Abstract Interpretation for Systems Biology

Part I: Hierarchy of Semantics

François Fages
INRIA Rocquencourt,France
http://contraintes.inria.fr
Francois.Fages@inria.fr

1. Theory of Abstract Interpretation
2. Syntactical Domain of SBML Reaction Rules
3. Stochastic Semantics Domain
4. Discrete Semantics Domain
5. Boolean Semantics Domain

1. Abstraction in Systems Biology

Models are built in Systems Biology with two contradictory perspectives :

1. Abstraction in Systems Biology

Models are built in Systems Biology with two contradictory perspectives :

1) Models for representing knowledge : the more concrete the better

detailed mechanistic reaction models (SBML), gene ontologies, protein functions, protein interactions, structures ...

1. Abstraction in Systems Biology

Models are built in Systems Biology with two contradictory perspectives :

1) Models for representing knowledge : the more concrete the better

detailed mechanistic reaction models (SBML), gene ontologies, protein functions, protein interactions, structures ...

2) Models for making predictions : the more abstract the better.

schematic reaction models (SBML), variable elimination, approximations, stationary states, influence graph ...

1. Abstraction in Systems Biology

Models are built in Systems Biology with two contradictory perspectives :

1) Models for representing knowledge : the more concrete the better

detailed mechanistic reaction models (SBML), gene ontologies, protein functions, protein interactions, structures ...

2) Models for making predictions : the more abstract the better.

schematic reaction models (SBML), variable elimination, approximations, stationary states, influence graph ...

These perspectives can be reconciled by organizing models into hierarchies of abstractions.

"To understand a system is not to know everything about it but to know abstraction levels that are sufficient for answering questions about it"

The Theory of Abstract Interpretation

In this setting [Cousot Cousot 77], a domain is a lattice $\mathcal{D}(\sqsubseteq, \perp, \top, \sqcup, \sqcap)$ where $\sqsubseteq$ is the "information loss" ordering.

Often just a power-set $\mathcal{P}(\mathcal{S})(\subseteq, \emptyset, \mathcal{S}, \cup, \cap)$ ordered by set inclusion.

The Theory of Abstract Interpretation

In this setting [Cousot Cousot 77], a domain is a lattice $\mathcal{D}(\sqsubseteq, \bot, \top, \sqcup, \sqcap)$ where $\sqsubseteq$ is the "information loss" ordering.

A Galois connection $\mathcal{C} \rightarrow_\alpha \mathcal{A}$ between two lattices $\mathcal{C}$ and $\mathcal{A}$ is defined by two abstraction and concretization functions $\alpha : \mathcal{C} \rightarrow \mathcal{A}$ and $\gamma : \mathcal{A} \rightarrow \mathcal{C}$ that are monotonic:

- $\forall\ x, y \in \mathcal{C}\ x \sqsubseteq_\mathcal{C} y\ \Rightarrow\ \alpha(x) \sqsubseteq_\mathcal{A} \alpha(y)$,
- $\forall\ x, y \in \mathcal{A}\ x \sqsubseteq_\mathcal{A} y\ \Rightarrow\ \gamma(x) \sqsubseteq_\mathcal{C} \gamma(y)$,

and are adjoint:

- $\forall c \in \mathcal{C}, \forall y \in \mathcal{A} : x \sqsubseteq_\mathcal{C} \gamma(y) \Leftrightarrow \alpha(x) \sqsubseteq_\mathcal{A} y$.

If $\gamma \circ \alpha$ is the identity, the abstraction α loses no information, and $\mathcal{C}$ and $\mathcal{A}$ are isomorphic from the information standpoint (although α may be not onto and γ not one-to-one).

Properties of Galois Connections

1. $\gamma \circ \alpha$ is extensive (i.e. $x \sqsubseteq_{\mathcal{C}} \gamma \circ \alpha(x)$) and represents the information lost by the abstraction;
2. $\alpha \circ \gamma$ is contracting (i.e. $\alpha \circ \gamma(y) \sqsubseteq_{\mathcal{A}} y$);
3. $\gamma \circ \alpha$ is the identity *iff* γ is onto *iff* α is one-to-one.
4. α preserves $\sqcup$, and γ preserves $\sqcap$;
5. $\gamma(a) = max\ \alpha^{-1}(\downarrow a) = \sqcup \alpha^{-1}(\downarrow a)$
6. $\alpha(c) = min\ \gamma^{-1}(\uparrow c) = \sqcap \gamma^{-1}(\uparrow c)$

where $\downarrow a = \{b \mid b \sqsubseteq a\}$ and $\uparrow a = \{b \mid a \sqsubseteq b\}$.

It is equivalent in the definition of Galois connections to replace the condition of adjointness by conditions 1 and 2,

or by condition 5 which also entails the monotonicity of γ.

Systems Biology Markup Language SBML Models

Formally, the concrete domain of reaction models is the powerset of all possible reaction rules ordered by set inclusion :

Def. 1 *Given a finite set $\mathcal{M}$ of molecule names, the universe of reactions is the set of rules*

$$\mathcal{R} = \{e \texttt{ for } S{=}{>}S' \mid \quad e \textit{ is a kinetic expression, and } S \textit{ and } S' \textit{ are solutions of molecules in } \mathcal{M}\}.$$

The domain of SBML reaction models is $\mathcal{C}_{\mathcal{R}} = (\mathcal{P}(\mathcal{R}), \subseteq)$.

Systems Biology Markup Language SBML Models

Formally, the concrete domain of reaction models is the powerset of all possible reaction rules ordered by set inclusion :

Def. 2 *Given a finite set $\mathcal{M}$ of molecule names, the universe of reactions is the set of rules*

$$\mathcal{R} = \{e \texttt{ for } S{=}{>}S' \mid \quad e \textit{ is a kinetic expression, and } S \textit{ and } S' \textit{ are solutions of molecules in } \mathcal{M}\}.$$

The domain of SBML reaction models is $\mathcal{C}_\mathcal{R} = (\mathcal{P}(\mathcal{R}), \subseteq)$.

In the SBML exchange format, no semantics are defined.

In BIOCHAM, three semantics are considered:

1. boolean : non-deterministic asynchronous transition system
2. differential : ODE (or hybrid system)
3. stochastic : continuous time Markov chain.

Stochastic Semantics

For a given volume V_k of the location where the compound x_k resides, a concentration C_k for a molecule is translated into a number of molecules $N_k = \lfloor C_k \times V_k \times N_A \rfloor$, where N_A is Avogadro's number.

The kinetic expression e_i for each reaction i evaluates on numbers of molecules for each compound, instead of concentrations, in a (positive) reaction weight τ_i.

An element s of the domain precisely defines a Markov chain, where the probability p_{ij} of transition from state S_i to S_j is obtained by normalizing the reaction rate $\tau_{i,j} = \sum_{(S_i,S_j,\tau)\in s} \tau$ in

$$p_{ij} = \frac{\tau_{ij}}{\sum_{(S_i,S_k,\tau_{ik})\in s} \tau_{ik}}$$

Stochastic Semantics Domain

Def. 3 *Let a discrete state be a vector of integers of dimension* $|\mathcal{M}|$*. The universe* $\mathcal{S}$ *of stochastic transitions is the set of triplets* (S_i, S_j, τ_{ij}) *where* S_i *and* S_j *are discrete states and* $\tau_{ij} \in \mathbb{R}^+$*.*

The domain of stochastic transitions is $\mathcal{D}_{\mathcal{S}} = (\mathcal{P}(\mathcal{S}), \subseteq)$*.*

Discrete states and solutions in reaction rules have the same mathematical structure, and can both be represented by $|\mathcal{M}|$-dimensional vectors of integers.

Galois Connection Syntactical → Stochastic Domain

Proposition 4 *Let $\alpha_{\mathcal{RS}} : \mathcal{C}_\mathcal{R} \to \mathcal{D}_\mathcal{S}$ be the function associating to a reaction model the state transition graph labelled with thte $\tau_{i,j}$'s. Let $\gamma_{\mathcal{RS}}(s) = \cup {\alpha_{\mathcal{RS}}}^{-1}(\downarrow s)$. $\mathcal{C}_\mathcal{R} \underset{\gamma_{\mathcal{RS}}}{\overset{\alpha_{\mathcal{RS}}}{\rightleftarrows}} \mathcal{D}_\mathcal{S}$ is a Galois connection.*

Proof: It is sufficient to show that $\alpha_{\mathcal{RS}}$ is monotonic and $\gamma_{\mathcal{RS}}(s) = max\ {\alpha_{\mathcal{RS}}}^{-1}(\downarrow s)$. $\alpha_{\mathcal{RS}}$ is monotonic as the addition of reaction rules cannot decrease the set of stochastic transitions. Let s be a set of stochastic transitions and $m = \gamma_{\mathcal{RS}}(s) = \cup {\alpha_{\mathcal{RS}}}^{-1}(\downarrow s)$, m is the model obtained by union of all the rules of models in ${\alpha_{\mathcal{RS}}}^{-1}(\downarrow s)$. We have to show that $m \in {\alpha_{\mathcal{RS}}}^{-1}(\downarrow s)$. Let us consider $\alpha_{\mathcal{RS}}(m)$, each of its edges comes from a rule of m, hence there exists a set of stochastic transitions $s' \subseteq s$ such that the rule belongs to a model m' with $\alpha_{\mathcal{RS}}(m') = s'$. The same edge is thus in s' and hence in s. Therefore $\alpha_{\mathcal{RS}}(m) \subseteq s$. □

Stochastic Semantics Domain

$\alpha_{\mathcal{RS}}$ is not one-to-one.

Stochastic Semantics Domain

$\alpha_{\mathcal{RS}}$ is not one-to-one.

For instance, the reaction models $m1 = \{$ `e for A => B`$\}$ and $m2 = m1 \cup \{$ `e for 2*A => A+B`$\}$ have the same set of stochastic transitions. $\gamma \circ \alpha$ is thus not the identity, the information lost by the stochastic abstraction is the elimination of redundant rules in the reaction model.

Stochastic Semantics Domain

$\alpha_{\mathcal{RS}}$ is not one-to-one.

For instance, the reaction models $m1 = \{$ `e for A => B`$\}$ and $m2 = m1 \cup \{$ `e for 2*A => A+B`$\}$ have the same set of stochastic transitions. $\gamma \circ \alpha$ is thus not the identity, the information lost by the stochastic abstraction is the elimination of redundant rules in the reaction model.

$\alpha_{\mathcal{RS}}$ is neither onto

Stochastic Semantics Domain

$\alpha_{\mathcal{RS}}$ is not one-to-one.

For instance, the reaction models $m1 = \{$ `e for A => B`$\}$ and $m2 = m1 \cup \{$ `e for 2*A => A+B`$\}$ have the same set of stochastic transitions. $\gamma \circ \alpha$ is thus not the identity, the information lost by the stochastic abstraction is the elimination of redundant rules in the reaction model.

$\alpha_{\mathcal{RS}}$ is neither onto as the stochastic transitions obtained from a reaction model enjoy some particular properties, such as for instance the following stability property w.r.t. the number of molecules in the states:

Proposition 5 *If two states S_1, S_2 are such that $S_1 \leq S_2$ pointwise, then for any model m and all transitions S_i, τ_i such that $(S_1, S_i, \tau_i) \in \alpha_{\mathcal{RS}}(m)$, there exist states $S_j = S_i + S_2 - S_1$ (pointwise) such that $(S_2, S_j, \tau_i) \in \alpha_{\mathcal{RS}}(m)$, i.e. all rules that apply in S_1 apply in S_2 with the same changes.*

PROOF: By definition of $\alpha_{\mathcal{RS}}$. $\square$

Discrete Semantics

Def. 6 *The universe $\mathcal{D}$ of discrete transitions is the set of pairs of discrete states. The domain of discrete transitions is $\mathcal{D}_{\mathcal{D}} = (\mathcal{P}(\mathcal{D}), \subseteq)$.*

The discrete semantics is the classical Petri net semantics of reaction models [RML93ismb,SHK06bmcbi,Chaouiya07bioinfo,GHL07cmsb].

Classical Petri net analysis tools can be used for the analysis of reaction models at this abstraction level.

For instance, the elementary mode analysis of metabolic networks [SPM02bioinfo] has been shown in [ZS03insilicobio] to be equivalent to the classical analysis of Petri nets by T-invariants.

Discrete Semantics

Proposition 7 *Let* $\alpha_{\mathcal{SD}} : \mathcal{D}_{\mathcal{S}} \to \mathcal{D}_{\mathcal{D}}$ *be the function associating to a set of stochastic transitions the discrete transitions obtained by projection on the two first components, and* $\gamma_{\mathcal{SD}}(d) = \cup {\alpha_{\mathcal{SD}}}^{-1}(\downarrow d)$. $\mathcal{D}_{\mathcal{S}} \underset{\gamma_{\mathcal{SD}}}{\overset{\alpha_{\mathcal{SD}}}{\rightleftarrows}} \mathcal{D}_{\mathcal{D}}$ *is a Galois connection.*

Proof: Here again, it suffices to show that $\alpha_{\mathcal{SD}}$ is monotonic and $\gamma_{\mathcal{SD}}(d) = max\ {\alpha_{\mathcal{SD}}}^{-1}(\downarrow d)$. Clearly $\alpha_{\mathcal{SD}}$ is monotonic as adding stochastic transitions will only increase the set of discrete transitions. Now let $s = \cup {\alpha_{\mathcal{SD}}}^{-1}(\downarrow d) = \cup {\alpha_{\mathcal{SD}}}^{-1}(\downarrow d)$, for all discrete transitions in $\alpha_{\mathcal{SD}}(s)$ there exists s' and $d' \subseteq d$ such that this transition corresponds to a stochastic transition in a s' and $d' = \alpha_{\mathcal{SD}}(s')$. The same transition is thus in d' and hence in d. Therefore $\alpha_{\mathcal{SD}}(s) \subseteq d$, i.e. $\alpha_{\mathcal{SD}}(s) \in \downarrow d$, and thus $s \in {\alpha_{\mathcal{SD}}}^{-1}(\downarrow d)$ q.e.d. $\square$

Remark that $\alpha_{\mathcal{SD}}$ is onto, but not one-to-one as the transition rates are simply forgotten.

Boolean Semantics

Def. 8 *Let a boolean state be a vector of booleans of dimension* $|\mathcal{M}|$ *indicating the presence of each molecule in the state. The universe* $\mathcal{B}$ *of boolean transitions is the set of pairs of boolean states.*

The domain of boolean transitions is $\mathcal{D}_{\mathcal{B}} = (\mathcal{P}(\mathcal{B}), \subseteq)$.

Boolean Semantics

Def. 9 *Let a boolean state be a vector of booleans of dimension* $|\mathcal{M}|$ *indicating the presence of each molecule in the state. The universe* $\mathcal{B}$ *of boolean transitions is the set of pairs of boolean states.*

The domain of boolean transitions is $\mathcal{D}_{\mathcal{B}} = (\mathcal{P}(\mathcal{B}), \subseteq)$.

Let $\alpha_{\mathcal{N}\mathcal{B}} : \mathbb{N}^{|\mathcal{M}|} \to \mathbb{B}^{|\mathcal{M}|}$ be the zero/non-zero abstraction (or threshold abstraction) from the integers to the booleans, and its pointwise extension from discrete states to boolean states.

Proposition 10 *Let* $\alpha_{\mathcal{D}\mathcal{B}} : \mathcal{D}_{\mathcal{D}} \to \mathcal{D}_{\mathcal{B}}$ *be the set extension of* $\alpha_{\mathcal{N}\mathcal{B}}$*. Let* $\gamma_{\mathcal{D}\mathcal{B}}(b) = \cup {\alpha_{\mathcal{D}\mathcal{B}}}^{-1}(\downarrow b)$*.* $\mathcal{D}_{\mathcal{D}} \underset{\gamma_{\mathcal{D}\mathcal{B}}}{\overset{\alpha_{\mathcal{D}\mathcal{B}}}{\rightleftarrows}} \mathcal{D}_{\mathcal{B}}$ *is a Galois connection.*

PROOF: $\alpha_{\mathcal{D}\mathcal{B}}$ is monotonic as the addition of discrete transitions can only augment the set of boolean transitions, and $\cup {\alpha_{\mathcal{D}\mathcal{B}}}^{-1}(\downarrow b) \in {\alpha_{\mathcal{D}\mathcal{B}}}^{-1}(\downarrow b)$ as all transitions in the image of $\gamma_{\mathcal{D}\mathcal{B}}(b)$ are in b. $\square$

BIOCHAM Boolean Semantics

Given a reaction model R, let us denote by S_{BB} the set of boolean transitions obtained by considering all pssible consumption of reactants.

For instance, a rule like `A+B=>C+D` is interpreted by four boolean transition rules :

- $A \wedge B \longrightarrow A \wedge B \wedge C \wedge D$
- $A \wedge B \longrightarrow \neg A \wedge B \wedge C \wedge D$
- $A \wedge B \longrightarrow A \wedge \neg B \wedge C \wedge D$
- $A \wedge B \longrightarrow \neg A \wedge \neg B \wedge C \wedge D$

Note that in Boolean Petri nets, or in Pathway Logic, complete consumption is always assumed.

Representing all possible consumptions is necessary for getting an over-approximation result.

BIOCHAM Boolean Semantics in the hierarchy of semantics

Proposition 11 *For any reaction model R, $\alpha_{\mathcal{DB}}(\alpha_{\mathcal{SD}}(\alpha_{\mathcal{RS}}(R))) \subseteq S_{BB}$.*

Proof: Since all our abstractions are defined pointwise, it is enough to prove it for only one rule in R. Let us consider e `for` S=>S'. By abuse of notation we will denote by S and S' the discrete states corresponding to solutions of same name. We have
$\alpha_{\mathcal{RS}}(R) = \{(S_i, S_j, e) | S_i \geq S,\ S_j = S_i - S + S'\}$ and thus
$\alpha_{\mathcal{SD}}(\alpha_{\mathcal{RS}}(R)) = \{(S_i, S_j) | S_i \geq S,\ S_j = S_i - S + S'\}$, which leads to
$\alpha_{\mathcal{DB}}(\alpha_{\mathcal{SD}}(\alpha_{\mathcal{RS}}(R))) = \{(S'_i, S'_j) | S_i \geq S,\ S_j = S_i - S + S',\ S'_i = \alpha_{\mathcal{NB}}(S_i),\ S'_j = \alpha_{\mathcal{NB}}(S_j)\}$. Since $S_{BB} = \{(T, T') | T \geq \alpha_{\mathcal{NB}}(S),\ \alpha_{\mathcal{NB}}(S') \vee (T \wedge \neg\alpha_{\mathcal{NB}}(S)) \leq T' \leq \alpha_{\mathcal{NB}}(T) \vee \alpha_{\mathcal{NB}}(S')\}$ the property holds as $S_i \geq S$ implies $S'_i \geq \alpha_{\mathcal{NB}}(S)$, and since $S_i \geq S$ we have
$S_j = S_i - S + S' \Rightarrow S_i - S + S' \leq S_j \leq S_i + S' \Rightarrow \alpha_{\mathcal{NB}}(S_i - S + S') = \alpha_{\mathcal{NB}}(S') \vee (\alpha_{\mathcal{NB}}(S_i) \wedge \neg\alpha_{\mathcal{NB}}(S)) \leq S'_j \leq \alpha_{\mathcal{NB}}(S_i + S') = \alpha_{\mathcal{NB}}(S_i) \vee \alpha_{\mathcal{NB}}(S')$
□

Differential Semantics ?

The differential semantics of reaction models interprets a set of reaction rules $\{e_i \texttt{ for } S_i\texttt{=>}S_i'\}_{i=1,\dots,n}$ over molecular concentration variables $\{x_1, ..., x_m\}$, by the following system of Ordinary Differential Equations (ODE):

$$dx_k/dt = \sum_{i=1}^{n} r_i(x_k) * e_i - \sum_{j=1}^{n} l_j(x_k) * e_j$$

where we recall that $r_i(x_k)$ (resp. l_i) is the stoichiometric coefficient of x_k in the right (resp. left) member of rule i.

- synchronous semantics (evolution of variables in parallel)
- deterministic semantics (average behavior)
- not compatible with the rule set inclusion ordering
- infinite number of molecules
- infinitesimal time steps

Abstract Interpretation for Systems Biology

Part II: Type Checking and Type Inference

François Fages

INRIA Rocquencourt,France

http://contraintes.inria.fr

Francois.Fages@inria.fr

1. Type Checking and Type Inference
2. Domain of Protein Functions
3. Domain of Protein Influences
4. Influence graph inferred from the syntactical domain
5. Influence graph inferred from the differential semantics

Type Checking/Inference by Abstract Interpretation

A type system $\mathcal{A}$ for a concrete domain $\mathcal{C}$ is a Galois connection $\mathcal{C} \rightarrow_{\alpha} \mathcal{A}$.

Type Checking/Inference by Abstract Interpretation

A type system $\mathcal{A}$ for a concrete domain C is a Galois connection $\mathcal{C} \rightarrow_{\alpha} \mathcal{A}$.

The type inference problem is

INPUT a concrete element $x \in \mathcal{C}$ (e.g. a reaction model)

OUTPUT its typing $\alpha(x)$ (e.g. the protein functions of the model).

Type Checking/Inference by Abstract Interpretation

A type system A for a concrete domain C is a Galois connection $\mathcal{C} \rightarrow_\alpha \mathcal{A}$.

The type inference problem is

INPUT a concrete element $x \in \mathcal{C}$ (e.g. a reaction model)

OUTPUT its typing $\alpha(x)$ (e.g. the protein functions of the model).

The type checking problem is,

INPUT $x \in \mathcal{C}$ (e.g. a reaction model)

and a typing $y \in \mathcal{A}$ (e.g. a set of protein functions),

OUTPUT determine whether $x \sqsubseteq_\mathcal{C} \gamma(y)$
(i.e. whether the reactions are compatible with the protein functions)

or equivalently $\alpha(x) \sqsubseteq_\mathcal{A} y$ (the typing contains the inferred types)

Type Checking/Inference by Abstract Interpretation

A type system A for a concrete domain C is a Galois connection $\mathcal{C} \rightarrow_{\alpha} \mathcal{A}$.

The type inference problem is

INPUT a concrete element $x \in \mathcal{C}$ (e.g. a reaction model)

OUTPUT its typing $\alpha(x)$ (e.g. the protein functions of the model).

The type checking problem is,

INPUT $x \in \mathcal{C}$ (e.g. a reaction model)

and a typing $y \in \mathcal{A}$ (e.g. a set of protein functions),

OUTPUT determine whether $x \sqsubseteq_{\mathcal{C}} \gamma(y)$
(i.e. whether the reactions are compatible with the protein functions)

or equivalently $\alpha(x) \sqsubseteq_{\mathcal{A}} y$ (the typing contains the inferred types)

Algorithms in $O(n)$ if the abstractions can be computed rule per rule.

Protein Functions as Types

Abstract domain $\mathcal{A}_{\mathcal{F}} = \mathcal{P}(\{\text{kinase}(A) | A \in \mathcal{M}\} \cup \{\text{phosphatase}(A) | A \in \mathcal{M}\})$

The typing of reactions by protein functions is defined by the abstraction :

$\alpha_{\mathcal{F}}$(`A =[B]=> C`) $= \{$`kinase(B)`$\}$ if `C` is strictly more phosphorylated than `A`

$\alpha_{\mathcal{F}}$(`A =[B]=> C`) $= \{$`phosphatase(B)`$\}$ if `C` is strictly less phosphorylated

$\alpha_{\mathcal{F}}$(`A + B => A-B, A-B => C + B`) $= \{$ `kinase(B)`$\}$

if `C` is strictly more phosphorylated than `A`

$\alpha_{\mathcal{F}}$(`A + B => A-B, A-B => C + B`) $= \{$ `phosphatase(B)`$\}$

if `C` is strictly less phosphorylated than `A`

Protein Functions as Types

Abstract domain $\mathcal{A}_{\mathcal{F}} = \mathcal{P}(\{\text{kinase}(A) | A \in \mathcal{M}\} \cup \{\text{phosphatase}(A) | A \in \mathcal{M}\})$

The typing of reactions by protein functions is defined by the abstraction :

$\alpha_{\mathcal{F}}(\texttt{A =[B]=> C}) = \{\texttt{kinase(B)}\}$ if `C` is strictly more phosphorylated than `A`

$\alpha_{\mathcal{F}}(\texttt{A =[B]=> C}) = \{\texttt{phosphatase(B)}\}$ if `C` is strictly less phosphorylated

$\alpha_{\mathcal{F}}(\texttt{A + B => A-B, A-B => C + B}) = \{\ \texttt{kinase(B)}\}$

if `C` is strictly more phosphorylated than `A`

$\alpha_{\mathcal{F}}(\texttt{A + B => A-B, A-B => C + B}) = \{\ \texttt{phosphatase(B)}\}$

if `C` is strictly less phosphorylated than `A`

Proposition 12 *Let* $\gamma_{\mathcal{F}}(f) = \cup {\alpha_{\mathcal{F}}}^{-1}(\downarrow f)$, $\mathcal{C}_{\mathcal{R}} \underset{\gamma_{\mathcal{F}}}{\overset{\alpha_{\mathcal{F}}}{\rightleftarrows}} \mathcal{A}_{\mathcal{F}}$ *is a Galois connection.*

More Precise Protein Function Typing

In SBML : no typing possible as there is no syntax for phosphorylation

In BIOCHAM : typing is possible but the syntax does not distinguish between phosphorylation, acetylation etc.

More precise protein function types:

$$\tau ::= kinase|phosphatase|kinase(\tau)|phosphatase(\tau)|T$$

where T denotes some basic types of proteins, with the subtyping relations $kinase(\tau) \preceq kinase$ and $phosphatase(\tau) \preceq phosphotase$.

Evaluation Results in BIOCHAM

- MAPK model [Levchenko et al. 00]

  the kinase function of `RAFK`, `RAF~{p1}` and `MEK~{p1,p2}` is inferred;

  the phosphatase function of `RAFPH`, `MEKPH` and `MAPKPH` is inferred;

  the kinase function of `MAPK~{p1,p2}` is not visible and not inferred.

Evaluation Results in BIOCHAM

- MAPK model [Levchenko et al. 00]

  the kinase function of `RAFK`, `RAF~{p1}` and `MEK~{p1,p2}` is inferred;

  the phosphatase function of `RAFPH`, `MEKPH` and `MAPKPH` is inferred;

  the kinase function of `MAPK~{p1,p2}` is not visible and not inferred.

- Model of the mammalian cell cycle control after [Kohn 99] 165 proteins and genes, 500 variables and 800 rules. Type inference in $<$ 1sec CPU :
  - No compound is both a kinase and a phosphatase;
  - `cdc25A` and `cdc25C` are the only phosphatases found together with the deacetylase `HDAC1`.
  - The `cdk` are inferred to be kinases only in complexes with cyclins;
  - the acetylases `pCAF`, `p300` are identified to kinases.

Use of Protein Functions Types

- Check the consistency of reaction models.
- Restrict the search space for reaction rules in model revision or network inference.
- Build modules according to protein functions

Influence Graphs as Types

$\mathcal{A}_{\mathcal{I}} = \mathcal{P}(\{A \text{ activates } B \mid A, B \in \mathcal{M}\} \cup \{A \text{ inhibits } B \mid A, B \in \mathcal{M}\})$.

The influence graph of a reaction model is defined by $\alpha_{\mathcal{R}\mathcal{I}} : \mathcal{C}_{\mathcal{R}} \rightarrow \mathcal{A}_{\mathcal{I}}$

$$
\begin{aligned}
\alpha_{\mathcal{R}\mathcal{I}}(x) = \quad & \{A \text{ inhibits } B && \mid \exists (e_i \text{ for } S_i \Rightarrow S'_i) \in x, \\
& && l_i(A) > 0 \text{ and } r_i(B) - l_i(B) < 0\} \\
& \cup \{A \text{ activates } B && \mid \exists (e_i \text{ for } S_i \Rightarrow S'_i) \in x, \\
& && l_i(A) > 0 \text{ and } r_i(B) - l_i(B) > 0\}
\end{aligned}
$$

Influence Graphs as Types

$\mathcal{A}_\mathcal{I} = \mathcal{P}(\{A \text{ activates } B \mid A, B \in \mathcal{M}\} \cup \{A \text{ inhibits } B \mid A, B \in \mathcal{M}\})$.

The influence graph of a reaction model is defined by $\alpha_{\mathcal{R}\mathcal{I}} : \mathcal{C}_\mathcal{R} \rightarrow \mathcal{A}_\mathcal{I}$

$$\begin{aligned}
\alpha_{\mathcal{R}\mathcal{I}}(x) = \quad & \{A \text{ inhibits } B && \mid \exists(e_i \text{ for } S_i \Rightarrow S_i') \in x, \\
& && l_i(A) > 0 \text{ and } r_i(B) - l_i(B) < 0\} \\
& \cup\{A \text{ activates } B && \mid \exists(e_i \text{ for } S_i \Rightarrow S_i') \in x, \\
& && l_i(A) > 0 \text{ and } r_i(B) - l_i(B) > 0\}
\end{aligned}$$

$$\begin{aligned}
\alpha_{\mathcal{R}\mathcal{I}}(\{\texttt{A} + \texttt{B} \texttt{=>} \texttt{C}\}) = \{ \quad & \text{A inhibits B, A inhibits A, B inhibits A,} \\
& \text{B inhibits B, A activates C, B activates C}\} \\
\alpha_{\mathcal{R}\mathcal{I}}(\{\texttt{A} = [\texttt{C}] \texttt{=>} \texttt{B}\}) = \{ \quad & \text{C inhibits A, A inhibits A, A activates B, C activates I}
\\
\alpha_{\mathcal{R}\mathcal{I}}(\{\texttt{A} = [\texttt{B}] \texttt{=>} _\}) = \{ \quad & \text{B inhibits A, A inhibits A}\} \\
\alpha_{\mathcal{R}\mathcal{I}}(\{_ = [\texttt{B}] \texttt{=>} \texttt{A}\}) = \{ \quad & \text{B activates A}\}
\end{aligned}$$

Influence Graphs as Types

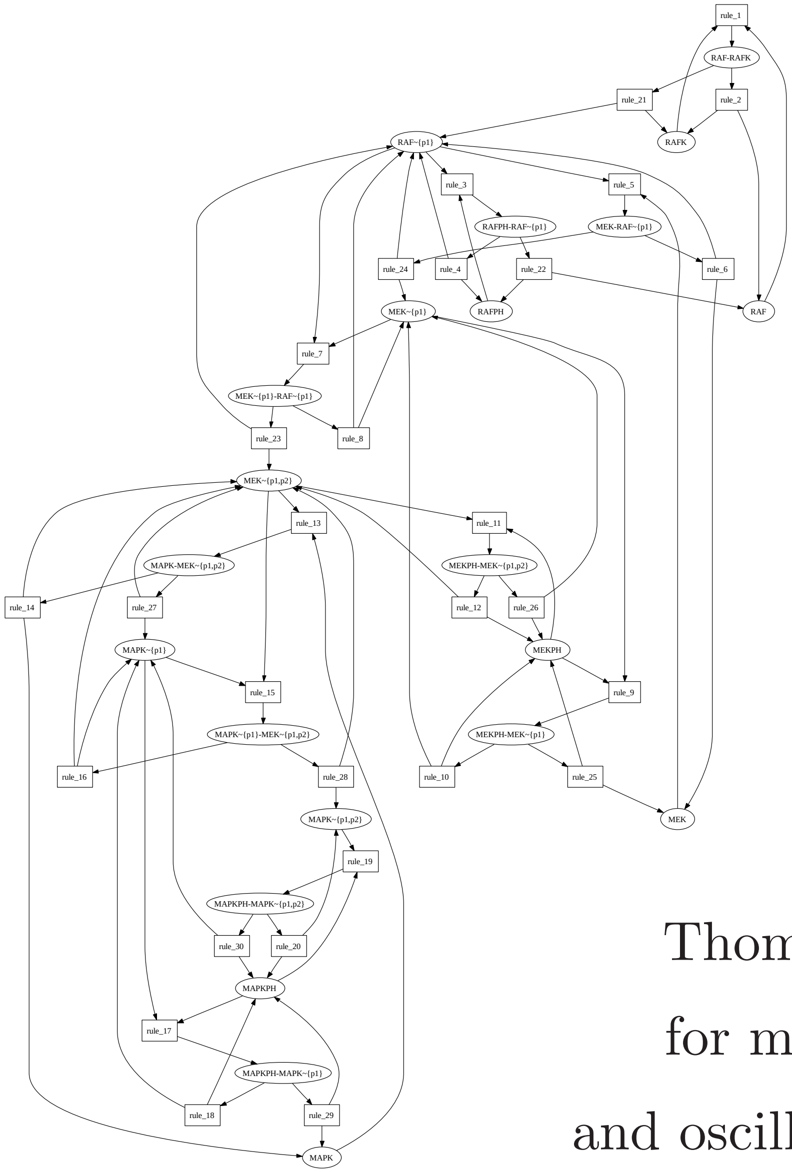
$\mathcal{A}_{\mathcal{I}} = \mathcal{P}(\{A \text{ activates } B \mid A, B \in \mathcal{M}\} \cup \{A \text{ inhibits } B \mid A, B \in \mathcal{M}\})$.

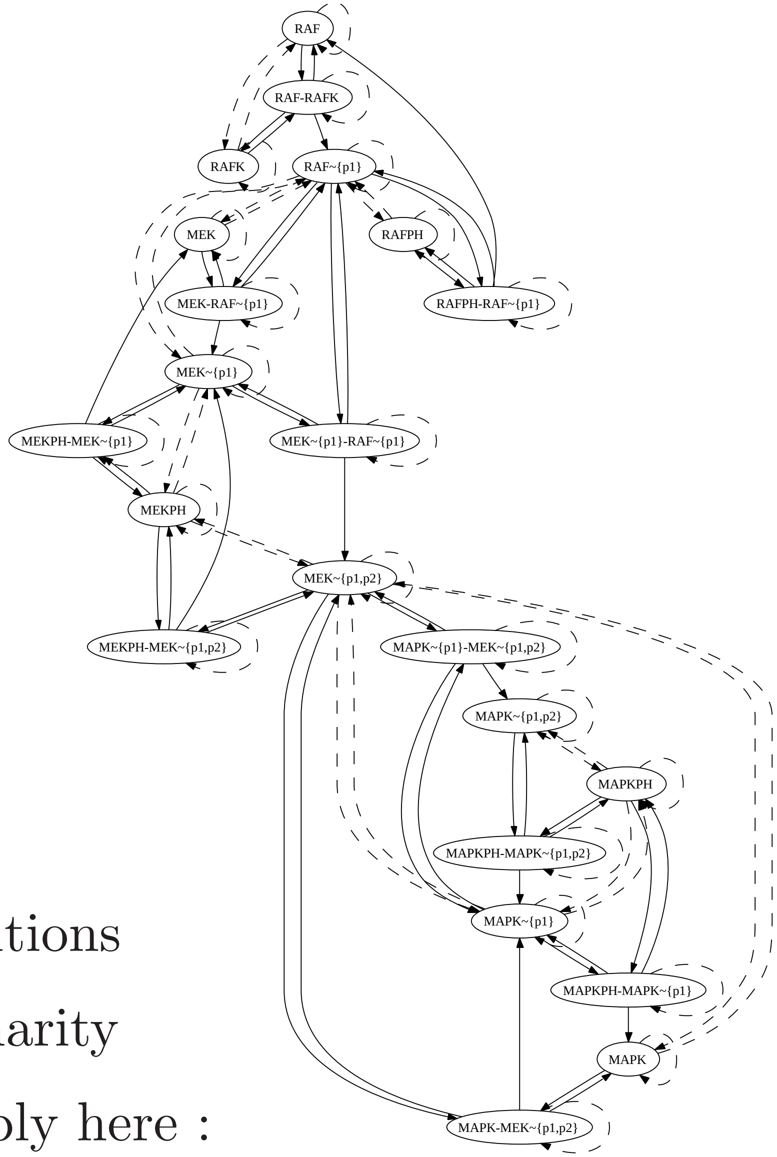
The influence graph of a reaction model is defined by $\alpha_{\mathcal{RI}} : \mathcal{C}_{\mathcal{R}} \rightarrow \mathcal{A}_{\mathcal{I}}$

$$\begin{aligned}
\alpha_{\mathcal{RI}}(x) = \quad & \{A \text{ inhibits } B && \mid \exists (e_i \text{ for } S_i \Rightarrow S'_i) \in x, \\
& && l_i(A) > 0 \text{ and } r_i(B) - l_i(B) < 0\} \\
& \cup\{A \text{ activates } B && \mid \exists (e_i \text{ for } S_i \Rightarrow S'_i) \in x, \\
& && l_i(A) > 0 \text{ and } r_i(B) - l_i(B) > 0\}
\end{aligned}$$

Proposition 13 *Let* $\gamma_{\mathcal{RI}}(f) = \cup {\alpha_{\mathcal{RI}}}^{-1}(\downarrow f)$, $\mathcal{C}_{\mathcal{R}} \underset{\gamma_{\mathcal{RI}}}{\overset{\alpha_{\mathcal{RI}}}{\rightleftarrows}} \mathcal{A}_{\mathcal{I}}$ *is a Galois connection.*

MAPK model: Reaction Graph $\rightarrow_\alpha$ Influence Graph



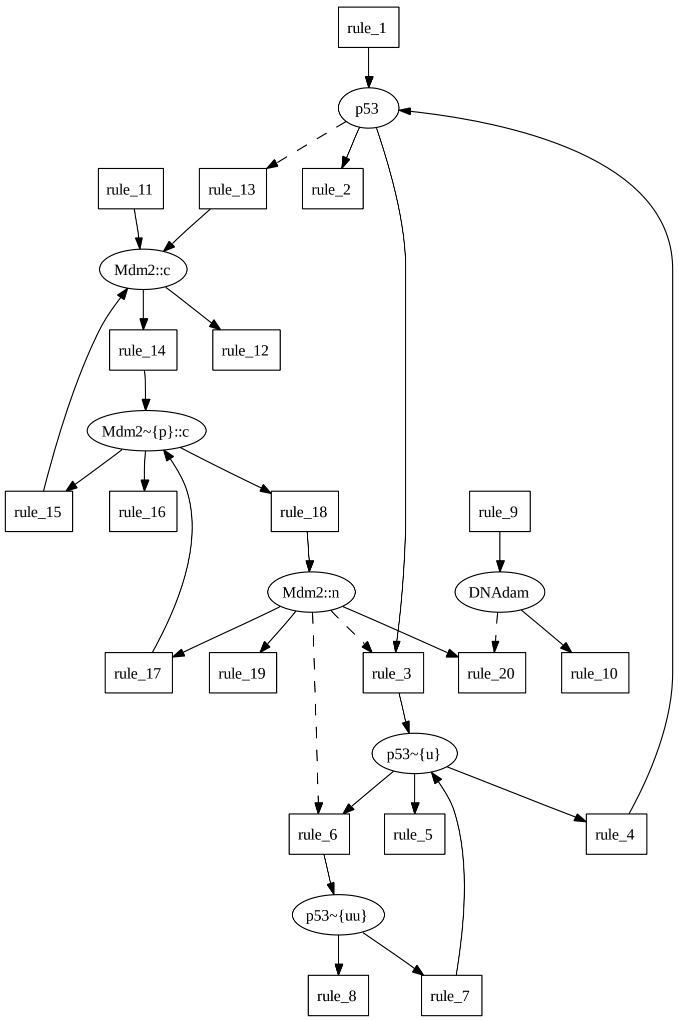


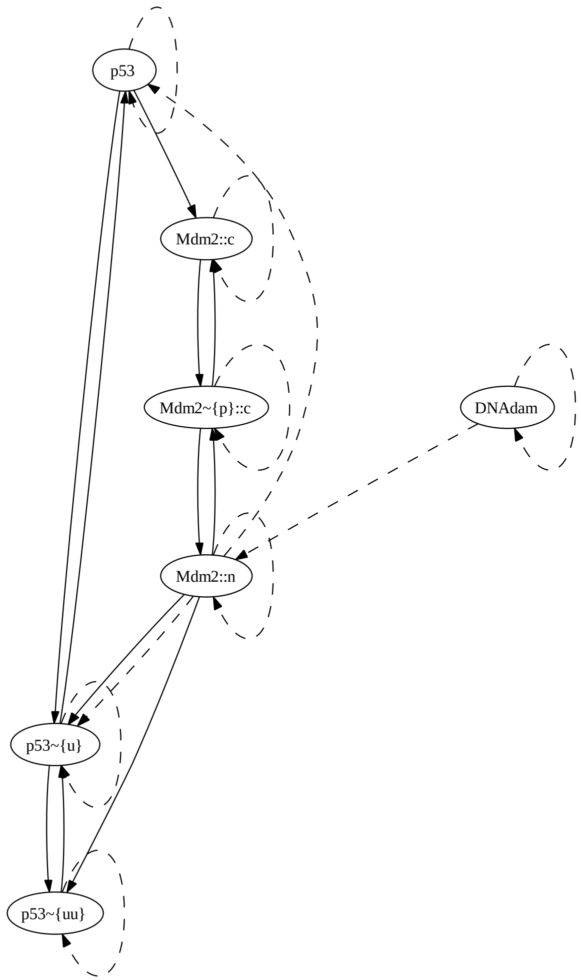
Thomas's conditions

for multistationarity

and oscillations apply here :

P53-Mdm2: Reaction Graph $\rightarrow_\alpha$ Influence Graph





Inhitions hidden in the kinetic expressions are missed

Use of Influence Types

- Check the consistency of reaction models
- Analyze the dynamics of the reaction model (multistationarity, oscillations, ...)
- Restrict the search space for reaction rules in model revision or network inference
- Build modules according to the influence graph

Influence Graph Abstraction from the Differential Semantics

Let us denote by β the mapping from $\mathcal{C}_{\mathcal{R}}$ to $\mathcal{D}_{\mathcal{J}}$ that extracts $\dot{x_k}$ and hence the Jacobian from the kinetic expressions in the reaction rules.

Def. 14 *The differential influence abstraction $\alpha_{\mathcal{J}\mathcal{I}} : \mathcal{D}_{\mathcal{J}} \rightarrow \mathcal{A}_{\mathcal{I}}$ is the function*

$$\begin{aligned}\alpha_{\mathcal{J}\mathcal{I}}(x) = \{&A \textit{ activates } B \mid \partial \dot{x_B} / \partial x_A > 0 \textit{ in some point of the phase space}\}\\ \cup\{&A \textit{ inhibits } B \mid \partial \dot{x_B} / \partial x_A < 0 \textit{ in some point of the phase space}\}\end{aligned}$$

defined purely from the kinetic expressions... compatibility with the rules ?

Monotonic Kinetics

Def. 15 *A kinetic expression e_i is monotonic w.r.t. a reaction model x iff for all molecules x_k we have*

1. *for all points of the phase space $\partial e_i/\partial x_k \geq 0$*
2. *if there exists a point in the phase space s.t. $\partial e_i/\partial x_k > 0$ then $l_i(x_k) > 0$*

The model x will be said to have monotonic kinetics if each of its reaction rules has a monotonic kinetic expression.

The mass action law kinetics, $e_i = k * \Pi {x_i}^{l_i}$, are monotonic

Hill's kinetics (and Michaelis-Menten kinetics when $n = 1$)
$e_i = V_m * {x_s}^n/(K_m + {x_s}^n)$ where $V_m = k * (x_e + x_e * x_s/K_m)$ for an enzymatic reaction $x_s = [x_e] => x_p$, are also monotonic.

Comparison to the Syntactical Influence Graph

Proposition 16 *For any reaction model x with monotonic kinetics,* $\alpha_{\mathcal{JI}} \circ \beta(x) \subseteq \alpha_{\mathcal{RI}}(x)$.

PROOF: If (A activates B) $\in \alpha_{\mathcal{JI}} \circ \beta(x)$ then $\partial \dot{B}/\partial A > 0$. Hence there exists a term $(r_i(B) - l_i(B)) * e_i$ in the ODE with $\partial e_i/\partial A$ of the same sign as $r_i(B) - l_i(B)$. Let us suppose that $r_i(B) - l_i(B) > 0$ then $\partial e_i/\partial A > 0$ and since e_i is monotonic we get that $l_i(A) > 0$ and thus that (A activates B) $\in \alpha_{\mathcal{RI}}(x)$. If on the contrary $r_i(B) - l_i(B) < 0$ then $\partial e_i/\partial A < 0$, impossible.

If (A inhibits B) $\in \alpha_{\mathcal{JI}} \circ \beta(x)$ then $\partial \dot{B}/\partial A < 0$. Hence there exists a term $(r_i(B) - l_i(B)) * e_i$ with $\partial e_i/\partial A$ of sign opposite to that of $r_i(B) - l_i(B)$. Let us suppose that $r_i(B) - l_i(B) > 0$ then $\partial e_i/\partial A < 0$, impossible. If on the contrary $r_i(B) - l_i(B) < 0$ then $\partial e_i/\partial A > 0$ and since e_i is monotonic we get that $l_i(A) > 0$ and thus that (A activates B) $\in \alpha_{\mathcal{RI}}(x)$.

$\square$

Comparison to the Syntactical Influence Graph

Even with mass action law kinetics, there is no equality between $\alpha_{\mathcal{JI}} \circ \beta$ and $\alpha_{\mathcal{RI}}$.

Comparison to the Syntactical Influence Graph

Even with mass action law kinetics, there is no equality between $\alpha_{\mathcal{JI}} \circ \beta$ and $\alpha_{\mathcal{RI}}$.

For instance let x be the following model :

$$\begin{array}{lll} k_1 * A & \text{for} & A => B \\ k_2 * A & \text{for} & \_ = [A] => A \end{array}$$

We have $\alpha_{\mathcal{RI}}(x) = \{$A activates B, A activates A, A inhibits A$\}$, however $\dot{A} = (k_2 - k_1) * A$, hence $\partial \dot{A} / \partial A$ can be made always positive or always negative or always null, resulting in the absence from $\alpha_{\mathcal{JI}} \circ \beta(x)$ of, respectively, A inhibits A, A activates A or both.

Non-monotonicity of β

β is not monotonic since adding rules can compensate an existing rule in the differential expression and eliminate terms in the differential equations.

The differential semantics is thus not an abstraction of the reaction models ordered by set inclusion in the sense of abstract interpretation.

The above case shows that $\alpha_{\mathcal{J}\mathcal{I}} \circ \beta$ applied to the first rule contains A inhibits A, whereas its application to the set of two rules (greater in $\mathcal{C}_{\mathcal{R}}$) may not.

A sufficient condition for β to be monotonic is that in the model no kinetic expression can compensate another one in the Jacobian. That is :
$\forall x_i, x_j \exists ? k$ s.t. $r_k(x_i) \neq l_k(x_i)$ and $\partial e_k / \partial x_j \neq 0$.

Precise Kinetics

Def. 17 *A kinetic expression e_i is precise w.r.t. a reaction model x iff for all molecules x_k we have*

1. *for all points of the phase space* $\partial e_i / \partial x_k \geq 0$
2. *there exists a point in the phase space s.t.* $\partial e_i / \partial x_k > 0$ *iff* $l_i(x_k) > 0$

Note that precise implies monotonic.

Proposition 18 *Mass action law, Michaelis Menten, and Hill kinetics are precise.*

Theorem 19 *If x has precise kinetics and no molecule is at the same time an activator and an inhibitor of the same target molecule, then* $\alpha_{\mathcal{RI}}(x) = \alpha_{\mathcal{JI}} \circ \beta(x)$.

Precise Kinetics

Proposition 20 *Let x be a model with precise kinetics, and A and B be two molecules.*

If A activates B is in $\alpha_{\mathcal{RI}}(x)$ but A inhibits B is not in $\alpha_{\mathcal{RI}}(x)$ then A activates B is in $\alpha_{\mathcal{JI}} \circ \beta(x)$ (and reciprocally for inhibitions).

PROOF: Since $\partial \dot{B}/\partial A = \sum_{i=1}^{n}(r_i(B) - l_i(B)) * \partial e_i/\partial A$ and all e_i are monotonic we get that $\partial \dot{B}/\partial A = \sum_{\{i \leq n | l_i(A)>0\}}(r_i(B) - l_i(B)) * \partial e_i/\partial A$.

Now if A activates B is in $\alpha_{\mathcal{RI}}(x)$ but A inhibits B is not in $\alpha_{\mathcal{RI}}(x)$ then all rule such that $l_i(A) > 0$ verify $r_i(B) - l_i(B) \geq 0$ and there is at least one rule for which the inequality is strict. We thus get that $\partial \dot{B}/\partial A$ is a sum of positive numbers, amongst which one is such that $r_i(B) - l_i(B) > 0$ and $l_i(A) > 0$ which, since x is precise, implies that there exists a point in the phase space for which $\partial e_i/\partial A > 0$ thus $\partial \dot{B}/\partial A > 0$ at that point and A activates B is in $\alpha_{\mathcal{JI}} \circ \beta(x)$.

□