

From abstract chemistry to biology

Ton Rullmann

13 september 2000

Background: a personal view

For years I have been involved with performing Molecular Dynamics and Monte Carlo type simulations of proteins and small molecules. What always struck me was the question how specificity arises, e.g. preferences for proteins to bind at specific DNA sites. How do “bound states” emanate from a multitude of thermally accessible configurations? A priori it is not even clear that specificity can be understood by attempting to delineate a bound state within the configuration space, since microstates belonging to the same bound state “basin” may cover a much larger energy scale than the (free) energy difference between two bound states. Obviously, one may attempt to calculate these macroscopic binding energies in a statistical mechanical approach, e.g. by using Molecular Dynamics simulations. These would have to cover an incredible amount of phase space; their results might be accurate, but would to some extent also be like a black box prediction. And one imagines that the overall outcome, such as the existence of one or a few bound states, can not depend on the precise values of all the atomistic details. Global aspects of the macroscopic picture must somehow be encoded in the energy landscape—otherwise the results would be very fragile. At least one can investigate this hypothesis, by using a simplified low-level description and varying its characteristics. Such attempts have been made for the problem of how proteins fold from a linear chain, leading to new conceptual images of this phenomenon which could be verified in experiments and in more detailed calculations.

Time and again I found myself fascinated by these topics. Inspiration came from writers such as Prigogine and Kauffman. A shift in perspective originated from reading the book “Mind and Nature” by the biologist Gregory Bateson. He analyses the importance of epistemological classification in our conversations about nature, i.e. the assignment of phenomena to classes and classes of classes. Homologies can be described, and homologies of homologies. Things are said to be “like” or “unlike” other things; we picture the relationship much more readily than the quantity. In botany e.g.: A stem is that which bears leaves. A leaf is that which has a bud in its angle. A stem is what was once a bud in that position.

This referential system is characteristic of our thinking, our mind. Fascinatingly, it may also reveal deep truths about nature itself. This is particularly clear in morphogenesis: the bud *is* a bud by being poised between the stem and the leaf, and by communicating with—i.e. differentiating itself from—its neighbours. In a very real way organisms develop by telling little stories, and stories upon stories, i.e. by providing (in Bateson’s words) a “little knot or complex of that species of connectedness which we call relevance”. The anterior side of a developing embryo provides relevance to the posterior side, and vice versa.

This brings back memories of doing Molecular Dynamics simulations: events never quite repeated themselves, and interesting things often happened just before the end. It probably is too far-fetched to compare watching a simulation of 1 ns with listening to a story. Nevertheless, it is in these minute and fleeting events at the molecular scale that the lives of cells and organisms begin to take shape. Consider a specific example: a transcription event where a regulating protein binds to a DNA operator site. This event is coupled to, e.g. a switch in the life cycle of a bacterium. Can such an event, where macromolecular concentrations may be close to a single copy per cell, really be described in terms of

equilibrium thermodynamics or compared to the situation of an isolated protein-DNA complex?

It seems that somehow in the progression from molecules to ensembles to catalytic networks to cells to organisms, the importance of the contingent starts to grow. Events become less predictable. They may be modelled, surely, but the variety of possible outcomes is increasing at each level—yet cells and organisms are quite obviously capable of maintaining stable states. It is not at all clear how to capture the interplay of the contingent and the recurrent in an integrated description. Authors such as Gould and Kauffman have vividly painted the problems in approaching evolving, complex systems from a deterministic standpoint. Gould emphasized the unique, historical aspects of species development. Kauffman has argued that besides random variation and selection a third factor is shaping evolution: self-organizing capacity.

Clearly my interest as a theoretical chemist lies with the transition between regimes: how can one link descriptions of the various levels? In particular I am interested to study the emergence of order and process in dynamic, molecular assemblies. At the one hand it should be possible to forge a link with statistical mechanical descriptions. At the other hand one has to conceptualize, and indeed formalize biology: what are the characteristics of macroscopic order and process? How does one distinguish between organization and non-organization? What is the basis for assigning “function” to an element or process? What is the interplay between function and context? These are powerful questions that go beyond attempts to quantitatively model e.g. an enzymatic network or a metabolizing cell. I believe that ultimately these questions are connected to epistemological issues, such as described by Bateson. Rather than intimidating, I find this exhilarating, because there is the prospect of reaching a better understanding of how we participate in the world we live in—and this can only be an utterly humane prospect.

Research ideas

We can use simulations to explore the behaviour of interacting sets of molecules, with a view to studying the emergence of large-scale order and process. In order to make this a practical proposition we must begin by abstracting the properties of these molecules. In particular, we begin by removing any specification of the mechanics of the interactions, i.e. by abstracting away the physics. We are left with the following characteristics:

- each molecule is a separate entity, having a particular set of properties
- the properties of molecules define their behaviour towards each other
- in reactions, one or more of the interaction partners may disappear, and new molecules can be produced [complexation and catalysis are special cases]

Of particular relevance is the fact that compounds act both as reactant and as operator, i.e. they are produced and consumed in reactions, while simultaneously their internal structure endows a particular function: a model of chemistry has to be a model of molecules “operating” on each other.¹ This is exactly the starting point of the AlChem (Algorithmic Chemistry) project developed by Fontana and Buss, which provided much inspiration for this

¹ One may speculate whether this is a major factor distinguishing the physical, chemical and biological levels: physical theory is about particle interactions, with function arising from collective properties; biological theories are about functions of complex assemblies, but function is not modelled in terms of the underlying matter (although matter can appear as an extraneous factor, such as food). Chemistry sits in the middle. Obviously this is an oversimplification.

note. They investigated a formalization of chemistry using λ -calculus which was chosen for its ability to express elements and functions on those elements as constructs within the same syntactical space. They showed that using λ -calculus as the constructive mechanics operating in a set of arbitrary valid expressions, they could obtain systems characterized by hypercyclic or more complex self-maintaining behaviour. These systems could be described in terms of a limited set of syntactical elements and algebraic structures (i.e. laws describing the transformation relations among the elements).

While these results are highly interesting, and suggest that self-maintaining behaviour may be common in abstract chemistries, the choice for a mathematical calculus as the analog of chemistry also has its limitations. In particular, it is difficult to account for the appearance of more than one product. Also, the high level of abstraction chosen makes it difficult to map to an actual chemistry.

Fontana and Buss aim to develop a formal theory of chemistry, without resorting to quantum mechanics. They note that this requires the formulation of a theory of object construction, i.e. a formal system capable of expressing the space of molecular objects and what is meant by the reaction arrow " \rightarrow ". A theory of object construction is the domain of a large body of computational science. Of course, the application of such a theory to chemistry, or rather casting chemistry in terms of an object construction theory, could distinguish between features of organization that derive from the underlying theory and features that are "merely" the consequence of a particular parametrization. Here Fontana and Buss sketch an intriguing panorama, linking theoretical chemistry to modern computational theory.

But I believe that a more pragmatic approach is not without merit. It strikes me that instead of developing a full-blown theory of object construction, one could take the construction as a given, and explore the consequences. Instead of defining a formal mechanics for object interconversion, one can supply a list, table or rule for all interactions. The crucial point is not the mechanics itself, but to have properties, and a *classification* of properties. The similarity of molecules is defined by similarity of their properties. Similar molecules have similar reactions (or not!). The relevant question then is, whether specific chemical typologies give rise to specific biological typologies [where I use the word biological to indicate some form of higher level organization or process, which can be captured in a reduced set of descriptors; it is the analogon of the algebraic structures of AlChem]. I do not propose to choose *a particular* simplified or abstract representation of chemistry, such as the models explored by Kauffman et al. What I propose is to generate a whole *series* of abstract representations, and systematically compare their outcomes when coupled to a kinetics scheme. This is not a case of "digital naturalism", but a resort to computational experimentation in the absence of a full-blown theory.

Sketch of simplistic model

- N molecules in well-stirred reactor.
- Each molecule belongs to one of K types.
- Molecules undergo bimolecular collisions, giving one of three outcomes:
 - i. $A + B \rightarrow C$
 - ii. $A + B \rightarrow C + D$
 - iii. $A + B \rightarrow C + D + E$ [special case: $A = C$]
- For every pair [A,B] the reaction result (type and identity of C, D, E) are pre-defined in a reaction list [note: C, D, E are also one of the K types]
- The K types are divided in L classes; the reaction list is set up in such a way that $A + B \rightarrow C$; $A + B' \rightarrow C'$; where C and C' are member of the same class if B and B'

are member of the same class; plus suitable extension to the other reaction types. It may not be possible to guarantee perfect class separation in this manner (“similar molecules have similar reactions”); a measure of class separation (or violation) can be developed. Alternatively, class separation can be degraded on purpose; in the extreme case, there is no relation between molecule class and reaction class.

- ❑ Study emergence of patterns in the molecular content and turn-over, for many different instantiations of the reaction list, and as a function of the variables: N; K; L; relative occurrence of reaction types i, ii, iii; “class separation”. What is changing, and what are the persistent features?

Extensions of the model

- ❑ Alternative, more detailed ways to construct reaction types, e.g. based on sub-molecular groups.
- ❑ Polymer reaction types.
- ❑ Regulatory type interactions ($A + B \rightarrow B'$; $B' \rightarrow A + B$; where B and B' differ in a subtle way).
- ❑ Effect of localization (non-homogeneous reactor).
- ❑ Quantitative effects (reaction probabilities, transport properties).

Long term goals

The primary goal is to study the emergence of “collective” properties of molecular networks, in other words: the origin of biological function. To what extent are function and order defined by logical-functional aspects of the networks, such as diversity, connectedness, classes and localization? What is the interplay between function and context? Pattern-recognition will be an important facet.

Secondly, I want to establish a link with quantitative treatments. Adding quantity to the simplistic model sketched above first of all has the effect of “specifying” the interactions and classifications². Furthermore, it opens up the machinery of statistical mechanics (entropy and energy concepts of the cell). I believe that the issue here is to try to understand the separation of scales (class, distance, energy, time,) which allows to make a transition from detailed molecular to averaged ensemble properties; and more importantly, how to combine these within one model. At this point a connection can be made to experimental studies of cellular networks and to whole cell models that are being built, thanks to advances in biochemical and genetic analysis. One may even speculate that all-molecule treatments of a cell may be within reach, if one can treat diffusion and flow in a macroscopic way: in a cell with a volume of $1 \mu\text{m}^3$, concentrations of 1 mM are equivalent to $\sim 10^6$ particles. Using some form of state-transition model, as described above, networks of this size could still be computationally tractable.

References

² Quantity introduces a measure of similarity, which can be the basis for classifying reactions and properties. It will be interesting to see whether a quantitative treatment gives different results than a purely “symbolic” model. The issue here is whether chemical identity will be modelled in a quantitative way. The second role for quantity is in defining kinetics and thermodynamics, i.e. modelling of physical chemistry. Obviously these topics are connected, particularly in biologically relevant areas such as complexation. E.g. one could use a quantitative model to describe allosteric and other regulatory effects, or one could try to capture the latter in a suitably refined classification scheme.

G. Bateson (1979) "Mind and Nature; a necessary unity". Originally published: E.P. Dutton, New York.

W. Fontana & L.W. Buss (1994) "What would be conserved if "the tape were played twice"?"

Proc. Natl. Acad. Sci. USA, **91**, 757-761.

W. Fontana & L.W. Buss (1996) "Towards a theory of biological organization: a programmatic statement". Available at:

<http://www.santafe.edu/~walter/AlChem/Statement/organization.html>

W. Fontana & L.W. Buss (1996) "The barrier of objects: from dynamical systems to bounded organizations". IIASA working paper WP-96-27. Available at:

<http://www.santafe.edu/~walter/publications.html>