Metabolic flux prediction via gene expression and metabolomics

Tomer Shlomi

Department of Computer Science, Technion, Haifa, Israel

October, 2009

Metabolic stuff we've been doing recently

Tomer Shlomi

Department of Computer Science, Technion, Haifa, Israel

October, 2009

Outline:

- Context-dependent Biomass Flux Balance Analysis (CB-FBA)
  - Predicting flux distributions by accounting for growth-associated demand for biomass production in a context-dependent manner
- RobustKnock: Predicting Metabolic Engineering Knockout Strategies for Chemical Production
  - Improving OptKnock by accounting for alternative pathways
- Predicting Metabolic Gene-Nutrient Interactions (GNIs) in yeast
  - Predicting constraints on nutrient availability in the growth media based on enzyme essentiality data
- Predicting Enzyme Sub-cellular Localization
  - Predicting enzymes' sub-cellular localization based on partial localization data for a subset of the enzymes in the network

Outline:

- Context-dependent Biomass Flux Balance Analysis (CB-FBA)
  - Predicting flux distributions by accounting for growth-associated demand for biomass production in a context-dependent manner
- RobustKnock: Predicting Metabolic Engineering Knockout
- st
- d
- Predicting enzymes' sub-cellular localization based on partial localization data for a subset of the enzymes in the network

Metabolic Flux Balance Analysis with Context-dependant Biomass.
T. Benyamini, O. Folger, E. Ruppin, T. Shlomi,
RECOMB, Systems Biology, 2009 (to appear)

NOT PUBLISHED YET

Outline:

- Context-dependent Biomass Flux Balance Analysis (CB-FBA)
  - Predicting flux distributions by accounting for growth-associated demand for biomass production in a context-dependent manner
- RobustKnock: Predicting Metabolic Engineering Knockout Strategies for Chemical Production
  - Improving OptKnock by accounting for alternative pathways
- Predicting Metabolic Gene-Nutrient Interactions in yeast
  - Predicting constraints on nutrient availability in the growth media based on enzyme essentiality data
- Predicting Enzyme Sub-cellular Localization
  - Predicting enzymes' sub-cellular localization based on partial localization data for a subset of the enzymes in the network

Outline:

- Context-dependent Biomass Flux Balance Analysis (CB-FBA)
  - Predicting flux distributions by accounting for growth-associated demand for biomass production in a context-dependent manner
- RobustKnock: Predicting Metabolic Engineering Knockout Strategies for Chemical Production
  - Improving OptKnock by accounting for alternative pathways
- Predicting Metabolic Gene-Nutrient Interactions in cancer
- Predicting Metabolic Engineering Knockout Strategies for Chemical Production: Accounting for Competing Pathways.
- N. Tepper, T. Shlomi (Submitted)

Predicting Metabolic Engineering Knockout Strategies for Chemical Production: Accounting for Competing Pathways.
N. Tepper, T. Shlomi (Submitted)

NOT PUBLISHED YET

Outline:

- Context-dependent Biomass Flux Balance Analysis (CB-FBA)
  - Predicting flux distributions by accounting for growth-associated demand for biomass production in a context-dependent manner
- RobustKnock: Predicting Metabolic Engineering Knockout Strategies for Chemical Production
  - Improving OptKnock by accounting for alternative pathways
- Predicting Metabolic Gene-Nutrient Interactions in yeast
  - Predicting constraints on nutrient availability in the growth media based on enzyme essentiality data
- Predicting Enzyme Sub-cellular Localization
  - Predicting enzymes' sub-cellular localization based on partial localization data for a subset of the enzymes in the network

Outline:

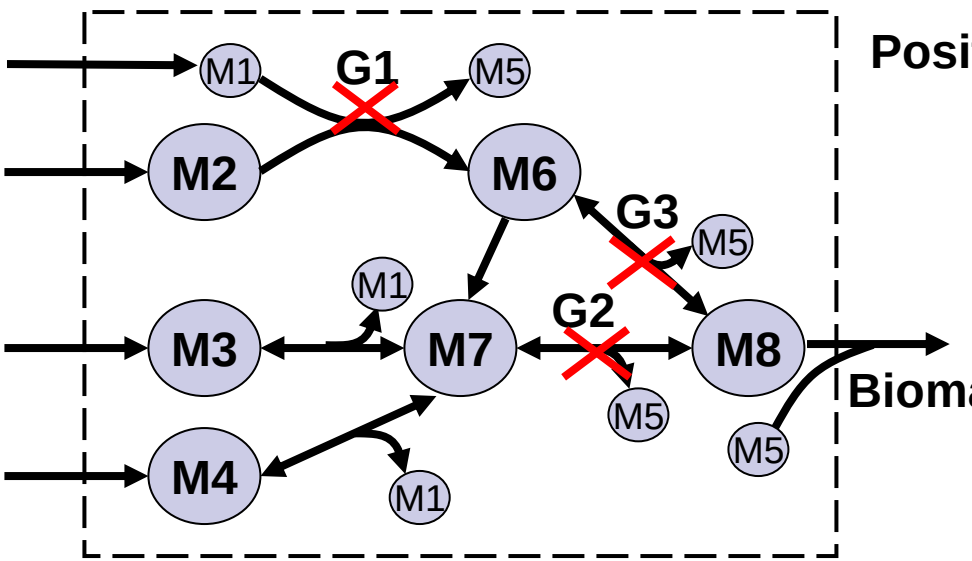
- Context-dependent Biomass Flux Balance Analysis (CB-FBA)
  - Improving OptKnock by accounting for alternative pathways
- Predicting Metabolic Gene-Nutrient Interactions in yeast
  - Predicting constraints on nutrient availability in the growth media based on enzyme essentiality data
- Predicting Enzyme Sub-cellular Localization
  - Predicting enzymes' sub-cellular localization based on partial localization data for a subset of the enzymes in the network

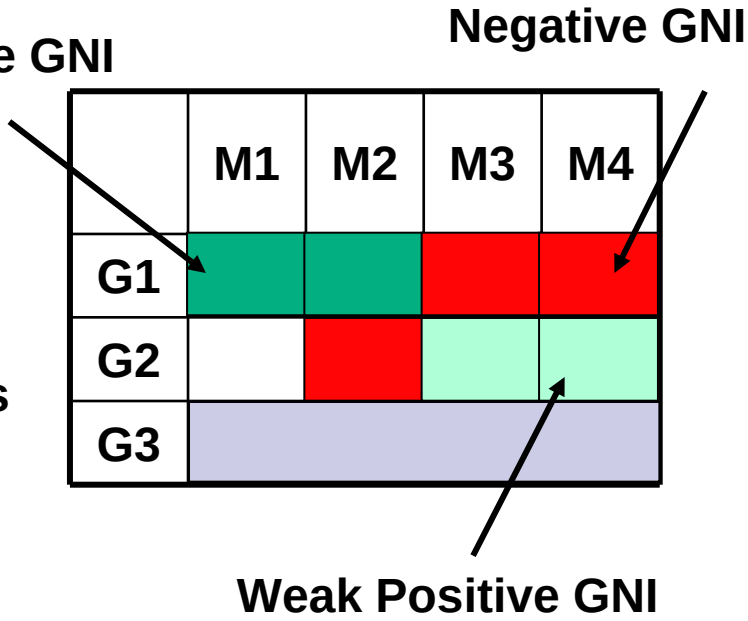
Metabolic Network-based Analysis of Yeast Gene-Nutrient Interactions.
I. Diamant, Y. Eldar, O. Rokhlenko, E. Ruppin, T. Shlomi.
Molecular BioSystems, DOI: 10.1039, 2009

Gene-Nutrient Interactions

- Under which growth media G1 is essential?

- A GNI represents a constraint on the presence/absence of a nutrient in the growth media under which a gene is essential
- A weak (vs. strong) GNI reflects a non-strict constraint





Diamant, et al., Molecular Biosystems, 2009

Predicting Gene-Nutrient Interactions

- Identified via a bi-level optimization problem
- Transformed into Mixed-Integer Linear Programming (MILP)

Maximize gene essentiality (wild-type vs. knockout growth-rate)
(over all growth media)

subject to

Maximize wild-type growth rate
(over fluxes)

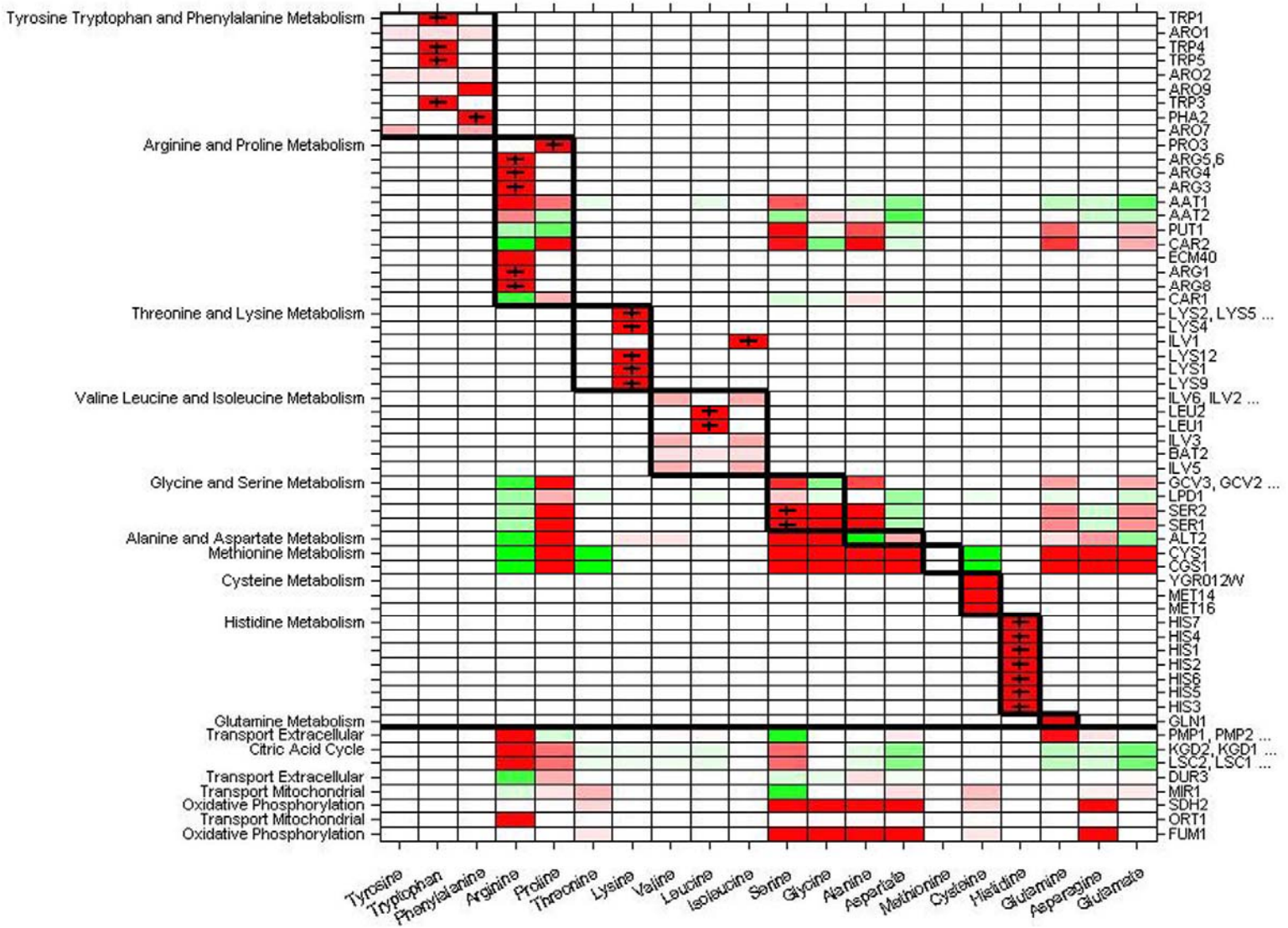
subject to
- uptake of media substrates
- network stoichiometry, thermodynamics and capacity constraints

Maximize knockout growth rate
(over fluxes)

subject to
- uptake of media substrates
- network stoichiometry, thermodynamics and capacity constraints
- Inactive knocked-out reactions

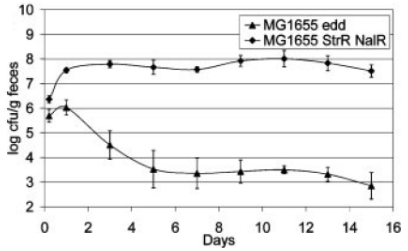
Diamant, et al., Molecular Biosystems, 2009

Gene-Nutrient Interactions in Yeast



GNI-based ‘Reverse Prediction’ of Growth Media Composition

- What is the natural growth environment of a pathogen within a host organism?
- Suppose we have in-vivo data on bacterial gene knockout essentiality

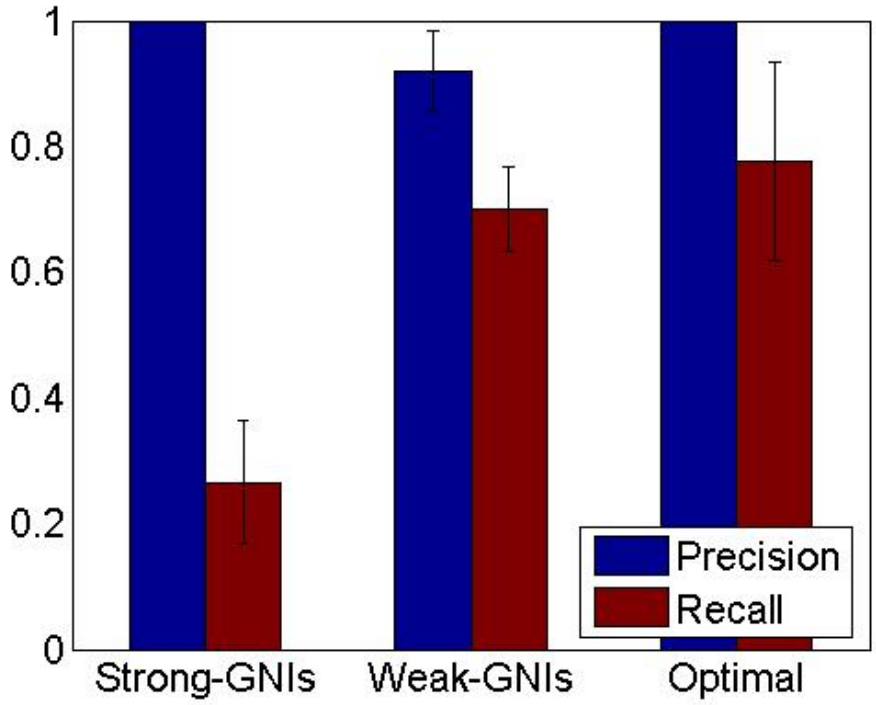


- Can we use the measured pattern of gene essentiality to predict constraints on the in-vivo growth environment of the bacteria?
- Unfortunately, we don’t have enough data of this kind. However…

Diamant, et al., Molecular Biosystems, 2009

GNI-based ‘Reverse Prediction’ of Growth Media Composition

- In simulations, GNI-based analysis provide accurate predictions of growth media composition based on gene essentiality data



Diamant, et al., Molecular Biosystems, 2009

Outline:

- Context-dependent Biomass Flux Balance Analysis (CB-FBA)
  - Predicting flux distributions by accounting for growth-associated demand for biomass production in a context-dependent manner
- RobustKnock: Predicting Metabolic Engineering Knockout Strategies for Chemical Production
  - Improving OptKnock by accounting for alternative pathways
- Predicting Metabolic Gene-Nutrient Interactions in yeast
  - Predicting constraints on nutrient availability in the growth media based on enzyme essentiality data
- Predicting Enzyme Sub-cellular Localization
  - Predicting enzymes' sub-cellular localization based on partial localization data for a subset of the enzymes in the network

Outline:

- Context-dependent Biomass Flux Balance Analysis (CB-FBA)
  - Predicting flux distributions by accounting for growth-associated demand for biomass production in a context-dependent manner
- RobustKnock: Predicting Metabolic Engineering Knockout
- Predicting Enzyme Sub-cellular Localization
  - Predicting enzymes' sub-cellular localization based on partial localization data for a subset of the enzymes in the network

Network-based Prediction of Metabolic Enzymes' Subcellular Localization.
S. Mintz, A. Aharoni, E. Ruppin, T. Shlomi.
Bioinformatics, 25(12): 247-252, 2009 (ISMB'09)

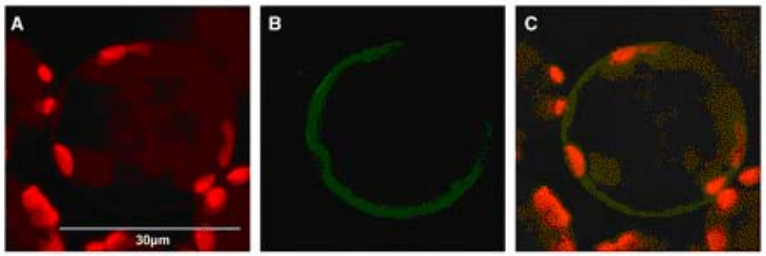
Derecting Protein Subcellular Localization

Experimental Methods:

- Green fluorescent protein (GFP) tagging
- Electron microscopy
- Subcellular fractionation + detection

Limitations:

- Costly
- Time-consuming



Wormit et.al, Plant Cell, 2006

Computational Methods:

- Sequence motifs
- Amino acid composition
- Homology
- PPI data

Limitations:

- Low number of compartments
- Performance varies across different organisms and compartments
- Relatively low availability of PPI networks

Mintz, et al., ISMB & Bioinformatics, 2009

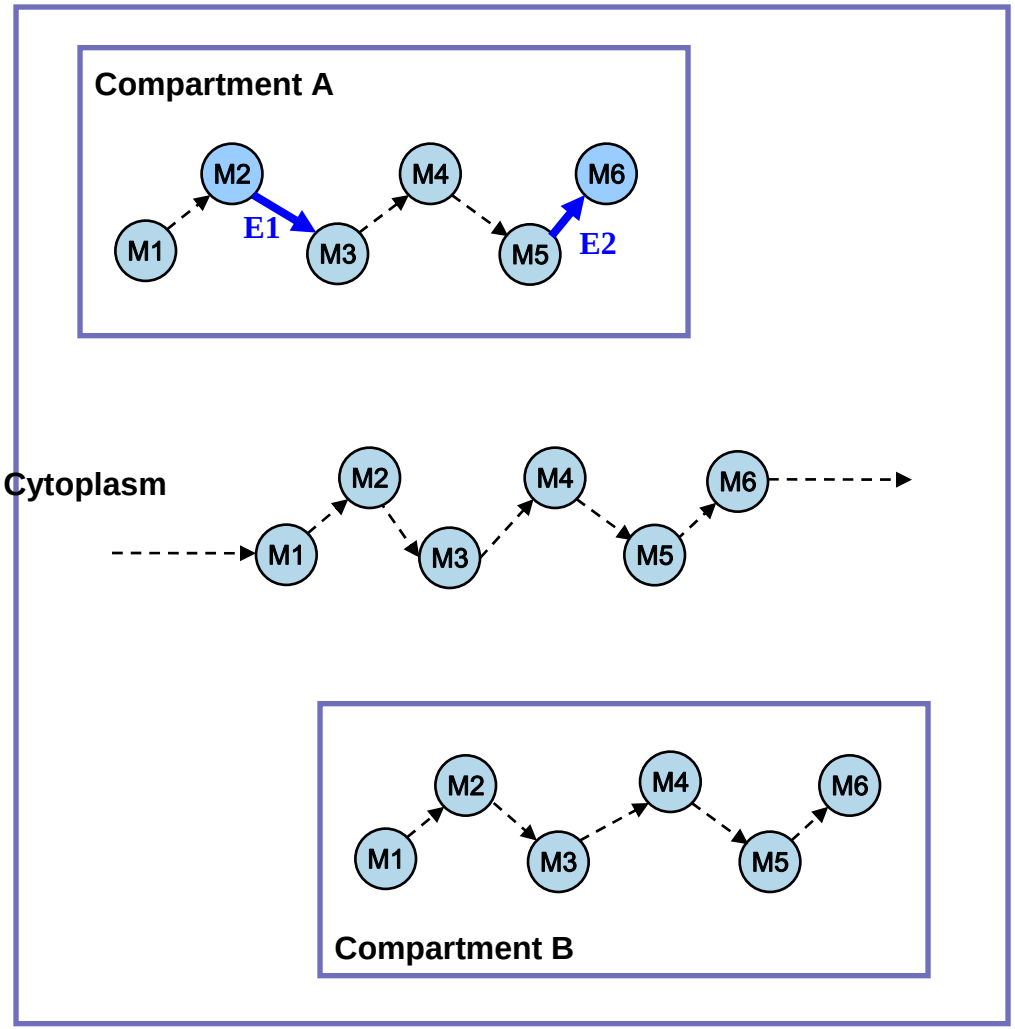
Research Objective

Predict metabolic enzymes' subcellular localization, based on:

- The organism's metabolic network
- Prior knowledge regarding localization of a subset of the enzymes
- Parsimonious assumption of minimal number of cross-membrane metabolite transports between compartments

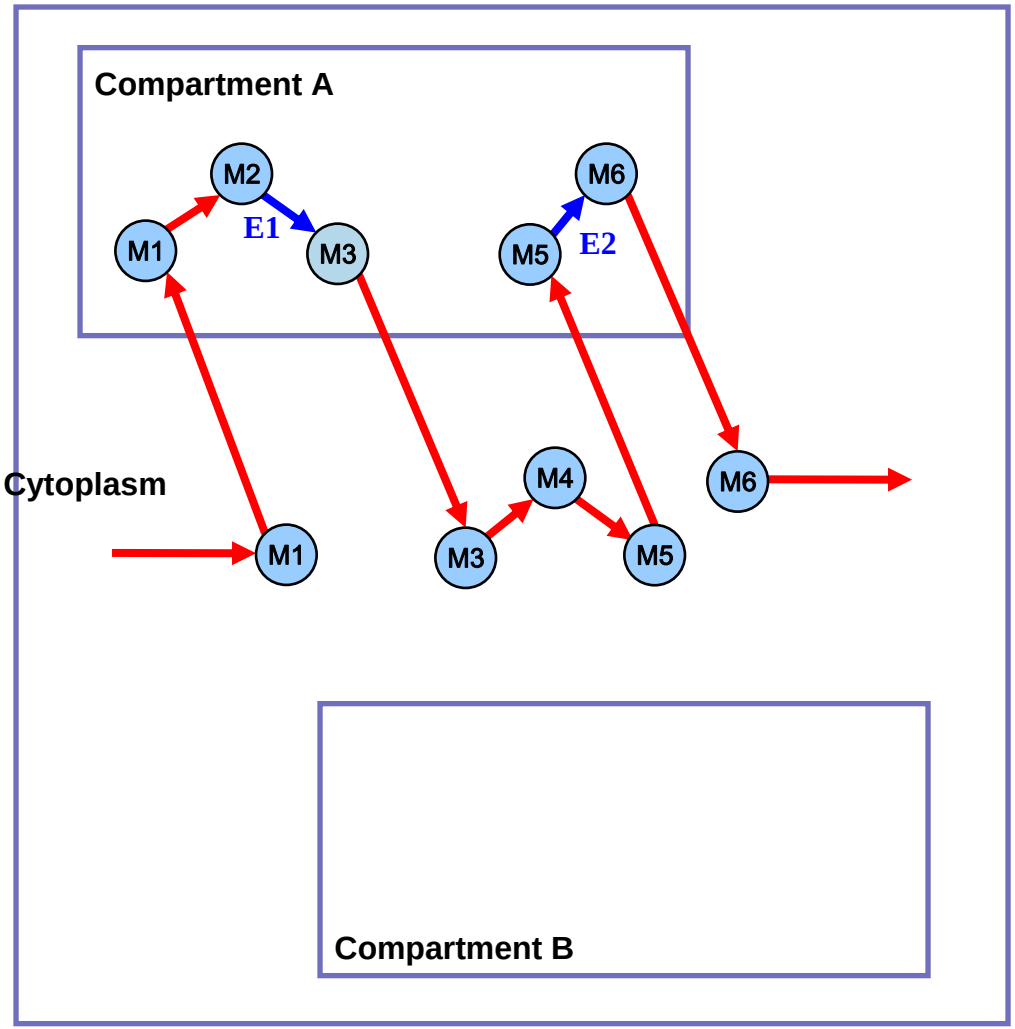
Mintz, et al., ISMB & Bioinformatics, 2009

Minimal Metabolic Transport Assumption



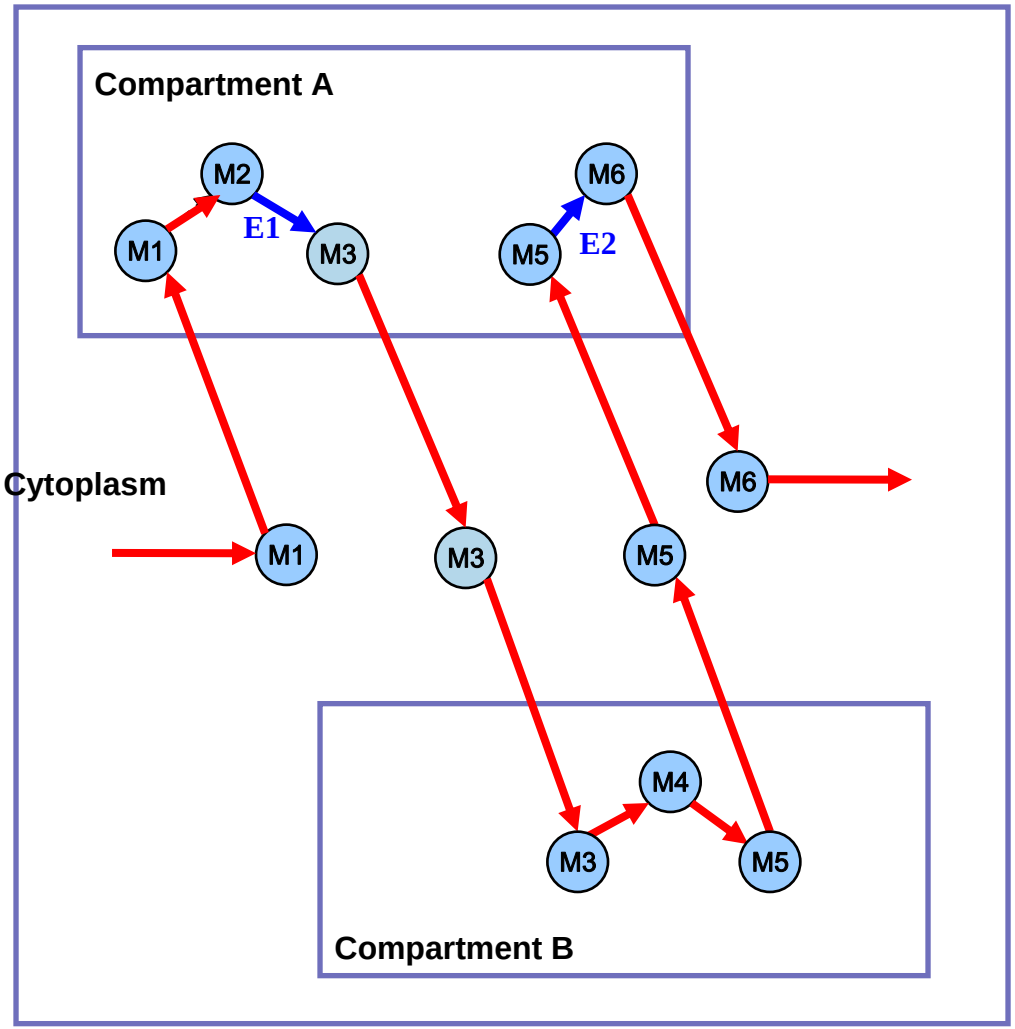
Mintz, et al., ISMB & Bioinformatics, 2009

Minimal Metabolic Transport Assumption



Mintz, et al., ISMB & Bioinformatics, 2009

Minimal Metabolic Transport Assumption

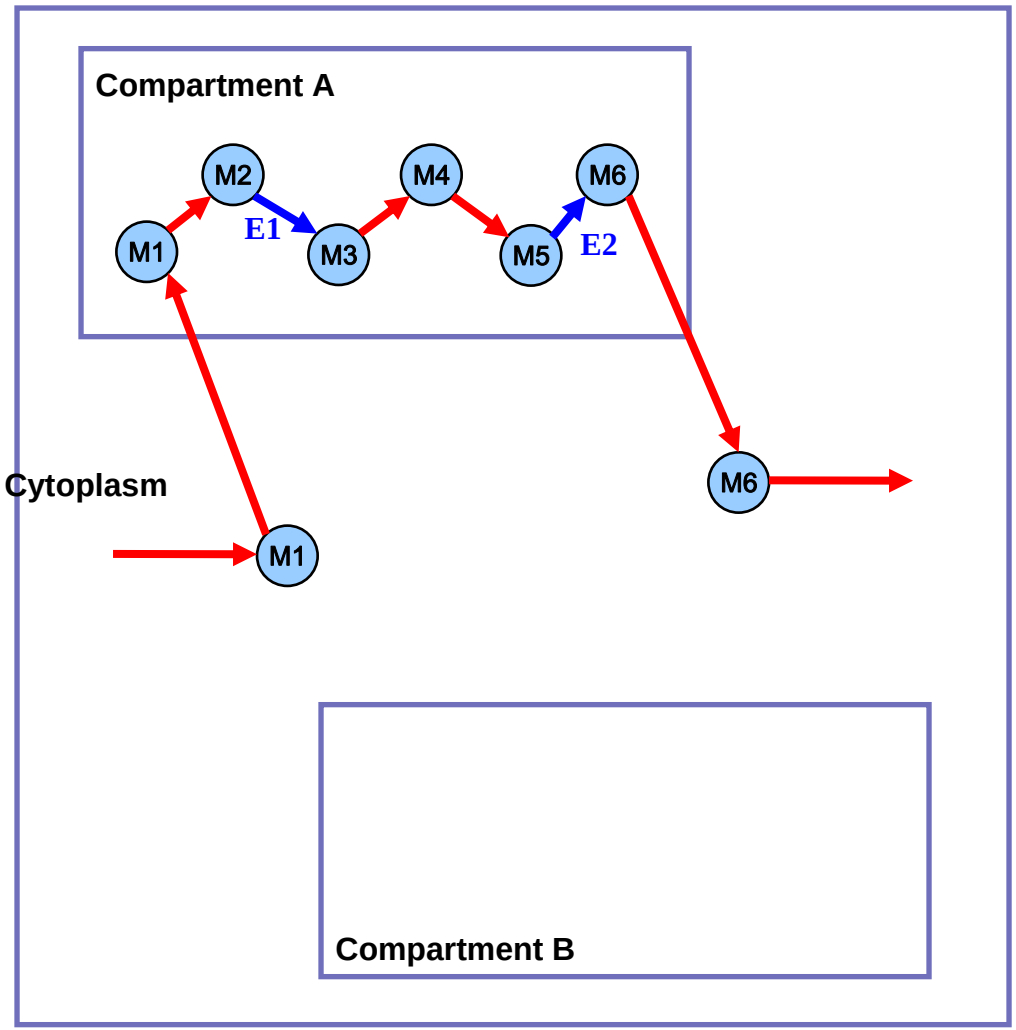


- Transport reactions depend on **transporter proteins**, imposing **energetic cost** or requiring the maintenance of a **membrane potential**

- **Minimize transport reactions**

Mintz, et al., ISMB & Bioinformatics, 2009

Minimal Metabolic Transport Assumption



- Match known localization data
- Assume minimal number of metabolite cross-membrane transports

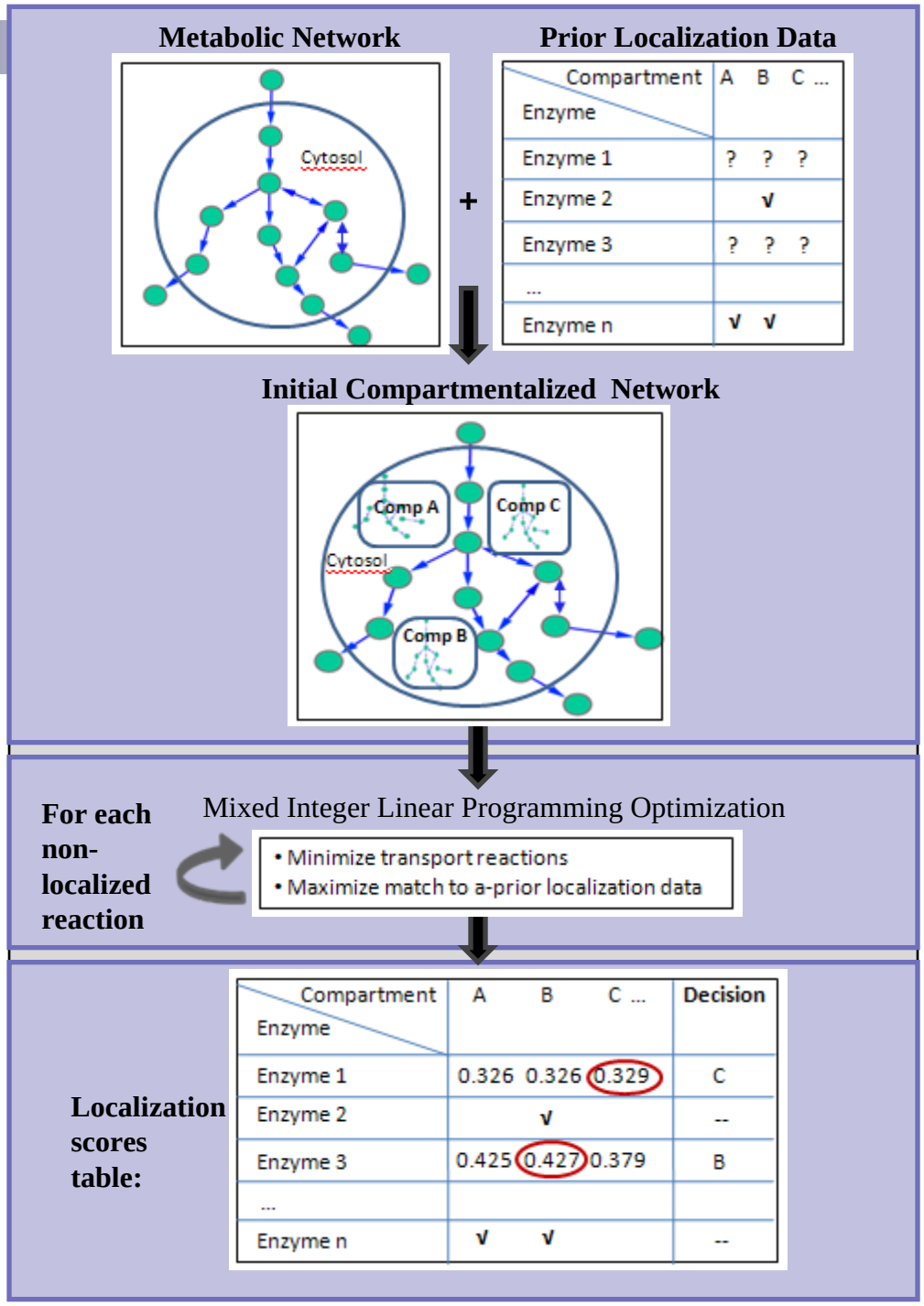
Mintz, et al., ISMB & Bioinformatics, 2009

CBM method for predicting localization:

Input:

Optimization process:

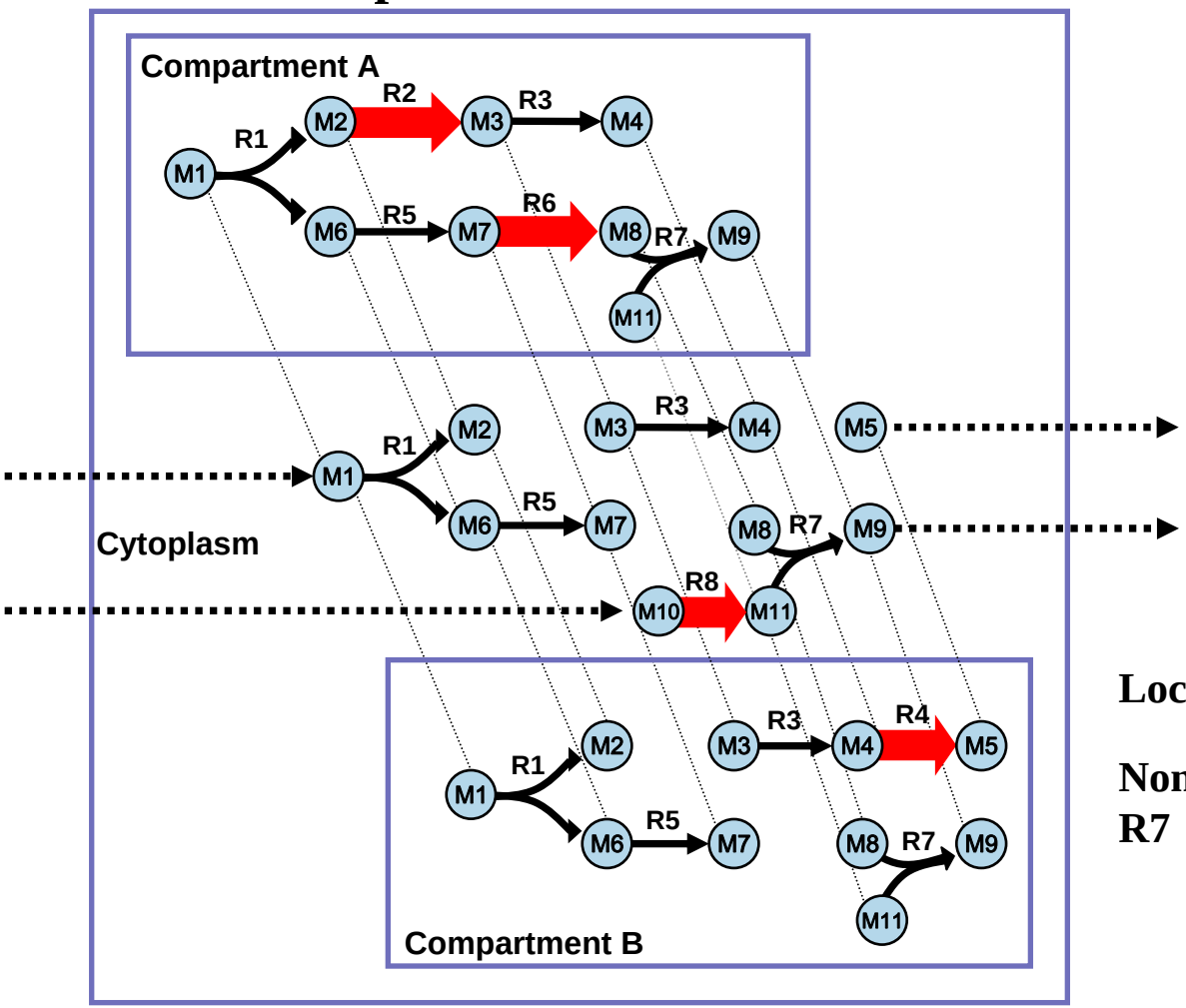
Output (prediction):



Mintz, et al., ISMB & Bioinformatics, 2009

Example

Initial compartmentalized network:



Localized reaction

Non-localized reaction

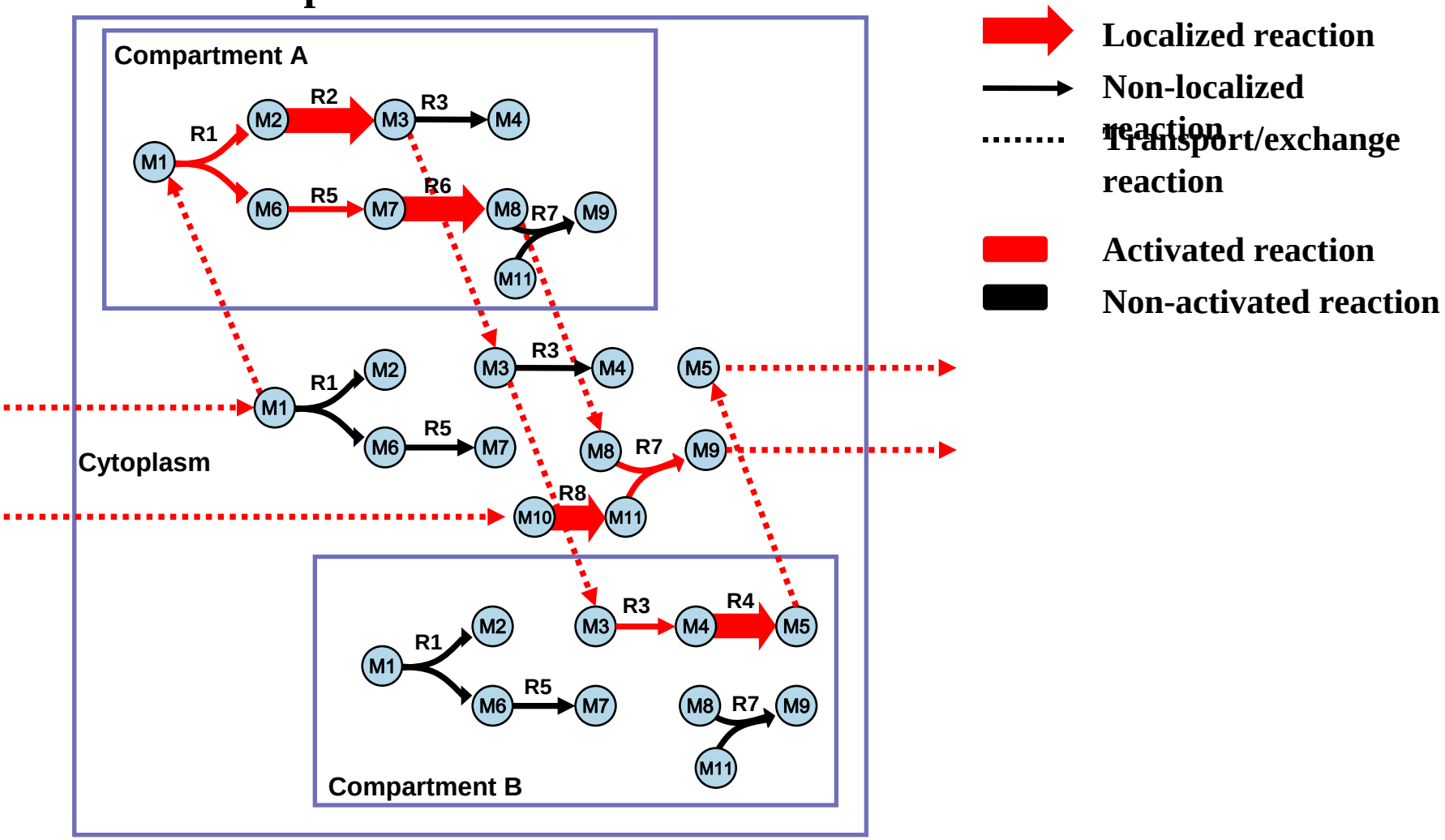
Transport/exchange reaction

Localized reactions – R2, R4, R6, R8

Non-localized reactions – R1, R3, R5, R7

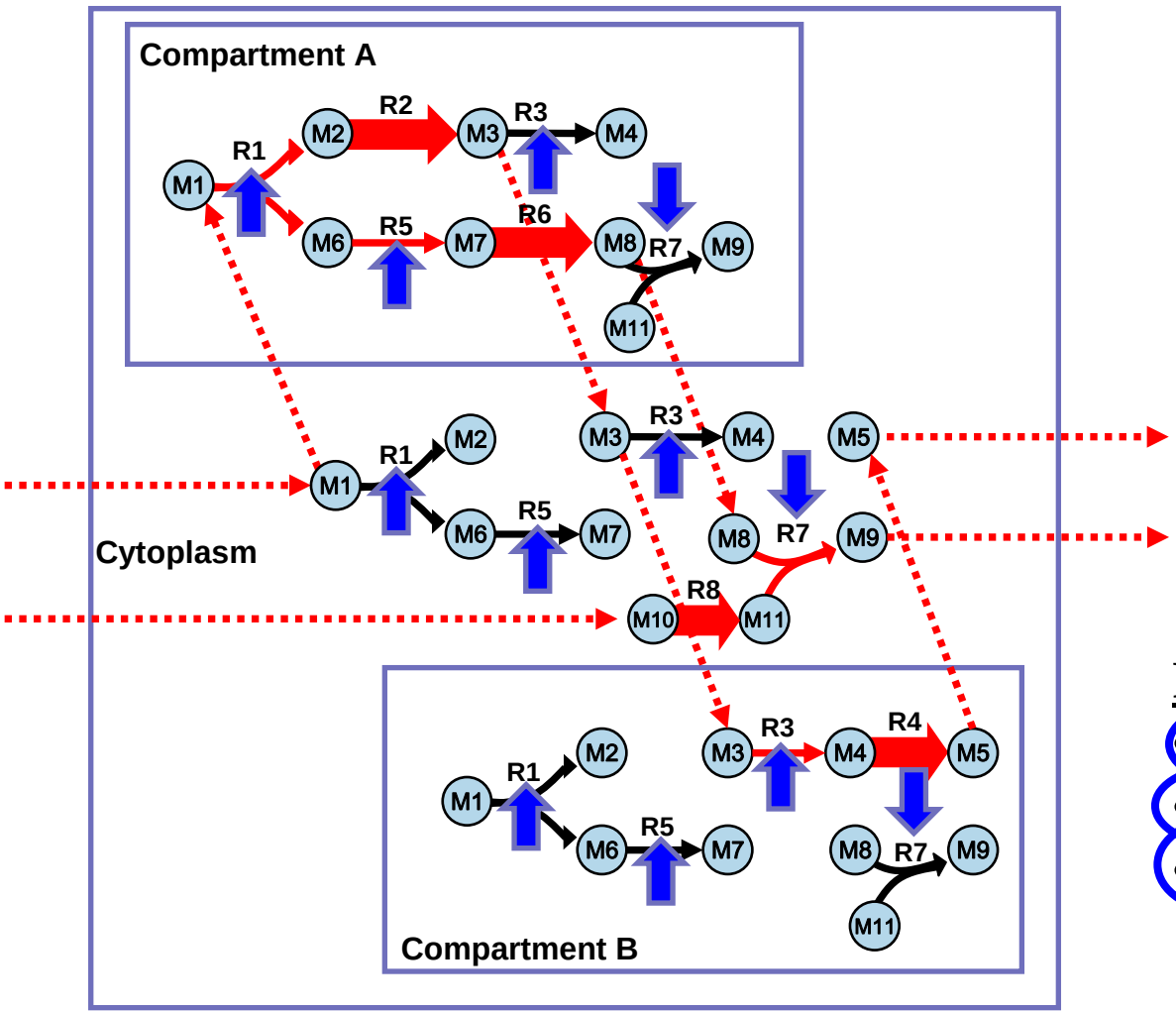
Example – Flux Distribution

Initial compartmentalized network:



Example – Results

Initial compartmentalized network:



Localized reaction

Non-localized reaction

Transport/exchange reaction

Activated reaction

Non-activated reaction

Predictions:

- R1, R5 - Compartment A
- R7 - Cytoplasm
- R3? - Compartment B

Validating Predictions via Metabolic Network of *S. cerevisiae*

- Genome-scale, fully compartmentalized metabolic network model of (Duarte et al, 2004)
- 1062 metabolites, 1149 reactions, 7 compartments

- To evaluate our method:
  1. Remove existing localization data
  2. **Cross validation test-** random localized vs. non-localized sets
  3. Apply our method
  4. Compute **- accuracy** (compared to experimental data)
     - coverage (portion of predictions with single predicted compartment)

Mintz, et al., ISMB & Bioinformatics, 2009

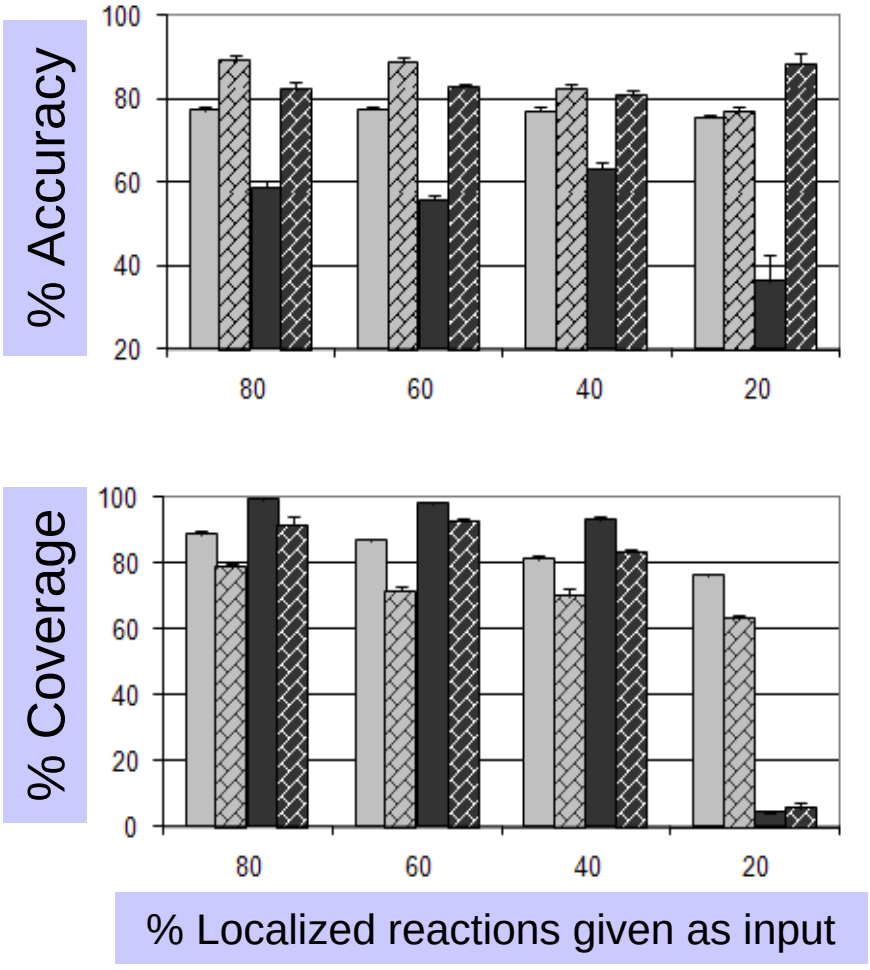
Comparison to Pathway Enrichment-Based Method

- Localization is determined based on the assignment of enzymes in pre-determined biochemical pathways
- For each pathway compute a set of hyper-geometric *p*-values reflecting the **pathway's enrichment for all compartments**, respectively
- Prediction based on compartment yielding the lowest *p*-value in the corresponding pathway

Mintz, et al., ISMB & Bioinformatics, 2009

Results

Accuracy and coverage for various fractions of localized



- **Robust accuracy**
- **Moderate coverage decline**

Collaborators

- My lab
  Naama Tepper
  Edward Vitkin
  Roi Adadi

- Eytan Ruppin's lab (Tel-Aviv)
  Tomer Benyamini
  Ori Folger
  Idit Diamant

- Asaph Aharoni's lab (Weizmann)
  Shira Mintz