Finding a Steady State on a Pathway via Translation into SAT Problem.

Takehide Soh[*1]　　Katsumi Inoue[*2,*1]

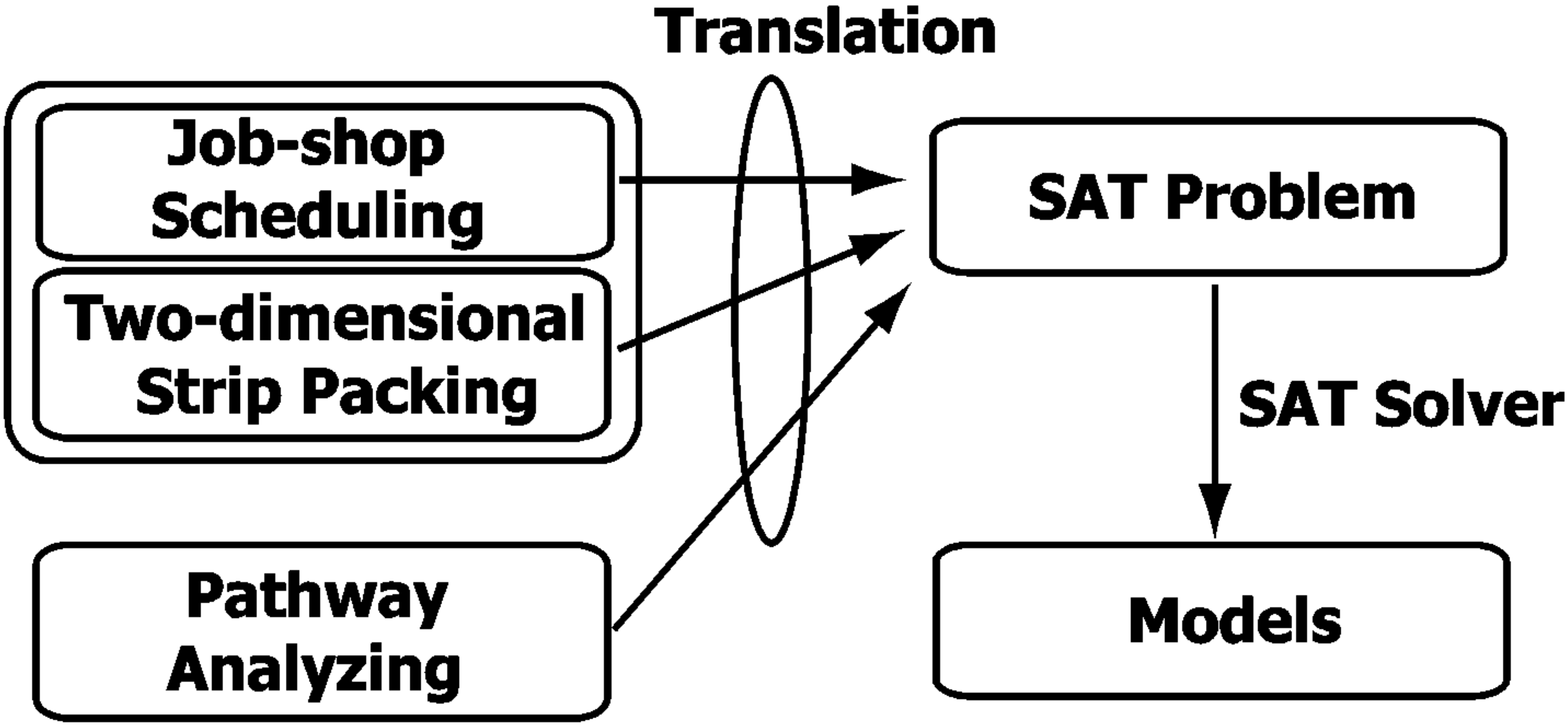
[*1]The Graduate University for Advanced Studies (SOKENDAI)

[*2] National Institute of Informatics

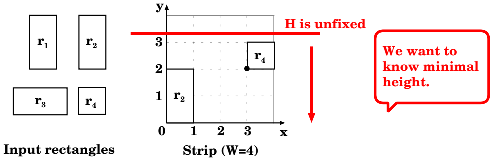
2009.9.29

Workshop@LRI

My previous and current research.



Two-dimensional strip packing problem (2SPP)



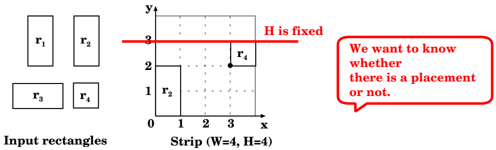
Definition of 2SPP

Input. A set $R = \{r_1, \ldots, r_n\}$ of n rectangles. Each rectangle $r_i \in R$ has a width w_i and a height h_i ($w_i, h_i \in \mathbb{N}$). A *Strip* of width $W \in \mathbb{N}$.

Constraints. Each rectangle cannot overlap with the others and the edges of the strip and parallel to the horizontal and the vertical axis.

Question. What is the minimum height such that the set of rectangles can be packed in the given strip?

Two-dimensional orthogonal packing problem (2OPP)



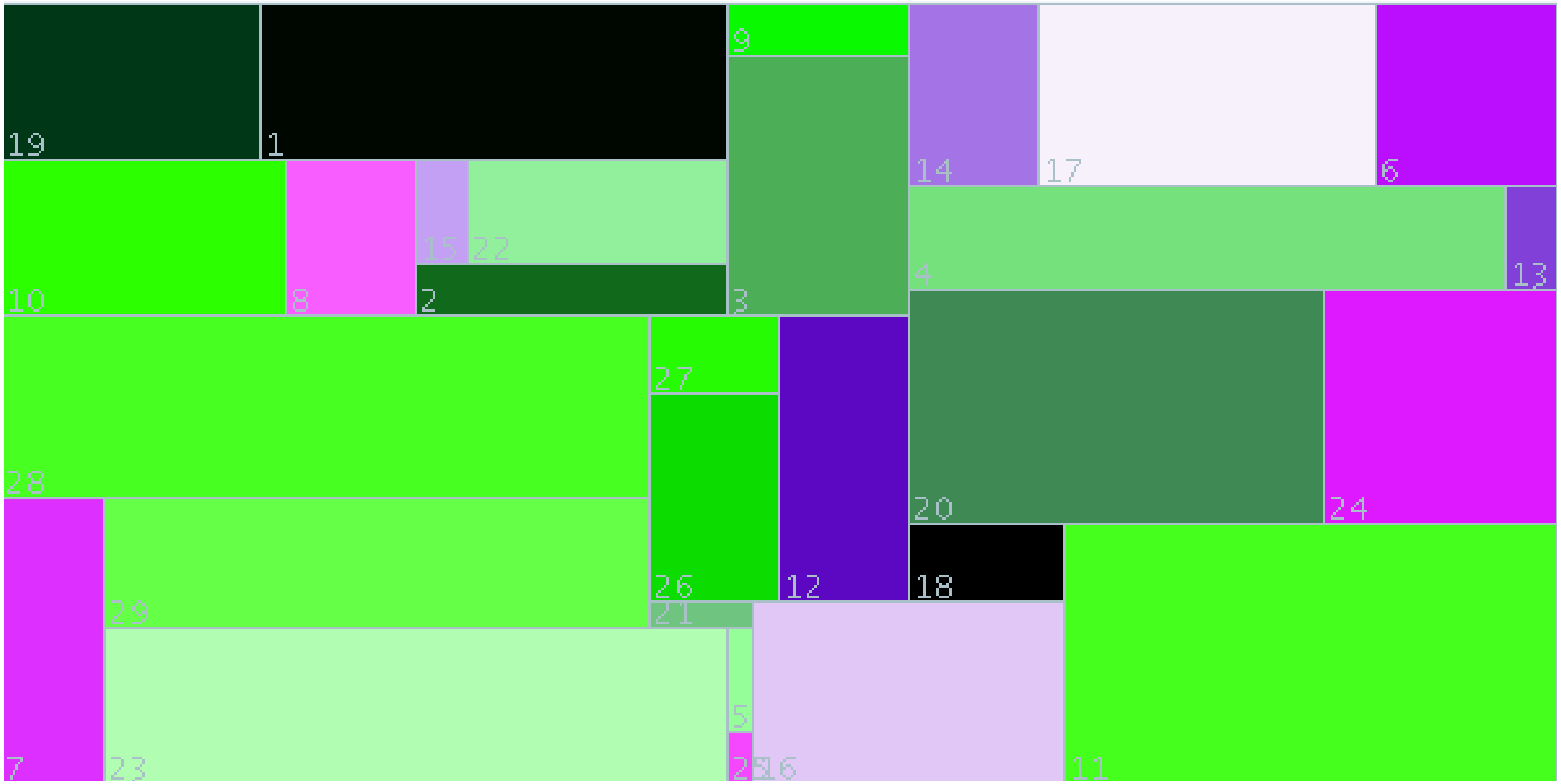
Definition of 2OPP

Input. A set $R = \{r_1, \ldots, r_n\}$ of n rectangles. Each rectangle $r_i \in R$ has width w_i and height h_i ($w_i, h_i \in \mathbb{N}$). A *Strip* of width W and height H ($W, H \in \mathbb{N}$).

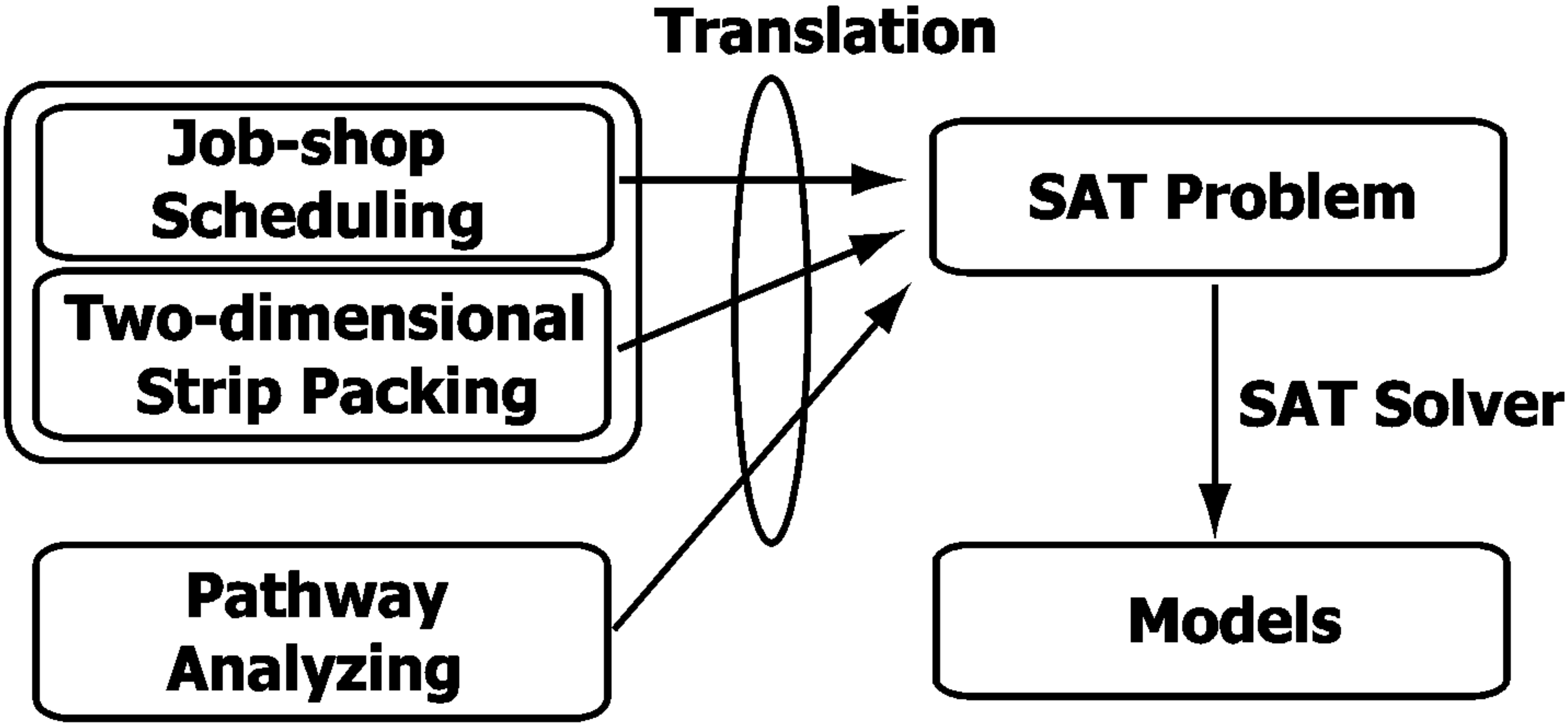
Constraints. Each rectangle cannot overlap with the others and the edges of the strip and parallel to the horizontal and the vertical axis.

Question. Can the set of rectangles be packed in the given strip?

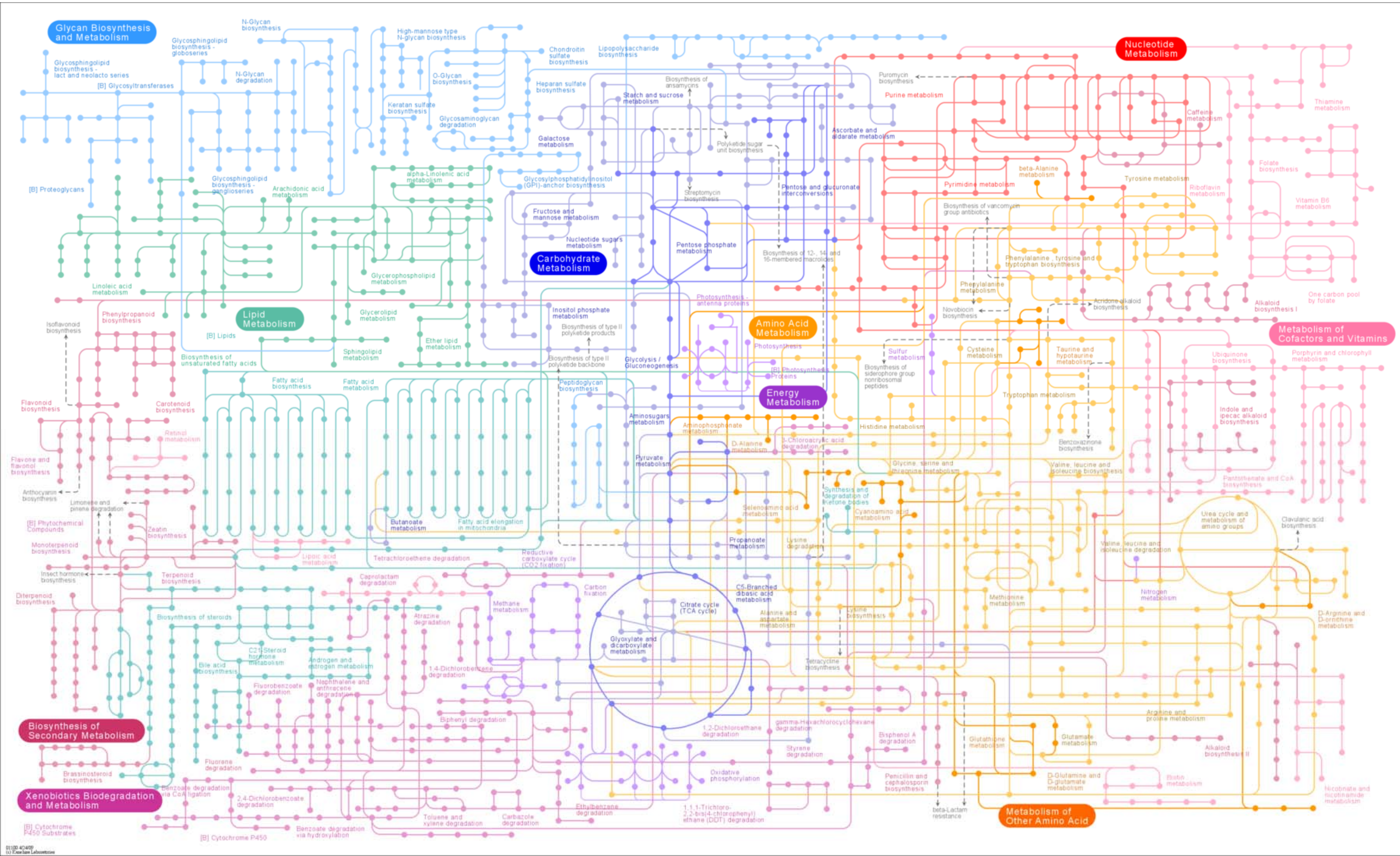
Ex. A result of 2SPP (HT08)

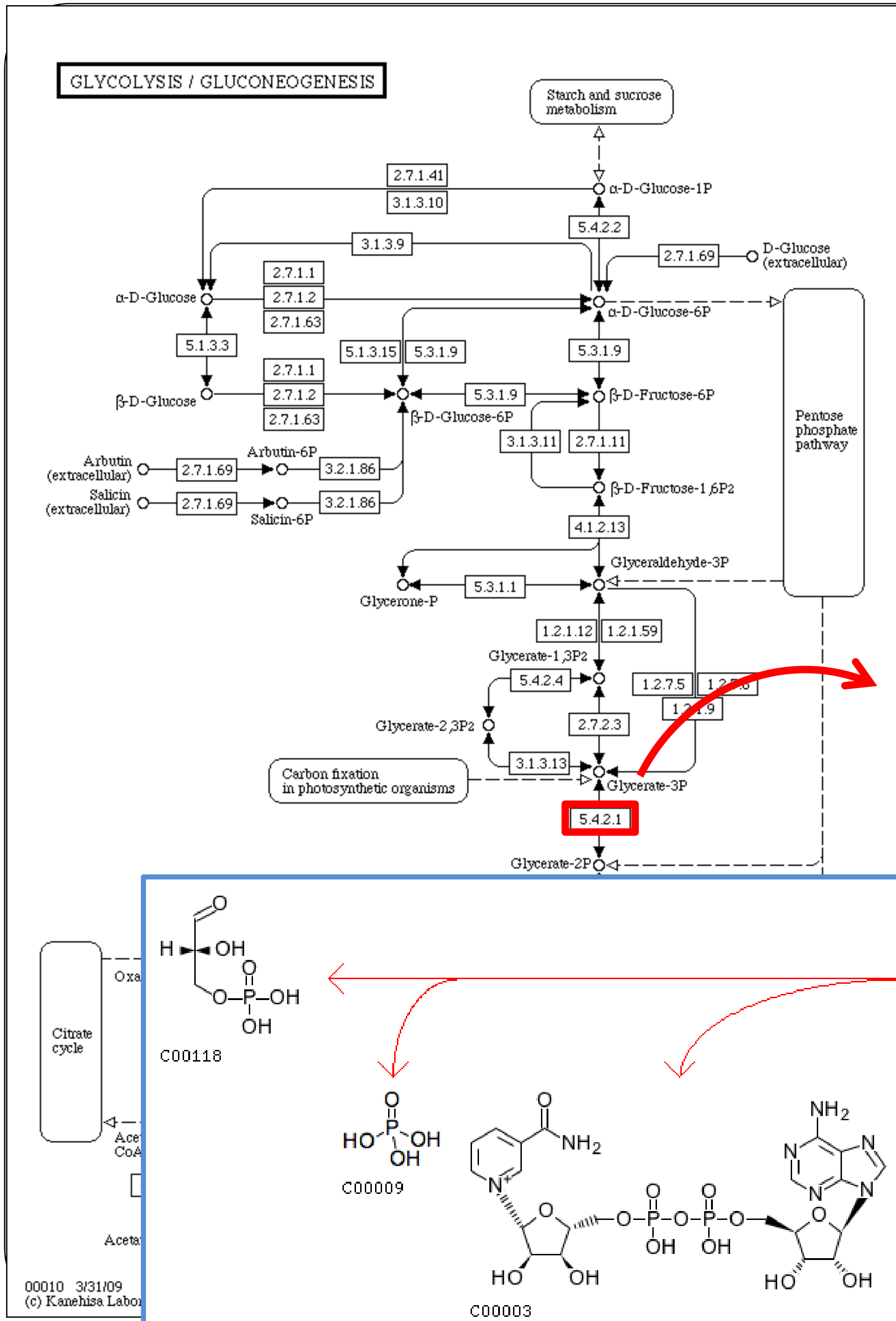


My previous and current research.



KEGG (http://www.genome.jp/kegg/) Pathway Data



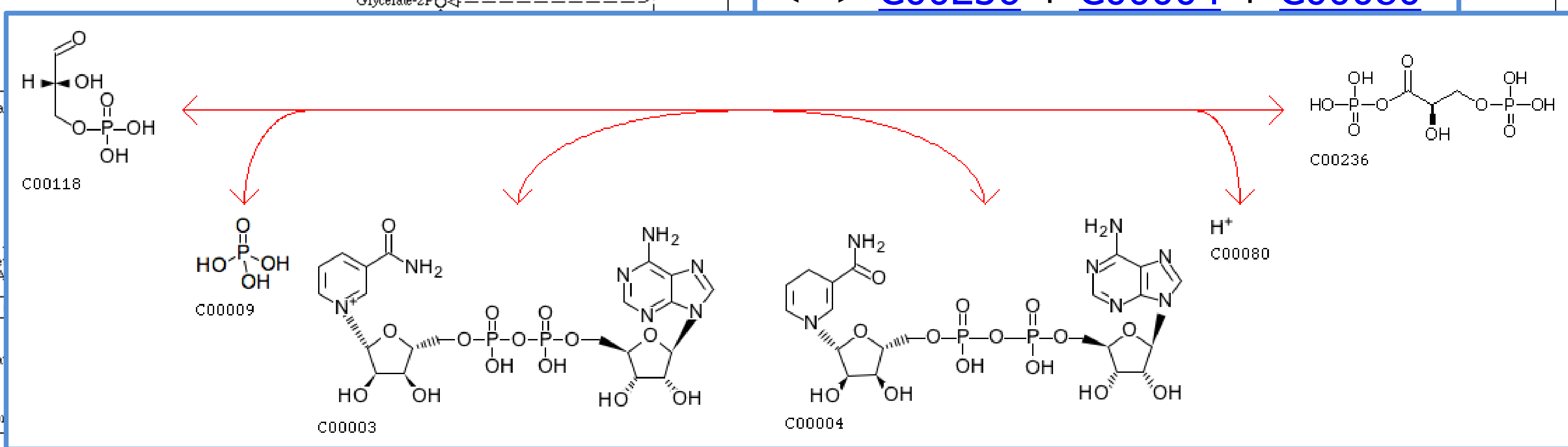


- Each node represents a chemical compound.
- Each edge represents a chemical reaction.

Chemical Equation

D-Glyceraldehyde 3-phosphate
+ Orthophosphate + NAD^+
<=>
3-Phospho-D-glyceroyl phosphate
+ NADH + H^+

C00118 + C00009 + C00003
<=> C00236 + C00004 + C00080



Previous Approaches

- A method which uses Weighted Max-SAT problem to find a steady states of a given pathway.
  - A. Tiwari, C. Talcott, M. Knapp, P. Lincoln, and K. Laderoute, "Analyzing Pathways using SAT-based Approaches", AB 2007.
- Pathway complement with Answer Set Programming
  - O. Ray and K. Whelan and R. King, "A nonmonotonic logical approach for modelling and revising metabolic network", CISIS, 2008.
  - T. Schaub and S. Thiele, "Metabolic Network Expansion with Answer Set Programming", ICLP 2009.
- Pathfinding method with a graph search.
  - D. Croes, F. Couche, S. J. Wodak and J. V. Helden, "Inferring Meaninguful Pathways in Weighted Metabolic Networks", J. Molecualr Biology, 2006.

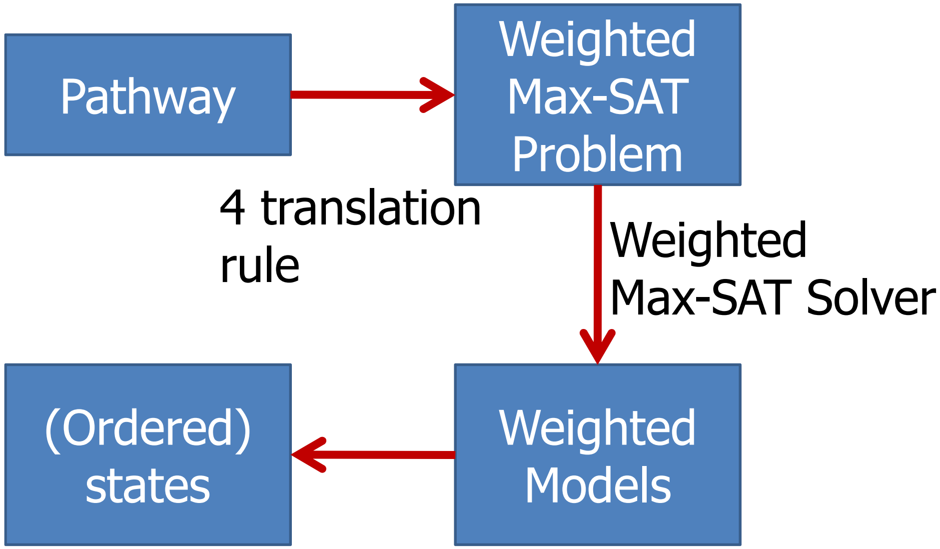
Previous Approaches

- A method which uses Weighted Max-SAT problem to find a steady states of a given pathway.
  - A. Tiwari, C. Talcott, M. Knapp, P. Lincoln, and K. Laderoute, "Analyzing Pathways using SAT-based Approaches", AB 2007.

Goal

- To find steady states which consume input compounds and provide target compounds.

Tiwari et al. 2007



[Tiwari *et al.* 2007]

- Each propositional variable represents whether the reaction is "activated" or "inactivated".
- Translate a given pathway into a Weighted Max-SAT problem.
- All translation rules are given by soft constraints. Sometimes it makes wrong ordering of outputs.

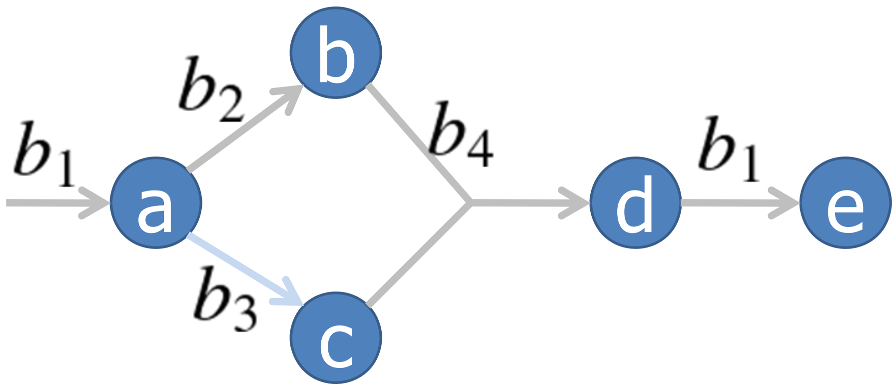
Tiwari et al. 2007

- Models of a propositional formula can be represented by a set of propositional variables.
- Each model is represented by the set of propositional variables to which it assigns true.

- Each propositional variable represents whether the reaction is "activated" or "inactivated".
- Translate a given pathway into a Weighted Max-SAT problem.
- All translation rules are given by soft constraints. Sometimes it makes wrong ordering of output paths.

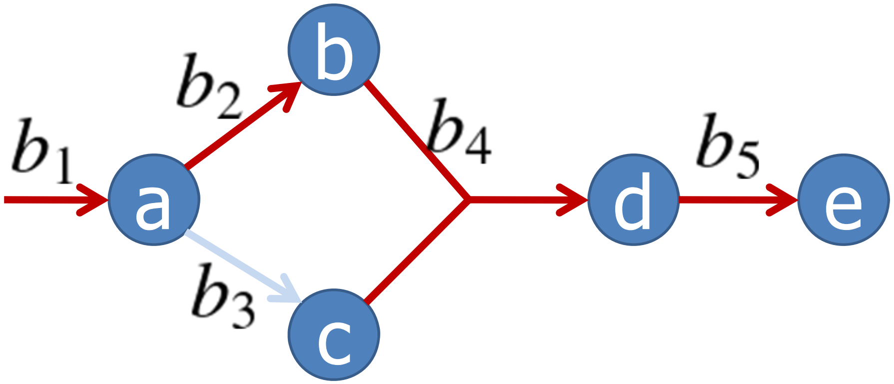
The Method of Tiwari *et al.*

Example



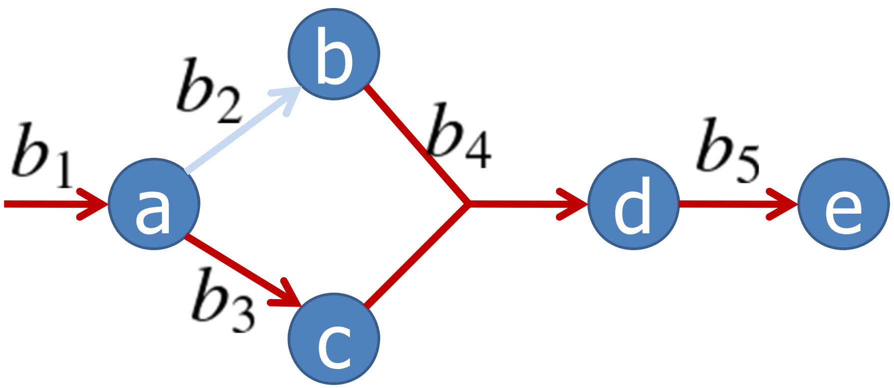
The Method of Tiwari *et al.*

- Obtained path:
  - b_1, b_2, b_4, b_5 (74)
  - b_1, b_3, b_4, b_5 (74)
  - b_1, b_2, b_5 (73)
  - b_1, b_3, b_5 (73)
  - b_1, b_2, b_3, b_4, b_5 (70)



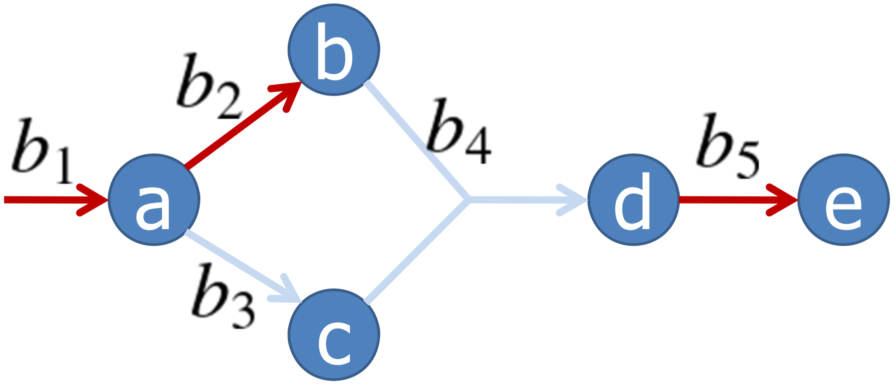
The Method of Tiwari *et al.*

- Obtained path:
  - b_1, b_2, b_4, b_5 (74)
  - b_1, b_3, b_4, b_5 (74)
  - b_1, b_2, b_5 (73)
  - b_1, b_3, b_5 (73)
  - b_1, b_2, b_3, b_4, b_5 (70)



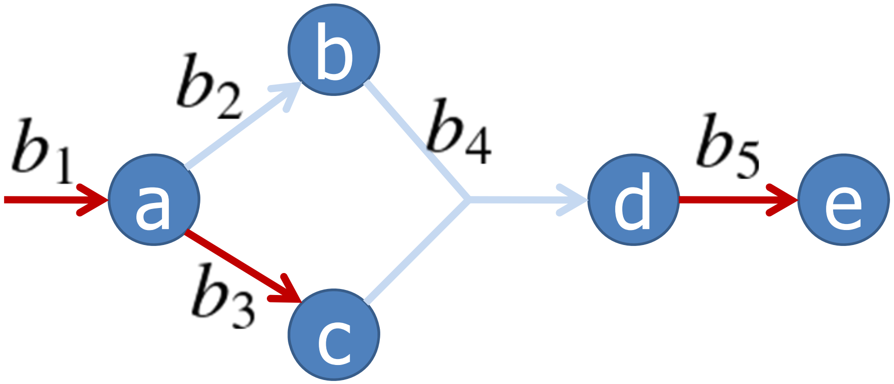
The Method of Tiwari *et al.*

- Obtained path:
  - b_1, b_2, b_4, b_5 (74)
  - b_1, b_3, b_4, b_5 (74)
  - b_1, b_2, b_5 (73)
  - b_1, b_3, b_5 (73)
  - b_1, b_2, b_3, b_4, b_5 (70)



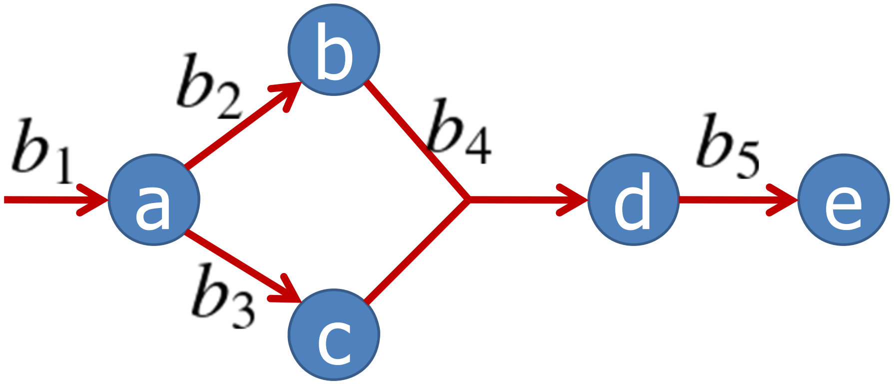
The Method of Tiwari *et al.*

- Obtained path:
  - b_1, b_2, b_4, b_5 (74)
  - b_1, b_3, b_4, b_5 (74)
  - b_1, b_2, b_5 (73)
  - b_1, b_3, b_5 (73)
  - b_1, b_2, b_3, b_4, b_5 (70)



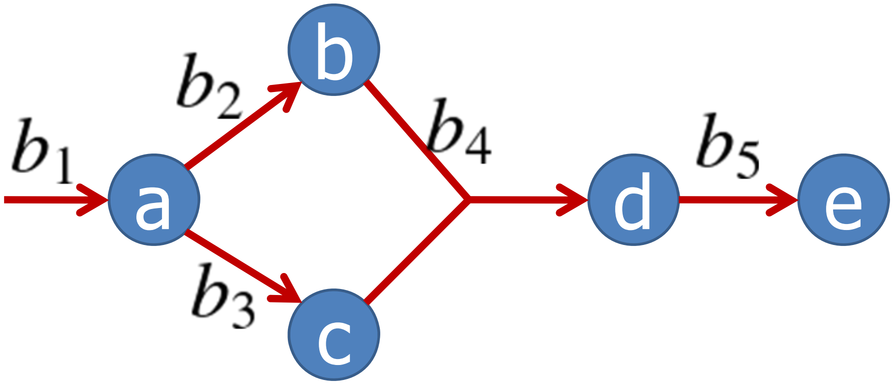
The Method of Tiwari *et al.*

- Obtained path:
  - b_1, b_2, b_4, b_5 (74)
  - b_1, b_3, b_4, b_5 (74)
  - b_1, b_2, b_5 (73)
  - b_1, b_3, b_5 (73)
  - b_1, b_2, b_3, b_4, b_5 (70)



Proposal

- Obtained Model
  - b_1, b_2, b_3, b_4, b_5



Preliminaries

- If a reaction $r_i \in \mathbf{R}$ is activated (or inactivated), then propositional variable $b_i \in \mathbf{B}$ is *True* (or *False*).
- $R(r_i)$ is a mapping from a reaction r_i to the set of chemical compounds which are needed to activate the reaction r_i.
- $P(r_i)$ is a mapping from a reaction r_i to the set of chemical compounds which are provided by the reaction r_i.
- $R^{-1}(s)$ is a mapping from a chemical compound **s** to the set of reactions which consume the chemical compound **s** as a reactant.
- $P^{-1}(s)$ is a mapping from a chemical compound **s**, to the set of reactions which provide the chemical compound **s** as a product.
- Let **F** be a mapping such that $F : \mathbf{B} \rightarrow \{true, false\}$.

Rules for the Translation

- **F** is called a valid assignment if it satisfies the following Rules:

- Rule 1

$$b_i \Rightarrow \bigwedge_{s \in R(r_i)} \bigvee_{r_j \in P^{-1}(s)} b_j$$

- Rule 2

$$\bigwedge_{s \in R(r_i)} \left(\bigvee_{r_j \in P^{-1}(s)} b_j \wedge \bigwedge_{r_j \in R^{-1}(s), j \neq i} \neg b_j \right) \Rightarrow b_i$$

- Input compound
- Target compound

Finding a steady state on a given pathway with SAT

- Let $r_{in} \in \mathbf{R}$ be a reaction which consumes the input compound, $r_{out} \in \mathbf{R}$ be a reaction which provide the target compound.
- Let Ψ be a conjunction of a formula obtained by the rule 1, 2 and an unit clause b_{in}.

$$\Psi = \Psi_1 \wedge \Psi_2 \wedge b_{in}$$

- We can find the steady state which consume b_{in} and provide b_{out} by the following procedure:
  1. To find minimal models which satisfy the formula Ψ.
  2. To pick up the model which include the propositional variable which corresponds to b_{out}.

Finding minimal models with SAT solver [Koshimura *et al.* 2009]

- Models of a propositional formula can be represented by a set of propositional variables.
- Each model is represented by the set of propositional variables to which it assigns true.

Definition 1

Let P, M_1 and M_2 be sets of propositional variables. Then, M_1 is said to be smaller than M_2 with respect to P if $M_1 \cap P$ is a proper subset of $M_2 \cap P$.

Definition 2

Let A be a propositional formula, P be a set of propositional variables, and M be a model of A. Then, M is said to be a minimal model of A with respect to P when there is no model smaller than M with respect to P

Finding a steady state on a given pathway with SAT

- Let r_{in} be a reaction which consumes the input compound, r_{out} be a reaction which provide the target compound.
- Let Ψ be a conjunction of a formula obtained by the rule 1, 2 and an unit clause b_{in}.

$$\Psi = \Psi_1 \wedge \Psi_2 \wedge b_{in}$$

- We can find the steady state which consume b_{in} and provide b_{out} by the following procedure:
  1. To find minimal models which satisfy the formula Ψ.
  2. To pick up the model which include the propositional variable which corresponds to b_{out}.

Finding minimal models with SAT solver [Koshimura *et al.* 2009]

Theorem 1

Let A be a propositional formula, P be an atom set, and M be a model of A. Then, M is a minimal model of A with respect to P *iff* a formula:

$$A \wedge \underbrace{\neg(a_1 \wedge \ldots \wedge a_m)}_{F_1} \wedge \underbrace{\neg b_1 \wedge \ldots \wedge \neg b_n}_{F_2}$$

is unsatifiable,
where $\{a_1, \ldots, a_m\} = M \cap P$, $\{b_1, \ldots, b_n\} = \overline{M} \cap P$.

Ex1.

$P=\{p_1, p_2, p_3, p_4\}$, $M=\{p_1, p_4, c, d\}$ are given,
$A \wedge \neg(p_1 \wedge p_4) \wedge \neg p_2 \wedge \neg p_3$

Procedure for finding a objective model.

begin

  while (`solve`(Ψ))

    $\Psi_{verify} = \Psi \wedge F_1 \wedge F_2;$

    if (!`solve`(Ψ_{verify}))

      if ($\{b_{out}\} \subset M$)

        Found objective model

      end

    end if

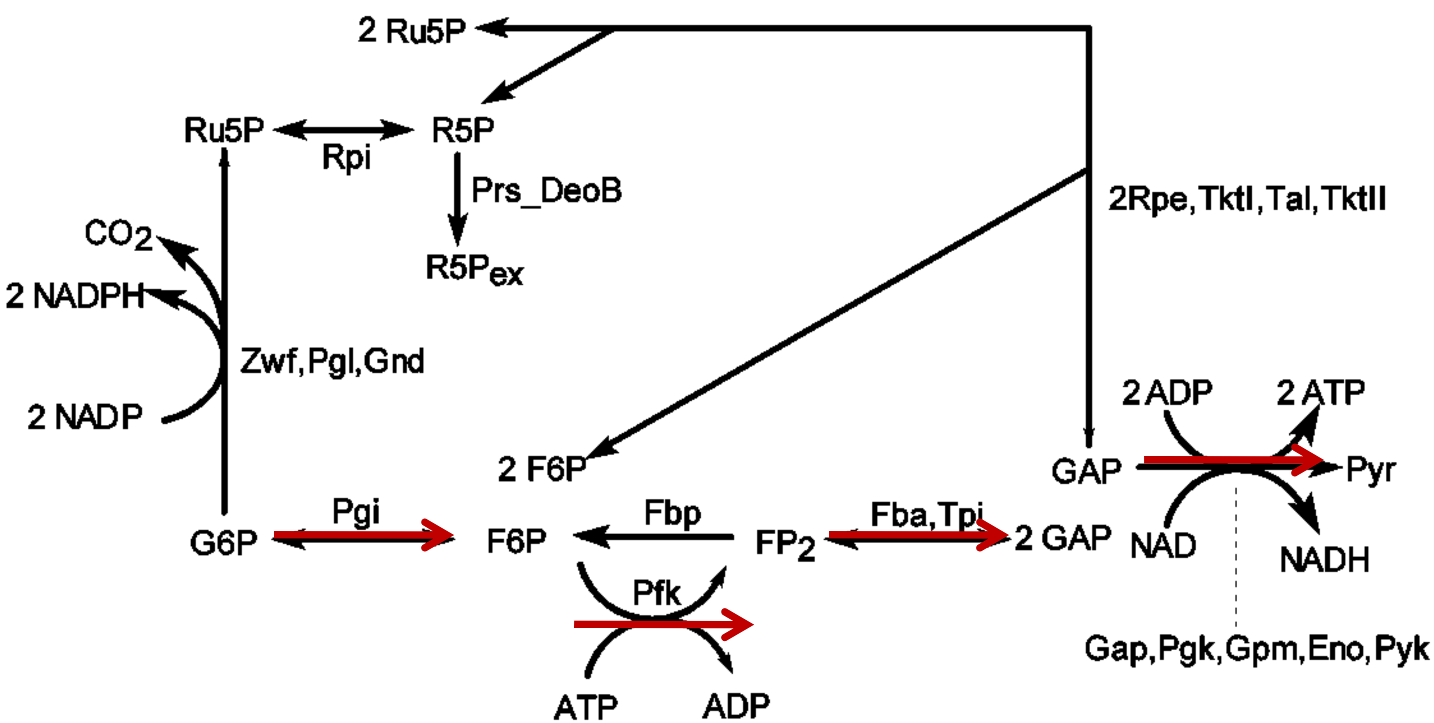
    $\Psi = \Psi \wedge F_1$

  end while

end

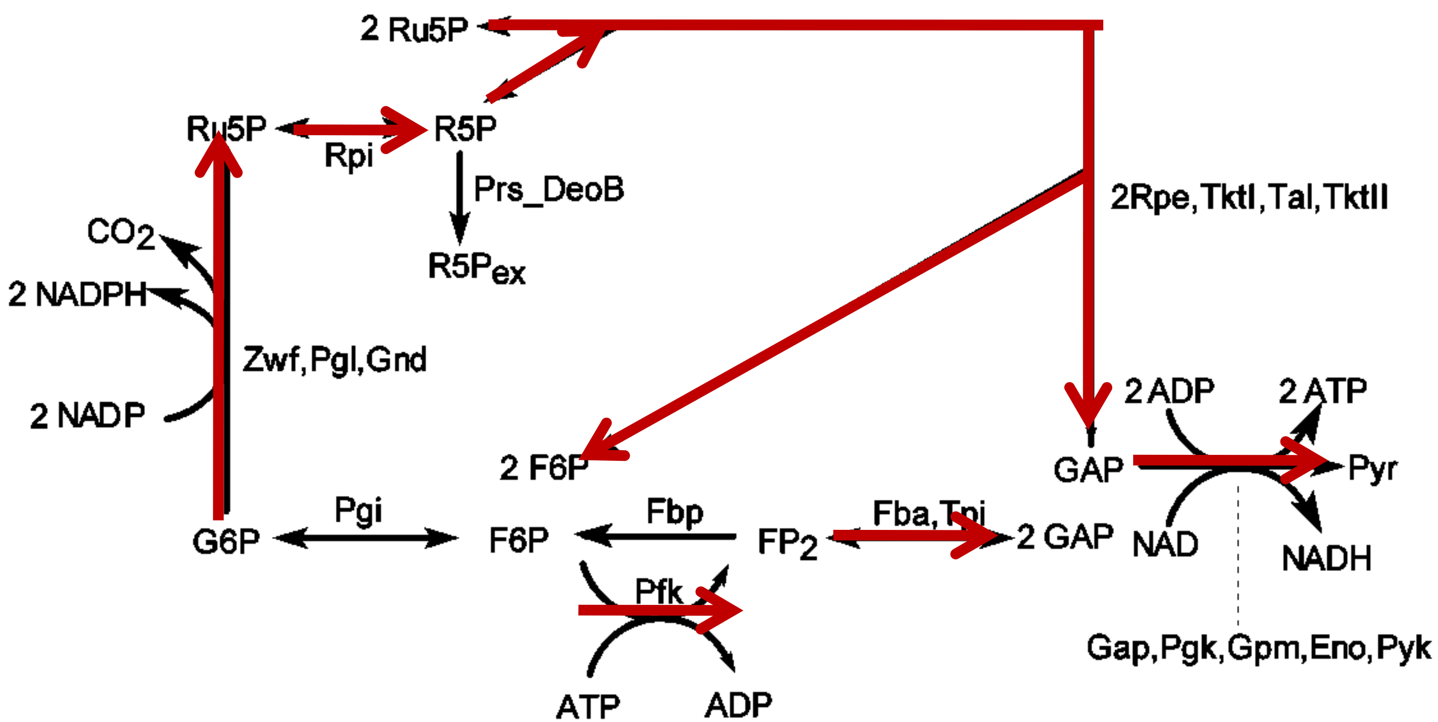
Experiment (1)

- Our method generate 3 steady states which correspond to the result by [Schuster *et al.*, 2000].



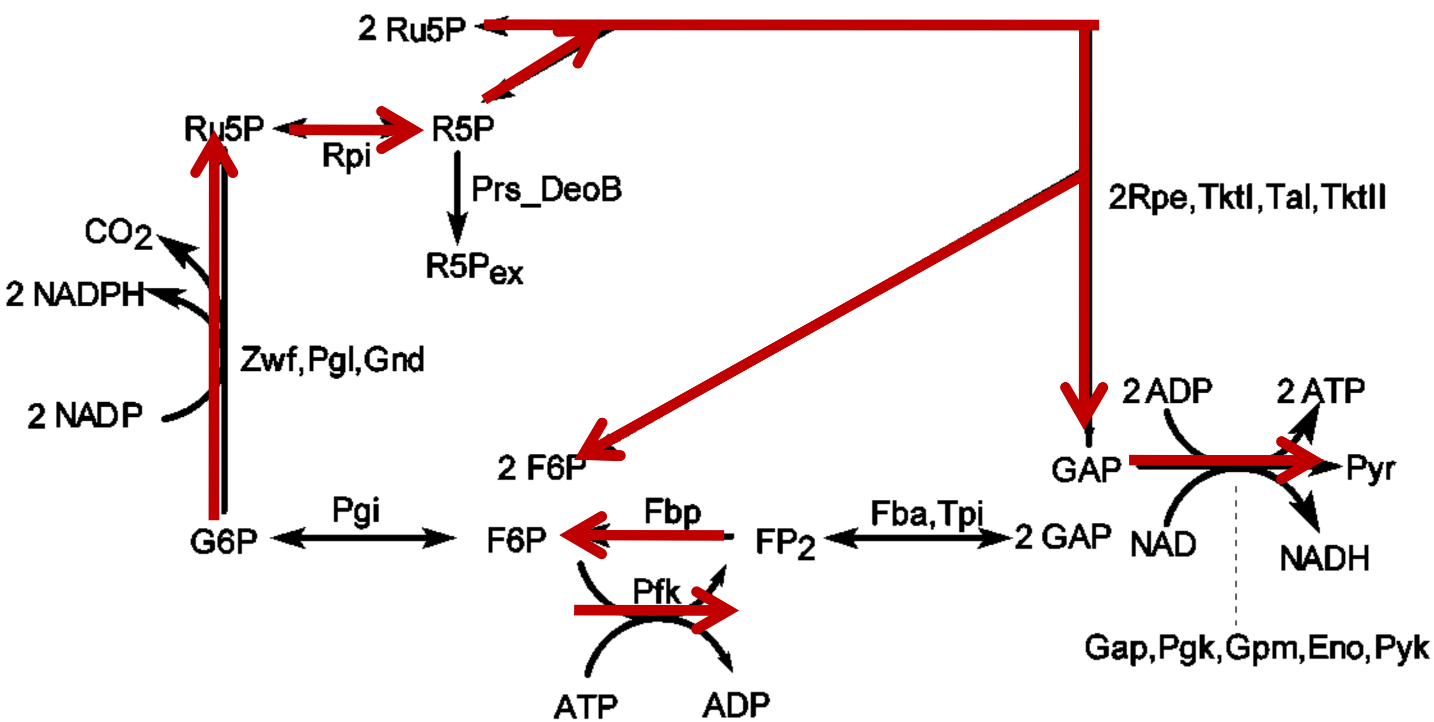
Experiment (2)

- Our method generate 3 steady states which correspond to the result by [Schuster *et al.,* 2000].



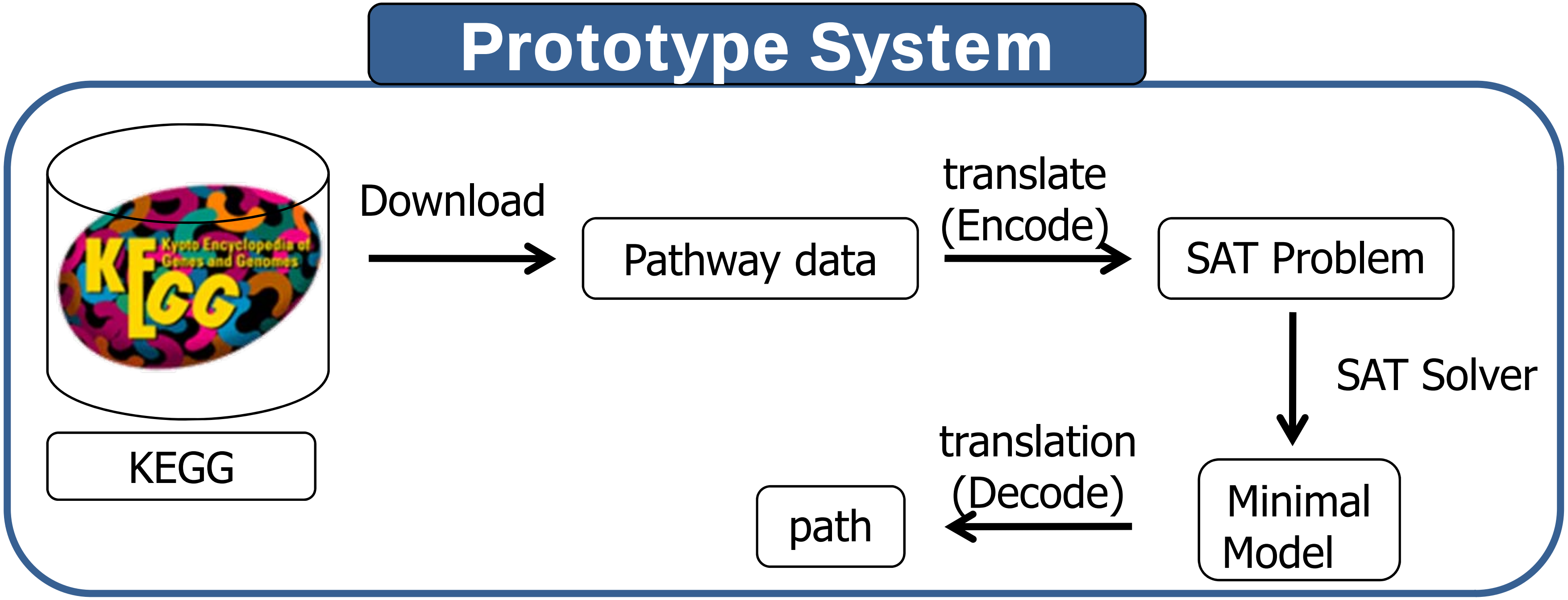
Experiment (3)

- Our method generate 3 steady states which correspond to the result by [Schuster *et al.*, 2000].



Experiment (4)

- Combining the method to the KEGG databese.
  - To evaluate our method with a more large network.
- Result
  - Found 4 steady states on “sce00010” which is available on KEGG Database.



Conclusion

- We found a steady state on glycolysis pathway via finding minimal models of the translated SAT problem.
  - One advantage of this method is to be able to flexibly add a biological rules compared to other search methods (ex. graph search).

- To apply this SAT-based method to other pathway analysis such as path-finding, pathway completion.