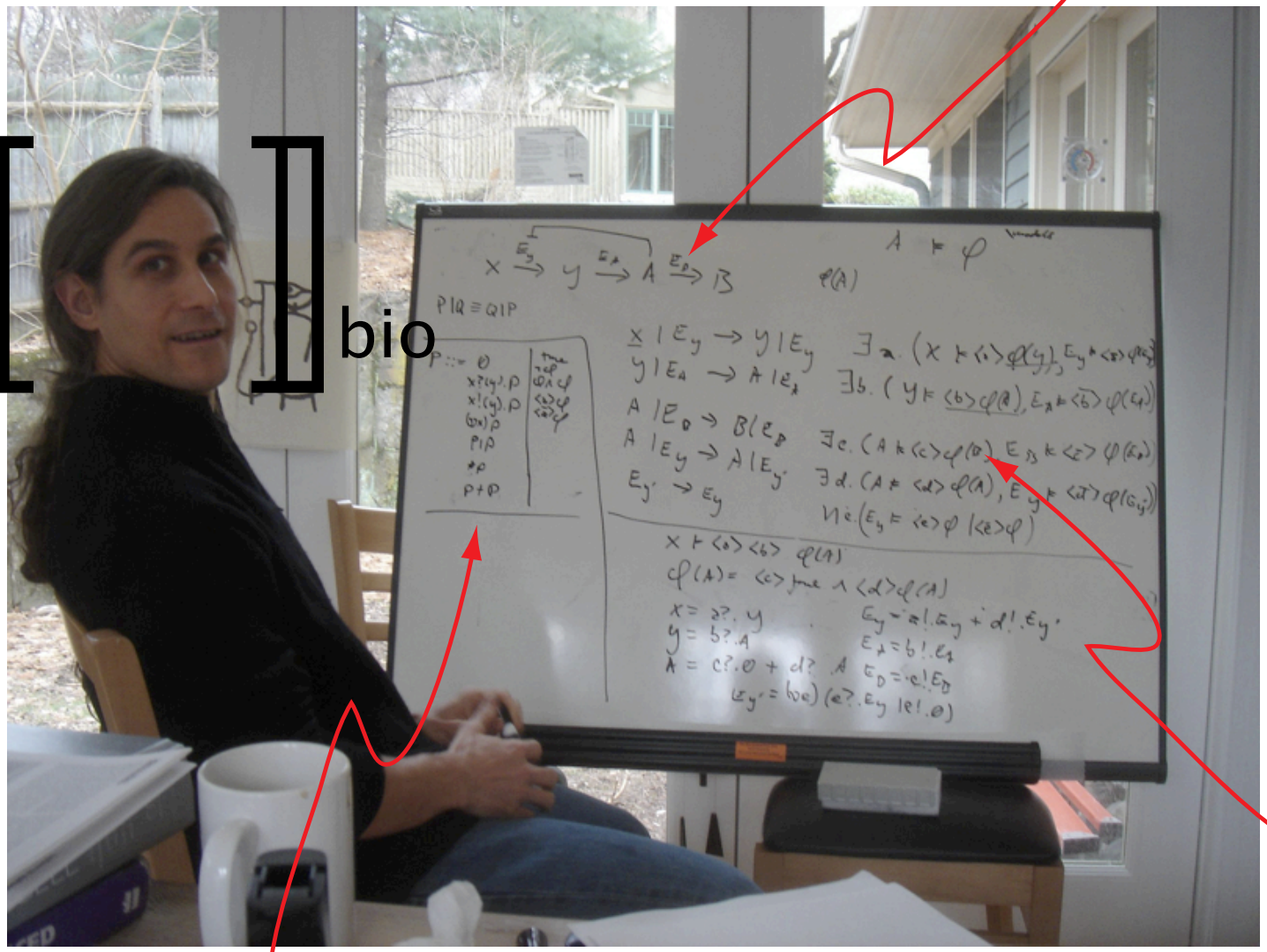
Logic and Systems Biology:

To Be Announced.

Walter Fontana
Department of Systems Biology
Harvard Medical School

LICS 2005, 06/29/05, Chicago

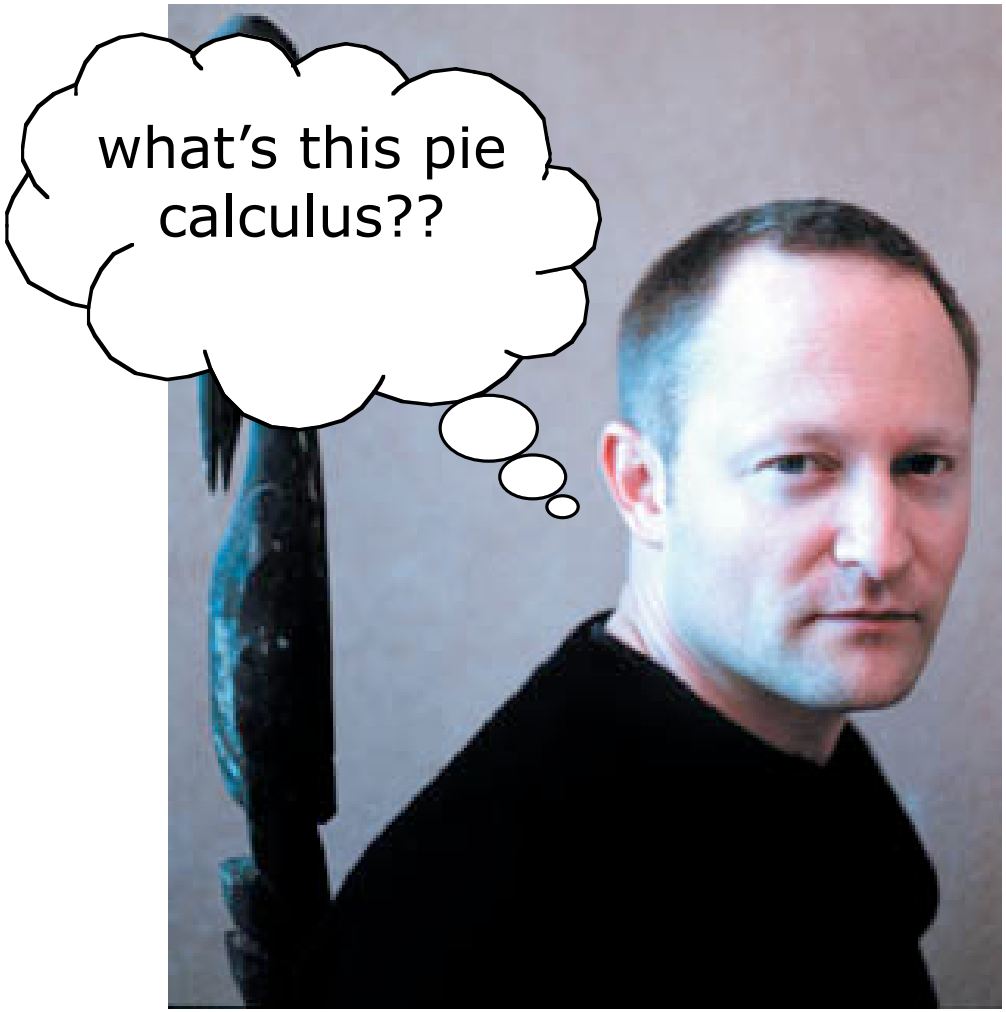
Greg Meredith, Djinnisys Corp.



Vincent Danos, Paris

Leo Buss, Yale

David Krakauer, Santa Fe Institute



slide intentionally left blank

the goal of systems biology is a system* for biology

(*) an intellectual framework with a theoretical foundation that engages experiment

At the least, such a foundation should enable

- an abstract characterization of "organization"
- a formal way of reasoning about the possible

Human Genome Project

Human Genome Organization (HUGO) formed

1988

Cystic-fibrosis gene identified by positional cloning

1989

1990

2003

Francis S. Collins *et al., Nature,* 422, 835-847 (2003)

whole genome sequencing had two fundamental consequences:

- it changed the conditions for gene identification
- it made the mechanisms of evolution observable

E.coli	5,533
mouse	30,000
mosquito	14,000
yeast (fission)	4,824
yeast (budding)	6,294
homo sapiens	25,000
arabidopsis	25,498
fly	14,100
worm	19,099
mycoplasma	468

gene numbers...

25,000 genes (ORFs)
10^{14} cells
> 200 cell types (??)

1,543 receptors (367 GPCRs)
518 protein kinases
150 protein phosphatases
1,850 transcription factors

source: Papin et al, Nature Reviews Molecular Cell Biology, 6, 99-111 (2005)

post-transcriptional processing: *alternative protein sequences*

- 40-60% of genes subject to alternative splicing
- 2.5 splice variants/gene on average

post-translational processing: *protein state*

- phosphorylation, methylation, acetylation, etc.
- 2.5 modifications / protein

complex formation: *assemblies*

- n proteins make n(n+1)/2 homo and heterodimers;
  224 proteins would generate enough "state" to control all genes

microRNA: *another kind of "gene"*

the external world

15 receptors per cell (1%) ?

... makes for 2^{15}=32,768 combinations (ligand bound/unbound)

the old picture of

one gene...one product...one function

is wrong.

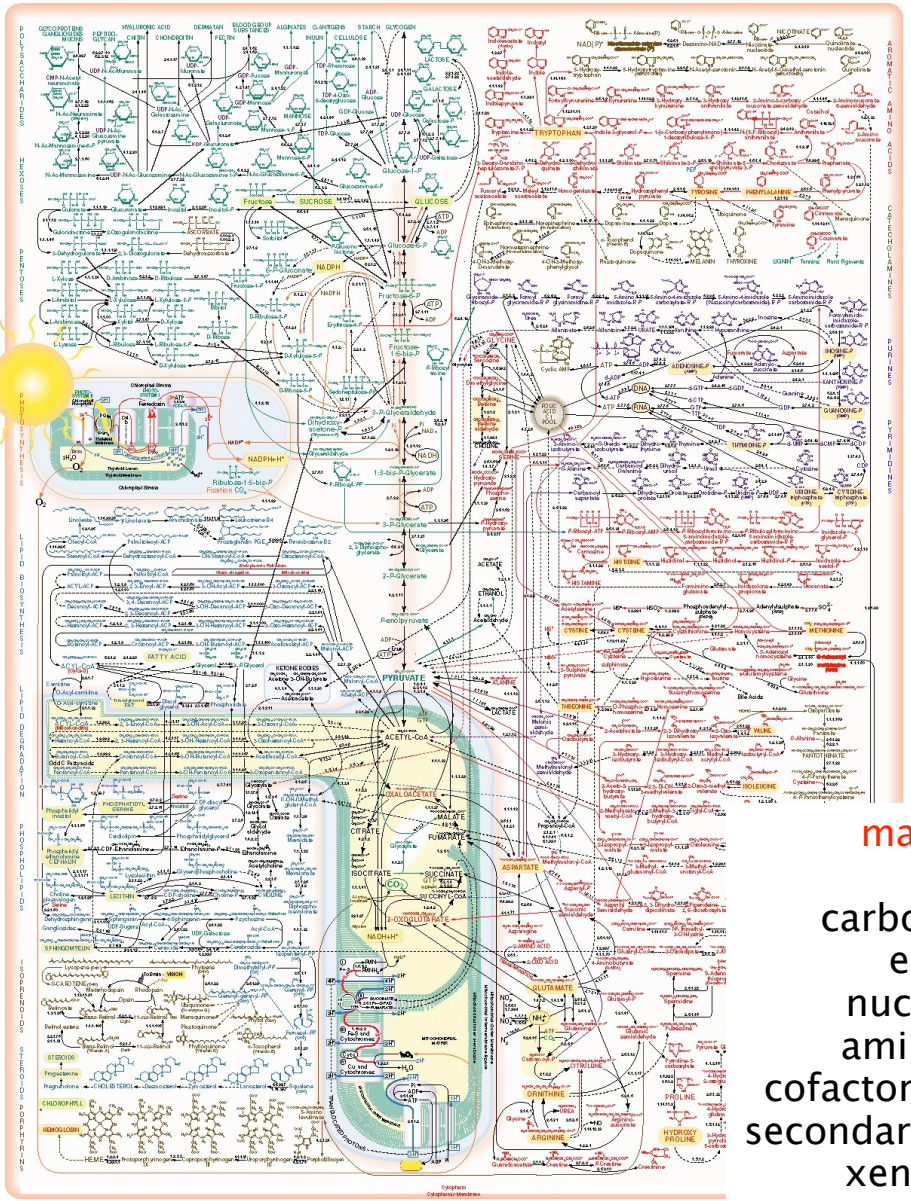
there is no molecular cell-biological version of “mentalese”

Systems biology emphasizes the dynamic and distributed nature of biological systems.

Be aware, however, that networks differ dramatically in kind.

For example,...

metabolic networks: plasticity of flow, rigidity of network



specialized enzymes transform substrates (synthesis and break-down)

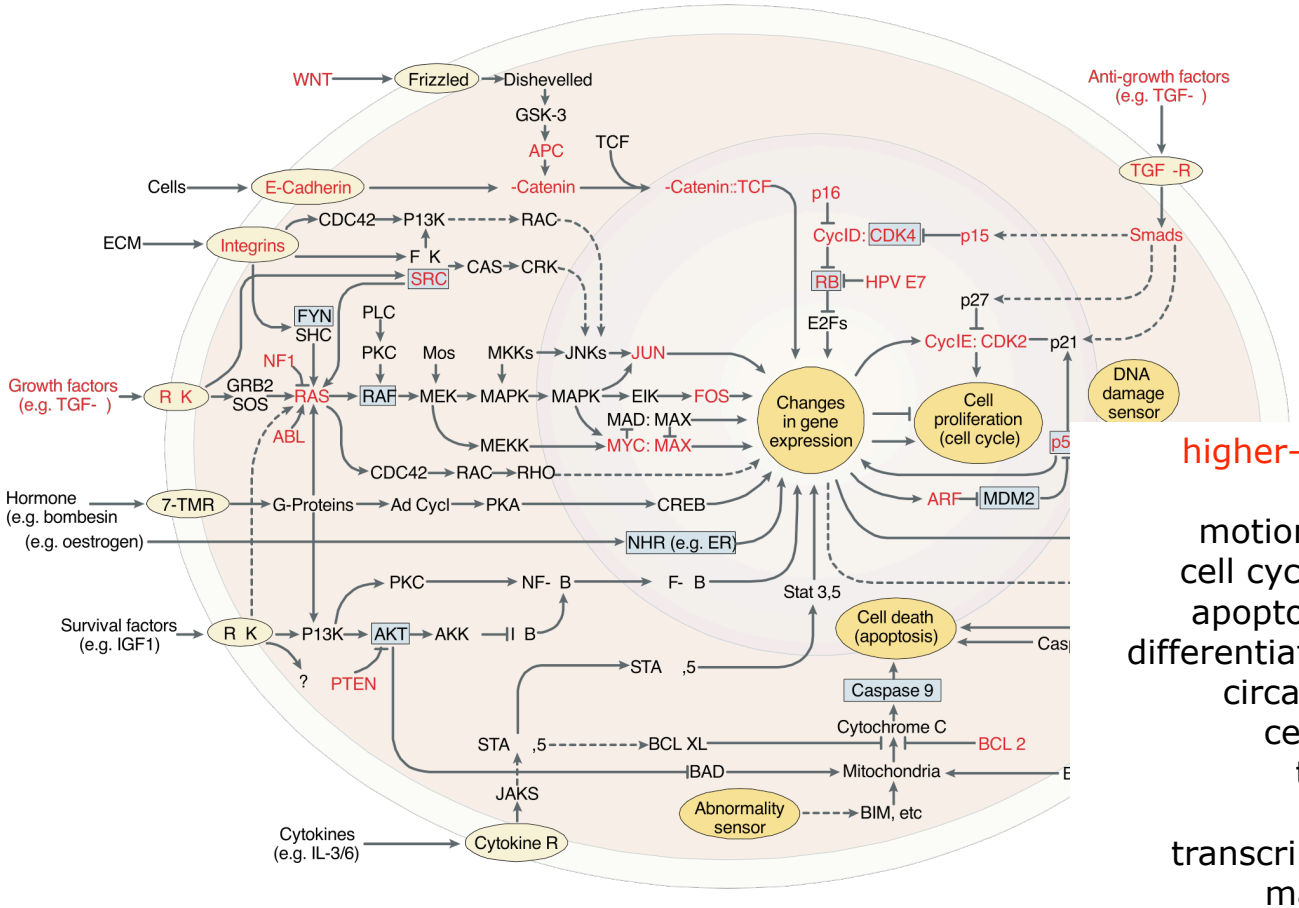
network *flows* are highly adaptive, but network structure is constrained by chemistry

materials:

carbohydrates
energy
nucleotides
amino acids
cofactors & vitamins
secondary metabolites
xenobiotics

no major chemical transformations.
proteins modify the state of other proteins.

networks are reconfigurable.
communication by physical contact: molecular recognition.



higher-order behavior:

motion (chemotaxis)
cell cycle (cell division)
apoptosis (cell death)
differentiation (development)
circadian rhythms
cell adhesion
transport
repair
transcription/translation
maintenance

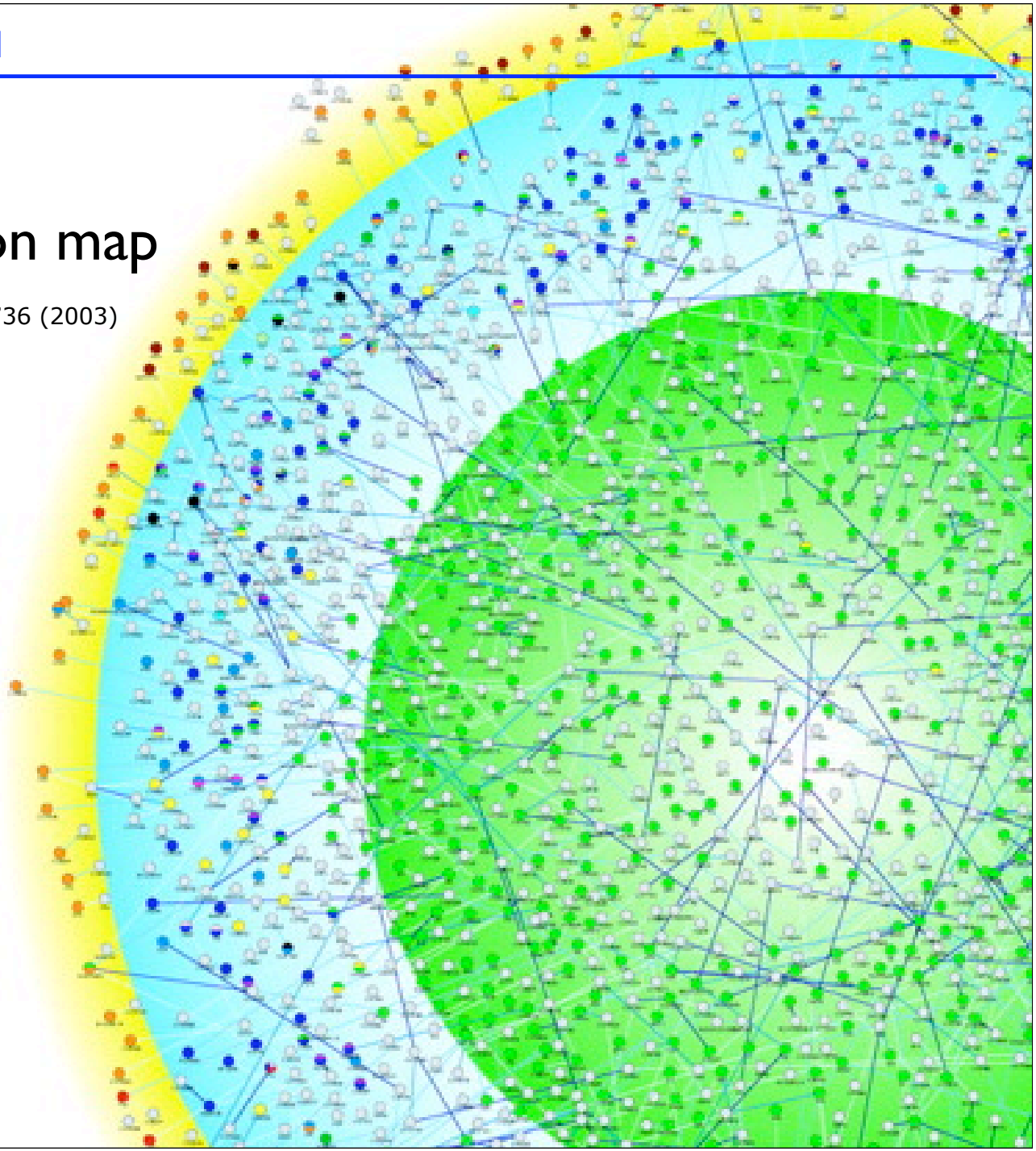
a systems biology hairball

Drosophila protein interaction map

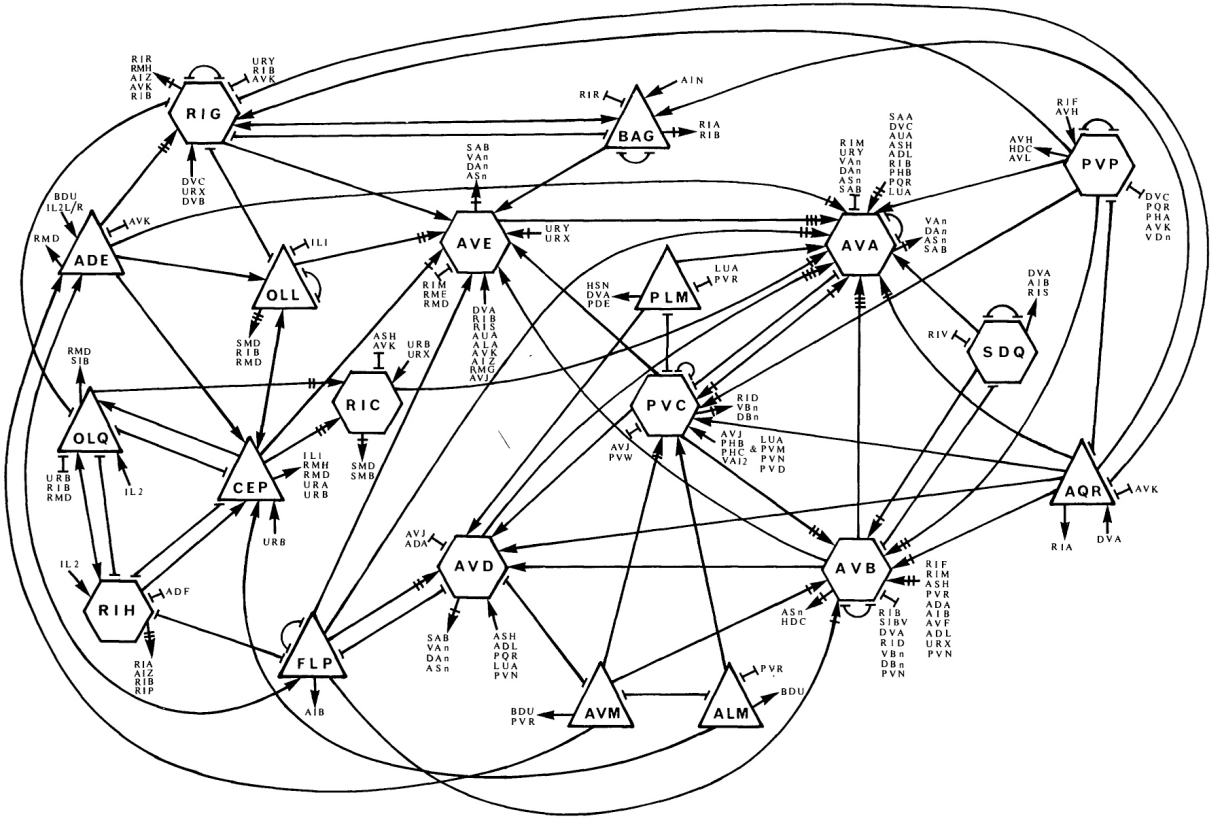
Giot et al, *Science*, **302**, 1727-1736 (2003)

total:
20,405 interactions
7,048 proteins
(of 13,656 coding loci)

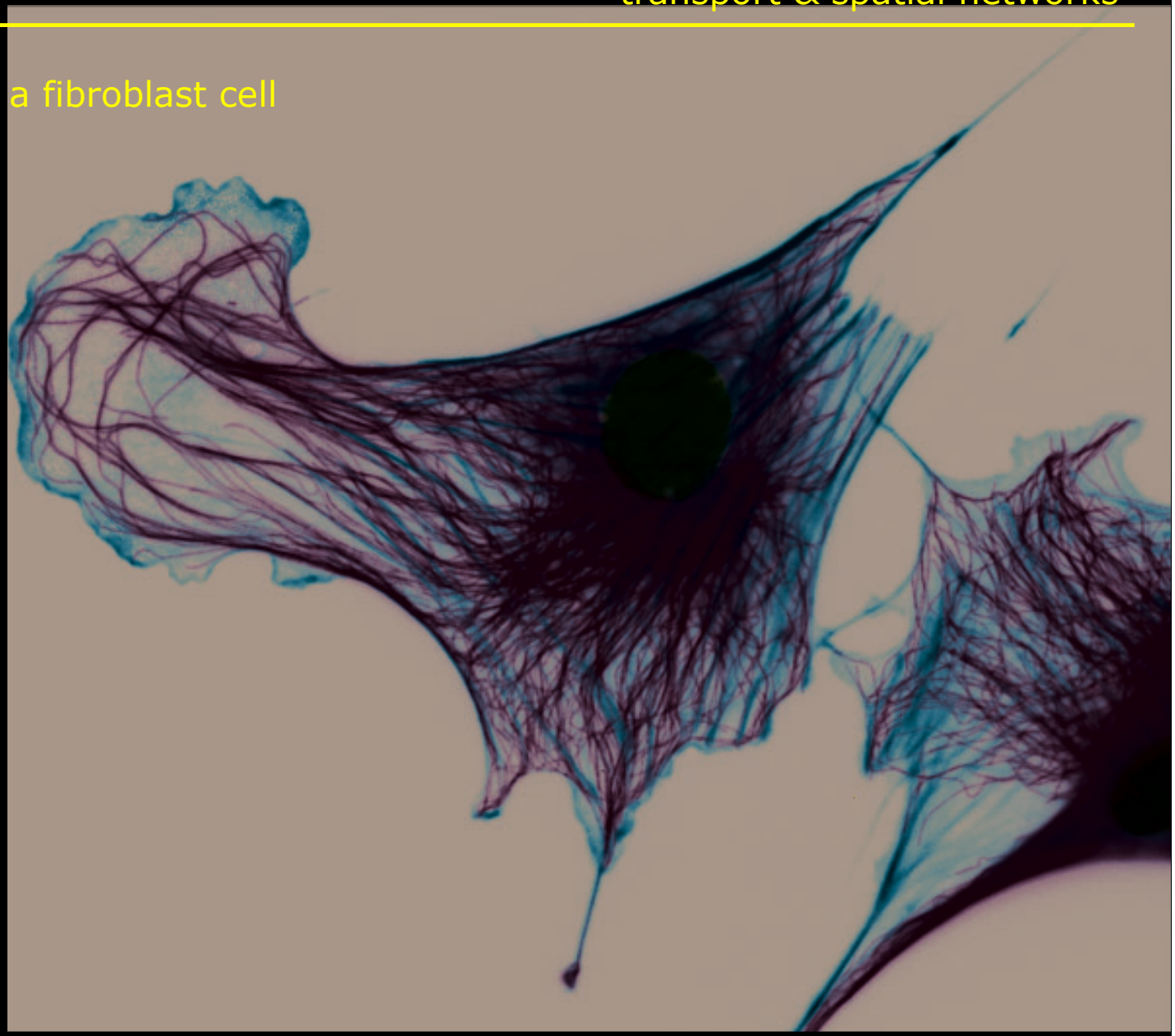
high confidence:
4,780 interactions
4,679 proteins



cells connect to cells in arbitrary point-to-point contacts.
network structure is extremely plastic.
(network structure might be constrained by packing the wires...)



the cytoskeleton in a fibroblast cell



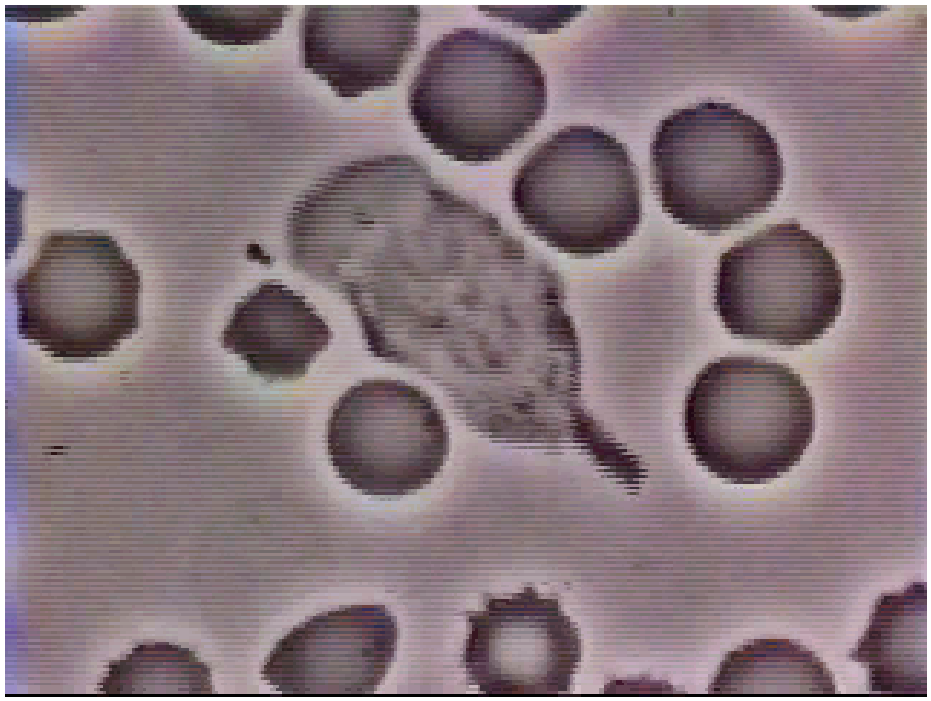
actin

microtubules

DNA

courtesy: Le Ma, Harvard Medical School

neutrophil chasing a bacterium



movie by Tim Stossel, Harvard Medical School

- we know a lot (but still not enough - and how do we know?) about structure at the component and the network level

- we know very little about "behavior" at the network level

- we know next to nothing about the "why?"
  (to invoke "survival" is useless for explaining why the networks we observe have the structure and behaviors they have.)

two powerful (yet misleading) reductionisms in biology:

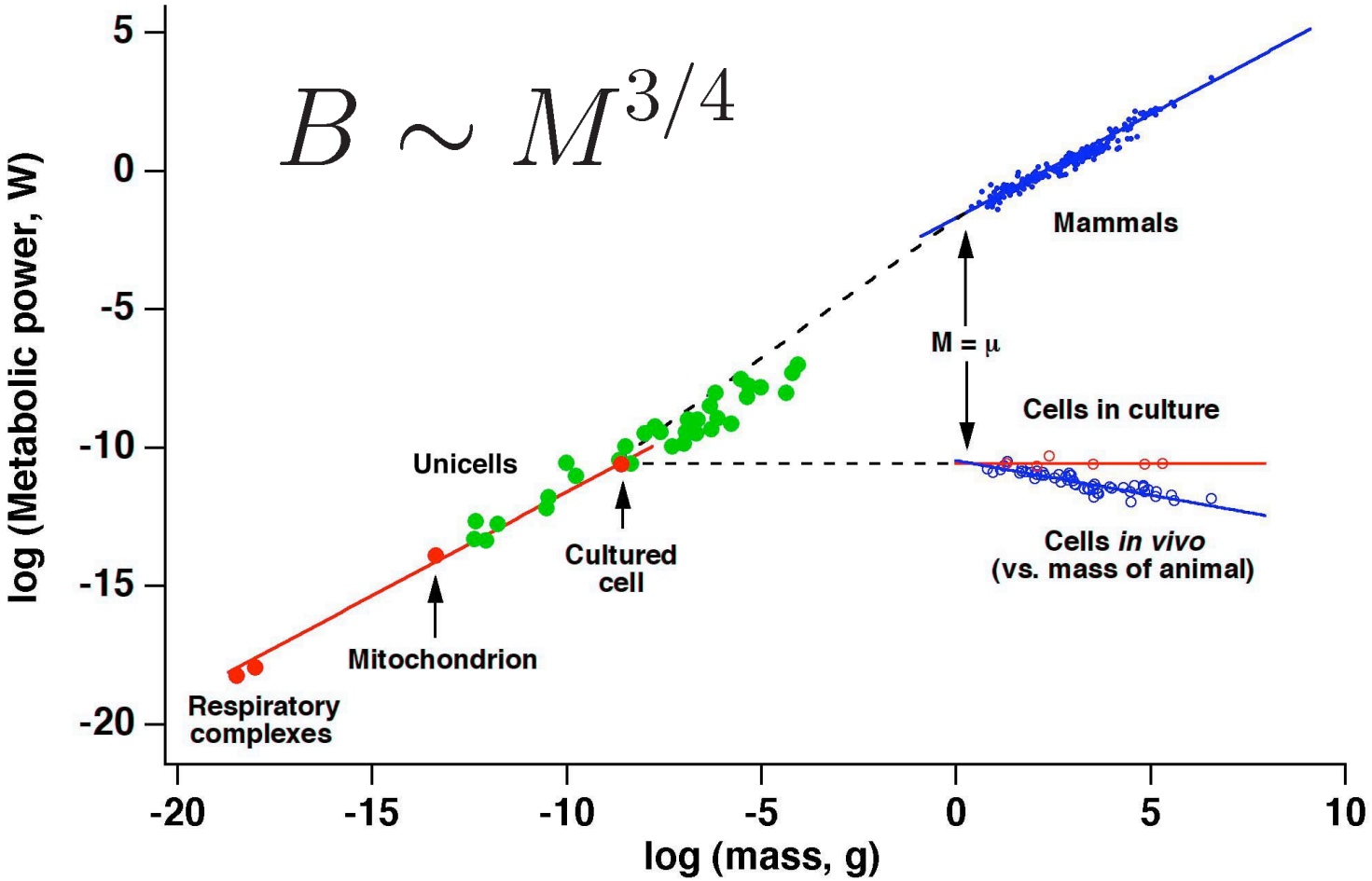
natural selection:
all aspects of biological organization are the outcome of selection

(selection is not generative: it dispenses of the unfit, but does not address the origin of novelty; the issue is "discoverability")

chemistry:
all biological properties are realized by vastly complex combinations of molecular properties

(the issue is about useful levels of description; think emergence; think coarse-graining.
organizations are often more stable than their components; autonomy)

Allometric scaling: 3/4 power law of energy consumption versus body mass holds over nearly 30 (!) orders of magnitude



G.West et al., Santa Fe / Van Savage, Bauer Center, Harvard

"Wir müssen wissen. Wir werden wissen."

memory

if everything turns over, how is persistence ensured?

robustness

what are its specifications and implementations?

evolvability

what about molecular systems promotes and constrains variation and innovation?

emergence

how can different levels of description of the same system co-exist?
how do new levels (observables or "contracted descriptions") arise?

aging

why do living systems lose informational integrity? is this unavoidable?

niche construction

living systems (or parts thereof) modify their environment, transforming natural selection pressures. how do we need to (re)think evolution?

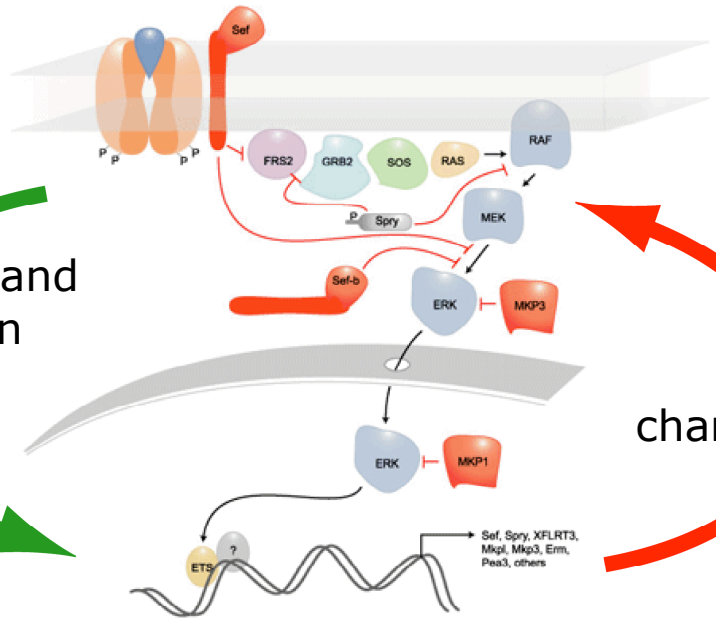
synthetic biology

the quest for programmable matter

molecular adaptation, short and long-term memory: signaling + gene expression

fast

network interprets signals and changes gene expression



gene expression changes network structure

slow

"learning" here means stably reconfiguring the signaling circuitry in response to signals

"Wir müssen wissen. Wir werden wissen."

memory

if everything turns over, how is persistence ensured?

robustness

what are its specifications and implementations?

evolvability

what about molecular systems promotes and constrains variation and innovation?

emergence

how can different levels of description of the same system co-exist?
how do new levels (observables or "contracted descriptions") arise?

aging

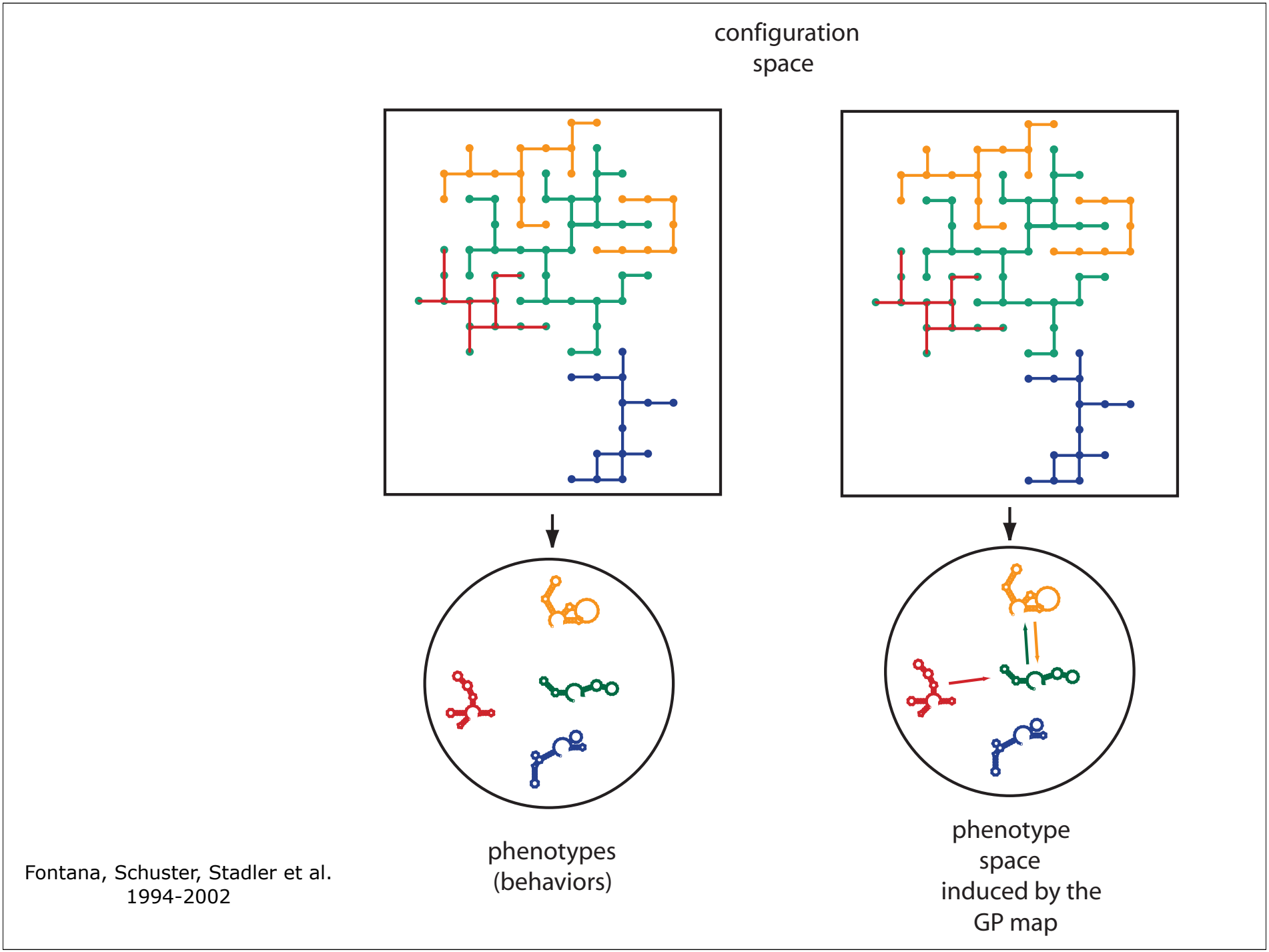
why do living systems lose informational integrity? is this unavoidable?

niche construction

living systems (or parts thereof) modify their environment, transforming natural selection pressures. how do we need to (re)think evolution?

synthetic biology

the quest for programmable matter


configuration
space
phenotypes
(behaviors)
phenotype
space
induced by the
GP map
Fontana, Schuster, Stadler et al.
1994-2002

"Wir müssen wissen. Wir werden wissen."

memory

if everything turns over, how is persistence ensured?

robustness

what are its specifications and implementations?

evolvability

what about molecular systems promotes and constrains variation and innovation?

emergence

how can different levels of description of the same system co-exist?
how do new levels (observables or "contracted descriptions") arise?

aging

why do living systems lose informational integrity? is this unavoidable?

niche construction

living systems (or parts thereof) modify their environment, transforming natural selection pressures. how do we need to (re)think evolution?

synthetic biology

the quest for programmable matter

What is a model?

Source: style.com
Photo: Marcio Madeira

Today, modeling in systems biology is heavily geared towards a pragmatic pursuit of quantitative prediction.
There is nothing wrong with that, but it remains to be seen whether this pursuit generates a productive interaction with a commitment to explanatory principles.

obvious challenges:

particle numbers: stochastic to deterministic regime
time scales: 10^{-3} s (kinases) to 10^{4} s (cell growth)
spatial aspects
too much state (not necessarily too many players)

[...] things, rather than obeying laws, create them through their modes of interaction. Alfred North Whitehead (1933)

the limits of dynamical systems

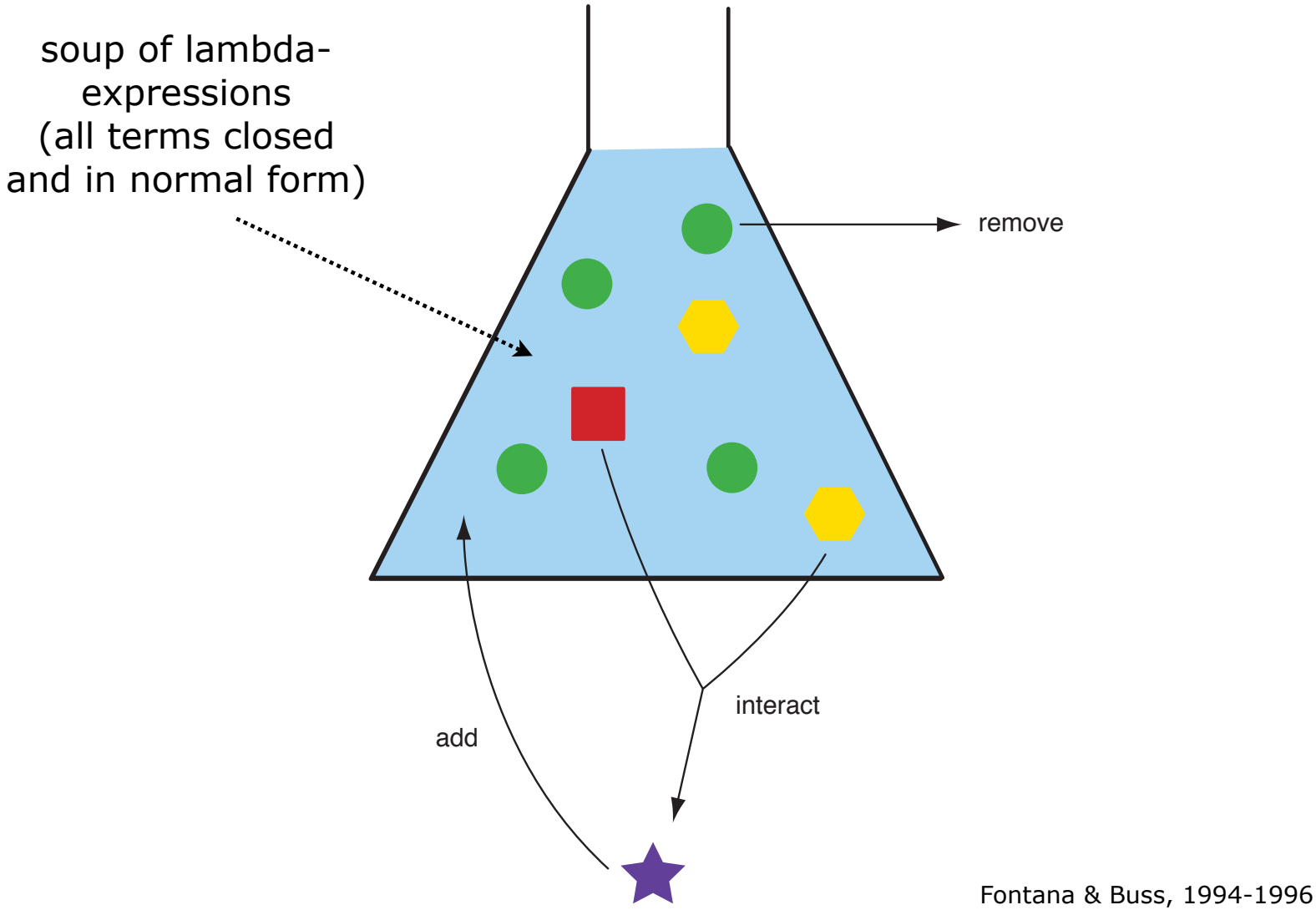
DS represent agents only in terms of the interactions that exist and are assumed relevant in a given context; they do *not* represent agents in terms of *potential* interactions that are not causally effective now, but would become so if context was changing.

To describe agents in these terms requires a representation of their structure, which holds the key to possible interactions that can become actualized in other contexts.

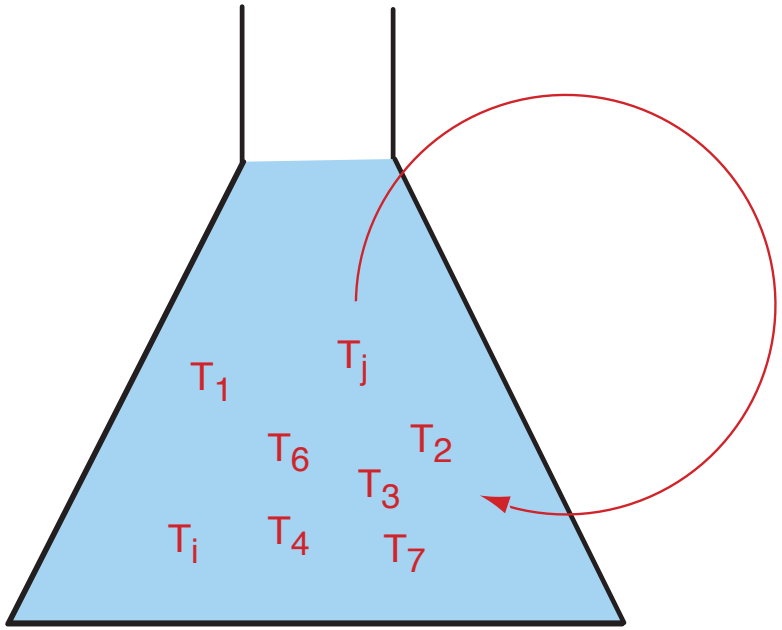
It is an open question to what extent *physical* structure and modes of interaction can be lifted into a formal logical representation that permits such reasoning in the molecular realm.

The bottleneck in biology is not quantification, but description. (ouch!)

Turing Gas or Church Soup or AlChemy

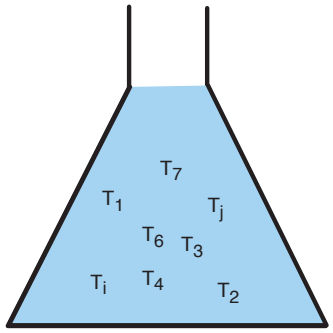


Fontana & Buss, 1994-1996



Self-maintenance is the consequence of a constructive feed-back loop:
it occurs when the construction processses induced by the constituents
of a system permit the continuous regeneration of these same constituents.

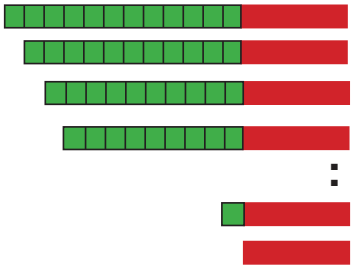
Immanuel Kant (Kritik der Urteilskraft, 1790): "[...] an organized product of nature is one in which all is end and, reciprocally, means too."



self-maintaining ensemble

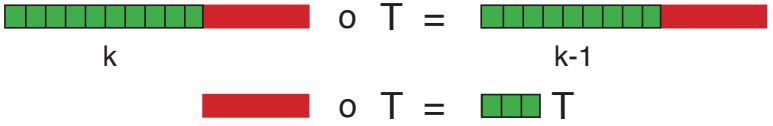
parse all terms into prefixes (not a λ-term)
and terminals (closed λ-term)

characteristic grammar

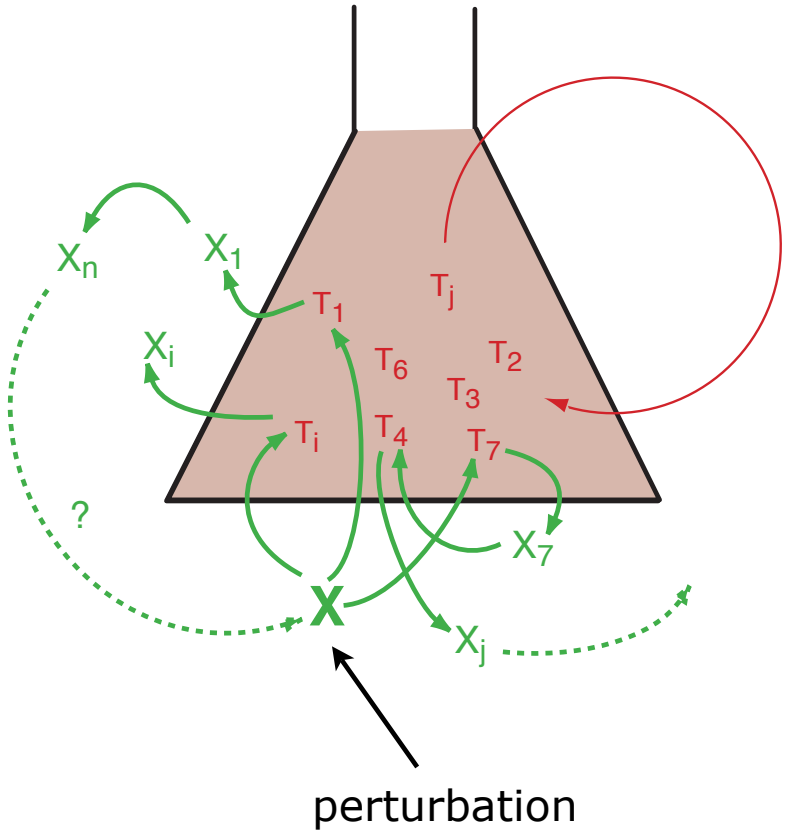


derive the transformational behavior of
building blocks and describe all interactions
in terms of a set of rewrite rules

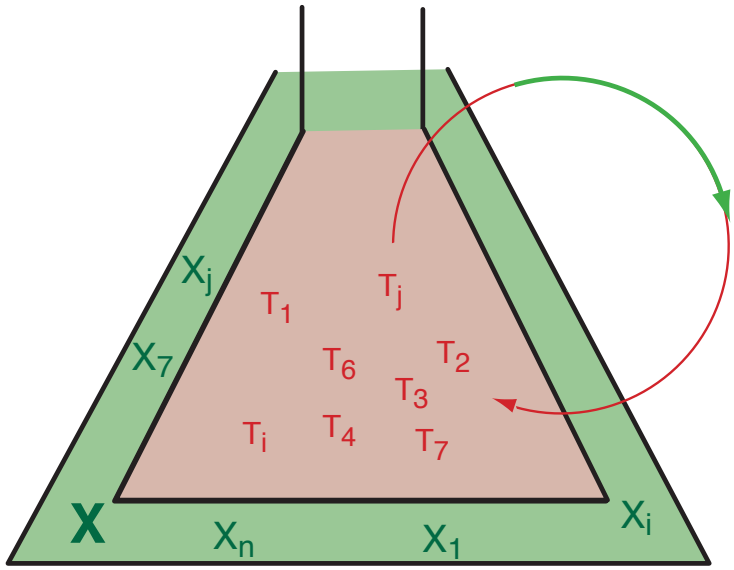
yields a specific algebraic structure



all reference to the underlying micro-mechanics
(λ-calculus) has been removed



perturbation



Applying a function to an argument in a typed calculus ...

$$\frac{A \longrightarrow B \qquad A}{B}$$

... corresponds to *modus ponens* in logic

This correspondence is known as the Curry-Howard isomorphism:

type σ $\longleftrightarrow$ logical formula (predicate) σ

λ-term of type σ $\longleftrightarrow$ proof of σ

so, our picture of chemistry becomes....

molecular shape	chemical predicate σ (statement about a chemical action)
molecule with that shape	proof of the validity of σ
chemical reaction	rule of inference
reactants / products	(proofs of) premises / conclusions

A more formal approach to agent-based systems is the π-calculus.
Its application to biology was pioneered by Shapiro, Regev, and Priami.
Its potential for biology was considerably deepened by Danos and Laneve.

The π-calculus
(Milner, Walker and Parrow 1989)

- a program specifies a network of interacting processes
- processes are defined by their potential communication activities
- communication occurs on complementary channels, identified by names
- message content: channel name

Processes and channels

Events

	$P,\ Q,\ \cdots$	process names (2.1)
	$x,\ y,\ \cdots$	channel names (2.2)
	$\overline{x},\ \overline{y},\ \cdots$	channel co-names (2.3)
$\pi ::=$	x	communication on channel name x (2.4)
	$\overline{x}$	communication on channel co-name x (2.5)
	$x(y)$	receive y along x (2.6)
	$\overline{x}\langle y\rangle$	send y along x (2.7)

Process syntax

$P ::=$	$P_1 \mid P_2$	parallel processes (2.8)
	$\pi\ .\ P_1$	sequential prefixing by communication (2.9)
	$\pi_1\ .\ P_1\ +\ \pi_2\ .\ P_2$	mutually exclusive communications (2.10)
	$(\mathsf{new}\ x)P$	new communication scope (2.11)
	0	inert process (2.12)

Structural congruence

$P \mid Q \equiv Q \mid P$	commutativity of PAR (2.13)
$(P \mid Q) \mid R \equiv P \mid (Q \mid R)$	associativity of PAR (2.14)
$P\ +\ Q \equiv Q\ +\ P$	commutativity of summation (2.15)
$(P\ +\ Q)\ +\ R \equiv P\ +\ (Q\ +\ R)$	associativity of summation (2.16)
$(\mathsf{new}\ x)0 \equiv 0$	scope of inert processes (2.17)
$(\mathsf{new}\ x)(\mathsf{new}\ y)P \equiv (\mathsf{new}\ y)(\mathsf{new}\ x)P$	multiple communication scopes (2.18)
$((\mathsf{new}\ x)P) \mid Q) \equiv (\mathsf{new}\ x)(P \mid Q)$ if $x \notin FN(Q)$	scope extrusion (2.19)
$A(\vec{y}) \equiv \{\vec{y}/\vec{x}\}Q_A$	recursive parametric definition (2.20)
$x(y).P = x(z).(\{z/y\}\ P)$ if $z \notin FN(P)$	renaming of input channel y (2.21)
$(\mathsf{new}\ y).P = (\mathsf{new}\ z).(\{z/y\}\ P)$ if $z \notin FN(P)$	renaming of restricted channel y (2.22)

Reaction rules

$(\cdots + \overline{x}\langle z\rangle.Q) \mid (\cdots + x(y).P) \rightarrow Q \mid P\{z/y\}$	communication (COMM)(2.23)
if $P \rightarrow P'$ then $P \mid Q \rightarrow P' \mid Q$	reaction under parallel composition (2.24)
if $P \rightarrow P'$ then $(\mathsf{new}\ x)P \rightarrow (\mathsf{new}\ x)P'$	reaction within restricted scope (2.25)
if $Q \equiv P, P \rightarrow P',$ and $P' \equiv Q'$ then $Q \rightarrow Q'$	reaction up to structural congruence (2.26)

$$E + S \rightleftharpoons ES \longrightarrow E + P$$

- Says nothing about internal structure of E, S, P, ES
- We want to encode the reaction scheme… but according to certain principles

$$[\![\ E + S \rightleftharpoons ES \longrightarrow E + P\ ]\!]_\pi$$

- $[\![- \rightarrow_{CHEM} -]\!]_\pi = - \rightarrow_\pi^* -$
- $[\![- +_{CHEM} -]\!]_\pi = - \mid -$

L.Greg Meredith (2005)

$$E + S \rightleftharpoons ES \longrightarrow E + P$$

from these we deduce

- $[\![E + S]\!]_\pi = [\![E]\!]_\pi \mid [\![S]\!]_\pi \rightarrow_\pi^* [\![ES]\!]_\pi$
- $[\![ES]\!]_\pi \rightarrow_\pi^* ([\![E]\!]_\pi \mid [\![S]\!]_\pi) + ([\![E]\!]_\pi \mid [\![P]\!]_\pi)$

from these we deduce

- $\exists x_0.([\![E]\!]_\pi \approx (\upsilon\, e)(x_0[e].[\![E]\!]_\pi' + X_E)) \;\&\; ([\![S]\!]_\pi \approx x_0(y).[\![S]\!]_\pi' + X_S)$
- $[\![ES]\!]_\pi \approx (\upsilon\, e)([\![E]\!]_\pi' \mid [\![S]\!]_\pi'\{e/y\})$

therefore

- $(\upsilon\, e)([\![E]\!]_\pi' \mid [\![S]\!]_\pi'\{e/y\}) \rightarrow_\pi^* ([\![E]\!]_\pi \mid [\![S]\!]_\pi) + ([\![E]\!]_\pi \mid [\![P]\!]_\pi)$

L.Greg Meredith (2005)

$$E + S \rightleftharpoons ES \longrightarrow E + P$$

since E is an enzyme, $[\![E]\!]_\pi$ is the future of $[\![E]\!]_\pi'$,
and $[\![S]\!]_\pi$ and $[\![P]\!]_\pi$ are the futures of $[\![S]\!]_\pi'\{e/y\}$

- $(\upsilon\ e)([\![E]\!]_\pi' \mid [\![S]\!]_\pi'\{e/y\}) \rightarrow_\pi^* ([\![E]\!]_\pi \mid [\![S]\!]_\pi) + ([\![E]\!]_\pi \mid [\![P]\!]_\pi)$

implies

- $\exists x_1 x_2.([\![E]\!]_\pi' \approx x_1(y).[\![E]\!]_\pi + x_2(y).[\![E]\!]_\pi + X_E) \,\&\, ([\![S]\!]_\pi' \approx x_1[e].[\![S]\!]_\pi + x_2[e].[\![P]\!]_\pi + X_S)$

setting ***X's*** to ***0*** and minimizing the number of $\rightarrow_\pi$ steps
we arrive at

- $[\![E]\!]_\pi = (\upsilon\ e)(x_0[e].(x_1(y).[\![E]\!]_\pi + x_2(y).[\![E]\!]_\pi))$
- $[\![S]\!]_\pi = x_0(y).(x_1[e].[\![S]\!]_\pi + x_2[e].[\![P]\!]_\pi)$

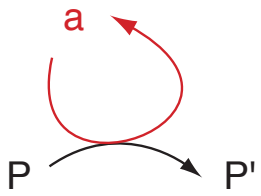
L.Greg Meredith (2005)

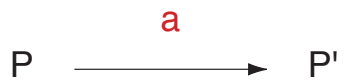
$$E + S \rightleftharpoons ES \longrightarrow E + P$$

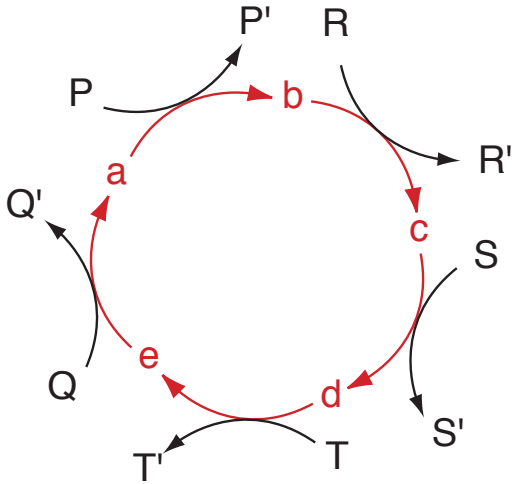
is really a reaction (network) type.

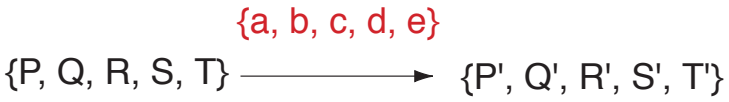
- use spatial logic (L.Caires) to capture the logical content (the characteristic formula F) of the process corresponding to this reaction
- translate biological networks into pi-processes x_i
- model-check F against the x_i
- thus identify networks with a (possibly dynamic) communication structure that behave like F (have that type)

The logic formula,
"the largest process X that behaves in some way and eventually becomes X",
describes the type "catalyst", which picks out the following red processes:

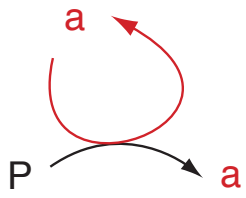


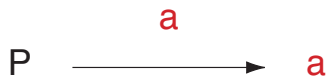


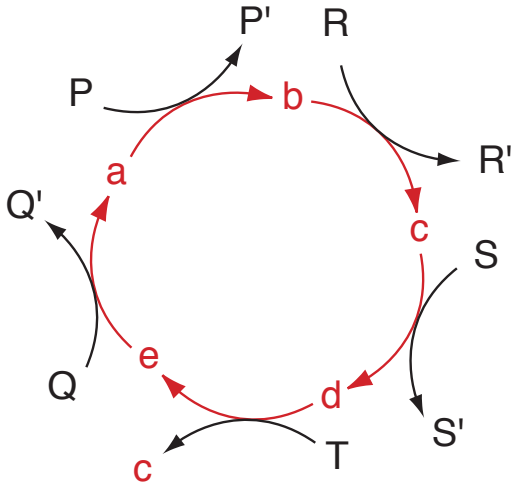


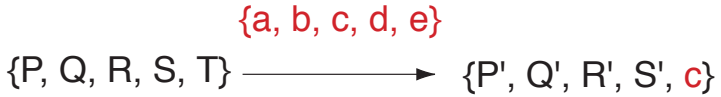


The spatial logic formula,
"the largest process X that behaves in some way and eventually becomes X|X",
describes the type "autocatalyst", which picks out the following red processes:

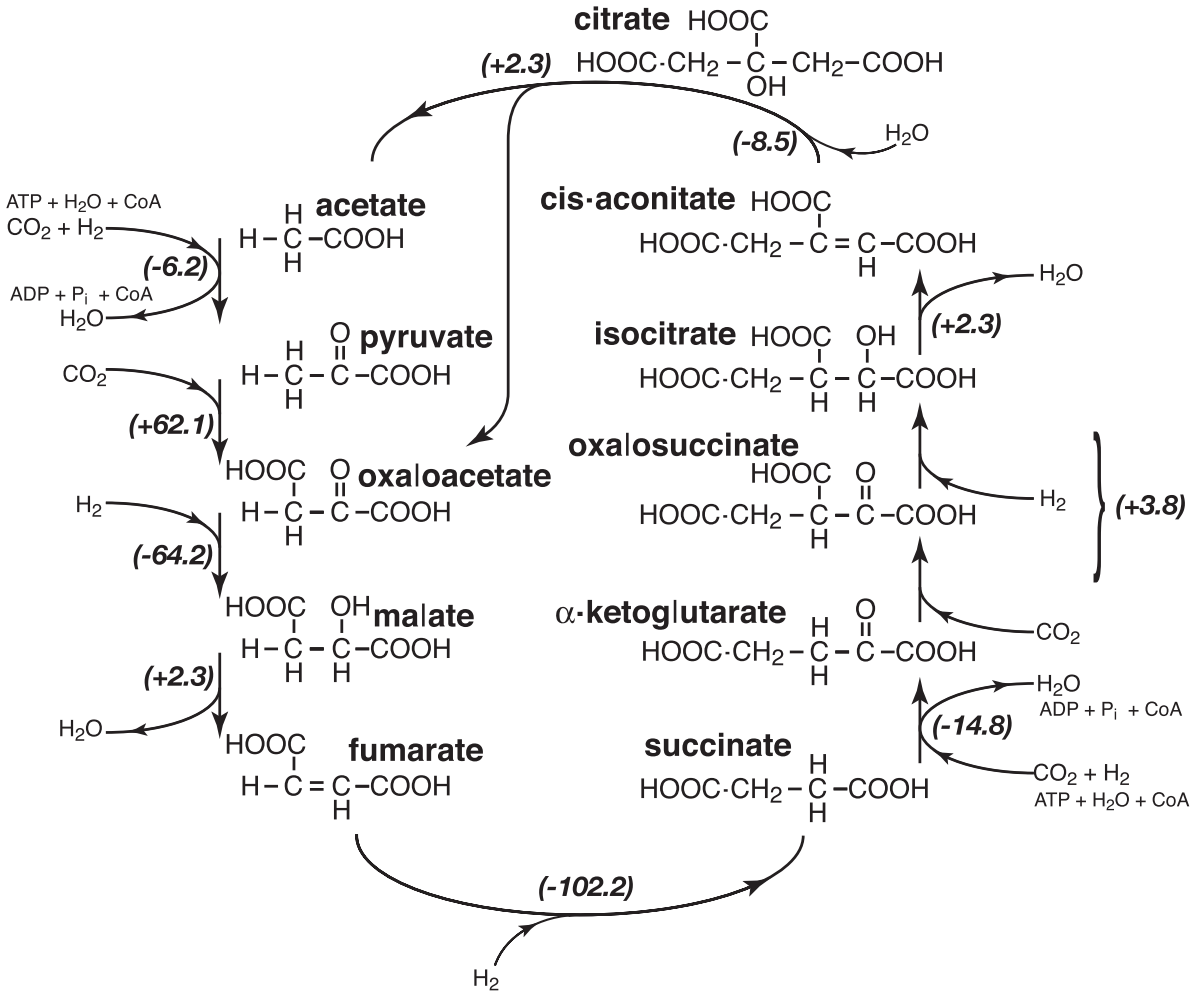








an autocatalytic network at the dawn of life ?

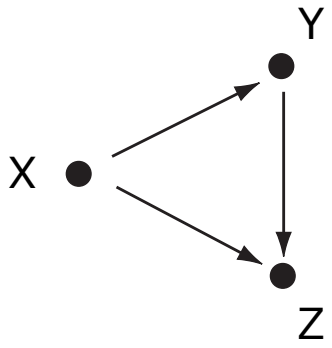


Eric Smith & Harold Morowitz, PNAS, 101, 13168-13173 (2004)

The search for networks that inhabit certain types is important, because it extends current efforts at detecting network motifs.

Such efforts focus on syntactical motifs, but network types are behavioral motifs!

- detect whether, in a network, certain subgraphs occur more frequently than expected (expectation means a suitably randomized control)
- those that do are presumably solutions to some problem(s)
- figure out the problem(s)

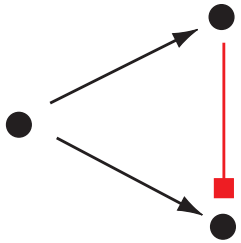


"coherent"

most abundant in E.coli and yeast

example:
feed-forward loop

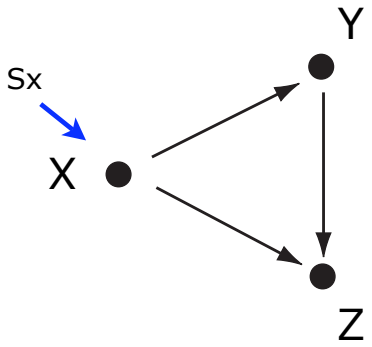
S. Shen-Orr, R. Milo, S. Mangan, U. Alon,
Nature Genetics, **31**, 64 (2002)

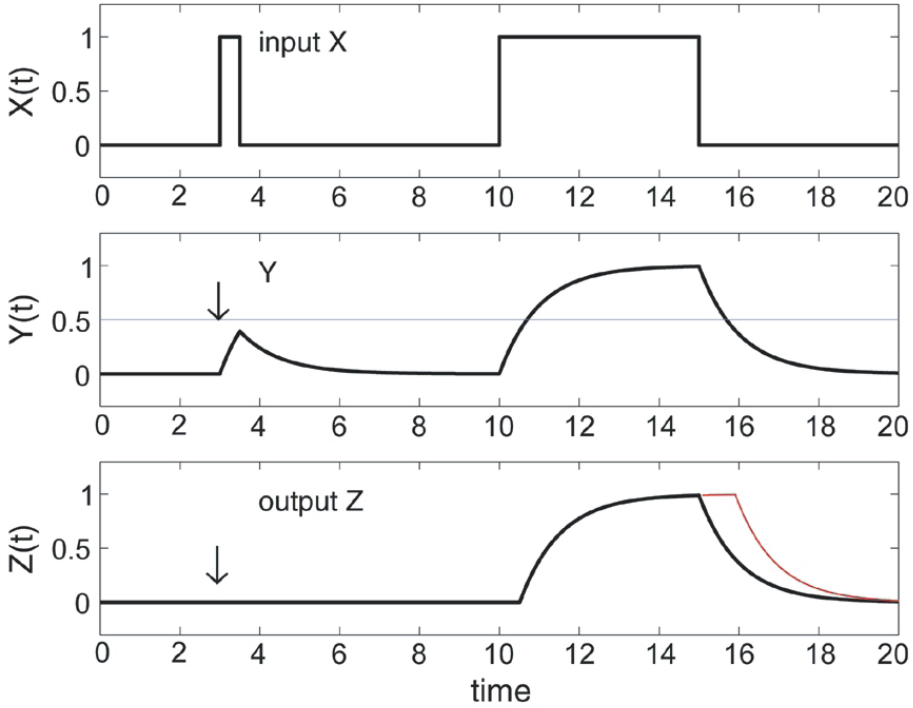


"incoherent"

fairly abundant in yeast

a delay mechanism...

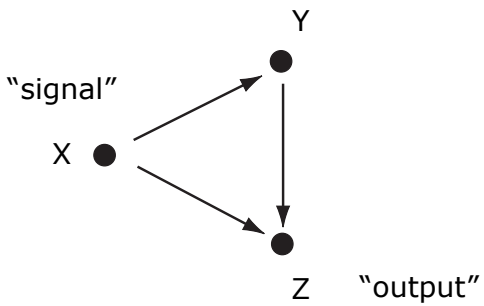




...implementing a pulse-filter

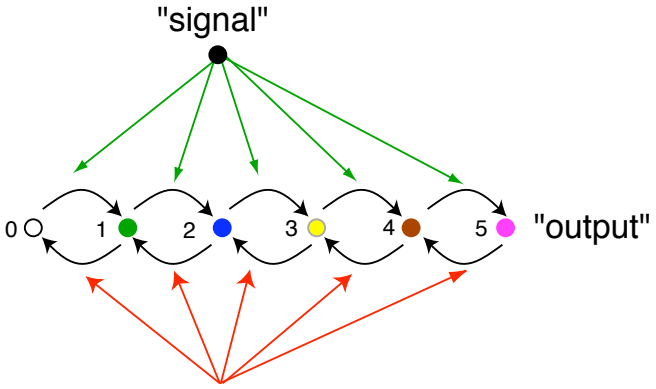
are these types expressible and checkable?

is "delay" a type?



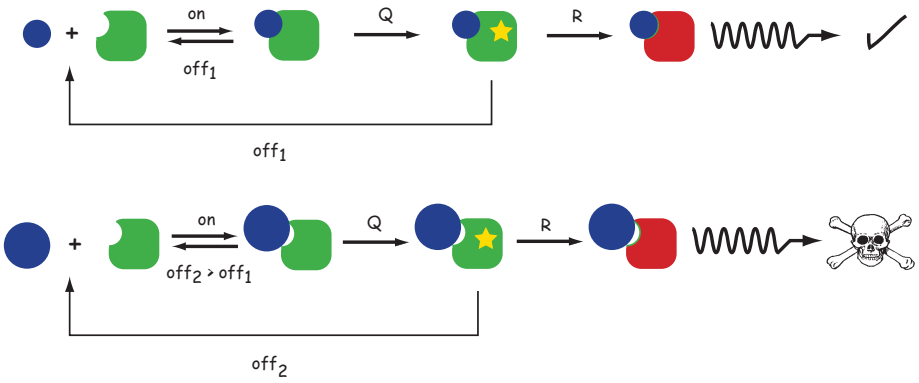
feed-forward

"delay" to filter



multisite phosphorylation

"delay" to postpone commitment

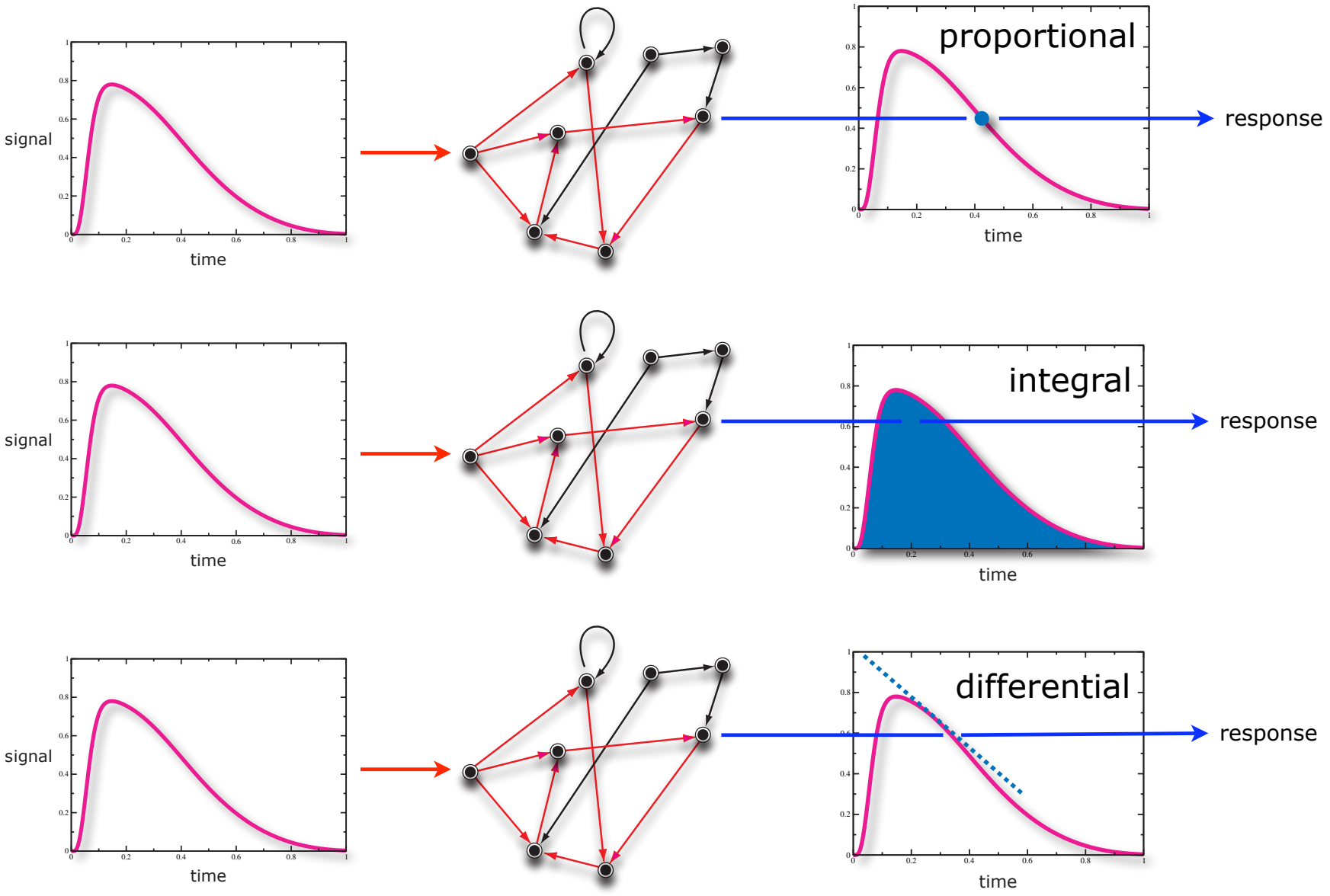


kinetic proofreading

"delay" to postpone commitment

are these types expressible and checkable?

types in control-engineering



from a physics of information to a biology of information

got guts?

i'm looking for a postdoc at the concurrency/biology interface of type:

must survive in a lab atmosphere & $\diamond$ talk to biologists & have some physics intuition.

(is this type inhabited?)