposób ciągły poprawiane i rozbudowywane dzięki wspólnemu wysiłkowi badaczy z całego świata. Jako przykład można podać tutaj bakterię *Escherichia coli*, której metabolizm w skali genomu został po raz pierwszy opisany w 2000 roku [*Edwards i Palsson, 2000*]. Od tego czasu opublikowano kilka różnych wersji modelu, z których najnowsza zawiera 2251 reakcji i 1136 metabolitów [*Orth i in., 2011*].

Mając do dyspozycji listę reakcji metabolicznych można przystąpić do konstrukcji modelu matematycznego. Należy wyznaczyć granicę systemu (odpowiadającą ścianie komórkowej), zdefiniować strumienie wejściowe (odpowiadające substancjom odżywczym obecnym w pożywce konsumowanej przez rosnące i dzielące się komórki) oraz wybrać strumienie wyjściowe (związki wydzielane przez komórki do cieczy hodowlanej). Jak wyznaczyć wektor wartości strumieni? Podejście dominujące we współczesnej inżynierii metabolizmu oparte jest na metodach optymalizacyjnych, w których kluczową rolę odgrywają zastosowane ograniczenia [*Park i in., 2009*].

Przyjmuje się najczęściej stan ustalony, w którym nie następuje akumulacja wewnątrzkomórkowych metabolitów. Dla każdego z nich powstaje równanie bilansu, w którym dopływ obejmujący reakcje produkujące dany związek równoważony jest przez odpływ, a więc reakcje zawierające ten związek w charakterze substratu. W efekcie powstaje układ równań liniowych, który w większości przypadków jest niedookreślony (liczba niewiadomych, czyli strumieni, jest większa od liczby równań odpowiadającej liczbie metabolitów) i posiada nieskończenie wiele rozwiązań. Wektor strumieni wyznaczyć można na drodze optymalizacji w przestrzeni wielowymiarowej. Nałożone ograniczenia, służące zredukowaniu przestrzeni możliwych rozwiązań, obejmują wartości strumieni uzyskane na drodze eksperymentalnej. Bierze się również pod uwagę jednokierunkowy przebieg niektórych reakcji w warunkach fizjologicznych, dane eksperymentalne z analizy X-omów [*Jensen i Papin, 2011*], prawa termodynamiki [*Hoppe i in., 2007*] oraz regulację genetyczną [*Shlomi i in., 2007*].

Biologiczną funkcją celu jest najczęściej równanie tworzenia biomasy komórkowej definiowane przez autorów modelu na podstawie danych eksperymentalnych. W podejściu tym zakłada się, że komórki zawsze dążą do wzrostu, co zostało ukształtowane w toku ewolucji na drodze naturalnej selekcji. Oznacza to, że strumienie przyjmują wartości, które

zagwarantują komórce maksymalną produkcję biomasy. Opisana metoda, znana jako analiza bilansu strumieni FBA (*Flux Balance Analysis*) [Orth i in., 2010], pozwala wyznaczyć wektor wartości strumieni zarówno wśród szczepów naturalnych, jak i poddanych delekcji (usunięciu) określonych genów związanych z metabolizmem (Rys. 1).



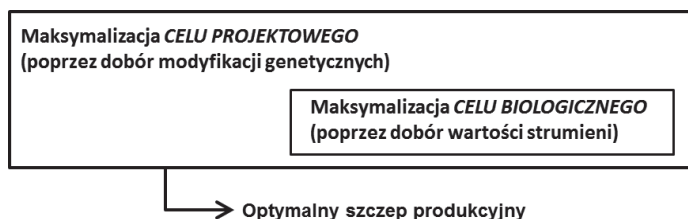
Rys. 1. Schemat ilustrujący wyznaczenie wartości strumieni metabolicznych zgodnie z metodą analizy bilansu strumieni (FBA). Maksymalizacja biologicznej funkcji celu (najczęściej równania tworzenia biomasy) prowadzi do wyznaczenia wektora wartości strumieni metabolicznych

Innym podejściem do obliczania strumieni jest MOMA (*Minimization of Metabolic Adjustment*) [Segre i in., 2002], zgodnie z którym szczepy modyfikowane genetycznie mogą zachowywać się inaczej niż szczepy dzikie, u których mechanizm maksymalizacji wzrostu powstawał na drodze ewolucji i selekcji. Według MOMA, wartości strumieni w komórce poddanej inżynierii genetycznej będą różnić się minimalnie w stosunku do wartości dla szczepu dzikiego (przed modyfikacją).

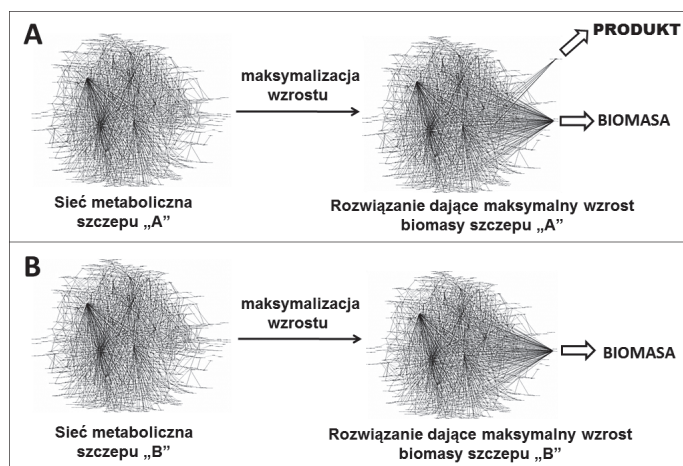
Inżynieria metaboliczna *in silico*

Początki inżynierii metabolicznej związane były z analizą pojedynczych szlaków i niezwykle prostymi modelami matematycznymi. Rekonstrukcja sieci biochemicznych na skalę genomu wyznaczyła nowy kierunek rozwoju dyscypliny, który opierał się na integracji metod komputerowych i laboratoryjnych. Pojawiło się zapotrzebowanie na narzędzia, które stanowiłyby wsparcie dla badaczy w planowaniu eksperymentów poprzez sugerowanie odpowiednich, często nieintuicyjnych, modyfikacji genetycznych prowadzących do celu projektowego. Obecnie tego typu narzędzia są często stosowane do wspomagania badań związanych z biosyntezą przez mikroorganizmy witamin, aminokwasów, biopaliw i innych cennych dla biotechnologii produktów [Park i in., 2009]. Pierwszym algorytmem optymalizacyjnym w inżynierii metabolicznej był *OptKnock* [Burgard i in., 2003], który stanowił wzór dla twórców narzędzi tego typu.

OptKnock pozwala wyznaczyć reakcje, których usunięcie z sieci doprowadzi do zwiększenia produkcji danego metabolitu (czyli do wzrostu wartości strumienia odpowiedzialnego za biosyntezę produktu). Oznacza to w praktyce usunięcie (delekcję) genu lub kilku genów, w których zapisana jest sekwencja enzymów odpowiedzialnych za daną reakcję metaboliczną. Algorytm oparty jest na dwóch poziomach optymalizacji (Rys. 2).



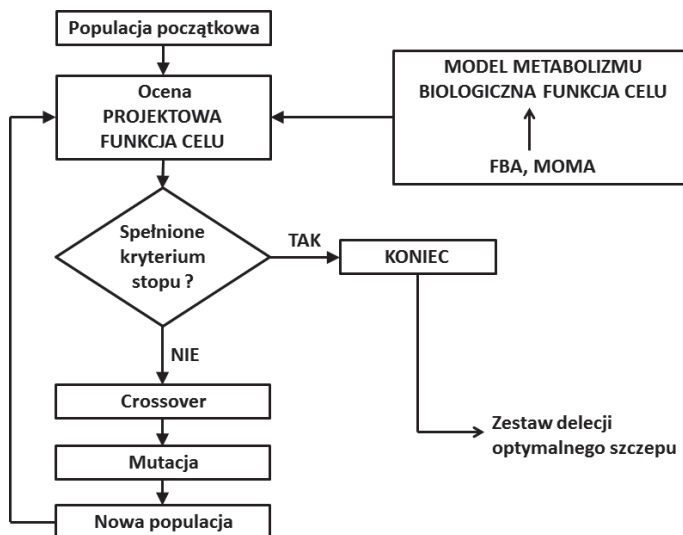
Rys. 2. Schemat działania algorytmu *OptKnock* [Burgard i in., 2003]. Każdy szczep maksymalizowany jest pod kątem biologicznej funkcji celu (najczęściej wzrostu). Wybrany jest szczep, który wykazuje przy tym maksymalną wartość projektowej funkcji celu (najczęściej strumienia biosyntezy produktu)



Rys. 3. Schemat ilustrujący zasadę wyboru najlepszych modyfikowanych genetycznie szczepów produkcyjnych *in silico*. A) Szczep „A” maksymalizuje wzrost biomasy i zachodzi przy tym biosynteza produktu. B) W przypadku szczepu „B” tworzenie produktu nie idzie w parze z maksymalnym wzrostem biomasy

OptKnock zakłada optymalizację pod kątem biologicznej funkcji celu (strumienia tworzenia biomasy), a następnie wybór mutantu o najwyższej wartości projektowej funkcji celu (strumienia tworzenia produktu). Wybrane jest takie rozwiązanie, przy którym biosynteza danego związku jest produktem ubocznym maksymalizacji wzrostu (Rys. 3).

OptKnock analizuje wszystkie możliwe kombinacje delekcji, co przy liczbie reakcji przekraczającej zwykle dwa tysiące wymaga długiego czasu obliczeniowego. Podjęto próbę rozwiązania tego problemu poprzez zastosowanie algorytmu ewolucyjnego *OptGene* [Patil i in., 2005], w którym najlepsze osobniki w populacji wymieniają się informacją genetyczną (w tym przypadku informacją dotyczącą obecności danej reakcji), dając początek nowym pokoleniom (Rys. 4).



Rys. 4. Schemat działania algorytmu *OptGene* [Patil i in., 2005]. Każdy szczep optymalizuje swój biologiczny cel (np. wzrost), po czym następuje ocena jego przydatności pod kątem tworzenia produktu, a więc celu projektowego. Szczepy dające najlepsze wyniki są kojarzone ze sobą na zasadzie *crossover*, czyli wzajemnej wymiany materiału genetycznego. Później wprowadzane są w sposób pseudolosowy mutacje i powstaje kolejne pokolenie szczepów *in silico*.

OptGene nie gwarantuje osiągnięcia globalnego maksimum funkcji celu, ale znacznie skraca czas obliczeniowy, co ma bardzo duże znaczenie przy większej liczbie delekcji i znacznie wzrastającej w takim przypadku liczbie kombinacji. Kolejną zmianą w stosunku do *OptKnock* jest przypisanie reakcji metabolicznych do genów. W niektórych bowiem przypadkach wyeliminowanie jednej reakcji wymaga w praktyce usunięcia w laboratorium kilku genów kodujących izoenzymy, czyli różne formy danego enzymu katalizujące tę samą reakcję.

Kolejnym godnym uwagi algorytmem optymalizacyjnym jest *OptStrain* [Pharkya i in., 2004]. Obejmuje on nie tylko analizę dotyczącą usuwania reakcji, ale również ich dodawania. *OptStrain* oblicza maksymalną możliwą wydajność biosyntezy produktu i wybiera szlak, który prowadzi do tworzenia produktu przy maksymalnej wydajności. Są wśród tych reakcji również takie, których nie ma w naturalnym szczepie. *OptStrain* minimalizuje liczbę obcych reakcji, a następnie szuka optymalnego zestawu delecji używając metody znanej z *OptKnock*.

Aktywacja i inhibicja reakcji została wzięta pod uwagę przez twórców algorytmu *OptReg* [Pharkya i Maranas, 2006]. Wpływ inhibicji bada się w tym przypadku poprzez ustalenie wartości strumienia dla danej reakcji znacznie poniżej poziomu ze stanu ustalonego, natomiast aktywacja wiąże się z wartością strumienia wyższą niż ta obliczona dla stanu ustalonego. Można dzięki temu ustalić nie tylko zestaw delecji, ale również zidentyfikować kluczowe punkty na mapie metabolicznej, w których warto by wzmocnić strumień lub go zredukować poprzez regulację na poziomie genetycznym.

Kierunek rozwoju inżynierii metabolicznej

Inżynieria metaboliczna osiągnęła poziom nauki o systemach i podchodzi do modyfikowanych organizmów w sposób charakterystyczny dla nauk inżynierskich, a więc holistycznie. Jej rozwój idzie w parze z rozkwitem biologii syntetycznej, która umożliwia modyfikowanie funkcji enzymów oraz projektowanie nowych szlaków biosyntezy produktów istotnych dla przemysłu i nauki wykorzystując podejście typowo inżynierskie [Keasling, 2010]. Na horyzoncie pojawia się idea projektowania i tworzenia zupełnie nowych syntetycznych szczepów zoptymalizowanych pod kątem biotechnologii przemysłowej. Jak daleko jesteśmy od zaprojektowania organizmu *de novo*? Opublikowano już przypadek stworzenia syntetycznego bakteryjnego genomu [Gibson i in., 2010], który nie został odrzucony przez komórki i przejął w nich funkcje naturalnego genomu. Sekwencja DNA została opracowana komputerowo, syntetyczny genom był jednak w większości oparty na sekwencji naturalnego szczepu. Nie można zatem mówić o stworzeniu nowej formy życia, a jedynie o przygotowaniu kopii naturalnego zestawu genów. Chociaż postęp w tej dziedzinie jest bez wątpienia ogromny, do zaprojektowania organizmu *in silico* konieczne jest pełne zrozumienie mechanizmów obecnych w komórkach idące w parze z rozwojem metod modelowania skomplikowanych biologicznych systemów. Wiedza i zrozumienie niekoniecznie idą jednak w parze z olbrzymią ilością generowanych danych powstałych w wyniku rozwoju technik eksperymentalnych. Brak jeszcze umiejętności zaprogramowania kodu DNA, który przyniósłby z góry założony zbiór cech organizmu. Mając również na uwadze brak akceptacji społecznej wobec tego typu praktyk, należy nadal ostrożnie podchodzić do idei konstruowania *de novo* syntetycznych szczepów produkcyjnych. Głównym przedmiotem badań inżynierii metabolicznej pozostaje zatem ukierunkowane optymalizowanie metabolizmu komórek, a nie projektowanie kompletnie nowych sieci biochemicznych. Można się jednak spodziewać, że postępy w dziedzinie biologii syntetycznej oraz modelowania sieci metabolicznych będą miały ogromny wpływ na kierunek rozwoju inżynierii metabolicznej.

LITERATURA

Bailey J.E., 1991. Toward a science of metabolic engineering. *Science*, 252, nr 5013, 1668-1675. DOI:10.1126/science.2047876

Becker J., Wittmann C., 2012. Systems and synthetic metabolic engineering for amino acid production – the heartbeat of industrial strain development. *Current Opinion in Biotechnology*, 23, 1-9. DOI: 10.1016/j.copbio.2011.12.025

Burgard A.P., Pharkya P., Maranas C.D., 2003. OptKnock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnology and Bioengineering*, 84, nr 6, 647-657. DOI: 10.1002/bit.10803

Chen J., Densmore D., Ham T.S., Keasling J.D., Hillson N.J., 2012. Device Editor visual biological CAD canvas. *Journal of Biological Engineering*, 6, nr 1. DOI: 10.1186/PREACCEPT-1967068768635731

Edwards J.S., Palsson B.O., 2000. The *Escherichia coli* MG1655 *in silico* metabolic genotype: its definition, characteristics, and capabilities. *Proceedings of the National Academy of Sciences of the United States of America*, 97, nr 10, 5528-5533. DOI: 10.1073/pnas.97.10.5528

Galdzicki M., Rodriguez C., Chandran D., Sauro H.M., Gennari J.H., 2011. Standard Biological Parts Knowledgebase. *PLoS ONE*, 6, nr 2, e17005. DOI: 10.1371/journal.pone.0017005

Gibson D.G., Glass J.I., Lartigue C., Noskov V.N., Chuang R.Y., Algire M.A., Benders G.A., Montague M.G., Ma L., Moodie M.M., Merryman C., Vashee S., Krishnakumar R., Assad-Garcia N., Andrews-Pfannkoch C., Denisova E.A., Young L., Qi Z.Q., Segall-Shapiro T.H., Calvey C.H., Parmar P.P., Hutchison C.A. 3rd, Smith H.O., Venter J.C., 2010. Creation of a bacterial cell controlled by a chemically synthesized genome. *Science*, 329, nr 5987, 52-56. DOI: 10.1126/science.1190719

Hoppe A., Hoffmann S., Holzhutter H.G., 2007. Including metabolite concentrations into flux balance analysis: thermodynamic realizability as a constraint on flux distributions in metabolic networks. *BMC Systems Biology*, 1, nr 23. DOI:10.1186/1752-0509-1-23

Jensen P.A., Papin J.A., 2011. Functional integration of a metabolic network model and expression data without arbitrary thresholding. *Bioinformatics*, 27, nr 4, 541-547. DOI: 10.1093/bioinformatics/btq702

Keasling J.D., 2010. Manufacturing molecules through metabolic engineering. *Science*, 330, nr 6009, 1355-1358. DOI: 10.1126/science.1193990

Kim I.K., Roldao A., Siewers V., Nielsen J., 2012. A systems-level approach for metabolic engineering of yeast cell factories. *FEMS Yeast Research*, 12, nr 2, 228-248. DOI: 10.1111/j.1567-1364.2011.00779.x

Marchisio M.A., Stelling J., 2011. Automatic design of digital synthetic gene circuits. *PLoS Computational Biology*, 7, nr 2, e1001083. DOI: 10.1371/journal.pcbi.1001083

Marchisio M.A., 2012. *In silico* implementation of synthetic gene networks. *Methods in Molecular Biology*, 813, nr 1, 3-21. DOI: 10.1007/978-1-61779-412-4_1

Metzker M.L., 2010. Sequencing technologies – the next generation. *Nature Reviews Genetics*, 11, nr 1, 31-46. DOI:10.1038/nrg2626

Nandagopal N., Elowitz M.B., 2011. Synthetic biology: integrated gene circuits. *Science*, 333, nr 6047, 1244-1248. DOI: 10.1126/science.1207084

Nielsen J., 2012. Translational and systems medicine. *Journal of Internal Medicine*, 271, nr, 108-110. DOI: 10.1111/j.1365-2796.2011.02490.x

Orth J.D., Conrad T.M., Na J., Lerman J.A., Nam H., Feist A.M., Palsson B.O., 2011. A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism–2011. *Molecular Systems Biology*, 7, nr 535. DOI:10.1038/msb.2011.65

Orth J.D., Thiele I., Palsson B.O., 2010. What is flux balance analysis? *Nature Biotechnology*, 28, nr 3, 245-248. DOI:10.1038/nbt.1614

Park J.M., Kim T.Y., Lee S.Y., 2009. Constraints-based genome-scale metabolic simulation for systems metabolic engineering. *Biotechnology Advances*, 27, nr 6, 979-988. DOI: 10.1016/j.biotechadv.2009.05.019

Patil K.R., Rocha I., Forster J., Nielsen J., 2005. Evolutionary programming as a platform for *in silico* metabolic engineering. *BMC Bioinformatics*, 6, nr 308. DOI:10.1186/1471-2105-6-308

Pharkya P., Burgard A.P., Maranas C.D., 2004. OptStrain: a computational framework for redesign of microbial production systems. *Genome Research*, 14, nr 11, 2367-2376. DOI: 10.1101/gr.2872004

Pharkya P., Maranas C.D., 2006. An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems. *Metabolic Engineering*, 8, nr 1, 1-13. DOI: 10.1016/j.mbs.2005.08.003

Segre D., Vitkup D., Church G.M., 2002. Analysis of optimality in natural and perturbed metabolic networks. *Proceedings of the National Academy of Sciences of the United States of America*, 99, nr 23, 15112-15117. DOI: 10.1073/pnas.232349399

Shlomi T., Eisenberg Y., Sharan R., Ruppin E., 2007. A genome-scale computational study of the interplay between transcriptional regulation and metabolism. *Molecular Systems Biology*, 3, nr 101. DOI: 10.1038/msb410014

Thiele I., Palsson B.O., 2010. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature Protocols*, 5, nr 1, 93-121. DOI:10.1038/nprot.2009.203

Zhang W., Li F., Nie L., 2010. Integrating multiple 'omics' analysis for microbial biology: application and methodologies. *Microbiology*, 156, nr 2, 287-301. DOI: 10.1099/mic.0.034793-0