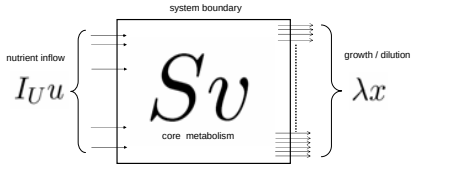
Systematic analysis of conservation relations in *E. coli* genome-scale metabolic network reveals novel growth media

Marcin Imielinski*, Calin Belta†, Adam Halasz*, and Harvey Rubin*

*University of Pennsylvania †Boston University

Abstract: Applying a fundamental result from linear programming theory, we identify a novel dual relationship between extreme semipositive conservation relations (ESCR) and production capabilities of a metabolic network under a given nutrient media. We exploit this relationship in an algorithm that systematically enumerates all minimal nutrient sets that render an objective species weakly producible (i.e. producible in the absence of irreversibility constraints) through a simple traversal of ESCR. We apply our results to a recent genome scale model of *E. coli* metabolism, in which we traverse the 51 anhydrous ESCR of the metabolic network to determine all 928 minimal aqueous nutrient media that render biomass weakly producible. We find 287 of these 928 species sets to be feasible as nutrients for biomass in the presence of irreversibility constraints. We also find that an additional 365 of these nutrient media render biomass producible in the presence of oxygen.

Theory



$S \in \mathbb{R}^{m \times n}$ Stoichiometry matrix of *m* species and *n* reactions

$x \in \mathbb{R}^m_+$ Vector of *m* species concentrations

$v \in \mathbb{R}^n$ Flux configuration through *n* reactions of network

$U \subset 1, \ldots, m$ Indices of nutrient species

$u \in \mathbb{R}^{|U|}$ Flux configuration through |*U*| nutrient inputs.

I_U Columns of *m* x *m* identity matrix corresponding to indices in set *U*

λ Growth (dilution) rate

$T \subseteq 1, \ldots, n$ Indices of irreversible reactions

Basic Formulation and Definitions

Dynamic equation of metabolic network in nutrient media *U*:

$$\dot{x} = I_U u + Sv - \lambda x,\ v_T \geq 0 \quad (1)$$

Quasi-steady state:

$$I_U \bar{u} + S\bar{v} = \bar{\lambda}\bar{x} \geq 0,\ \bar{v}_T \geq 0 \quad (2)$$

Producibility of species *i* under nutrient media *U [2]*:

Feasibility of $\bar{x}_i > 0$ *in Eq. 2 during growth.*

Weak producibility of species *i* under nutrient media *U*

Producibility in absence of irreversibility constraints (i.e. setting $T = \emptyset$ *in Eq. 2)*

Cone of semipositive conservation relations of *S*

$$G = \mathcal{N}(S^T) \bigcap \mathbb{R}^m_+$$

Extreme semipositive conservation relations (ESCR) of *S* [1]

Vectors that lie on an extreme ray (i.e. an edge) of $\mathbf{G}$*. These are the "simplest" semipositive conservation relations obeyed by* $\mathbf{S}$*. They correspond to pools of the largest chemical subunits conserved by all reactions in* $\mathbf{S}$*. [1]*

Theoretical results

Theorem 1: A species *i* is weakly producible in a nutrient media *U* if and only if every *i*-containing ESCR contains a species from *U*.

Biochemical translation: a species *i* is weakly producible if and only if the nutrient media *U* supplies every moiety pool to which species *i* belongs

Corollary 1: A species *i* is weakly producible in a *W*-containing nutrient media *U* if and only if every *i*-containing, non-*W*-containing ESCR contains a species from *U* (where $W \subset U$).

Algorithm

Objective: Determine all minimal *weak* (*W*-containing) *nutrient sets U* for a set of "objective" species *J*. A species set *U* is a *weak nutrient set* for *J* if *every* species in *J* is weakly producible under *U*. Such a set is minimal if none of its (*W*-containing) subsets render *every* species in *J* weakly producible.

Schematic: Let *E* be the collection of *J*-containing (non-*W* containing) ESCR of S. According to the theorem to the left, each minimal *weak* (*W*-containing) *nutrient set U* is a minimal set of species that contributes to every ESCR in E. Viewing each ESCR as a subset of species, *U* is called a *minimal hitting set* for E. We can thus systematically construct the collection of minimal weak (W-containing) nutrient sets as the collection M of minimal hitting sets of E, computed using the following algorithm:

```
M = minimal_hitting_sets(E)
{
  M = { ∅ \\ initialize M with empty set
  for each E in E
     P = positive components of E
     for each M in M disjoint with P
      remove M from M
        for each p in P
           if M does not contain subset of M ∪{p} add M ∪{p} to M
           endfor
     endfor
  endfor
}
```

Model: *E. coli* iJR904 [4]

- 762 species (143 extracellular) involved in 932 reactions.
- Biomass reaction consumes ~50 intracellular species and produces pseudo-species biomass

Results

- We find only 51 non-water-containing (i.e. anhydrous) ESCR associated with *E. coli* iJR904!
- Application of Corollary 1 to these anhydrous ESCR accurately predicts weak producibility under any aqueous nutrient media.
- Biomass-containing anhydrous ESCR induce 11 classes of extracellular species that are equivalent as weak nutrients for biomass in any aqueous media.
- Application of the minimal hitting set algorithm to the set of biomass-containing anhydrous ESCR yields 928 minimal weak aqueous nutrient sets for biomass.
- 287 of these 928 sets render biomass producible in the context of irreversibility constraints.
- An additional 365 of these nutrient sets render biomass producible in the presence of oxygen.

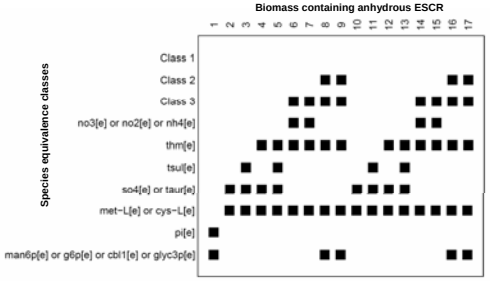


Fig. 1: Biomass-containing anhydrous ESCR for *E. coli* iJR904
Sparsity patterns of 17 biomass-containing anhydrous ESCR associated with the E. coli iJR904 genome scale metabolic model, projected to 143 extracellular species. Rows correspond to 11 extracellular species equivalence classes, induced by membership in identical anhydrous biomass-containing ESCR. Each column represents the positive extracellular species components of an ESCR as the union of species equivalence classes indicated by the black squares.

Class 1	Class 2	Class 3
AMP	(S)-Propane-1,2-diol	1,5-Diaminopentane
Choline	2-Dehydro-3-deoxy-D-gluconate	meso-2,6-Diaminoheptanedioate
L-Carnitine	3-hydroxycinnamic acid	4-Aminobutanoate
Dimethyl sulfide	3-(3-hydroxy-phenyl)propionate	N-Acetyl-D-glucosamine
Dimethyl sulfoxide	Acetate	N-Acetyl-D-mannosamine
Fe^{2+}	Acetoacetate	N-Acetylneuraminate
L-Fucose 1-phosphate	Acetaldehyde	Adenine
gamma-butyrobetaine	2-Oxoglutarate	Adenosine
Glycine betaine	L-Arabinose	D-Alanine
H^+	Butyrate (n-C4:0)	L-Alanine
K^+	Citrate	Allantoin
D-Methionine	CO_2	L-Arginine
Sodium	Dihydroxyacetone	L-Asparagine
O_2	Ethanol	L-Aspartate
Trimethylamine	Formate	Cytosine+C68
Trimethylamine N-oxide	D-Fructose	Cyanate
	L-Fucose	Cytidine
	Fumarate	Deoxyadenosine
	D-Galactose	Deoxycytidine
	D-Galactarate	Deoxyguanosine
	D-Galactonate	Deoxyinosine
	Galactitol	Deoxyuridine
	D-Galacturonate	D-Glucosamine
	D-Glucose	L-Glutamine
	D-Gluconate	L-Glutamate
	D-Glucarate	Glycine
	D-Glucuronate	Guanosine
	D-Glyceraldehyde	Guanine
	Glycerol	L-Histidine
	Glycolate	Hypoxanthine
	Hexadecanoate (n-C16:0)	L-Isoleucine
	L-Idonate	Indole
	D-Lactate	Inosine
	L-Lactate	L-Leucine
	Lactose	L-Lysine
	L-Malate	Nicotinate
	Maltose	Ornithine
	Maltohexaose	L-Phenylalanine
	Maltopentaose	(R)-Pantothenate
	Maltotriose	L-Proline
	Maltotetraose	Putrescine
	D-Mannose	D-Serine
	Melibiose	L-Serine
	D-Mannitol	Spermidine
	octadecanoate (n-C18:0)	L-Threonine
	Phenylpropanoate	Thymidine
	Pyruvate	L-Tryptophan
	D-Ribose	L-Tyrosine
	L-Rhamnose	Uracil
	D-Sorbitol	Urea
	Succinate	Uridine
	Sucrose	L-Valine
	L-tartrate	Xanthine
	Trehalose	Xanthosine
	tetradecanoate (n-C14:0)	
	D-Xylose	

Fig 2: Extracellular species equivalence classes
Species in each of these equivalence classes belong to the same set of biomass-containing anhydrous ESCR of E. coli iJR904. According to Corollary 1, this means that they are interchangeable in any weak aqueous nutrient set for biomass. Classes 1, 2 and 3 appear respectively to be species that do not share anhydrous chemical moieties with biomass, carbon sources, and carbon-nitrogen sources. Note: these classes arise purely from network analysis.

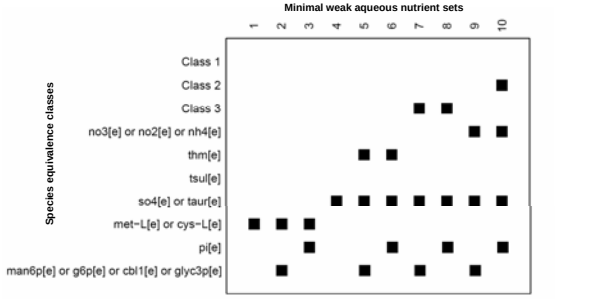


Fig. 3: Minimal weak aqueous nutrient sets for biomass in *E. coli* iJR904 [4]
928 minimal weak aqueous nutrient sets for biomass, expressed as conjunctions of 11 species equivalence classes. Each column j represents a collection of minimal weak nutrient sets, each formed by choosing one species from each species equivalence class denoted by a black square in the column.

cbl1[e]	Cob(I)alamin	**met-L[e]**	L-Methionine	**pi[e]**	Phosphate
cys-L[e]	L-Cysteine	**nad[e]**	NAD	**so4[e]**	Sulfate
g6p[e]	D-Glucose 6-phosphate	**nh4[e]**	ammonium	**taur[e]**	Taurine
glyc3p[e]	Glycerol 3-phosphate	**nmn[e]**	NMN	**thm[e]**	Thiamin
glyc3p[e]	Glycerol 3-phosphate	**no2[e]**	Nitrite	**tsul[e]**	Thiosulfate
man6p[e]	D-Mannose 6-phosphate	**no3[e]**	Nitrate		

Table 1: Species abbreviations
Note: The '[e]' suffix represents the extracellular compartment.

Conclusions / Observations

- Theorem 1 generalizes the notion of "weak connectivity" in a graph to a metabolic network, which is a *directed hypergraph*. (ESCR ~ weak component, weak producibility of *i* under U ~ weak connectivity of *i* and *j*)
- Minimal weak nutrient sets are related to ESCR exactly as minimal cut sets are related to elementary modes [3].
- The small number of anhydrous ESCR associated with this network and large equivalence classes of species that emerge from these ESCR suggest that the primary functional / compositional difference between species lies in their hydration state.
- Predictions generated by our approach are direct results of network annotation and the particular growth model. They suggest *in vivo* experiments that can further iterative model building and refinement.
- Our approach is extensible to alternative models of growth and essentiality, including models that go "beyond biomass".

References:

1. Heinrich R, and Schuster S. 1996. The Regulation of Cellular Systems. Chapman and Hall, New York.
2. Imielinski M, Belta C, Halasz A, Rubin H. 2005. Investigating metabolite essentiality through genome-scale analysis of Escherichia coli production capabilities. Bioinformatics. 21(9):2008-16.
3. Klamt S, Gilles ED. Minimal cut sets in biochemical reaction networks. Bioinformatics. 2004 Jan 22;20(2):226-34.
4. Reed JL, Vo TD, Schilling CH, Palsson BO. 2003. An expanded genome-scale model of Escherichia coli K-12 (IJR904 GSM/GPR). Genome Biol. 4(9):R54.

Acknowledgements:

M.I. is supported by the BioAdvance Fellow Award in Bioinformatics. A. H. is supported by NIH-NLM IBIF award 1-F37-LM-008343-1. C. B. is supported by NSF CCF 0432070. H.R. is supported by DARPA BioComp.