A model's model

Vincent Danos
CNRS, Paris 7

Vision

— Models —not data-bases— will be the new currency for biological knowledge.

— How do we do that ?

We use an hitherto unexploited technology in a bio context: observational logics (more about this later).

What's a model

A model breakdown:

— a set of interactions (IL), rates, distinguished observables and scenarios

— generates curves (SIM)

— and actual observations (OL) (data fit, behaviour in various scenarios)

— the biology: system of interest, biological implications

Model as a knowledge engine

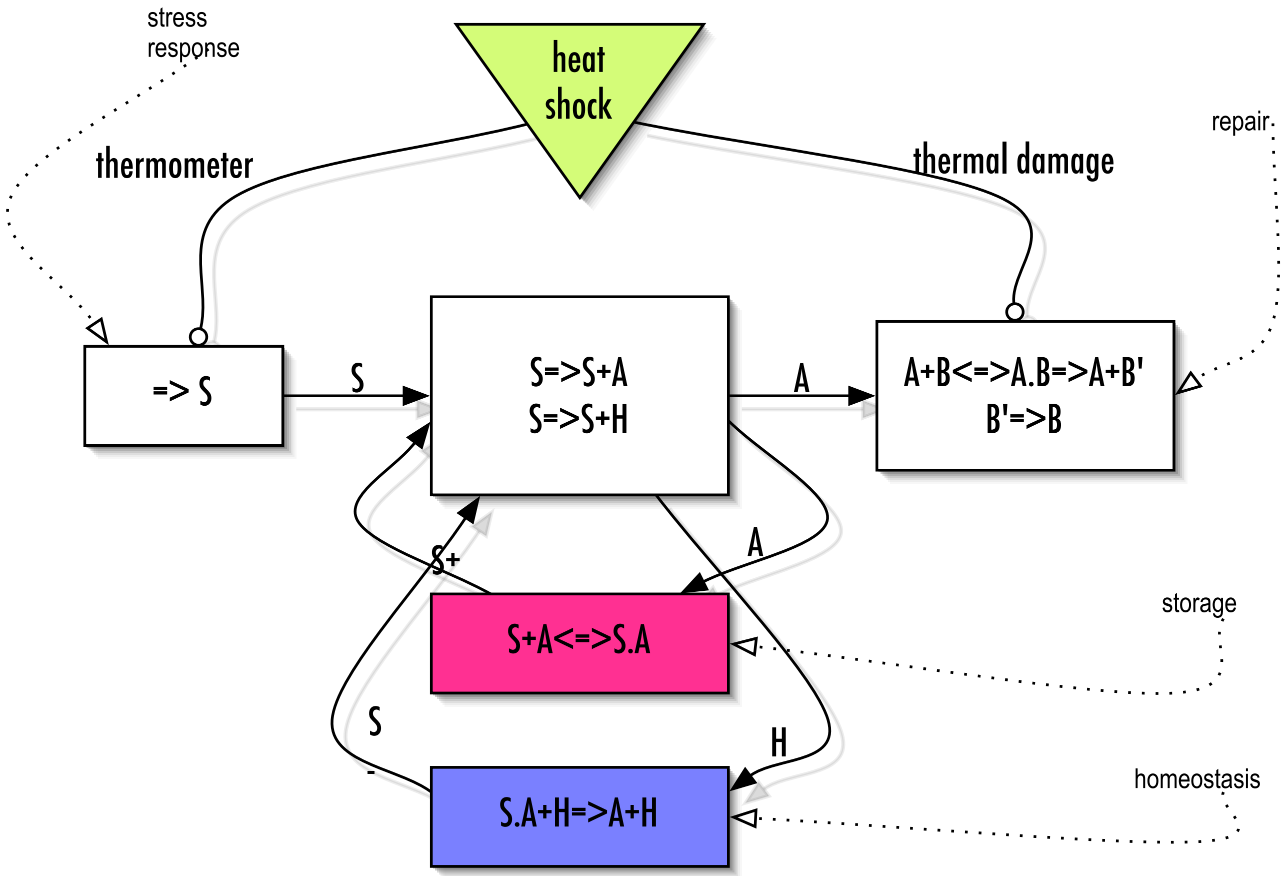
A model is a knowledge engine.

$\Rightarrow$ What we do is: formalising observations (OL).

In effect this will run the engine faster.

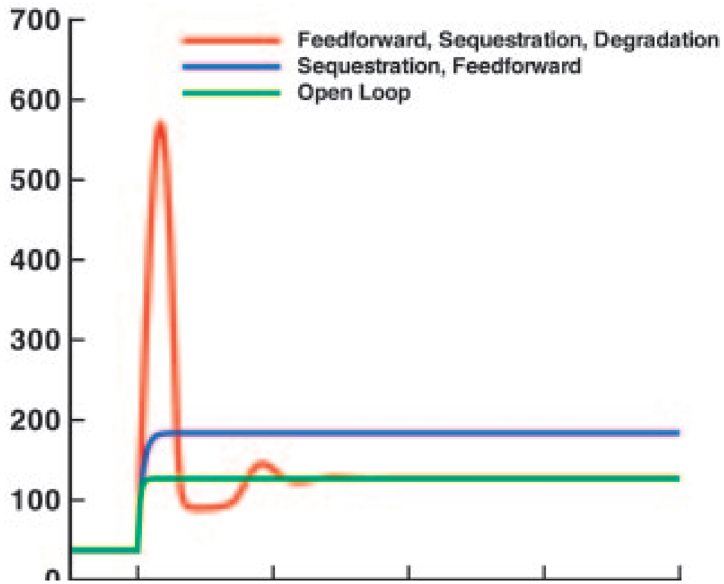
1. E. Coli heat shock response

A heat response model:



SIM

Obtaining curves (Doyle'05) for various observables (here the S response) in various scenarios (with/without storage and H-controlled S-degradation module)



OL: Steady states and Fit

Steady state (unstressed), τ_0 is S basal translation rate:

$$\mathcal{M}\langle\tau_0, 0\rangle \models \mathbf{S}_{>.99}(SA = 7 \wedge S = 1 \wedge J = 2000)$$

Model fits the S observations:

$$\mathcal{M}\langle 20\tau_0, 0\rangle \models \mathbf{P}_{>.95}\{\pi \mid \pi(0) = s_0 \wedge S(3') \in [600 \pm 100]\}$$

where:

— π is a trajectory of the model;

— s_0 is the initial no stress steady state;

— $20\tau_0$ is the boosted heat shock translation rate, and

— $x = 0$ corresponds to no interaction with repair module.

OL: Scenarios and Transients

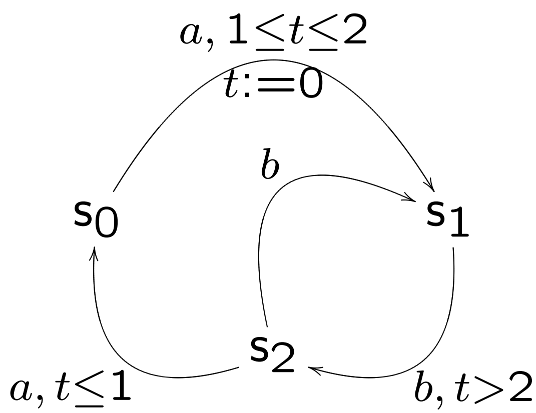
The simulation transients can be formalised as (with $\Diamond^I\phi := \top\mathbf{U}^I\phi$):

$$\begin{aligned}
\mathcal{M}(\tau^+, x^+) &\models \mathbf{P}_{>.99}(\Diamond^{\leq 10'}(S \geq 2S(0))) \\
\mathcal{M}(\tau^-, x^+) &\models \mathbf{P}_{>.99}(\Diamond^{\leq 15'}(S \geq 2S(0))) \\
\mathcal{M}(\tau^\pm, x^+) &\models \mathbf{P}_{>.99}(\Diamond^{\leq 6'}(\mathsf{J}.S = 0)) \\
\mathcal{M}(\tau^\pm, x^-) &\models \mathbf{P}_{>.99}(\Diamond^{\leq 30'}(\mathsf{J}.S \leq 1))
\end{aligned}$$

Lab notebook

Define a function (or macroscopic organization) of $\mathcal{S}$ as a finite prob/time automaton $\mathcal{A}$ (the states of which are called macro-states).

Observations can be collated:



OL so what ?

What becomes possible if one formalises OL:
— this is needed to make models vehicles of knowledge
— to stylise the model output and what are the features of interest
— this is the only means to communicate to a machine (iteration)
— this is a basis for model calibration and comparison (model equivalence wrt to a function, model metrics)
— this is a element for searching libraries (IT/executable annotation)

OL more reasons why

Fundamentally:

— observation are keys because the phenotypic map is not injective

— experiments don't speak the curve languages

— neither is selection

— nor are the curves really existing:

- individual variation/robustness,
- noisy signals/denoising

2. What's a model

A model's model

A set of reactions defines a system $\mathcal{S} = (S, P, \rightarrow_a^t)$, with state space S, parameter space P, and associated transition system $\rightarrow_a^t \in G(S)$ (G the behaviour functor depends on the chosen operational semantics: ODE, MP, ND) where:

— $t \in T$ is a time,

— $a \in L$ is an action (intervention, external events, perturbation, signal, stimulus, control, scheduler, adversary, context), with $L := (T \rightarrow P) + (S \rightarrow S)$.

Actions

Examples of actions (taken from the Srivastava and Doyle heat shock papers):

— step-changing some rate $T \to P$:
$l_1 = 2.5 \times \tau$, $l_2 = 20 \times \tau$ (thermal boosts),
$l_3 = \beta_{15} =: 2 \times \beta_{10}$, $l_4 = \beta_{15} =: 1/2 \times \beta_{10}$ (repair coupling)

— step-modifying some concentration $S \to S$:
$m_1 = 100 \times [unfolded]/[folded]$ (damaged proteins)

whatever one wants to scenarise in the model (adding in a new reaction, although this can always be seen as changing a rate; see below).

NBs

Above, l_1 is supposed to fix all other parameters to some default value (as is the case in the hs example). So that each rate-action fully determines a choice of rates (hence defines a determinising scheduler with no residual non determinism, and hence given an l and an s one can run the system).

Note that: 1) the model family has to be complete wrt interventions, *i.e.*, L has to be interpretable as changes to the model either on P the parameter space, or S the state space; 2) rate-actions are naturally having a continuous topology and one could ask whether the system is continuous wrt to actions; 3) examples of rate-actions are step functions (no transient rate-change as would be needed in feed-forward motive).

Scenario

Note that actions have a positive linear structure (stable under positive real linear combinations).

Define a *scenario* (or an experiment) as a state $init$, an action $a \in L$, and an observable $\eta : S \times T \to \mathbb{R}^n_+$.

Examples:
— initial state (steady for non-stress), $a = m_1 + l_2$, $\eta_1 = S, A, B$,
— initial state, $a = m_1 + l_4 + l_1$, $\eta_2 = S, AS, ES$.

Observations

Queries (observations, properties) to the model under a scenario $\langle init, a, \eta \rangle$ are then expressed as CSL(η) formulas (where η is made part of the CSL observables, the so-called *atomic propositions*; or CTL in the deterministic case), *e.g.*:

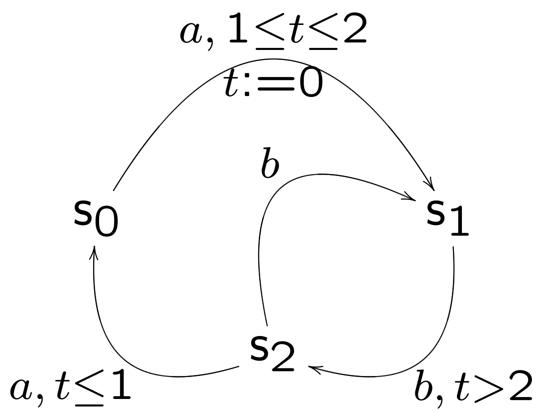
$$\mathcal{S}\langle init, l_2 + l_3, \eta_1 \rangle \models \mathbf{P}_{>.9}(\Diamond^{\leq 10'} S \geq 2S(0))$$

This is saying that in the said scenario with high probability the S-level will rise above twice the original level with 10 minutes.

3. What's a behaviour

A model specification model

Define a function (or macroscopic organization) of $\mathcal{S}$ as a finite prob/time automaton $\mathcal{A}$ (the states of which are called macro-states). Here is an example:



It is understood that actions have to be defined for a duration that exceeds the transition-time (so far we took safely the duration to be infinite). Resets can be useful to define pseudo-periodic behaviour (for oscillators for instance, think of the repressilator motive for instance).

CSL

That $\mathcal{A}$ can be seen as a generalised CSL formula glued to $\mathcal{S}$ via an rigid explicit map from $S(\mathcal{A}) \to \wp(S(\mathcal{S}))$ connecting macro-states to sets of microstates:

$$
\begin{aligned}
\mathsf{s}_0 &:= x_0 \in [0.01, 0.03] \wedge y_1 \leq 0.002 \wedge \overline{x}_1 \leq 0.01 \\
\mathsf{s}_1 &:= x_1 < 0.001 \wedge v_2 = v_1 \\
\mathsf{s}_2 &:= y_1 > 0.2 \wedge v_2 > 400 min^{-1}
\end{aligned}
$$

defined for instance in terms of intervals of concentrations, concentration derivatives, reaction speeds (jumps, threshholds) and their probabilistic variants (reaction scores, stochastic gradients). This may or may not be a partition.

In this reading macrostates are explicitely sets of microstates, and one can ask whether the obtained predicate is true, and whether the case studies lay wholly within this kind of logic.

What's a model

— building, tuning, analysing, predicting, interpreting

4. Steady states

Heat shock

$$
\begin{aligned}
&\xrightarrow{\beta_1} S && (1)\\
S &\xrightarrow{\beta_2} S + A && (2)\\
S &\xrightarrow{\beta_3} S + H && (3)\\
S &\xrightarrow{\beta_4} && (4)\\
A &\xrightarrow{\beta_5} && (5)\\
H &\xrightarrow{\beta_6} && (6)\\
S + A &\xrightarrow{\beta_7}_R SA && (7)\\
SA &\xrightarrow{\beta_8}_R S + A && (8)\\
SA + H &\xrightarrow{\beta_9} A + H && (9)\\
A + B &\xrightarrow{\beta_{10}}_R AB && (10)\\
AB &\xrightarrow{\beta_{11}}_R A + B && (11)
\end{aligned}
$$

$$AB \xrightarrow{\beta_{12}} A + B' \tag{12}$$

$$B' \xrightarrow{\beta_{13}} B \tag{13}$$

Steady state study

We turn now to a parameter agnostic study of the steady states of the system. One has, starting with the simpler linear relations:

$$\begin{array}{ll}
0 = \beta_3[S] - \beta_6[H] & [H]' = 0 \\
0 = \beta_2[S] - \beta_5[A] & [A_{tot}]' = 0 \\
0 = \beta_{12}[AB] - \beta_{13}[B'] & [B']' = 0 \\
0 = \beta_{10}[A][B] - (\beta_{11} + \beta_{12})[AB] & [AB]' = 0 \\
0 = \beta_7[S][A] - \beta_8[SA] - \beta_9[SA][H] & [SA]' = 0 \\
0 = \beta_{13}[B'] + \beta_{11}[AB] - \beta_{10}[A][B] & [B]' = 0 \\
0 = \beta_1 - \beta_4[S] - \beta_7[S][A] + \beta_8[SA] & [S]' = 0
\end{array}$$

Steady States (cont')

So

$$
\begin{aligned}
&[H] = \beta_3/\beta_6[S]\\
&[A] = \beta_2/\beta_5[S]\\
&[AB] = \beta_{13}/\beta_{12}[B']\\
&[B']/[B] = \beta_{12}\beta_{10}\beta_2/\beta_5\beta_{13}(\beta_{11} + \beta_{12})[S]\\
&[SA] = 1/\beta_8(\beta_4[S] + \beta_7\beta_2/\beta_5[S]^2 - \beta_1)\\
&\beta_7\beta_2/\beta_5[S]^2 = (\beta_8 + \beta_9\beta_3/\beta_6[S])[SA]
\end{aligned}
$$

replacing the penultimate line into the last obtains a cubic equation in $[S]$ which fixes the value of S, H, A, SA and the ratio B'/B but not the absolute value of B, B', AB which depends on the initial amount of $B_{tot} = B + B' + AB$ (since $[B_{tot}]' = 0$).

New needed modality for equilibrium.

5. heat shock lab book

hsp repair module analysis

In the repair module $\mathcal{R} : A \to B'$, A is representing a chaperone, B is an unfolded protein, and B' a properly folded one.

Default rates are $\alpha_6, \alpha_7, \alpha_8, \alpha_9^\star = 150, 0, 100, 10$ (starred rates are susceptible to change, *i.e.*, they are parameters, part of scenarios), reactions:

$$
\begin{aligned}
&A, B \leftrightarrow^{6,150}_{7,0} AB \\
&AB \to^{8,100} A, B' \\
&B' \to^{9,10^\star} B
\end{aligned}
$$

Heat shock scenarios are defined as:

— (Doyle) α_9 : $40 \to 200$, (thermal damage to folding), α_0 : $75 \to 100$ (S synthesis boost).

— (Srivastava) $\alpha_9 := 2\alpha_9$, $\alpha_0 := 20\tau_0$

Observables

Observables: B', and perhaps ancillaries S, A.

One has $\mathcal{R}/\mathcal{M} = A$, that is to say within the complete model $\mathcal{M}$, only A is visible; B' is made visible by fiat, because we declare it as an observable; AB is invisible, besides $d/dtA + AB = 0$.

At steady state, $v_6 = v_7 = v_8 = v_9$ all speeds are equal (0 if $v_7 = 0$ as above), and could give a measure of the turnover cost.

Response $\rho(t) = B'/B+B'$ depends on the control variable A; can we optimise ρ against some notion of cost (such as the turn-over of the model, defines as the sum of the reaction speeds ?)

Physiological ranges for states and rates.

Steady state experiments

Define (arbitrary) state $q_0 :=$ A=0,B=3000,AB=1000,B'=1000, one has (approximate values for states, have to be read as 10% intervals):

$$
\begin{array}{llllll}
e_1 & = & q_0 & \rightarrow^{4s}_{\alpha} & \text{A=500,B=0,AB=500,B'=4500} =: q_1 & \\
e_2 & = & q_0 & \rightarrow^{6s}_{\alpha_9=2\alpha_9} & \text{A=100,B=0,AB=900,B'=4100} =: q_2 & \alpha_9 \uparrow \\
e_3 & = & q_1 & \rightarrow^{6s}_{\alpha_9=2\alpha_9} & q_2 & \\
e_4 & = & q_0 & \rightarrow^{2s}_{\alpha_9=2\alpha_9, A:=+1500} & \text{A=2000,B=0,AB=500,B'=4500} & \\
 & & & & = q_1 + 1500A & A \uparrow
\end{array}
$$

Comments

— $\alpha_9 \uparrow$: $A, B' \downarrow$ steady state is less favourable reached more slowly;
— $A \uparrow$: $A \uparrow$, $B' =$, same B' steady state value reached faster; linear dependence in A (see analytic steady state study below);
— about the residual $A = 500$ in q_1: once coupled with the storage module $\mathcal{S}(A) = A, S \leftrightarrow^{10,?}_{11,?} AS$, superfluous As (which were intrumental in a fast response) will catch free Ss and therefore $d/dtA \downarrow$ (one can't take these superfluous A off the system, this is not digital programming!)

Note that all these are steady state (as opposed to transient ones) observations, rewritable in CTL:

$$
\begin{array}{lcll}
\mathcal{R}\langle q_0; \alpha\rangle & \models & \mathbf{S}^{4s}_{>.9} q_1 & \\
\mathcal{R}\langle q_0; \alpha_9 = 2\alpha_9\rangle & \models & \mathbf{S}^{6s}_{>.9} q_2 & \alpha_9 \text{ control} \\
\mathcal{R}\langle q_0; A := +1500\rangle & \models & \mathbf{S}^{2s}_{>.9} q_1 + 1500A & A \text{ control}
\end{array}
$$

Test the storage idea

Repair module has now to be coupled with $Synth(S, A, H)$ [0–5] and $Store(S, A)$ [10–11], *e.g.*, $e_5 = e_1 + \text{SA=20} + \alpha_{10}, \alpha_{11} \neq 0$ (though this is not going to make a big difference, unless one adds $S \to S, A$; note the linear structure on experiments).

The *storage idea* to test: reactions 6, 10 are competing for A:

$$\alpha_9 \uparrow \to A \downarrow \to v_{11}/v_{10} \uparrow \to S, A, AB \uparrow$$

but it is not clear whether the is significant.

Notebook

This obtains a formal experiment notebook.

Note that in this specific example we have been recording only steady state information but there is no need to restrict to these, as follows from the general model.

Conclusion

Publish the model:

— with actions, observations, and scenarios (the model's model)

— (OL) with behaviour (the model's specification and connecting map)

perhaps even with:

— with notebook (the series of experiments building the model, scanning the parameter space)