# Walter Fontana - Research

---

**1** – formalization and emergence of functional organization

The formal structure of evolutionary theory is based upon the dynamics of (populations of) "individuals." It therefore assumes the entities whose existence it is supposed to explain. At the heart of the existence problem is determining how biological organizations arise in ontogeny and in phylogeny. The research theme we called *Algorithmic Chemistry* (AlChemy) was an attempt at constructing a formal framework for thinking about molecular organization through an abstraction of chemistry. Our stance was to view molecules as rules of transformation and to exploit a mathematical theory of functions (λ-calculus) to represent abstract "molecules" that act upon one another generating new molecules or rules of transformation. Under suitable boundary conditions this model generates self-maintaining collectives of rules whose mutual transformations permit the continuous regeneration of these same rules. The "organization" of such a system is specified by the relationships of transformation that enable self-maintenance, i.e. the algebraic structure of the system. This framework permits to address the robustness of organizations with respect to the elimination of components (self-repair), the addition of components not belonging to the organization (constrained extension) and the merger of autonomous organizations into higher-order structures (integration).

$$A \equiv A_i^n \equiv \lambda^{i-1}.A_1^{n-i+1}$$

$$A_1^{n-i+1} \circ Y \Rightarrow \lambda^{n-i}.Y$$

$$\lambda^k.X \circ Y \Rightarrow \lambda^{k-1}.X \quad (k>0)$$

$$\mathrm{A} = \{A\}$$

$$\mathrm{A}^* = \{\lambda^k.A_1^1 \mid 0 \le k \le n-1\} \quad (i=n)$$

$$\mathrm{A}^* = \{\lambda^k.A_1^{n-i+1} \mid k \ge 0\} \quad (i<n)$$

The algebraic structure of a simple self-maintaining set of λ-expressions

Although this thread has slipped into the background of my portfolio, its basic vision remains intact. I'm gravitating back to it from a wider and updated perspective in the pursuit of a research center on "biology and computation."

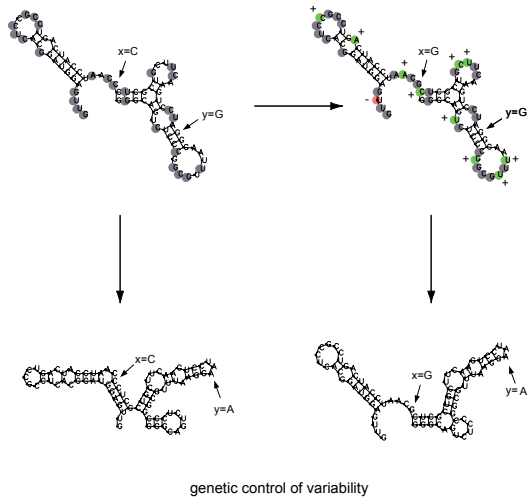
Key collaborator: **Leo Buss** (Yale)

Selected papers:

- W.Fontana and L.W.Buss, The Barrier of Objects: From Dynamical Systems to Bounded Organizations, in: *Boundaries and Barriers*, J.Casti and A.Karlqvist (eds.), pp.56–116, Addison-Wesley, 1996 [pdf]
- W.Fontana and L.W.Buss, 'The Arrival of the Fittest': Toward a Theory of Biological Organization *Bull. Math. Biol.*, **56**, 1-64 (1994)
- W.Fontana and L.W.Buss , What would be conserved if 'the tape were played twice'?, *Proc. Natl. Acad. Sci. USA*, **91**, 757–761 (1994) [pdf]

## **2** – genotype-phenotype mappings

The heritable modification of biological phenotypes occurs by mutation of the genotype. Accessing one phenotype from another is therefore an indirect process mediated by the mapping from genotype to phenotype (development). Evolutionary dynamics and innovation therefore depend on the statistical features of this mapping. I studied this mapping at the level of a single type of molecule: RNA. An RNA molecule is a sequence (genotype) that folds into a shape (phenotype). The statistical analysis of RNA folding has produced insights that may generalize to more complex systems. The most consequential feature is the notion of a *neutral network*: a mutationally connected set of sequences folding into the same shape and spanning a web through sequence space. Neutral networks allow populations to drift across genotype space without losing their current phenotype. In this way, populations can access many more novel phenotypes than if they were confined to a small region of genetic space. Neutral networks dispel the dichotomy of robustness versus evolvability by demonstrating that robustness enables change. Prompted by this line of work, researchers at MIT's Whitehead Institute discovered neutral networks in RNA test tube experiments (Schultes and Bartel, *Science*, 289, 448-452, 2000).




The context dependent effects of mutations on RNA structure.

The adjacency of neutral networks can be used to define a notion of *phenotype space* based on the *accessibility* of phenotypes through genetic mutation rather than phenotypic similarity. Other concepts that emerged from this line of research are *shape space covering* (that all frequent shapes occur within a relatively small neighborhood of any random sequence) and *plasto-genetic congruence* (that the genetic variability of a shape on a given sequence correlates with the alternative structures accessible by thermal fluctuations). Plasto-genetic congruence suggests a trade-off between genetic robustness and phenotypic plasticity.

Key collaborators: **Peter Schuster** (Vienna), **Peter Stadler** (Vienna), **Lauren Ancel-Meyers** (Austin)

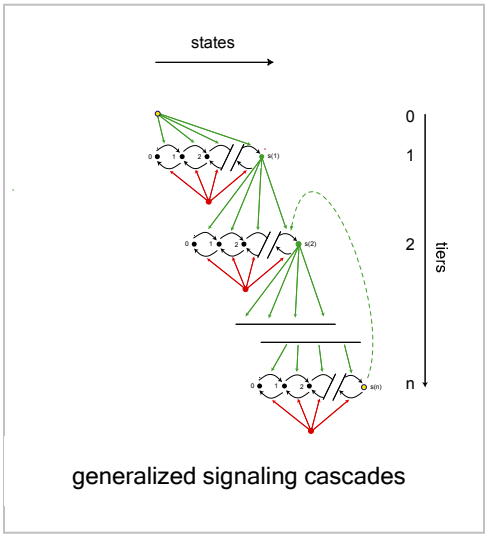
Selected papers:


- W.Fontana, Modelling 'Evo-Devo' with RNA, *BioEssays*, **24**, 1164–1177 (2002) [pdf]
- L.W.Ancel and W.Fontana , Plasticity, Evolvability and Modularity in RNA, *J.Exp.Zool. (Mol.Dev.Evol.),* **288**, 242–283 (2000) [pdf]
- W.Fontana and P.Schuster, Continuity in Evolution: On the Nature of Transitions, *Science*, **280**, 1451–1455 (1998) [pdf]
- P.Schuster, W.Fontana, P.F.Stadler and I.Hofacker, From Sequences to Shapes and Back: A Case Study in RNA Secondary Structures, *Proc. Roy. Soc. (London) B*, **255**, 279–284 (1994) [pdf]

## **3** – distributed molecular control

Organization is deeply shaped by mechanisms and processes of regulation and control. In collaboration with David Krakauer, I'm developing and analyzing models of molecular signal transduction networks. These networks generate cellular behavior in response to signals intercepted at the cell surface. We study simple molecular communication mechanisms that underlie functional behavior and how networks might reconfigure themselves dynamically to "learn," "memorize," "represent," "parse" and "integrate" biological information. Specific projects include:

*Switching and stochasticity in phosphorylation chains.* Molecular signaling often involves the concurrent phosphorylation and dephosphorylation of a target molecule at multiple sites. A bistable switch can result when the fully phosphorylated target molecule feeds back positively on the phosphorylation of its precursors. In molecular systems, unlike in electronic circuits, the shape of a "potential surface" is often maintained by the same processes whose dynamics that potential governs. This is the province of statistical many-body theory. We use analytical and numerical techniques to study the existence of switches as a function of particle number.



generalized signaling cascades

*Generalized signaling cascades*. Signaling pathways contain cascades comprising several tiers of sequentially activated kinases. Within each tier, kinases may be phosphorylated at multiple sites. This project is about understanding the (deterministic and stochastic) dynamics of multiple phosphorylation in the context of cascade depth.

*Dynamically reconfigurable networks.* Signaling networks can modulate the expression of genes. Some of these genes may code for components of the signaling network that controls their expression. Such a feed-back loop enables a signaling network to reconfigure itself in response to signals. This project is about the dynamical characterization of such networks and their potential for simple forms of learning.

Key collaborators: **David Krakauer** (Santa Fe), **Supriya Krishnamurty** (Stockholm), **Eric Smith** (Santa Fe)

Selected papers:

- S.Krishnamurthy, E.Smith, D.Krakauer and W.Fontana, Non-equilibrium phase transitions in a cellular signaling chain, submitted (2003) [pdf]

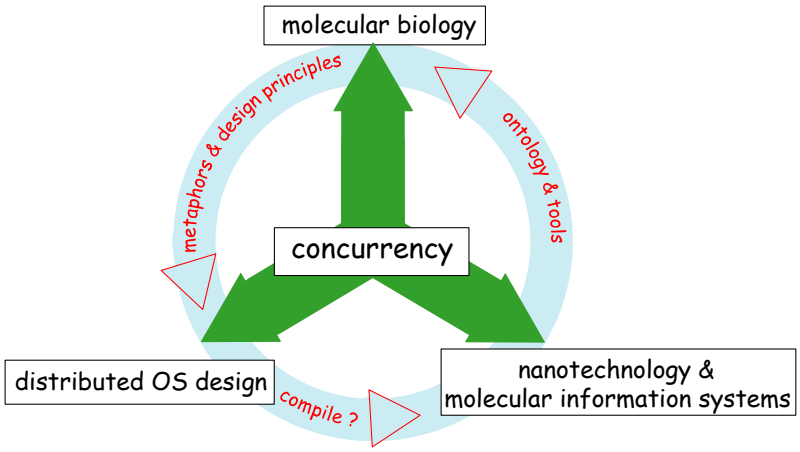
Papers in preparation:

- Generalized Signaling Cascades (with D.Krakauer)

The computational (or informational) perspective has become commonplace in thinking about biological systems. It is far from obvious, however, what kind of computational system would make such a metaphor meaningful. That system is not a Turing machine. But if the notion of computation that Turing, Church and Hilbert developed in the 1930s is inadequate for molecular systems, what *are* we talking about?

The interaction between computation and biology is not limited to the implementation of traditional parallel algorithms with novel materials (as in today's notion of "DNA computing") or the clever coding of biological algorithms (such as "genetic algorithms" or "neural networks") in traditional programming languages. The mind-bending aspects arise when we try to understand what kind of "evolvable language" has evolved within a medium of macromolecules that control, destroy and manufacture one another. Has this chemical language a distillable abstract content or is it inextricably buried in its material substrate? What programming discipline does it entail? To ask such questions, language designers need biologists. To answer such questions, biologists need language designers.

In the traditional notion of computation, a "function" is an algorithm to which an input is supplied and from which an output is retrieved. Other than that, a function does not interact with the outside world. Dropping this restriction leads to the notion of "process." The output of a process is not specified by its internal structure alone, but depends on its interactions with other processes. A process implements behavior, not an input/output relation. The distinction between process and function is similar in kind and consequences to the distinction between open and closed systems in thermodynamics. Since the 70s, computer scientists have been developing theories of computation based on interaction, creating a field known as "concurrency." Its main thrust is the construction of formal languages for reasoning about systems of non-deterministically communicating processes. Such a framework is of concern to systems biologists and designers of distributed operating systems alike. Indeed, concurrency is now being applied in both domains.




Concurrency as a lingua franca

I believe this development to be significant in a more radical sense. Consider a typical agent-based simulation of a biomolecular system. The behavior of a molecule is represented with a piece of program written in some traditional language, such as C or Java. We know that a simulation is not a theory. But the question is not: Why should it? The question is rather: Why couldn't it? What would it take for it to actually be a theory?

A computer language is a formal system. Yet, the relation between syntax and semantics in traditional programming languages does not conform with the structure-behavior relation of biomolecules. The syntax of a program does not represent a molecule in such a way as to induce a correspondence between behaviors resulting from the syntactical modification of program and molecule. The answer to this mismatch must be a programming language that constitutes a theory of the desired domain of chemistry. We don't know whether or for which domains such a language exists, but concurrency frameworks like π-calculus are a starting point. Aviv Regev, Ehud Shapiro, Corrado Priami and others have shown that such a calculus can be used "as is" for simulating signal transduction pathways. The issue now is whether these calculi can be developed into a *mathematical method* adequate for producing theory that could not be achieved by other means. If this was the case, a program for simulating a biomolecular system would be at the same time a formula for reasoning about it abstractly.

A language of this kind might well become a *lingua franca*. A lingua franca permits "idioms" developed or discovered in one domain to be immediately available to other domains. (The theory of dynamical systems may serve as an example of a lingua franca.) The lingua franca envisioned here enables a conceptual, mathematical and technological trading zone between systems biology, concurrency theory and the design of operating systems and programmable (or evolvable) synthetic molecular information systems. My hope is to help bring about a research center dedicated to the development of such a lingua franca. Its primary motivation is the search for the foundations of a theoretical molecular biology, the notion(s) of computation it implies and the technologies it enables.

This way of thinking is a joint effort with **Greg Meredith** (Microsoft Redmond), **Luca Cardelli** (Microsoft Cambridge), **Cosimo Laneve** (Bologna), **Corrado Priami** (Trento), **Vincent Danos** (Paris), **David Krakauer** (Santa Fe), **Leo Buss** (Yale) and many others.