Inductive Logic Programming Applied to Systems Biology

Gabriel Synnaeve

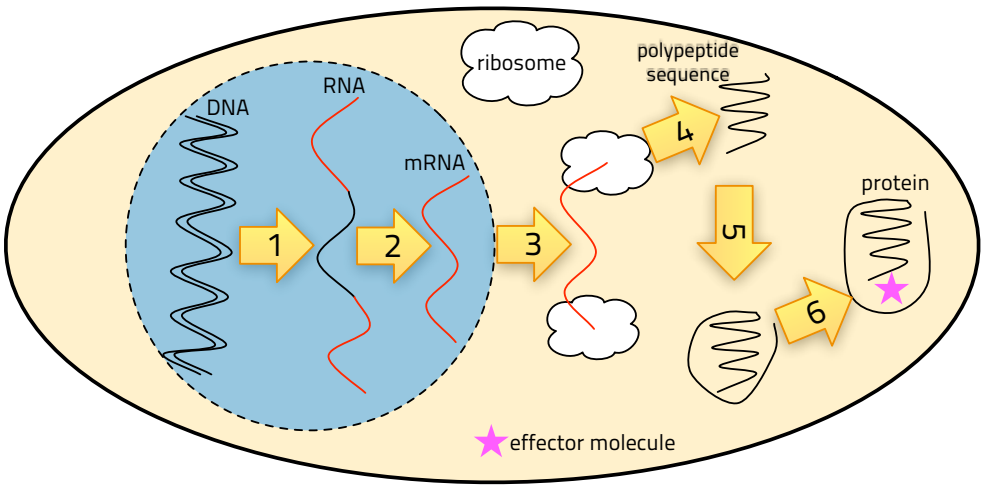
internship at NII, University of Grenoble and ENSIMAG, 20min speech

September 29, 2009

Introduction

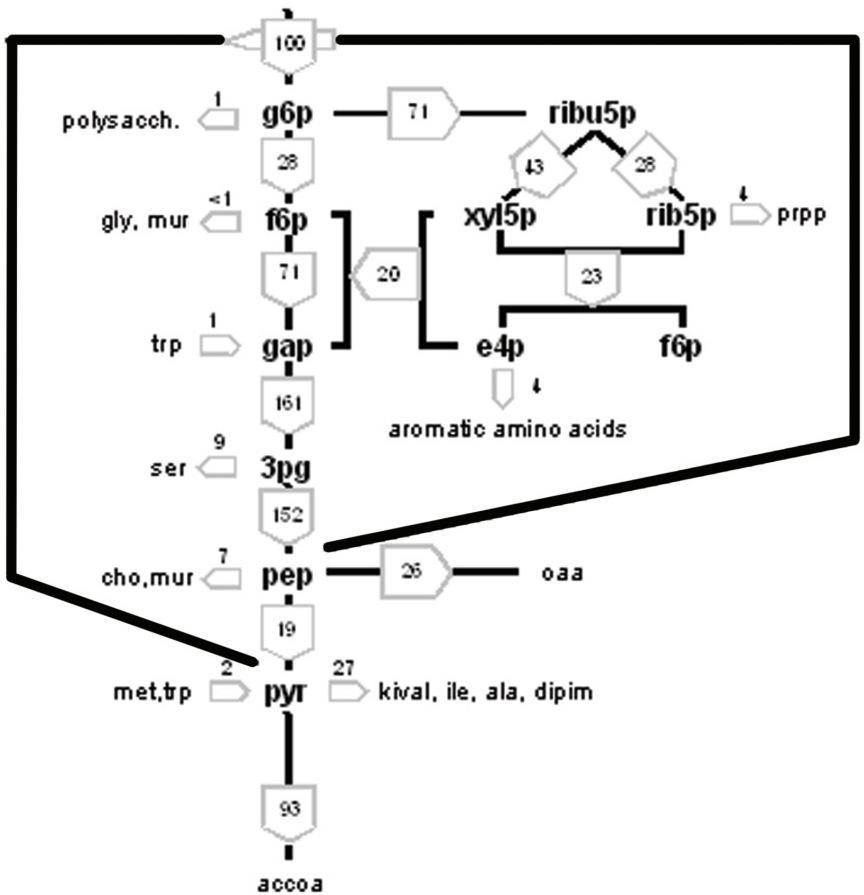
- National Institute of Informatics, Tokyo
- Supervising: Katsumi Inoue, Taisuke Sato, Pierre Bessière, Augustin Lux
- Head (first 2 weeks)
- Tail (last 4 months)

The Cell



1. Transcription
2. Post-transcription
3. Migration of the mRNA
4. Translation
5. Post-transl./Folding
6. Binding of an effector

Metabolic Pathways



A metabolic pathway: Glucolysis and Pentose Phosphate of *Escherichia Coli*

- Graphs of interconnected reactions
- Glucose $\text{enters}_{ATP} \Rightarrow_{ADP}$ 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)

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 $\neg\mathcal{H}$. In Inoue's lab, it is now done with `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.: `SOLAR`: A Consequence Finding System for Advanced Reasoning. TABLEAUX 2003, LNAI, Vol. 2796, pp. 257-263, Springer (2003).

Previous Works

Why? Inhibitionary effects of toxins on chemical reactions.

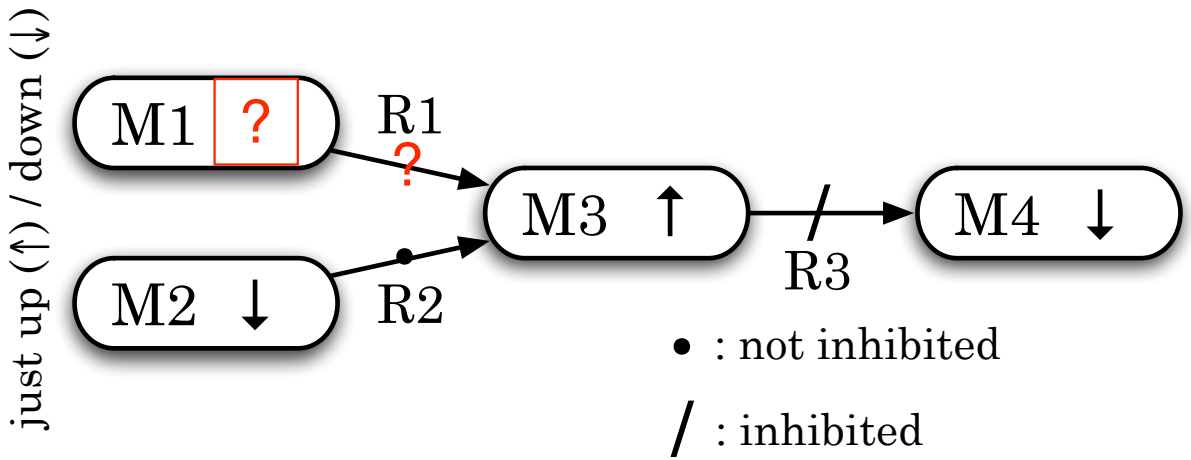
How? Metabolic flux analysis through induction with rules that explain the concentration changes (up or down) between 2 experiments, with and w/o toxin.


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

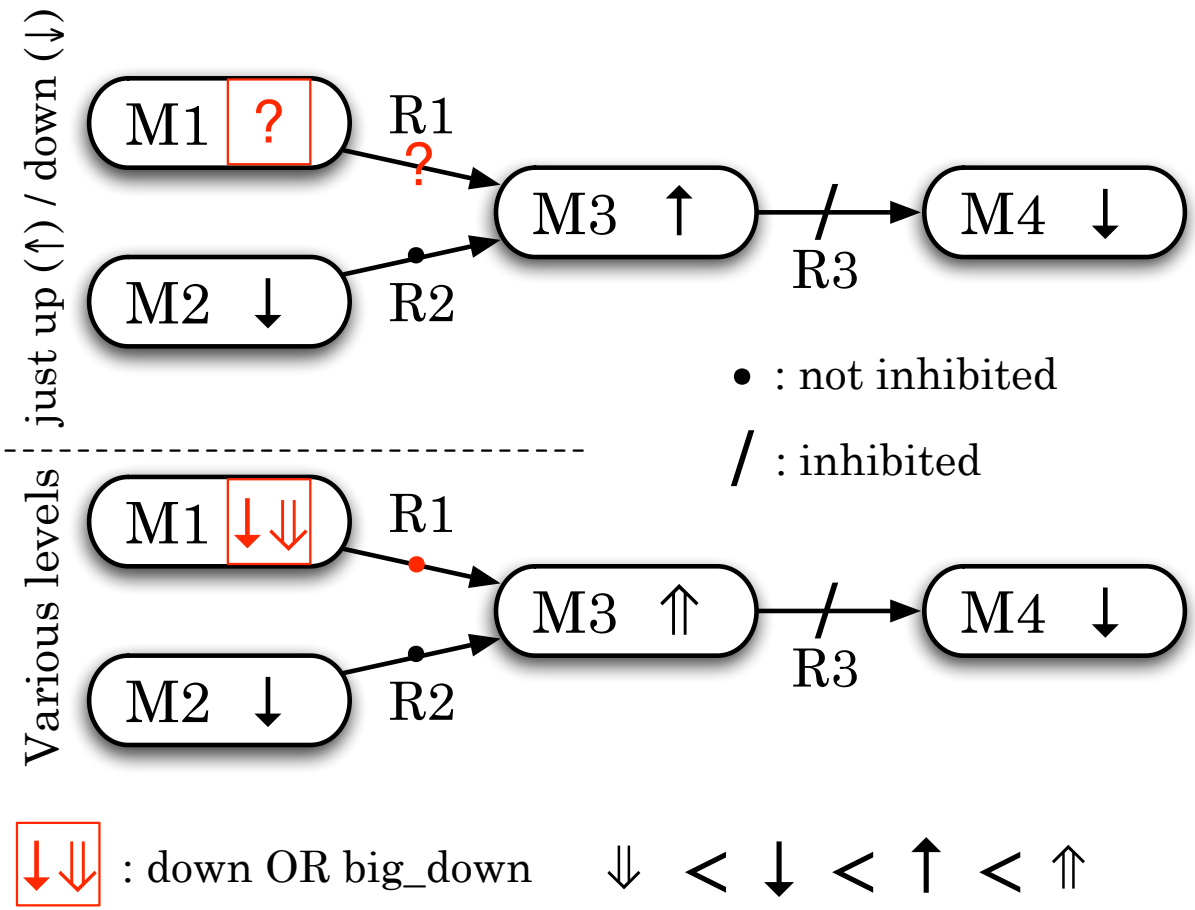

King R.D., Whelan K.E., Jones F.M., Reiser P.G.K., Bryant C.H., Muggleton S.H., Kell D.B. & Oliver S.G.: Functional genomic hypothesis generation and experimentation by a robot scientist. Nature 427, 247-252 (2004).

Limits of The Previous Models

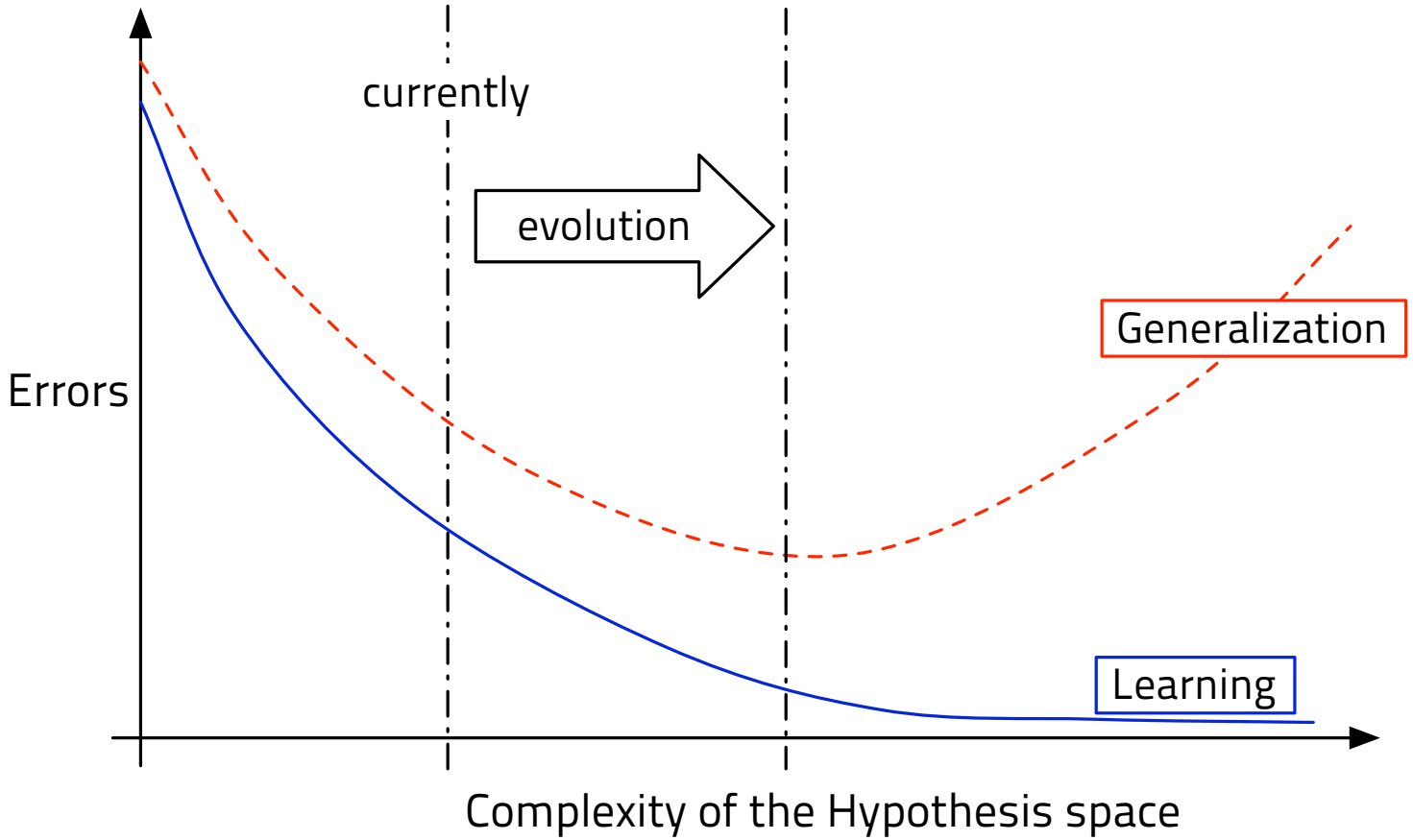
- No models for dynamic transitions
- Not enough information to be precise enough:



Dealing With More Knowledge



Be More Precise, Avoid Overfitting

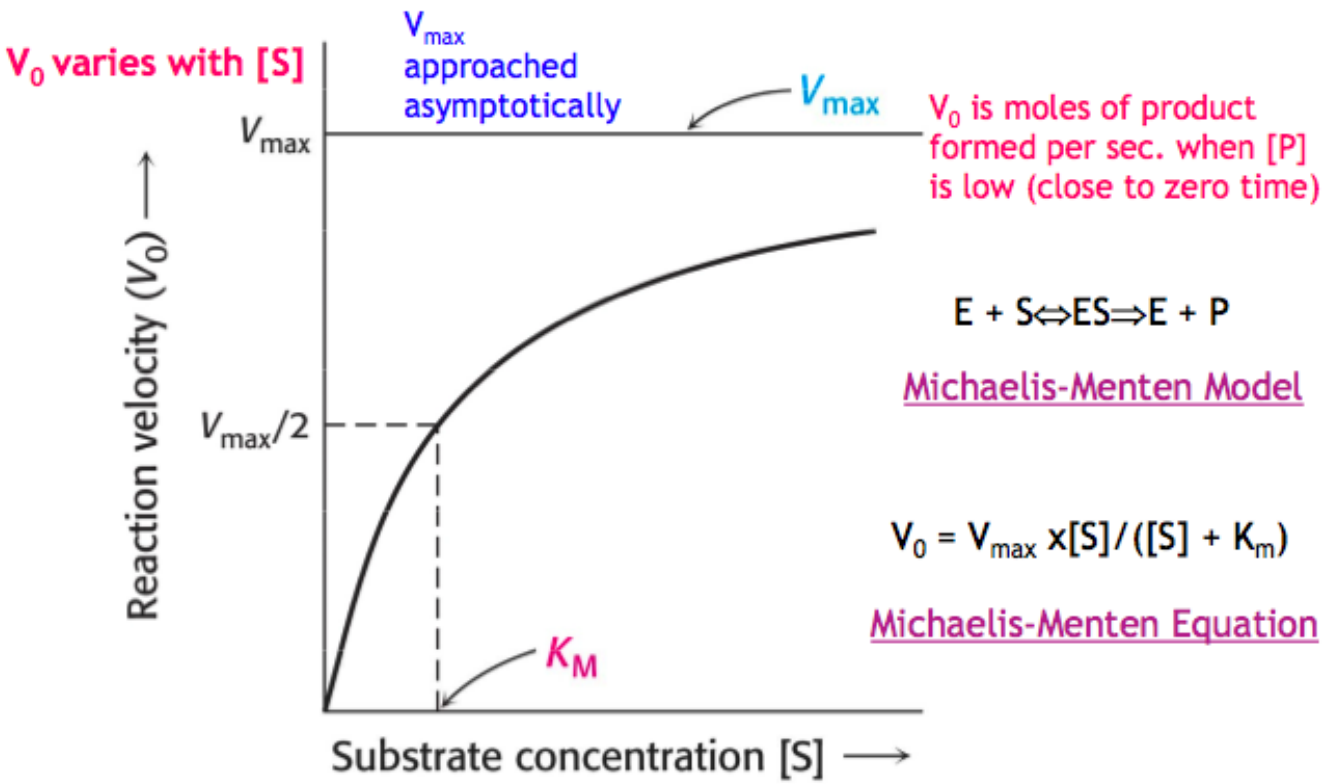


Chemical Kinetic and Michaelis-Menten Equation

Speed of a one-way reaction

v = #{reactions per second catalyzed 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

$$E + S \rightleftharpoons_{k_{-1}}^{k_1} ES \rightarrow^{k_2} E + P$$

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

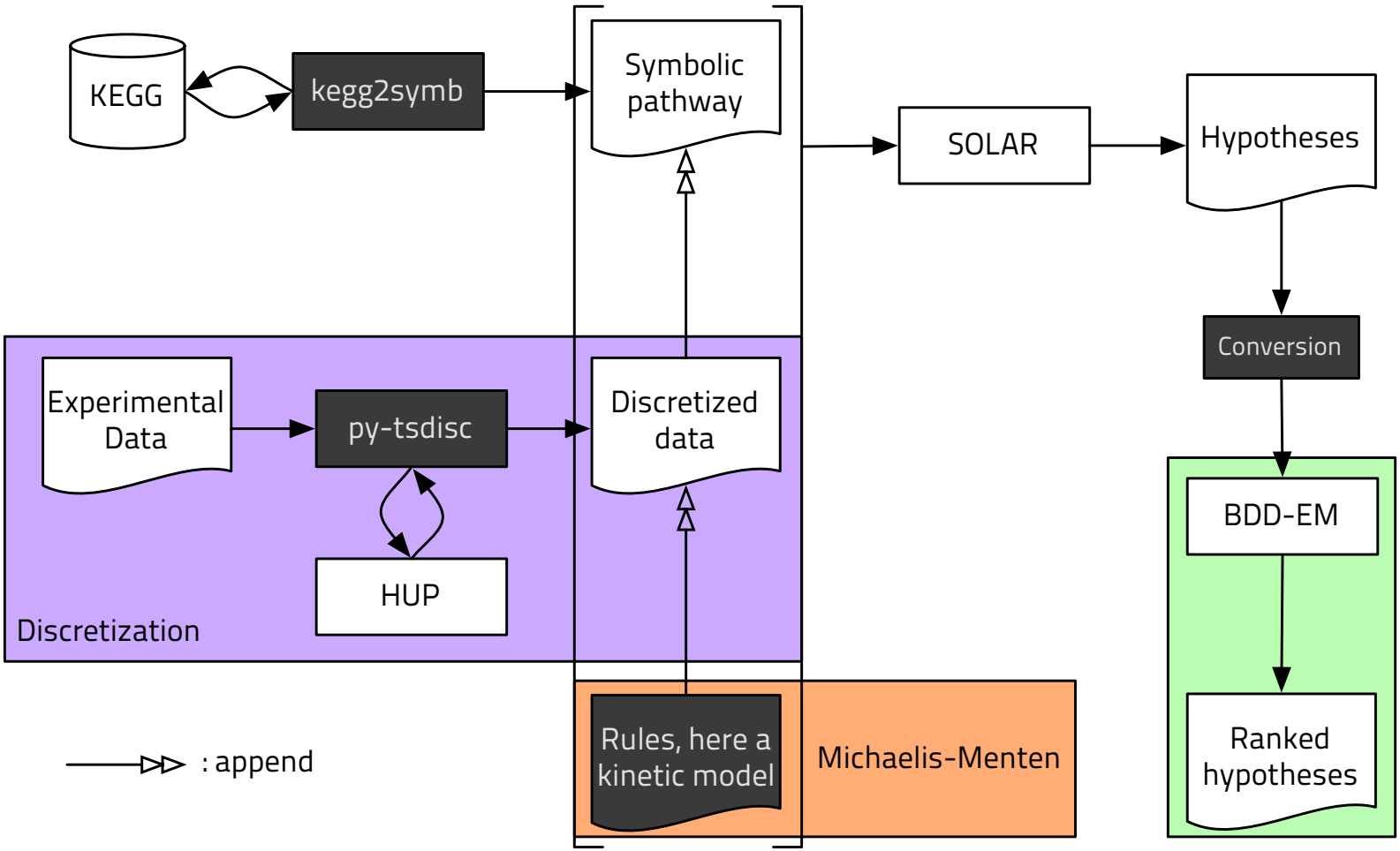
$$\frac{d[P]}{dt} \longrightarrow_{disc.time} \frac{[P]_{T+timestep} - [P]_T}{(T + timestep) - T} \quad (2)$$

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

We chose to work with a constant timestep :

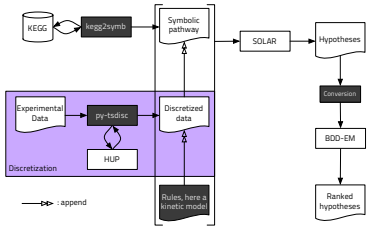
$$\implies [P]_{T+1} = V_m \frac{[S]_T}{[S]_T + K_m} + [P]_T \quad (3)$$

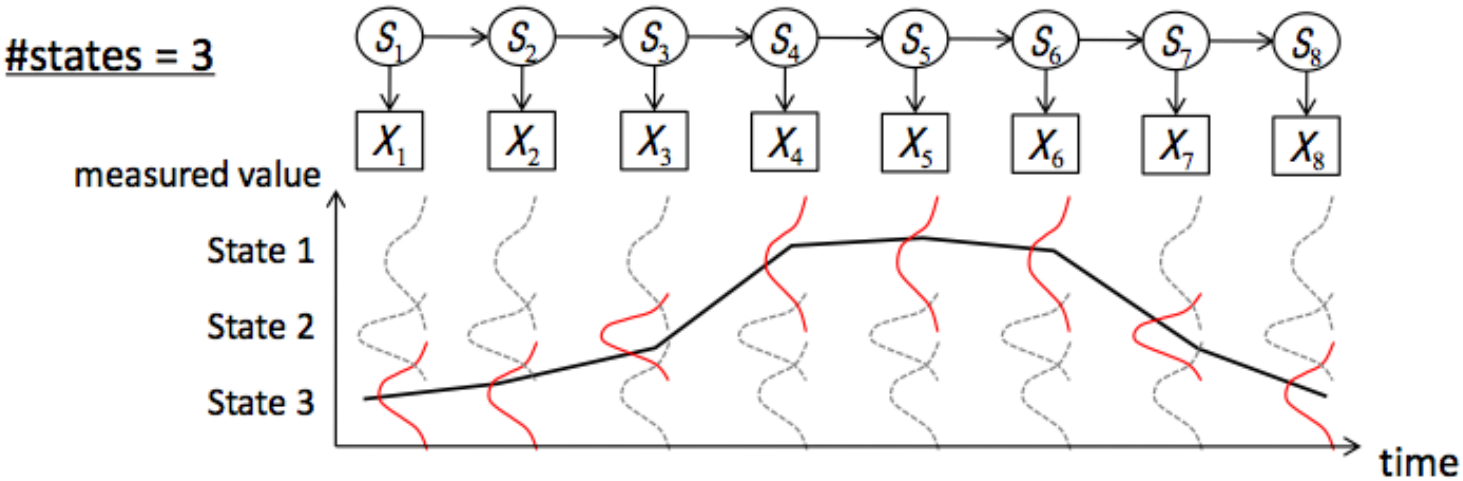
My Work

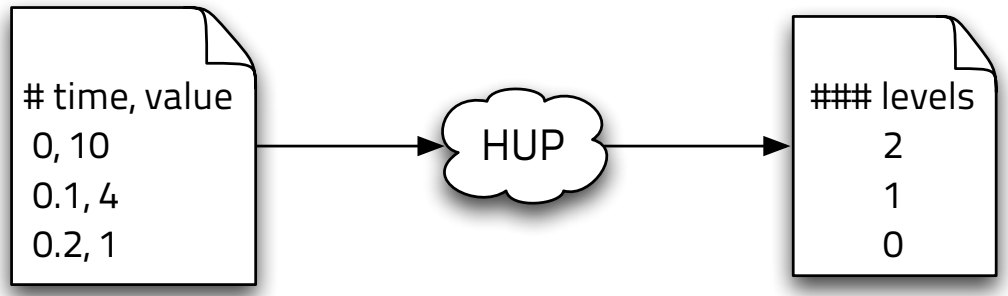

KEGG
kegg2symb
Symbolic pathway
SOLAR
Hypotheses
Conversion
Experimental Data
py-tsdisc
Discretized data
HUP
Discretization
BDD-EM
Rules, here a kinetic model
Michaelis-Menten
Ranked hypotheses
: append

HUP: HMM Utility Program

A work from Yoshitaka Kameya.
Clustering method that uses Continuous Hidden Markov Model + Bayesian Score:

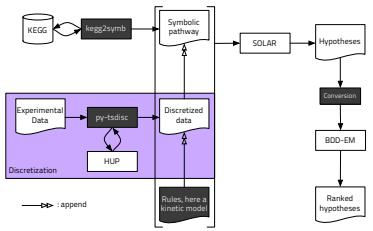


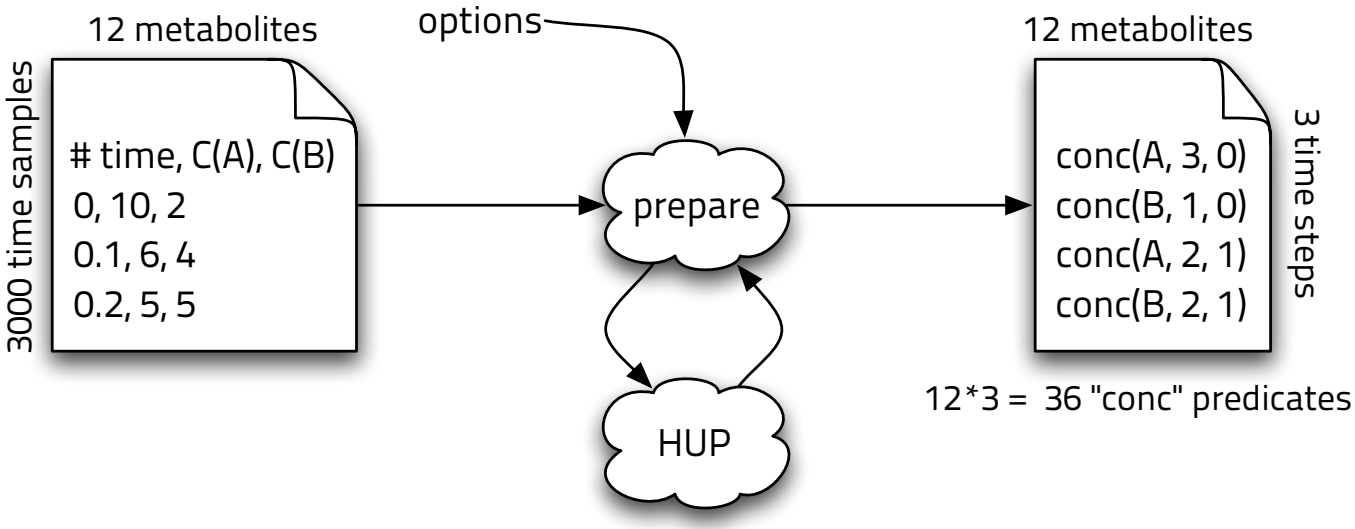




"prepare.py" - Wrapping HUP

Adapts the entry file (concentration, speed, acceleration) and feed it to HUP with the right options, calculates subsets for time discretization and then write the solar-like file.



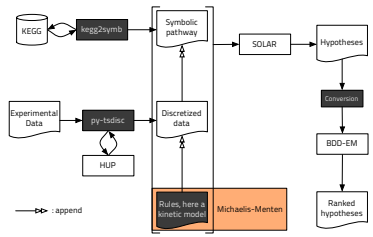


For the presented results on *E.Coli*, I remarked that the values of K_m and concentration were very sparse $\Rightarrow$ we took the log.

Logic Kinetic Modeling

If we make the approximations for extreme values in:

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



With only 3 levels, as we have in our discretization of *E.Coli* experiments, we will get the following simple rules:

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

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

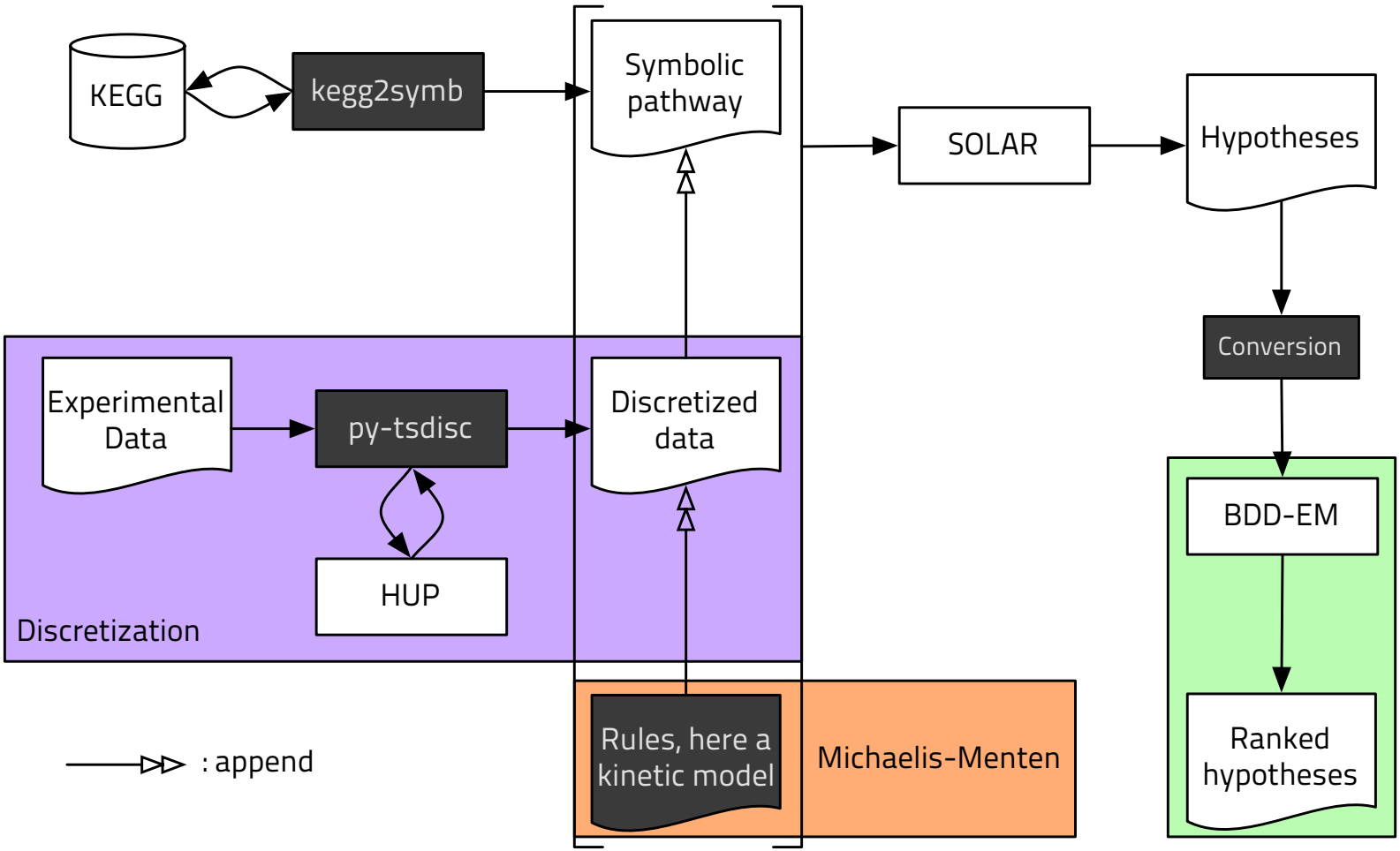
$$\underline{[S] \simeq K_m}$$

`reaction(S, P, Km)` $\wedge$ `concentration(S, 1, 0)` $\wedge$ `concentration(Km, 1, 0)` $\wedge$ `concentration(P, L, 0)` $\rightarrow$ `concentration(P, L, 1)`

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

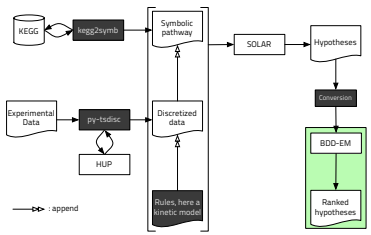
`reaction(S, P, Km)` $\wedge$ `concentration(S, 2, 0)` $\wedge$ `concentration(Km, 0, 0)` $\rightarrow$ `concentration(P, 2, 1)`

The Whole Process

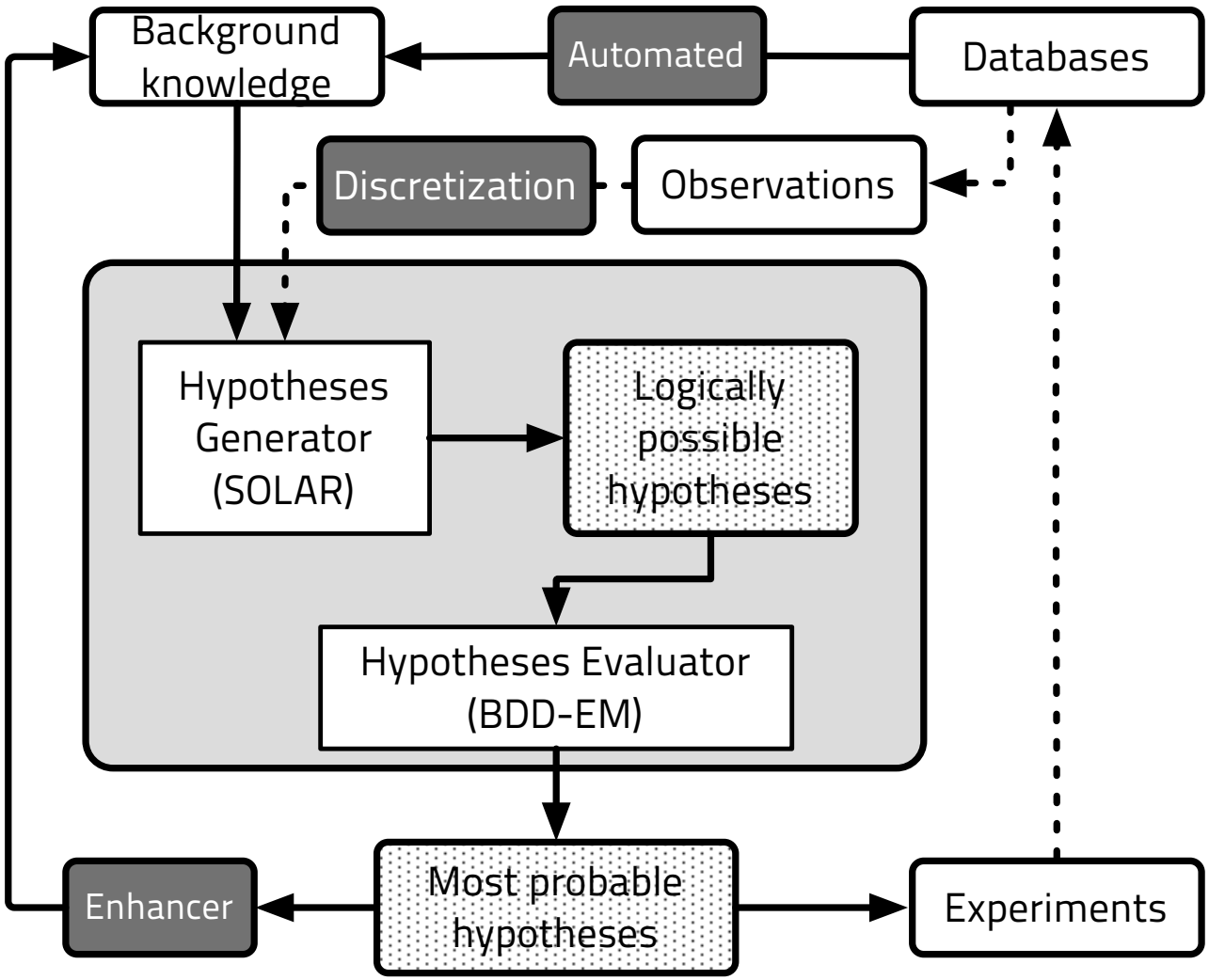


Ranking the hypotheses with BDD-EM

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



Conclusion: My Ideas



Conclusion: What's left?

To be able to deal with *any* pathways for which we would have concentrations and michaelis-menten constants. 2 problems:

1. more than 3 levels
   - A proof of concept has been made with a `compute` predicate working in the Java part of `SOLAR`
   - Kinetic rules can be automatically generated
2. reactions with more than "1 substrate - 1 product"
   - `kegg2symb` already deals with all kind of reactions
   - Allosteric kinetics

1 accepted poster @ ILP'09 (Leuven), 6 pages
1 submission (rejected) @ DS'09 (Porto), 14 pages

Thanks

Many thanks go to `shuffle(thanks)`:
(>'-')> (Katsumi Inoue, Andrei Doncescu, Taisuke Sato, Pierre Bessière, Yoshitaka Kameya, Yoshitaka Yamamoto, Takehide Soh, Nicolas Dumazet, Elsa Prieto, (all_my_friends)).

Thank you for your attention.

Arigato gozaimasu.

Any questions?

Any questions?