



Motivation: Huge number of EFMs in genome-scale metabolic network reconstructions \rightsquigarrow targeted search

Problem statement

Input:

- ▷ Metabolic network $\mathcal{N} = (\text{Met}, \text{Reac}, S)$ (assume $\text{Irr} = \text{Reac}$)
- ▷ Set of t target reactions $\{r_1, \dots, r_t\} \subseteq \text{Reac}$,
- ▷ Natural number $k \geq 1$.

Output:

- ▷ Set E of EFMs in \mathcal{N} , $|E| = k$,
- ▷ $\text{supp}(e) \supseteq \{r_1, \dots, r_t\}$, for all $e \in E$.



Acuña et al. 09 and 10

Theorem

1. Computing an EFM containing one given target reaction ($t = 1$) can be done in polynomial time.
2. Deciding whether there exists an EFM containing $t \geq 2$ target reactions is NP-complete (even for $t = 2$).



David/Bockmayr 13

Iteratively, compute a sequence

$$\mathcal{N}^1, e^1, \mathcal{N}^2, e^2, \dots, \mathcal{N}^k, e^k,$$

of subnetworks $\mathcal{N}^i = (\text{Met}, \text{Reac}^i)$ and EFMs e^i of \mathcal{N} such that for all $i \in \{1, \dots, k\}$:

- ▷ The target reactions r_1, \dots, r_t belong to \mathcal{N}^i .
- ▷ e^i is an EFM in \mathcal{N}^i involving r_1 .
- ▷ r_2, \dots, r_t are directionally coupled to r_1 in \mathcal{N}^i , i.e., $v_{r_1} \neq 0$ implies $v_{r_2} \neq 0, \dots, v_{r_t} \neq 0$, for all $v \in C^i$.
- ▷ None of e^1, \dots, e^{i-1} is a flux mode in \mathcal{N}^i .



Acuña et al. 09

Linear optimisation problem

$$\begin{aligned} \text{LP}(\mathcal{N}): \min & 0 \\ \text{s.t. } Sv & = 0, \\ & v_{r_1} \geq 1, \\ & v_r \geq 0, \quad \forall r \in \text{Reac}. \end{aligned}$$

\rightsquigarrow use Simplex method to compute a non-zero vertex (basic feasible solution)



One target reaction, k EFMs

de Figueiredo et al. 09

- ▷ k -shortest EFMs
- ▷ Sequence of mixed-integer linear optimisation problems

$$\text{MILP1}(E) : \min \sum_{r \in \text{Reac}} a_r$$

s.t.

$$Sv = 0,$$

$$v_{r_1} \geq 1,$$

$$a_r \leq v_r \leq M a_r, \quad \forall r \in \text{Reac},$$

$$v_r \geq 0, \quad \forall r \in \text{Reac},$$

$$a_r \in \{0, 1\}, \quad \forall r \in \text{Reac},$$

$$\sum_{r \in \text{supp}(e)} a_r \leq |\text{supp}(e)| - 1, \quad \forall e \in E.$$



Computational results: one target reaction

- ▷ 100 EFMs, *E. coli* iAF1260
- ▷ 20 EFMs, *S. cerevisiae* iND750

Method	NoR	LI	AHD
Shortest EFMs	54	25-26	12.79
Algorithm 1	272	25-57	26.082

Method	Integer variables		Continuous variables	
	Length	Time	Length	Time
M = 10	Shortest	6-10	1719s	2074s
	Algo 1	6-15	16s	15s
M = 100	Shortest	6-10	8158s	3421s
	Algo 1	6-21	21s	18s
M = 1000	Shortest	6-10	14362s	7780s
	Algo 1	6-31	16s	29s



Generic approach: one target reaction

David/Bockmayr 13

- ▷ Compute **feasible** solution of (MILP1) \rightsquigarrow subnetwork \mathcal{N}^i
- ▷ Solve LP(\mathcal{N}^i) to obtain EFM e^i

Algorithm 1

Step Action

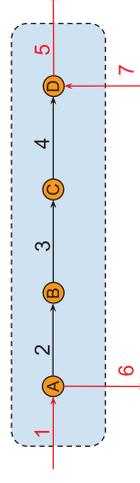
0. Initialize $i := 1, E := \emptyset$.
1. Try to find a feasible solution (v^i, a^i) of MILP1(E).
2. If MILP1(E) is infeasible, then STOP.
3. Otherwise, use (v^i, a^i) to derive subnetwork \mathcal{N}^i .
4. Find a basic feasible solution e^i of LP(\mathcal{N}^i).
5. Let $E := E \cup \{e^i\}$ and $i := i + 1$.
6. If $i > k$ then STOP.
7. Go to Step 1.



Two target reactions

First attempt

$$\begin{aligned} \text{(MILP2)} : \min \quad & \sum_{r \in \text{Reac}} a_r \\ \text{s.t.} \quad & Sv = 0, \\ & v_{r_1} \geq 1, \\ & v_{r_2} \geq 1, \\ & a_r \leq v_r \leq M a_r, \quad \forall r \in \text{Reac}, \\ & v_r \geq 0, \quad \forall r \in \text{Reac}, \\ & a_r \in \{0, 1\}, \quad \forall r \in \text{Reac}. \end{aligned}$$





David/Bockmayr 13

Proposition

- ▷ An optimal solution v^* of (MILP2) is either an EFM or the sum of two EFMs.
- ▷ In the subnetwork \mathcal{N}^* defined by (v^*, a^*) , the reactions r_1 and r_2 are either fully coupled or mutually exclusive.

Consequences

- ▷ Refine (MILP2) by excluding the second case.
- ▷ Require that r_1 is directionally coupled to r_2 (or vice versa).
- ▷ Use **Farkas' Lemma** to express this condition by a set of linear constraints in the dual space.



$$\begin{aligned}
 \text{MILP3}(E) : \min \quad & 0 \\
 \text{s.t.} \quad & Sv = 0 \\
 & v_{r_1} \geq 1, \\
 & v_{r_2} \geq 1, \\
 & a_r \leq v_r \leq M_0 a_r, \quad \forall r \in \text{Reac}, \\
 & S^T y + u^T x \geq M_1 (a - \mathbf{1} - u^{i_2}), \\
 & \quad \quad \quad -x \geq \mathbf{1}, \quad \text{DirC}(r_1, r_2) \\
 & \sum_{r \in \text{supp}(e)} a_r \leq |\text{supp}(e)| - 1, \quad \forall e \in E, \\
 & v_r \geq 0, \quad \forall r \in \text{Reac}, \\
 & a_r \in \{0, 1\}, \quad \forall r \in \text{Reac}, \\
 & x, y_m \in \mathbb{R}, \quad \forall m \in \text{Met}.
 \end{aligned}$$

(u^i denotes the r -unit vector)



Theorem

Let $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^n$.

The system of linear inequalities

$$Ax \leq b$$

has no solution $x \in \mathbb{R}^n$ if and only if the system

$$u^T A = 0, \quad u^T b = -1, \quad u \geq 0 \quad \Leftrightarrow \quad \begin{pmatrix} A^T \\ b^T \end{pmatrix} u = \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \quad u \geq 0$$

has a solution $u \in \mathbb{R}^m$.



Step	Action	Algorithm 2
0.	Initialize $i := 1, E := \emptyset$.	
1.	Try to find a feasible solution (v', a') of MILP3(E).	
2.	If MILP3(E) is infeasible, then STOP.	
3.	From (v', a') derive subnetwork \mathcal{N}^i .	
4.	Find a basic feasible solution e^i of $\text{LP}(\mathcal{N}^i)$.	
5.	Let $E := E \cup \{e^i\}$ and $i := i + 1$.	
6.	If $i > k$ then STOP.	
7.	Go to Step 1.	



- ▷ For t target reactions, $t > 2$, it is enough to add directional coupling constraints $\text{DirC}(r_1, r_2)$, $\text{DirC}(r_1, r_3)$, ..., $\text{DirC}(r_1, r_t)$.
- ▷ **Prototype software available:**
<https://sourceforge.net/projects/caefm>



- ▷ Compute EFMs for every pair of reactions
- ▷ Time out of 60 secs per pair

Network	# Reac.	# Pairs	EFMs found	Mean length	No EFM exists	No answer
<i>E. coli</i> central	90	8010	7691	24.36	176	143
<i>H. pylori</i>	269	72092	66749	46.57	1862	3481

↪ for most pairs, there exists an EFM, which can be found by the algorithm.



Shortest flux modes vs. EFMs

- ▷ Two target reactions
- ▷ Computing shortest flux modes with MILP2 vs. computing EFMs with MILP3
- ▷ Time out of 60 secs per pair

Network	# Reac.	# Pairs	EFMs found with MILP3	EFMs found with MILP2	False pos with MILP2
<i>E. coli</i> central	90	8010	7691	5212	2686
<i>H. pylori</i>	269	72092	66749	206	9213

↪ MILP2 in general not sufficient.