

# Integracija modelov metabolizma in genskega regulatornega omrežja

## Pregled področja

Jernej Puc

IŠRM – UNI-LJ

Marec 2019

# Motivacija po integraciji bioloških modelov

Metabolni, regulatorni, signalni idr. modeli lahko izpostavljajo različne vidike in lastnosti istega večjega sistema.

Čeprav so posamezno pomembno pripomogli k razumevanju bioloških sistemov, pogosto niso zadostni pri opisovanju pripadajočih celičnih procesov. Od tu izvira motivacija po njihovem združevanju oz. konstrukciji bolj holističnega modela, ki bi predstavljal boljši približek dejanski biološki sliki.

V okviru takih integracijskih modelov se npr. povzema vpliv regulatornih mehanizmov na metabolizem in obratno.

# Pristopi v literaturi

Članki v tej domeni so običajno dokaj skopi in nejasni glede metodologije – namesto splošne uporabnosti je v ospredju besedna interpretacija rezultatov specifičnih poskusov.

Večji del pristopov, izpostavljenih v nadaljevanju, je bil že povzet v delu [Imam et al., 2015](#):

- ▶ rFBA (regulatory FBA), [Covert et al., 2002](#)
- ▶ SR-FBA (steady-state rFBA), [Shlomi et al., 2007](#)
- ▶ iFBA (integrated FBA), [Covert et al., 2008](#)
- ▶ PROM (probabilistic regulation of metabolism), [Chandrasekaran et al., 2010](#)
- ▶ Whole-cell model, [Karr et al, 2012](#)
- ▶ ME (metabolic and expression) model, [O'Brien et al., 2013](#)
- ▶ Cost-benefit model, [Carrera et al., 2014](#)
- ▶ TPM (time-point model) segmentation, [Sarkar et al., 2019](#)

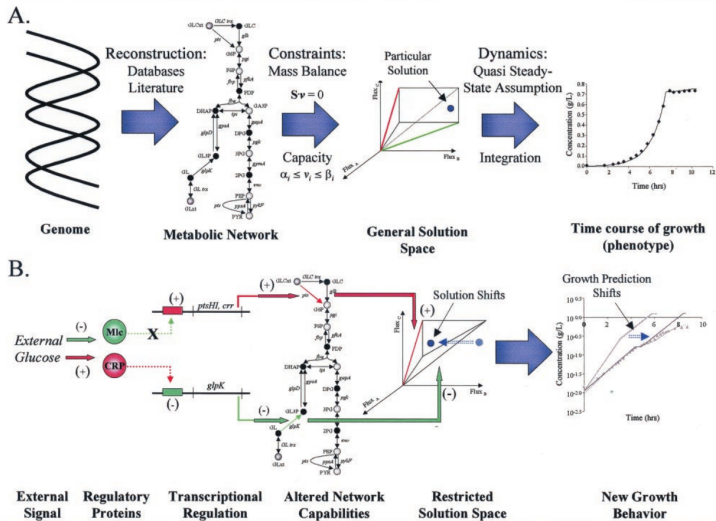
# rFBA (Covert et al., 2002)

- ▶ Uporablja logična pravila za aproksimacijo transkripcijskega nadzora nad metabolnim omrežjem:
  - Dovoljeni samo dve stanji (ON/OFF) za omrežne reakcije.
  - Stanje reakcij je določeno z (ne)izraženostjo genov v danih okoliščinah:
    - Regulatorna logika je pridobljena na osnovi literature.
    - Regulatorna logika lahko opisuje zunanje in notranje celične pogoje ali aktivnost.
  - Stanje reakcij določa, ali je pretok za reakcijo v FBA neomejen ali omejen na 0.
- ▶ Predpostavljeno je, da so metabolni procesi dosti hitrejši od regulatornih:
  - Tako so stanja kvazi-ravnovesna in se lahko izvaja dinamične simulacije v diskretnih časovnih korakih.

# rFBA – postopek

1. Začenši s  $t = 0$  se izvaja simulacija v majhnih časovnih korakih ( $\Delta t = 3s$  za primer v članku).
2. Na vsakem koraku se ovrednotijo trenutne okoljske in notranje razmere celice, da se določi smer regulacije za vsak reguliran gen:
  - V primeru regulacije navzgor, tj. količina komponente se poveča, je reakcija v metabolnem sistemu neomejena (s strani genske regulacije).
  - V primeru regulacije navzdol, tj. količina komponente se zmanjša, je reakcija v metabolnem sistemu omejena na 0.
  - \* Omejitve se upoštevajo po ocenjenem časovnem zamiku (v prispevku pomembno zaradi numerične integracije).
3. Ko so regulatorne omejitve za določen korak vnešene v sistem, se izračuna optimalna porazdelitev pretokov.
4. Porazdelitev pretokov v vsakem časovnem koraku se uporabi za določitev poteka obravnavane količine (npr. celične rasti).

# rFBA – primer



**Slika:** Transkripcijska regulacija omeji prostor rešitev za dano stanje.

## rFBA – primer

*For example, external glucose is sensed by various regulatory proteins in the cell, among them CRP, which is activated (red arrow), and Mlc, which is inactivated (green arrow) by the glucose signal. As a result, transcription of glucose ABC transporter operon *ptsHI-crr* is not repressed by Mlc, whereas transcription of a glycerol kinase gene, *glpK*, is repressed by CRP.*

*The presence or absence of expression of these genes leads to the availability or unavailability of the respective reactions or transport processes in the metabolic network and possibly to the removal of available extreme pathway basis vectors from the steady-state solution space.*

*The result is a time-dependent solution space defined by temporary regulatory constraints in addition to the invariant constraints mentioned previously, which may exhibit a new behavior if the previous solution is no longer in the space.*

# rFBA – regulatorna logika

Reaction	Regulatory Logic	Refs.
ICIT -> GLX + SUCC ACCOA + GLX -> COA + MAL	IF not (IclR) IF not (ArcA or IclR)	1,2,31 1,2,23,31
PYR + COA + NAD -> NADH + CO2 + ACCOA	IF (not PdhR)	1,3
ACTP + ADP <=> ATP + AC		
CIT <=> ICIT	IF ( <i>GLCxt or LCTSxt or RIBxt or GLxt or LACxt or PYRxt or SUCCxt or ETHxt or ACxt or FORxt</i> )	4
CIT <=> ICIT	IF ( <i>GLCxt or LCTSxt or RIBxt or GLxt or LACxt or PYRxt or SUCCxt or ETHxt or ACxt or FORxt</i> )	4
ATP + AC + COA -> AMP + PPI + ACCOA	IF not ( <i>GLCxt or LCTSxt or RIBxt or GLxt or LACxt or PYRxt or SUCxt or ETHxt</i> ) and (not IdR)	5
ACCOA + 2 NADH <=> ETH + 2 NAD + COA	IF not (O2xt) or not (O2xt and Cra)	6,7
ATP + AMP <=> 2 ADP ATP <=> ADP + PI + 4 HEXT		
QH2 + .5 O2 -> Q + 2 HEXT	IF (not FNR) or ArcA	25
QH2 + .5 O2 -> Q + 2.5 HEXT	IF not (ArcA or FNR)	25
PYR + NADH <=> NAD + LAC LAC + Q -> PYR + QH2 2PG <=> PEP FDP <=> T3P1 + T3P2 FDP -> F6P + PI		

**Slika:** Ozadje za pridobitvijo regulatorne logike.



# SR-FBA (Shlomi et al., 2007)

- ▶ Rešuje arbitrarnost rešitev rFBA.
- ▶ Vpeljava metabolno-regulatornih ravnovesnih stanj (MRS).
  - Stanje sistema v danem konstantnem okolju, ki ga opisujejo pretoki in logične vrednosti regulatorne logike glede na pripadajoče omejitve.
- ▶ MRS v danem koraku je najden z reševanjem mešanega celoštevilskega linearne programa (MILP).
  - Formuliran glede na regulatorno logiko.

# SR-FBA – metabolna regulacija

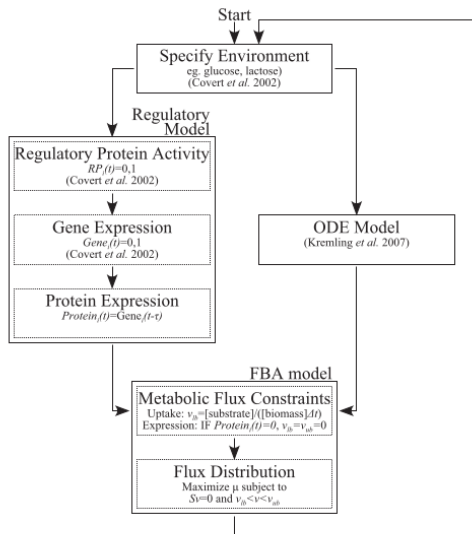
- ▶ Verjetno del že same rFBA, a v Covert et al. (2002) ni bila omenjena.
- ▶ Predstavljena z aktivacijo ali represijo genov glede na vrednosti pretokov.
  - Oblika logičnih predikatov.
- ▶ Povraten vpliv metabolnega omrežja na gensko regulatorno.

Omejitve mešanega linearnega programa določajo:

- ▶ meje fluksa v metabolnem omrežju,
- ▶ regulatorne omejitve:
  - omejitve, ki izhajajo iz genske regulatorne logike,
  - omejitve, ki izhajajo iz metabolne regulatorne logike,
- ▶ preslikave od genov do reakcij:
  - pogojevanje razpoložljivosti proteinov z izraženostjo genov,
  - pogojevanje aktivnosti reakcij s prisotnostjo proteinov,
- ▶ pogojevanje fluksa z aktivnostjo pripadajočih reakcij.

## iFBA (Covert et al., 2008)

- ▶ Združuje FBA metabolnega omrežja, logični opis transkripcijskega regulatornega omrežja in kinetični (ODE) model.
- ▶ rFBA in ODE modela združujejo skupni metaboliti in pretoki.
- ▶ Izbrana dolžina koraka  $\Delta t$  mora biti dovolj velika, da je lahko privzeto ravnovesje metabolnega stanja, in dovolj majhna, da je numerična integracija ODE modela dovolj točna.
  - Rečeno je, da za vrednosti intervala  $\Delta t \in [30s, 5min]$  ni bilo opazne razlike v rezultatih.



**Fig. 2.** Schematic of the simulation algorithm.

Omejitve linearnega programa določajo:

- ▶ okoljske omejitve izmenjevalnih pretokov  $v_{ex}$  (1) in transportne omejitve (2):
  1. *Where the maximum flux through an exchange reaction is limited by the amount of substrate in the culture medium.*
  2. *Represented as a maximum substrate uptake or by-product secretion rate.*
  - \* Okoljske omejitve so primerjane s transportnimi – za omejitev izmenjavelnega pretoka v danem koraku se uporabi najbolj omejujoče vrednosti.
- ▶ regulatorne omejitve (izpeljane iz predikatov),
- ▶ ujemanje med modeloma:
  - Spodnjo in zgornjo mejo skupnih pretokov posodablja ODE model.

# iFBA – postopek

1. Začenši pri začetnih pogojih (literatura) ali pogojih predhodnega koraka se ODE model numerično integrira.
2. Z izhodom ODE modela in logičnim regulatornim modelom se posodobi omejitve za FBA.
3. Po optimizaciji s FBA se z izračunanimi pretoki posodobijo biomasa in koncentracije zunanjih metabolitov:

► Po literaturi:

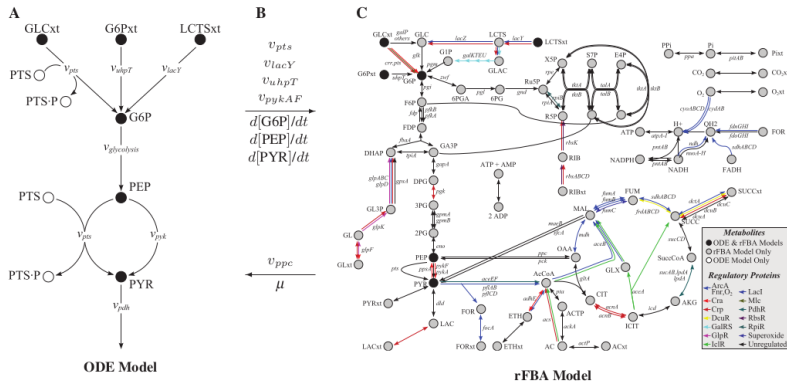
$$[\text{biomass}](t + \Delta t) = \beta [\text{biomass}](t) e^{\mu \Delta t}$$

$$[\text{met}_i](t + \Delta t) = [\text{met}_i] + \frac{v_{\text{ex}}}{\mu} [\text{biomass}](t) (1 - e^{\mu \Delta t})$$

\*  $\beta$  – faktor za ujemanje z eksperimentalnimi vrednostmi

\*  $\mu$  – pretok “reakcije” rasti

4. Biomasa in konc. zun. metabolitov se nato uporabijo pri num. integraciji ODE modela v naslednjem koraku.



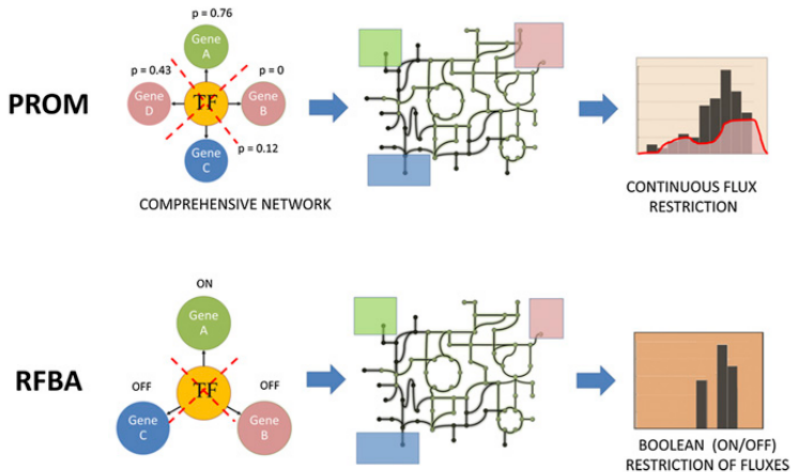
**Fig. 1.** Diagram of the metabolic, regulatory and signaling networks used to build the iFBA model. (A) A schematic of the complete ODE model (Kremling *et al.*, 2007). (B) A list of the variables passed between the models as part of iFBA. Arrows indicate whether a value is being passed from the ODE model to rFBA or vice versa, and metabolites common to both the rFBA and ODE models are colored black. (C) A schematic of the complete rFBA model (Covert and Palsson, 2002). Regulated fluxes are indicated by the key (lower right).



# PROM (Chandrasekaran et al., 2010)

- ▶ Regulacija namesto Boolove logike uporablja verjetnostne uteži.
  - Verjetnosti so uporabljene pri omejitvah pretokov za posamezne reakcije.
- ▶ Uporaba verjetnosti omogoča večji razpon potencialnih odzivov (v smislu aktivacije in represije).

# PROM – primerjava z rFBA



**Slika:** Uporaba verjetnosti omogoča večji razpon potencialnih odzivov.

# PROM – določitev verjetnosti interakcije gen-TF

Potrebni so podatki o izraženosti genov glede na transkripcijske faktorje (TF).

Za vsak gen-TF par:

1. Izraženost gena je binarizirana z nizkim pragom.
  - ▶ Pod pragom  $\rightarrow 0$  (OFF), nad pragom  $\rightarrow 1$  (ON).
  - ▶ Prag je fiksni in konst. za vse gene.
2. Verjetnost je določena s številom vzorcev v naboru podatkov, pri katerih je gen aktiviran, ko TF ni aktiviran, npr. za gen  $A$  in TF  $B$ :

$$\Pr(A = 1|B = 0) = \frac{N(A = 1|B = 0)}{N(B = 0)}$$

- ▶ Če je nabor podatkov dovolj velik, se lahko pridobi robustno oceno verjetnosti interakcije.
- ▶ TF so odvisni od okoljskih razmer.

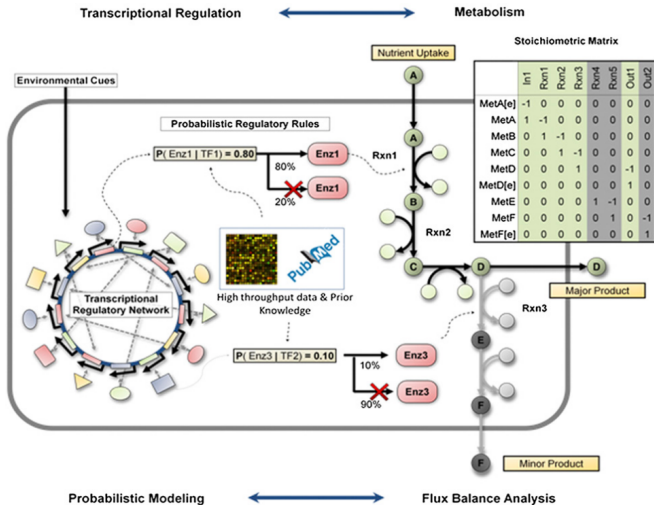
# PROM – dodatne omejitve LP

Če je  $\Pr(A = 1|B = 0) = p$ , je lahko vrednost pretoka reakcije, ki jo regulira gen  $A$  kvečjemu  $p \cdot V_A^{\max}$ .

- ▶  $V_A^{\max}$  je ocenjen s FVA (flux variability analysis) na nereguliranem metabolnem modelu.

Ko so vse omejitve določene, se LP rešuje kot običajno.

# PROM

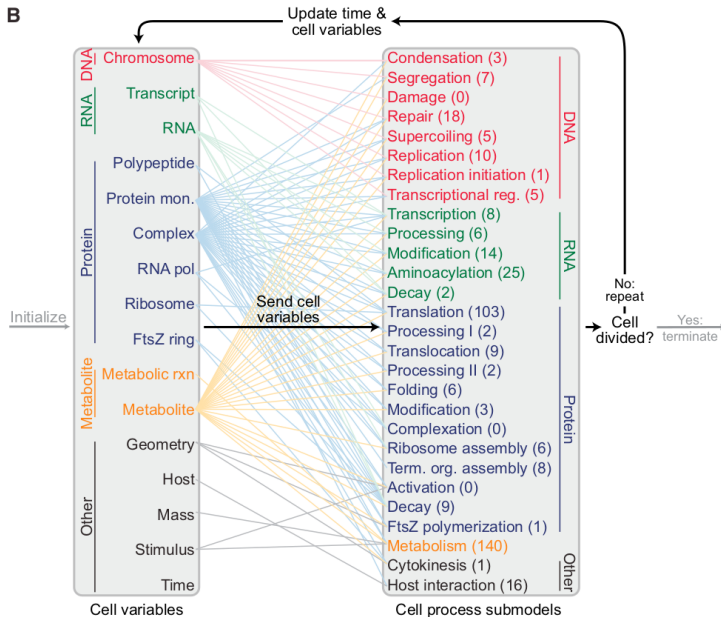


Slika: Model PROM.

# Whole-cell model (Karr, 2012)

*The whole-cell model was constructed by integrating submodels of 28 essential cellular processes. The model includes over 1,900 quantitative parameters and is based on over 900 primary research articles, reviews, books, and databases. Most of these parameters were implemented as originally reported. However, several other parameters were carefully reconciled.*

# Whole-cell model



# Whole-cell model

*Submodels are connected through common metabolites, RNA, protein, and the chromosome. The model integrates cellular function submodels through 16 cell variables.*

*The whole-cell model is simulated using an algorithm comparable to those used to numerically integrate ODEs.*

*First, the cell variables are initialized. Next, for each 1s time step, the submodels retrieve the current values of the cellular variables, calculate their contributions to the temporal evolution of the cell variables, and update the values of the cellular variables.*

*This is repeated thousands of times during the course of each simulation. Finally, the simulation terminates when either the cell divides or the time reaches a predefined maximum value.*



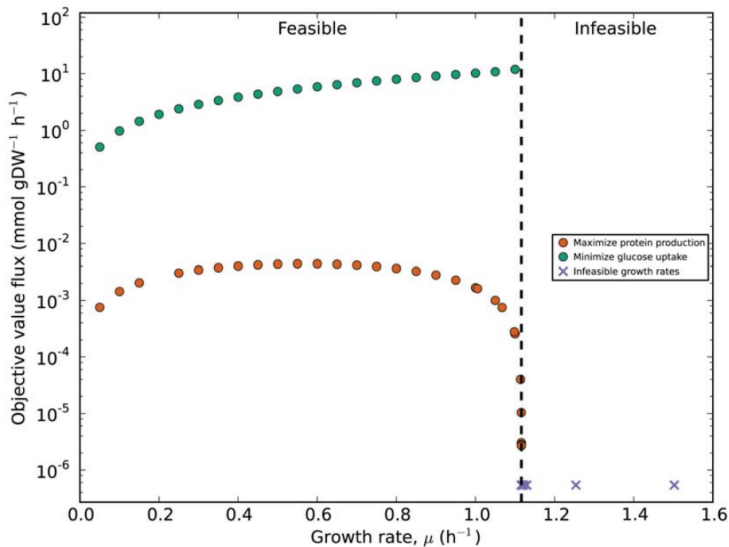
# ME model (O'Brien et al., 2013)

- ▶ Mikrobni rastni model – išče optimalno celično stanje za rast v danem ravnovesnem okolju.
- ▶ Vključuje stehiometrično predstavitev genskega regulatornega omrežja:
  - Povzema pomembne dele mehanizmov makromolekularne sinteze:
    - Od sinteze RNA in proteinov do transkripcije, RNA procesnih reakcij idr.
  - Določa globalne, z rastjo povezane omejitve.
  - Omogoča združevanje z metabolnim modelom:
    - Oba podajata omejitve istega linearnega programa.
- ▶ Ne vsebujejo od stanja oz. okolja odvisnih dinamičnih regulatornih omejitev.
  - Te nimajo nujno preproste stehiometrične predstavitve.

# ME-model – optimizacijski proces

1. Določijo se omejitve metabolnega modela (razpoložljivost in koriščenje metabolitov).
2. Določijo se omejitve genskega regulatornega modela.
3. Določijo se z rastjo (growth rate,  $\mu$ ) pogojene omejitve:
  - ▶ Ker so nekatere omejitve genskega regulatornega modela funkcije rasti, optimizacija rasti ne zadostuje pogojem LP.
  - ▶ Optimizacija zato poteka iterativno:
    - Po zaporedju bisekcije se išče pretok največje rasti ( $\mu$ ), od koder naprej so LP nedopustni.
4. Izvede se optimizacija LP in določi rast v novi iteraciji.

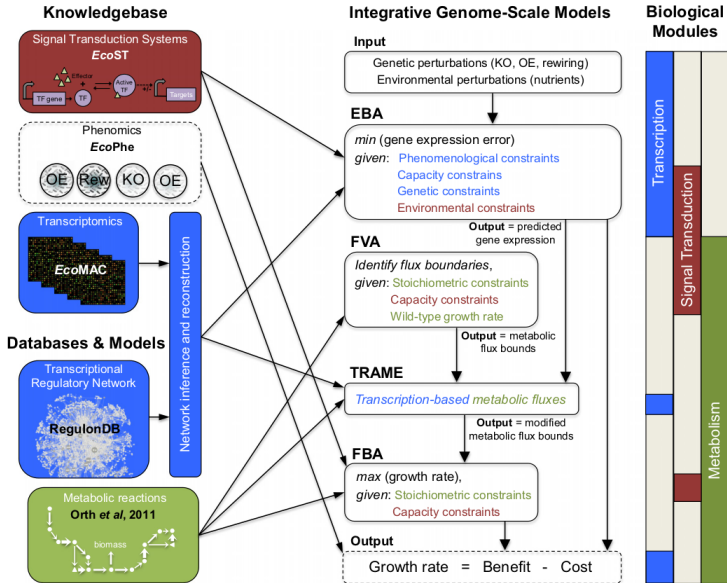
# ME-model – optimizacijski proces



# Cost-benefit model (Carrera et al., 2014)

- ▶ Metodologija, ki poenoti različne biološke procese in njihove interakcije skozi več slojev v skupen okvir.
  - Združuje informacije transkripcijskega regulatornega, signalnega in metabolnega omrežja s transkriptomskimi in fenomskimi podatki.
- ▶ Posluži se nekaj na novo ustvarjenih metod:
  - EBA (expression balance analysis)
  - TRAME (transcription-based metabolic flux enrichment)
- ▶ Izhod ( $\mu$ ) določen s cost-benefit formulacijo.

# Cost-benefit model – integracijski okvir



## Cost-benefit model – EBA

*We have constructed a [large] gene expression compendium. We model the mRNA dynamics of all genes in the compendium as a function of the TF concentration by linear ordinary differential equations (ODEs).*

*Then, we developed [EBA] to predict the global gene expression profile of *E. coli* under genetic modifications and environmental changes. EBA formulates an optimization problem to find the gene expression profile subject to four sets of constraints (phenomenological, capacity, environmental, and genetic constraints). Specifically, we used a fitness function that minimizes the gene expression errors of the TFs.*

# Cost-benefit model – TRAME

*We created a transcription-based metabolic flux enrichment (TRAME) method to integrate metabolic and transcriptional regulatory networks modifying the  $V_{min}$  and  $V_{max}$  calculated from Flux Variability Analysis (FVA) for each metabolic flux in the *E. coli* metabolic model. This approach determines the new values of the flux bounds  $V_{min}$  and  $V_{max}$  for a given enzyme.*

## Cost-benefit model – output

*Integration of the various cellular and environmental components to phenotypic changes was performed through a cost-benefit model. As such, we compute the growth burden due to the production and maintenance of all proteins (cost), as well as the growth advantage due to the energy uptake of the metabolic pathways in each environment (benefit).*

*In this cost-benefit model, the genetic cost is defined as the relative reduction in growth rate ( $\mu$ ) due to the production of essential proteins.*

*To measure the cost  $c$ , we computed the deviation between the [reference] and [EBA-predicted] gene expression profiles. To compute the metabolic benefit  $B$ , we used the metabolic sub-model. The growth rate is given by the difference between the benefit and the cost:*

$$\mu = B - c$$



# TPM segmentacija (Sarkar et al., 2019)

- ▶ Primer vključitve časovnega pogojevanja v FBA na podlagi transkriptomskih podatkov o genskem izražanju.
- ▶ Glede na te podatke se v modelu omejuje zgornje meje reakcijskih pretokov, s čimer se razlikuje med dnevnimi in nočnimi časovnimi modeli.

# TPM segmentacija

*The diurnal rhythm of metabolic gene expression driven by the circadian clock and its metabolic consequences is explored.*

*[...] a periodic model of Synechocystis sp. PCC 6803 metabolism [...] spans a 12-hr light/12-hr dark cycle by segmenting it into 12 Time Point Models (TPMs) with a uniform duration of two hours.*

*The 12 TPMs allow for the incorporation of time-dependent constraints that capture the cyclic nature of cellular processes.*

*The developed framework allows for the flow of metabolites across TPMs while inventorying metabolite levels and only allowing for the utilization of currently or previously produced compounds.*

*Imposing bounds on reactions informed by temporally-segmented transcriptomic data enables simulation of phototrophic growth as a single linear programming (LP) problem.*

# TMP segmentacija – grafično

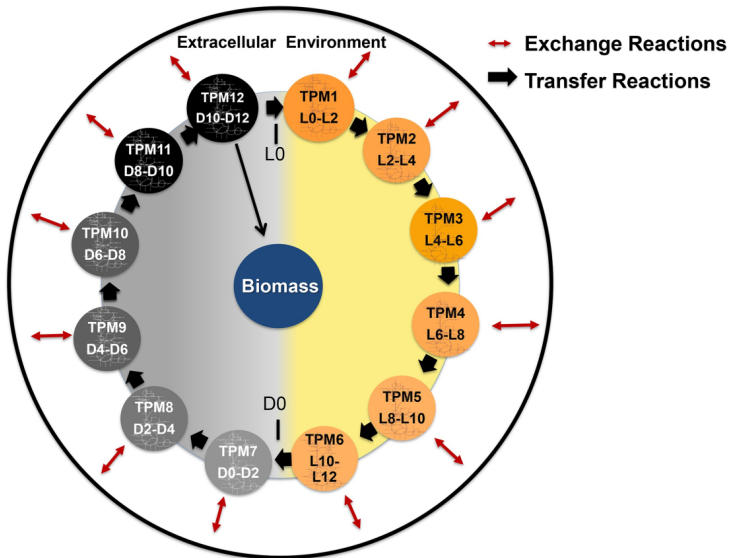


Fig 1. An overview of CycleSyn. There are 12 individual TPMs, with metabolites being transferred through them unidirectionally in the direction indicated. Photons are only supplied to the first six TPMs, and biomass is sequestered in TPM12, i.e., the last dark TPM. Exchange reactions move nutrients and other metabolites in and out of the cell, while transfer reactions shuttle metabolites from one TPM to the next.

# TMP segmentacija – LP

The optimization formulation used to determine maximum biomass production flux is

$$\text{maximize } v_{\text{biomass, TPM12}}$$

subject to

$$\sum_{j \in J} S_{ijk} v_{jk} = 0, \quad \forall i \in I, k \in K$$

$$v_{\text{CO}_2 \text{ uptake}, k} \leq 1.1, \quad k = 1, \dots, 6$$

$$v_{\text{Photon uptake}, k} \leq 60, \quad k = 1, \dots, 6$$

$$v_{\text{ATP maintenance}, k} \geq 10, \quad \forall k \in K$$

$$v_{jk}^{LB} \leq v_{jk} \leq v_{jk}^{UB}, \quad \forall j \in J, k \in K$$

$$0 \leq v_{j,k} \leq 10,000, \quad \forall j \in J^{\text{Transfer}}, k \in K$$

$$v_{\text{Photosynthesis}, k} \leq \frac{1}{2} (v_{\text{Pigment Transfer}, k \text{ to } k+1} - v_{\text{pigment transfer}, k-1 \text{ to } k}) M_C, \quad \forall k \in K$$

where  $S_{ijk}$  is the stoichiometric coefficient for metabolite  $i$  in reaction  $j$  and TPM  $k$ ,  $v_{jk}^{LB}$  and  $v_{jk}^{UB}$  are the upper and lower flux bounds for reaction  $j$  in TPM  $k$ ,  $I, J$ , and  $K$  denote the sets of total metabolites, reactions, and Time Point Models (TPMs), respectively, and  $J^{\text{Transfer}}$  the set of all transfer reactions.

