Kinetic Models and Qualitative Abstraction for Relational Learning in Systems Biology

Gabriel Synnaeve,
Katsumi Inoue, Andrei Doncescu, Taisuke Sato
Yoshitaka Kameya, Masakazu Ishihata, Hidetomo Nabeshima

2009 internship at NII (Tokyo),
now at E-Motion@INRIA / Collège de France (Paris)

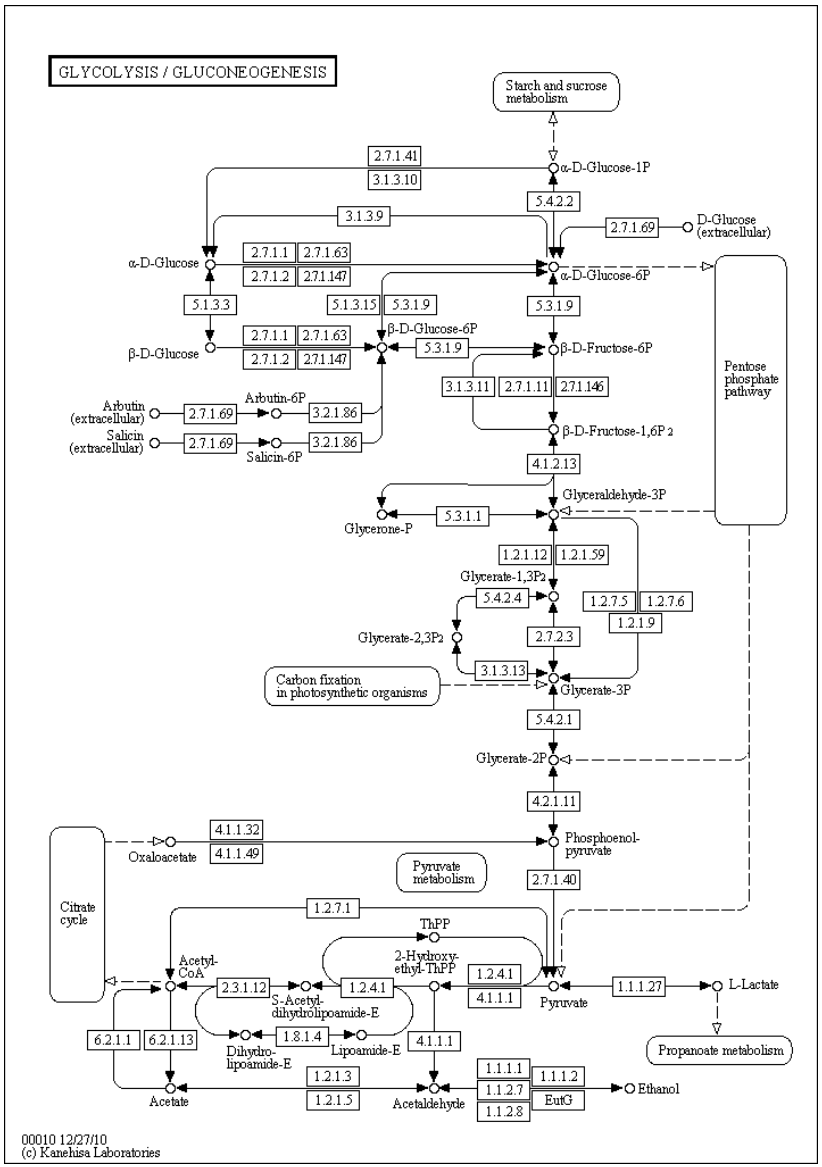
January 27, 2011

Introduction

Problems: **exploit** experimental data, **learn** what rules the cell.

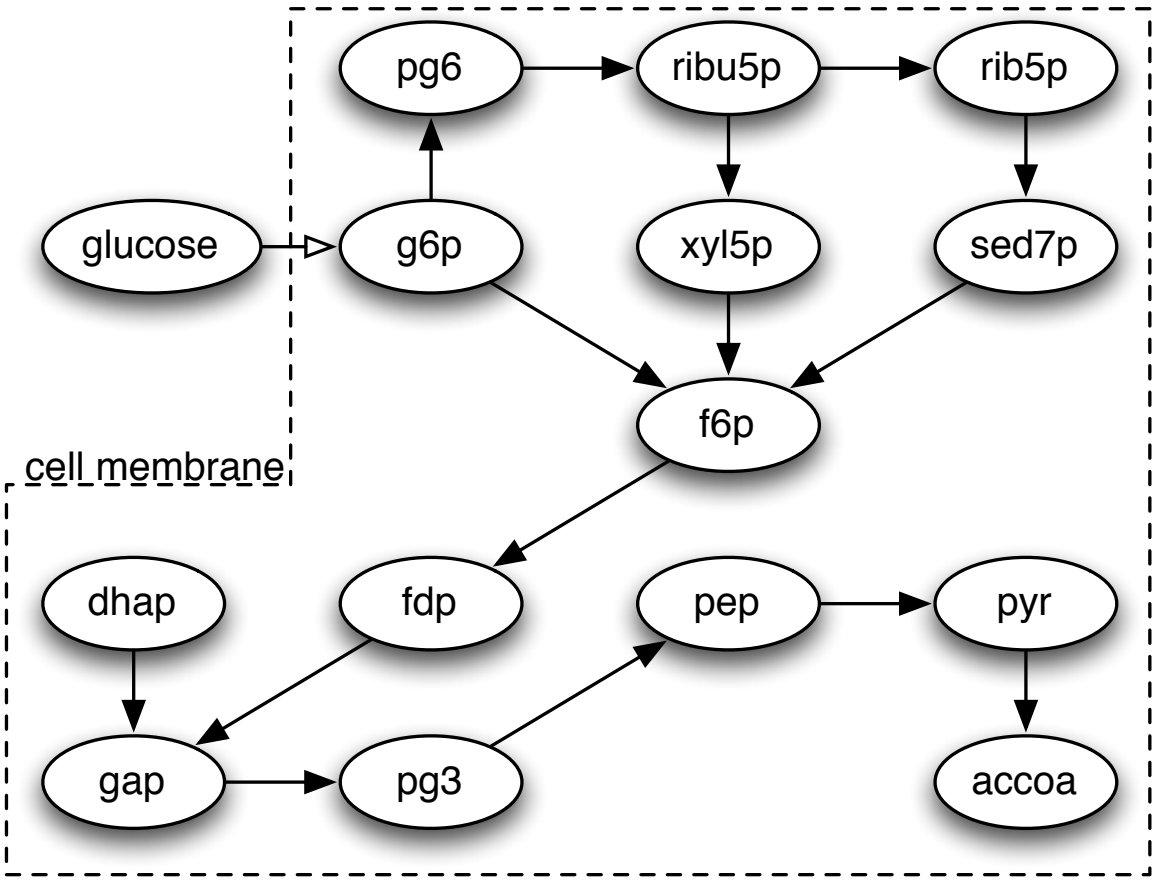
Method: **discretize** metabolites concentration, **combine** with existing pathways structures, use **kinetic models** with **inductive logic programming**.

(Metabolic) Pathways



- Graphs of interconnected reactions
- Glucose enters $ATP \Rightarrow ADP$ G6P
- Chain of reactions to take energy and store it in ATP/NADH (2 per molecule of Glucose)
- Acetyl CoA is at the origin of the Krebs cycle (part of cellular respiration)

Glycolysis and Pentose Phosphate of *E. Coli*



Abduction & Induction

ILP strength lies in the fact that learnt rules/clauses are directly useable in a logical program.

Induction & Abduction

From $\mathcal{B}$ackground Knowledge $\wedge$ $\mathcal{E}$xamples
$\diamond$ Find $\mathcal{H}$ypotheses satisfying $\mathcal{B} \wedge \mathcal{H} \models \mathcal{E}$ and $\mathcal{B} \cup \mathcal{H} \not\models \bot$
Abduction: ground (or $\exists$ quant.) formulaes, direct causes of observations that are called explanations.
Induction: universally ($\forall$) quantified formulaes (small $\mathcal{B}$), more general hypotheses.

 R. J. Mooney: Integrating abduction and induction in machine learning. IJCAI97 Workshop on Abduction and Induction in AI, 37–42 (1997).

 Flach P. A., Kakas A. C.: Abduction and induction: Essays on their relation and integration. Kluwer (2000).

Inverse Entailment (Consequence Finding)

ILP is interested in the formulas derived from $\mathcal{B} \wedge \neg\mathcal{E}$ that are not derived from $\mathcal{B}$ alone.

Inverse Entailment

The previous definition is equivalent to $\mathcal{B} \wedge \neg\mathcal{E} \models \neg\mathcal{H}$ and $\mathcal{B} \not\models \neg\mathcal{H}$.

We can then use a consequence finding procedure (resolution, tableaux) to find $\neg\mathcal{H}$ (SOLAR).


Inoue, K.: Linear resolution for consequence finding. Artificial Intelligence 56:301-353 (1992).


Inoue K.: Induction as consequence finding. Machine Learning, 55:109–135 (2004).


Nabeshima H., Iwanuma K., and Inoue K.: S0LAR: A Consequence Finding System for Advanced Reasoning. TABLEAUX 2003, LNAI, Vol. 2796, pp. 257-263, Springer (2003).

New age began there

The point for automatic **qualitative reasoning** through **ILP** has been made.


King, R., Whelan, K., Jones, F., Reiser, P., Bryant, C., Muggleton, S., Kell, D., and Olivier, S. (2004). Functional genomic hypothesis generation and experimentation by a robot scientist. Nature, 427:247–252.


King, R., Garrett, S., and Coghill, G. (2005). On the use of qualitative reasoning to simulate and identify metabolic pathways. Bioinformatics, 21(9):2017–2026.

Inhibitionary effect of toxins

Metabolic flux analysis through **induction**: rules that explain the concentration changes (up or down) between 2 experiments, with and w/o toxin.

Doncescu, A., Inoue K., Yamamoto Y.: Knowledge Based Discovery in Systems Biology Using CF-Induction. LNCS N.4570, pages 395-404 (2007).

Dealing with kinetics?

Main complete models use **ordinary differential equations**.

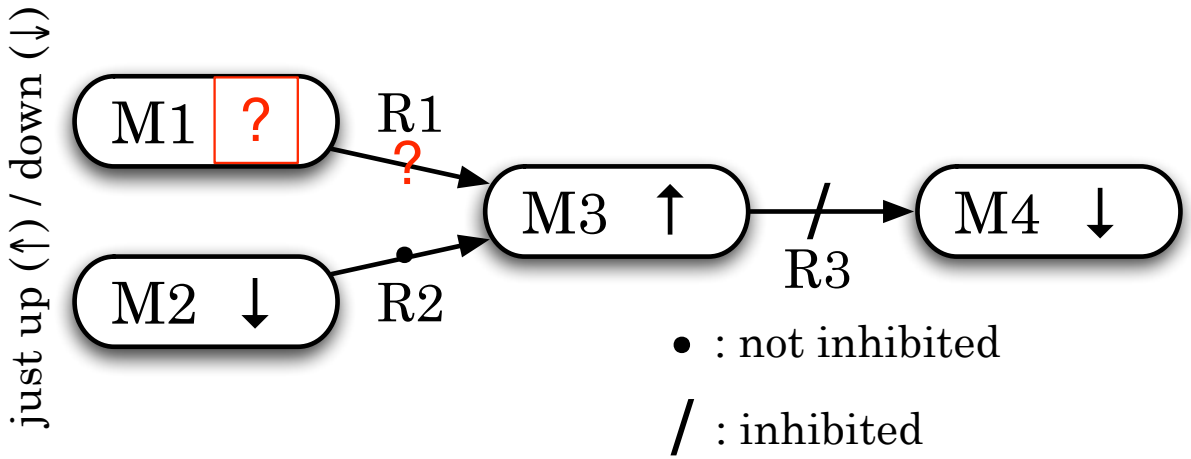
Temporal logic combined with **stochastic logic programming ⇒ kinetic models**.


Franco R. and Canela E.: Computer simulation of purine metabolism. Eu- ropean Journal of Biochemistry, 144:305-315 (1984).

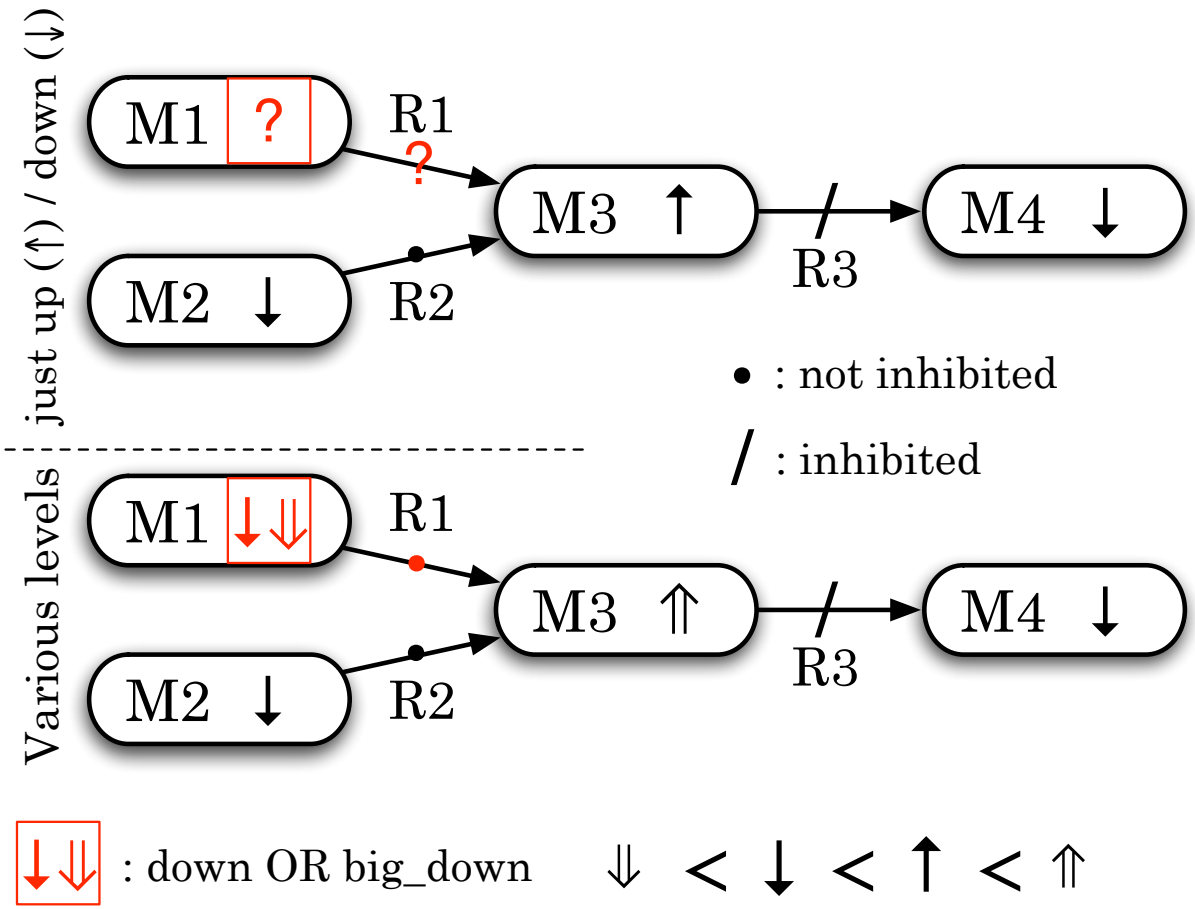

Fages, F., Soliman, S., and France, I. R. (2008). Model revision from temporal logic properties in systems biology. In: Probabilistic Inductive Logic Programming. LNAI, volume 4911, pages 287–304.

Limits of The Previous Models

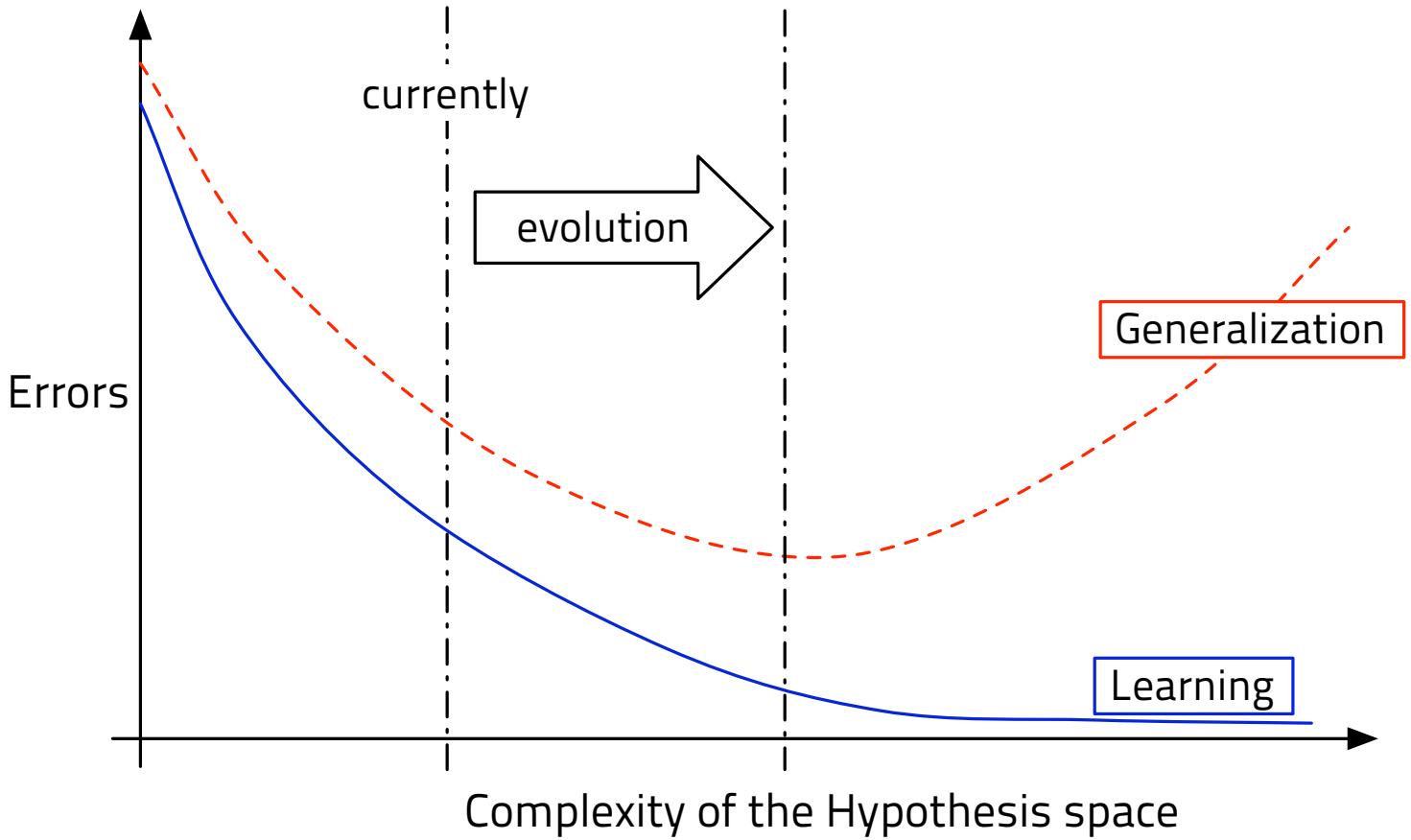
- No models for **dynamic transitions**
- Not enough information to be **precise** enough:



Dealing With More Knowledge



Be More Precise but Avoid Overfitting

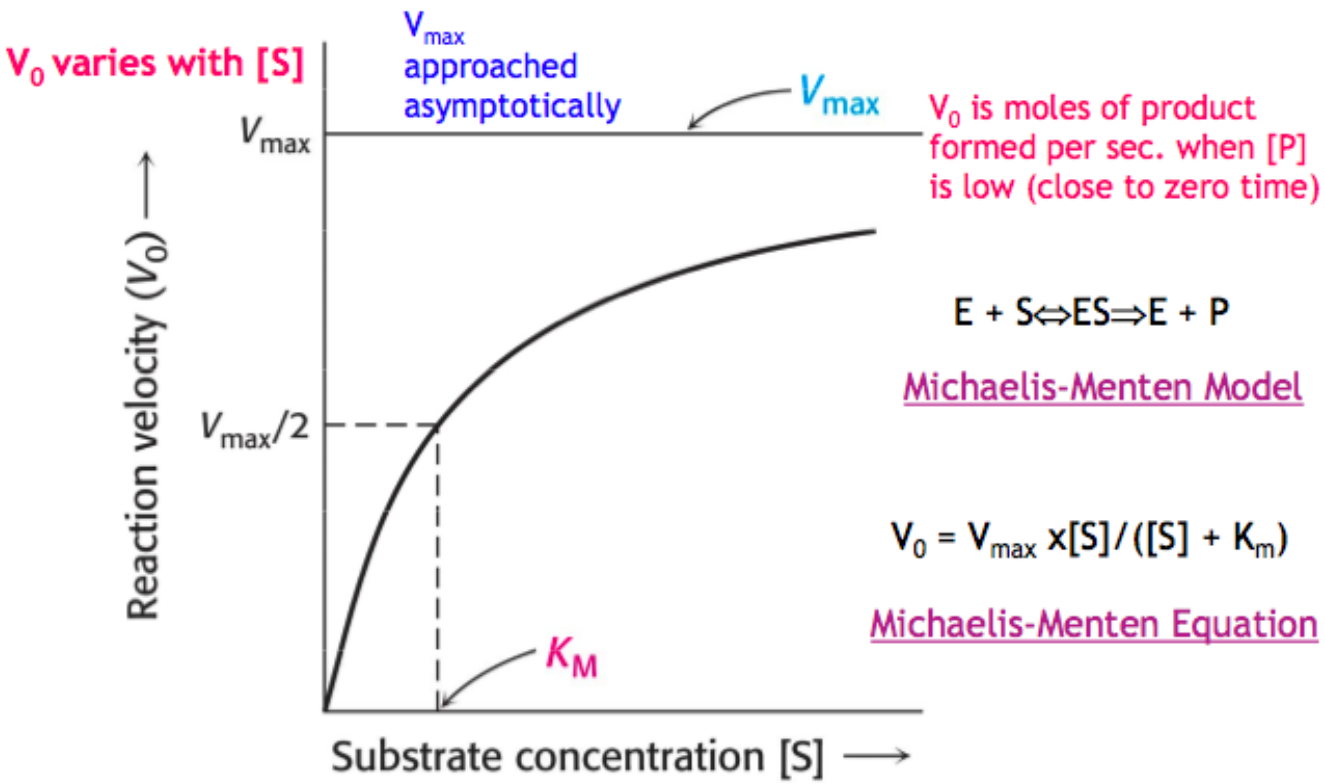


Michaelis-Menten Kinetics

Speed of a one-way reaction

v = #{products per second per mole of the enzyme}

$$sS \rightarrow pP \;\Rightarrow\; v = -\frac{1}{s}\frac{d[S]}{dt} = \frac{1}{p}\frac{d[P]}{dt}$$



Simplification of Michaelis-Menten Equation

$$E + S \underset{k_{-1}}{\overset{k_1}{\leftrightharpoons}} ES \overset{k_2}{\rightarrow} E + P$$

$$\text{Michaelis} - \text{Menten equation} : \frac{d[P]}{dt} = V_m \frac{[S]}{[S] + K_m} \quad (1)$$

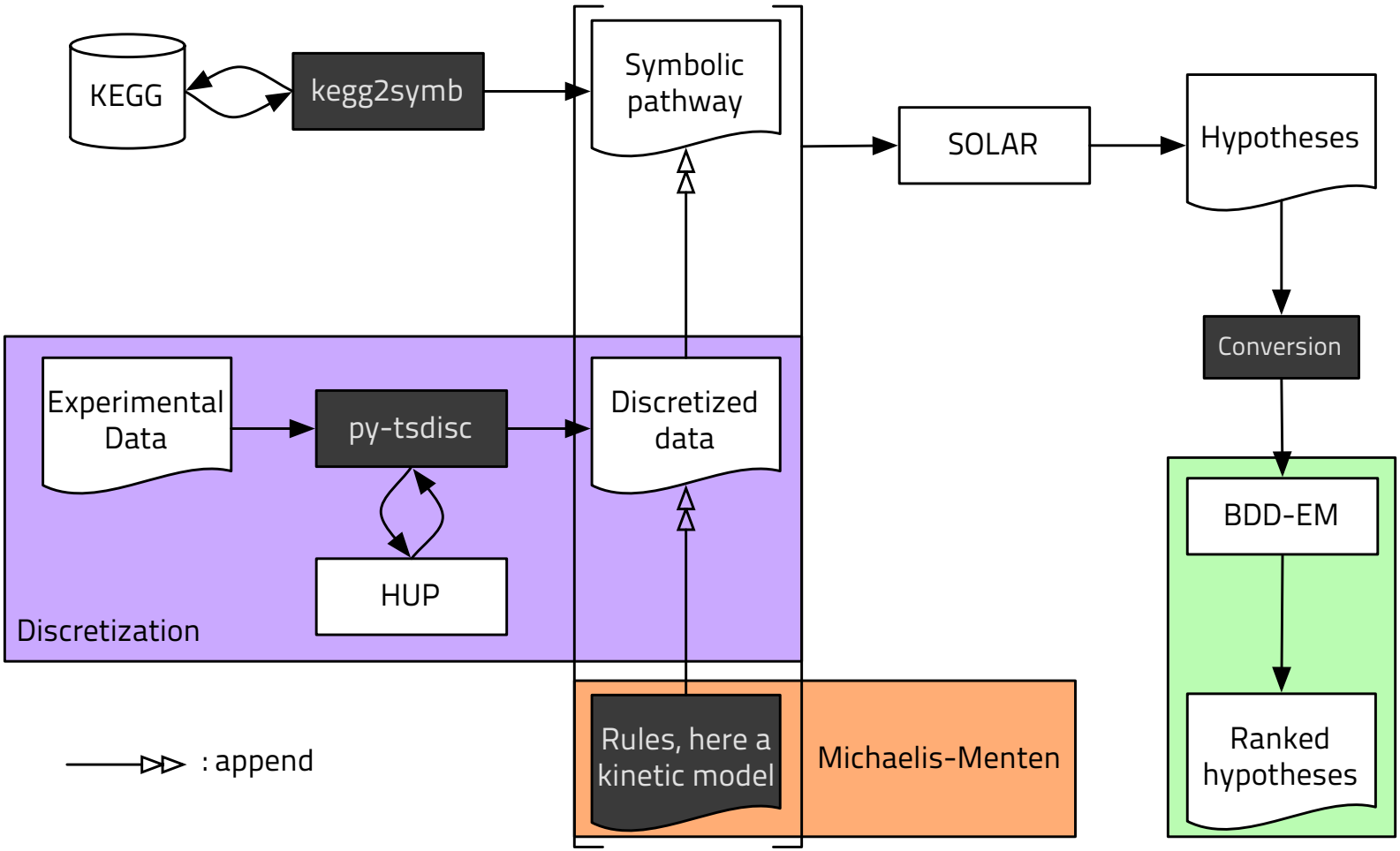
$$\frac{d[P]}{dt} \overset{disc.time}{\longrightarrow} \frac{[P]_{T+timestep} - [P]_T}{(T + timestep) - T} \quad (2)$$

$$(1)\ and\ (2) \implies V_m \frac{[S]_T}{[S]_T + K_m} \approx \frac{[P]_{T+timestep} - [P]_T}{(T + timestep) - T}$$

We chose to work with a constant timestep :

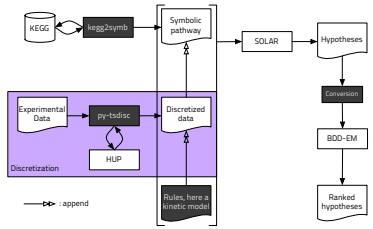
$$\implies [P]_{T+1} = V_m \frac{[S]_T}{[S]_T + K_m} + [P]_T \quad (3)$$

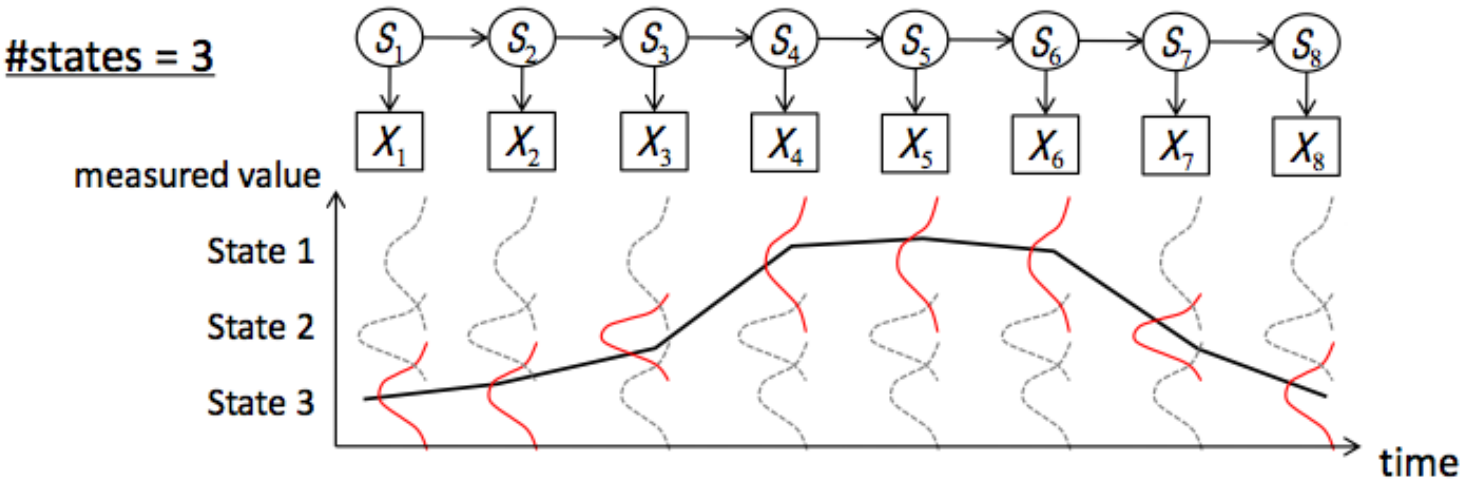
Additional work and tools

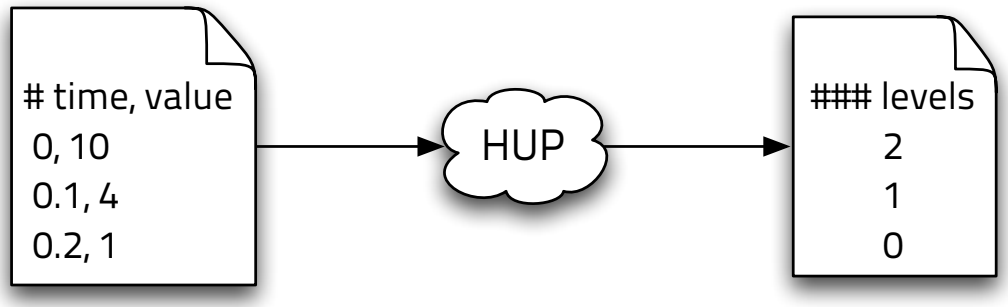


HUP: HMM Utility Program

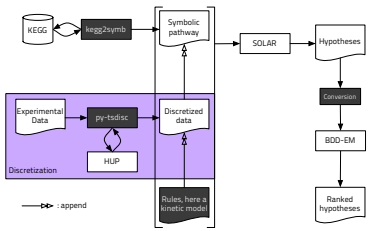
Clustering method that uses Continuous Hidden Markov Model + Bayesian Score:

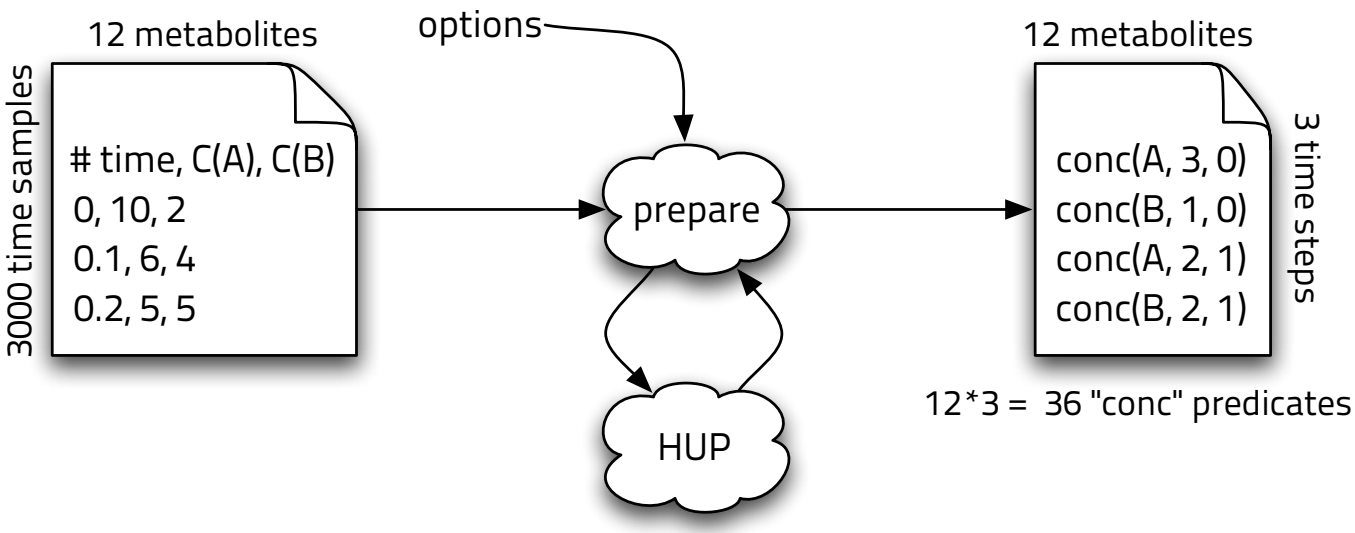




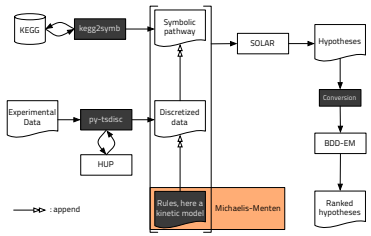


Wrapping HUP





Logical Kinetic Modeling



Approximations for extreme values in:

$$[P]_{T+1} = V_m \frac{[S]_T}{[S]_T + K_m} + [P]_T$$

Logical Kinetic Modeling (rules)

$$\underline{[S] \ll K_m}$$

$$\texttt{reaction(S, P, Km)} \wedge \texttt{concentration(S, 0, 0)} \wedge \texttt{concentration (Km, 2, 0)} \wedge \texttt{concentration(P, L, 0)} \rightarrow \texttt{concentration(P, L, 1)}$$

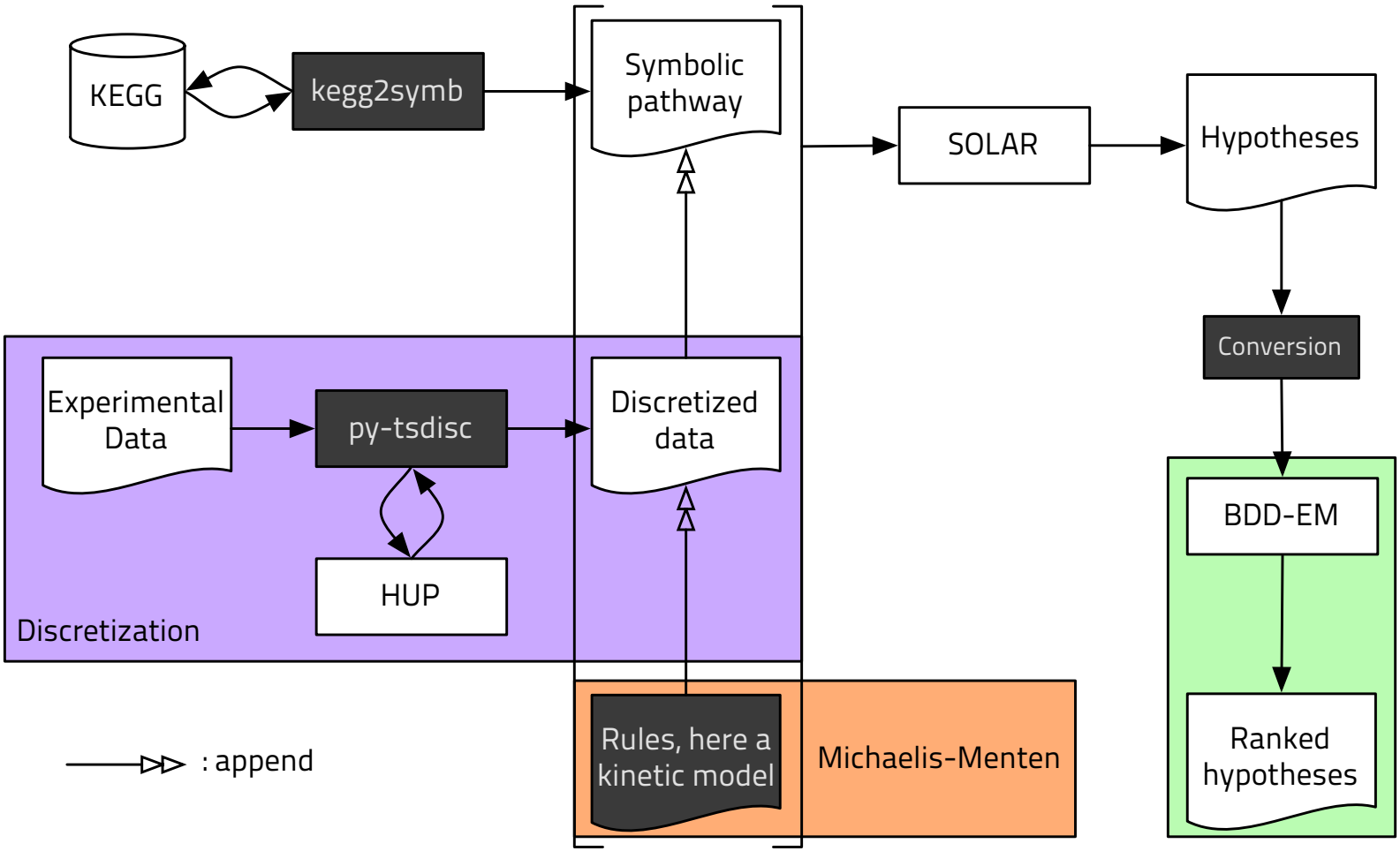
$$\underline{[S] \simeq K_m}$$

$$\texttt{reaction(S, P, Km)} \wedge \texttt{concentration(S, 1, 0)} \wedge \texttt{concentration(Km, 1, 0)} \wedge \texttt{concentration(P, L, 0)} \rightarrow \texttt{concentration(P, L, 1)}$$

$$\underline{[S] \gg K_m}$$

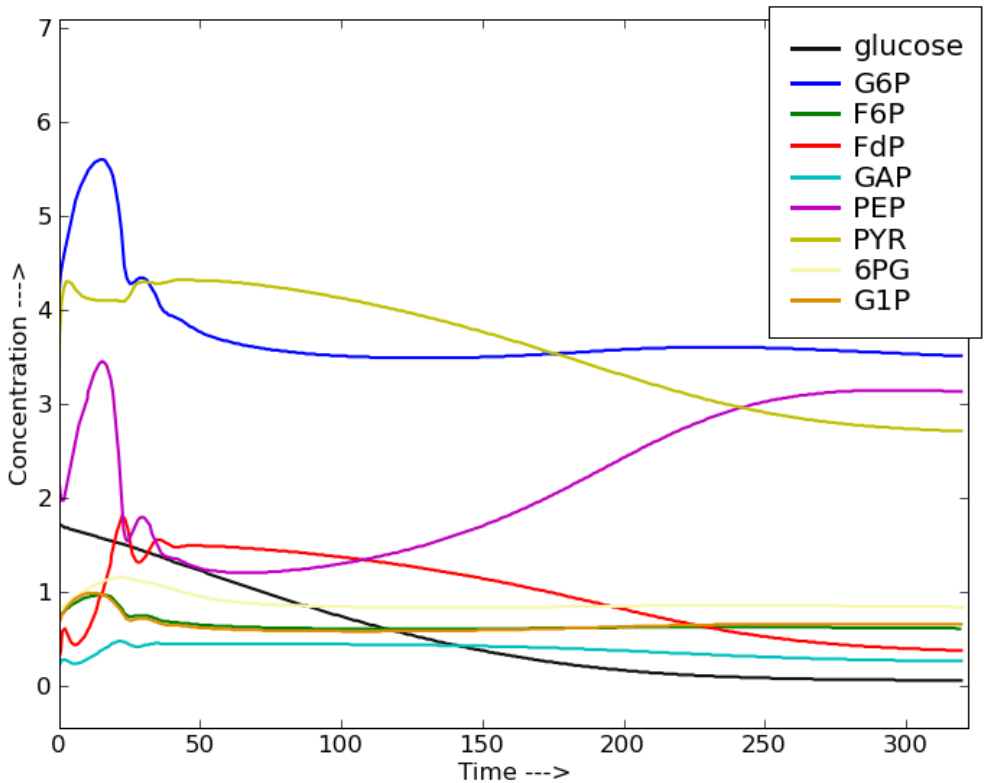
$$\texttt{reaction(S, P, Km)} \wedge \texttt{concentration(S, 2, 0)} \wedge \texttt{concentration(Km, 0, 0)} \rightarrow \texttt{concentration(P, 2, 1)}$$

Data-centric schema



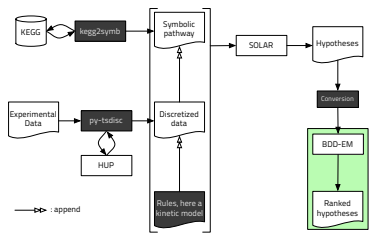
Inputs

Structure of the pathway(s) + Background knowledge (MM) + Metabolites concentrations



Ranking the hypotheses with BDD-EM

Hyp. no.	Probability	Abducted concentrations levels at T=0
H130	≈ 1.0	pg3: 2, adp: 0
H392	$4.879.E^{-1}$	sed7p: 0, e4p: 2, f6p: 0, pg3: 2, adp: 0
H216	$7.567.E^{-2}$	pg3: 2, adp: 0, pep: 0, atp: 2, pyr: 2
H196	$6.930.E^{-2}$	fdp: 0, dhap: 2, gap: 0, pg3: 2, adp: 0
H356	$5.621.E^{-2}$	pg3: 2, adp: 0, g6p: 1, nadph: 1
H94	$3.692.E^{-2}$	sed7p: 0, e4p: 2, f6p: 0, pg3: 2, adp: 0, pep: 0, atp: 2, pyr: 2
H251	$3.497.E^{-2}$	glucose: 2, adp: 0, pg3: 2
H286	$3.382.E^{-2}$	sed7p: 0, e4p: 2, f6p: 0, fdp: 0, dhap: 2, gap: 0, pg3: 2, adp: 0
H405	$2.796.E^{-2}$	pg3: 2, adp: 0, pep: 2, atp: 0
H167	$2.743.E^{-2}$	sed7p: 0, e4p: 2, f6p: 0, pg3: 2, adp: 0, g6p: 1, nadph: 1
.		
.		
H378	$1.974.E^{-8}$	glucose: 2, adp: 0, sed7p: 0, e4p: 2, f6p: 0, fdp: 0, dhap: 2, gap: 0, pg3: 2, pep: 0, atp: 2, pyr: 2, g6p: 0, nadph: 2, pg6: 1



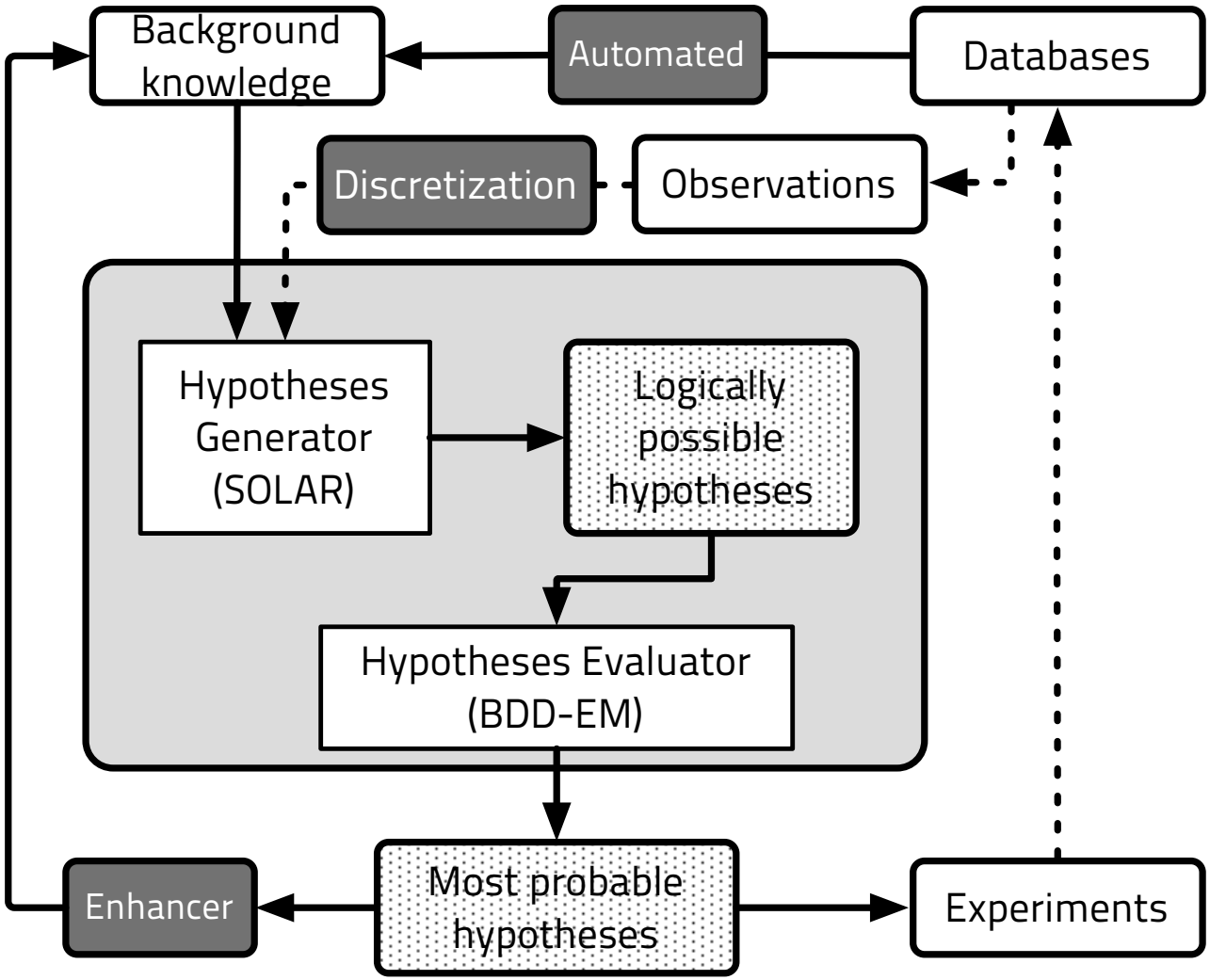
Is is correct?

Ok with: $glucose + 2ADP + 2P + 2NAD^+ \rightarrow$
$2\ pyruvate + 2ATP + 2(NADH, H^+) + 2H_2O$

Agree with:

Peters-Wendisch, P., Schiel, B., Wendisch, V., and et al., E. K. (2001).Pyruvate carboxylase is a major bottleneck for glutamate and lysine production by corynebacterium glutamicum. Molecular Microbiol. Biotechnol., 3(2).

Full system


Background knowledge
Automated
Databases
Discretization
Observations
Hypotheses Generator (SOLAR)
Logically possible hypotheses
Hypotheses Evaluator (BDD-EM)
Enhancer
Most probable hypotheses
Experiments

Future tracks:

- **Enhancing of the knowledge base**, 2 simple and sound algorithms (one in the paper):
  - most probable hypotheses first
  - smallest number of hypothesis additions (biggest abducibles coverage first)
- **Finer discretization**: trivial with our **continuous HMMs with parameter tying**.
- **Automatic** generation of MM rules (for orders > 3).

Conclusion

We presented and validated a **method** and **tools** to work on **real data**.

Working with **other experiments** on **more complex** organisms and pathways (for instance *S. Ce*) will require:

- Enhancing of the KB
- Finer discretization
- Kinetic rules

Thanks

Thank you for your attention.

Any questions?

Any questions?

Me: gabriel.synnaeve@gmail.com
Prof. Inoue: ki@nii.ac.jp