

Molecular Computing: From Conformational Pattern Recognition to Complex Processing Networks

Michael Conrad and Klaus-Peter Zauner

Wayne State University, Dept. of Computer Science,
Detroit MI 48202, USA
e-mail: biocomputing@cs.wayne.edu

Abstract. Natural biomolecular systems process information in a radically different manner than programmable machines. Conformational interactions, the basis of specificity and self-assembly, are of key importance. A gedanken device is presented that illustrates how the fusion of information through conformational self-organization can serve to enhance pattern processing at the cellular level. The device is used to highlight general features of biomolecular information processing. We briefly outline a simulation system designed to address the manner in which conformational processing interacts with kinetic and higher level structural dynamics in complex biochemical networks. Virtual models that capture features of biomolecular information processing can in some instances have artificial intelligence value in their own right and should serve as design tools for future computers built from real molecules.

1 Introduction

Molecular computers are natural or artificial systems in which individual macromolecules play a critical information processing role. A real molecular computer is a natural or artificial system built up out of molecules. A virtual molecular computer is a simulation system residing on a digital base machine, used either for the purpose of elucidating principles of natural biological information processing or for the purpose of designing a "real" artificial system.

The capabilities of natural molecular computers are radically different than the capabilities of digital machines, due to the very different structure-function relations. The purpose here is to draw attention to essential operative principles and then to briefly outline a modeling system, to be referred to as the CKSD simulator, that addresses the co-action of conformational, kinetic, structural and dynamic processes in complex biochemical reaction networks. Even briefer mention will be made of a hierarchical model, called the artificial neuromolecular (ANM) brain, that puts subcellular processing into a larger (vertical) framework and that illustrates how ideas derived from natural computing systems can be used to expand computer science.

2 General Principles

Shape (or conformation) based interactions among macromolecules are of central importance, since these provide for a vast spectrum of intrinsic pattern recognition operators [5]. The gedanken device pictured in Fig. 1, to be referred to as the self-assembly model, illustrates how such molecular pattern recognition can contribute to higher level symbolic pattern processing [6]. The self-assembly device corresponds to a cell in a natural biological system and networks of such devices would correspond to cellular networks such as the immune system or the brain.

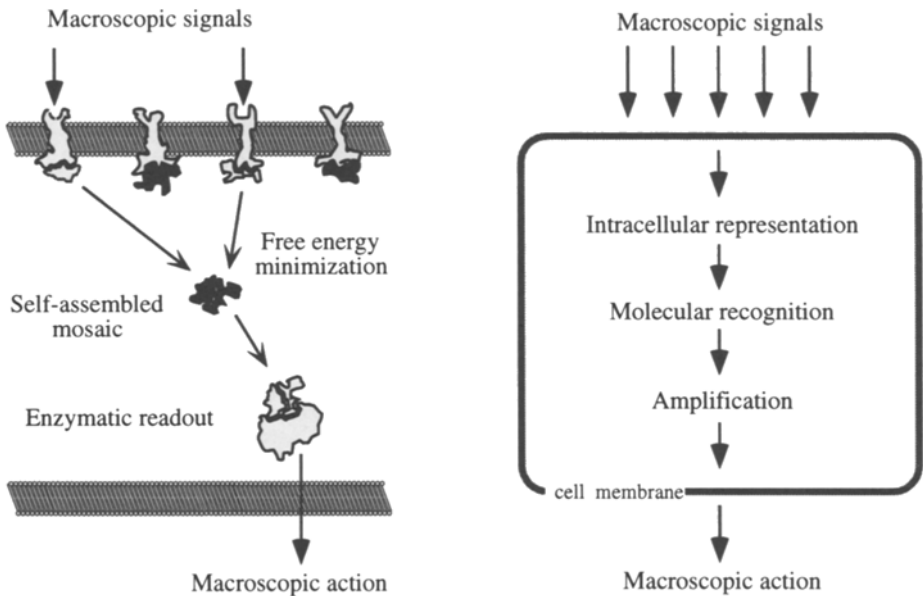


Fig. 1. Self-assembly thought device and transduction-amplification process (from [11]).

The self-assembly device receives patterns of input signals on its external membrane. If an input line is activated a protein is released, otherwise not. Thus the input signals are recoded as molecular shapes. These molecular shapes self-assemble, like self-organizing jigsaw puzzle pieces, to form a complex. Different complexes will be formed in response to different input patterns. Shape features common to subsets of possible complexes will be associated with different groupings of input patterns. Enzymes that recognize these shape features can then trigger an output action appropriate to the input pattern.

The self-assembly device can be thought of as an advanced biosensor. Inputs are transduced to a physiochemical state. Proteins (or other biological elements)

are then used to detect this state in a highly discriminating way. If the proteins are activated they trigger an output. The feature added by the self-assembly model is this: it would use the self-organization of protein complexes to selectively fuse different pieces of information, including pieces arriving at different locations in space and different points in time.

Clearly the self-assembly device is an extreme abstraction of a biological cell. Nevertheless it provides a paradigm whose features suggest the following general principles:

1. *Percolation of molecular pattern recognition.* Molecular pattern recognition makes a direct contribution to the capacity of cells to recognize and process patterns of inputs. The essential point is the conversion of a symbolic pattern recognition problem into a free energy minimization process. The clever physics of self-assembly is brought to bear on the solution of the problem. The self-assembly device falls to a solution as naturally as salt crystallizes.
2. *Percolation of quantum parallelism.* The device points to the manner in which quantum features, through conformational-electronic interactions that enhance the specificity and speed of macromolecular pattern recognition, percolate across scale to enhance the real time capabilities of cells and organisms [7]. The effect utilizes the classical-nonclassical interface afforded by the coupling of nuclear and electronic degrees of freedom. Conformation, as defined by the positions of the atomic nuclei, is a classical concept. The mass of electrons is small enough, however, for the wave properties of matter to play a role. Some of the electrons in a protein are sufficiently delocalized so that their motions cannot be wrapped together with the motions of the nuclei to which they are bound (the usual approximation). The electronic system will then be in a superposition of states even at normal temperatures, leading to interference effects that serve to funnel thermal energy to selected nuclear degrees of freedom.
3. *Distribution of molecular pattern recognition.* The control properties of biochemical reaction networks cannot be understood in terms of models that collapse the pattern processing power of enzymes into simple switching behavior. Whether and how the enzyme acts can depend on a variety of milieu influences. The conformational self-organization of the enzyme and its substrate serves to fuse these milieu signals, just as in the self-assembly model, except that the contribution of molecular pattern recognition to cellular control is distributed over many reactions.
4. *Integrative role of subcellular fine structure.* The cytoskeleton, a self-assembled system of fibers that play an important role in controlling cell form and form changes, provides a plausible medium for fast signal integration [12, 13]. The picture is this: signals arriving at the cell surface at points separated in time and space are brought into coincidence at specific macromolecular or poly-macromolecular nodes within the cell by the integrative action of the subcellular medium. The signals are fused by conformational reorganizations at these nodes, which then initiate new signals, eventually culminating in the output action taken by the cell.

5. *Adaptive self-stabilization.* The input-output transform performed by systems operating on the basis of the self-assembly principle possesses intrinsic adaptive properties [9]. If the action taken by the system in response to some input is unsuitable its internal structure will be disturbed. Effects of the disturbance will persist (exhibit hysteresis) due to the self-assembly properties of the components. If the input-output behavior of the restructured system is suitable the error signal will be small and the structure will then "earn the right" to stay in its new form. Adaptive self-stabilization is an error feedback process, with the extra feature that the structure of the system is brought into the feedback loop.
6. *High evolutionary adaptability.* Proteins are well suited for adaptation through variation and selection due to their high structure-function plasticity. Adaptation of this type is a generalization of adaptive self-stabilization in which the error (variation) and correction (selection) processes are mediated by populations of organisms. The key fact is this: proteins have two representations, one in terms of the linear sequence of amino acids and the second in terms of folded shape. The two representations are connected by a conformational self-organization (or folding) that can be gradually distorted. Single mutations are much more likely to yield a small change in conformation and hence an acceptable change of function than single mutations in a purely syntactical system, provided that they do not affect a crucial site. Redundant amino acids can always be added that buffer the effect of mutation on functionally critical shape features. Adding redundancy increases the dimensionality of the fitness surface, thereby increasing the probability of saddle points as opposed to isolated peaks [8]. The evolution process is thus less likely to stagnate.
7. *Moldable building blocks.* Multicellular systems, such as the brain, are built up from cells with powerful and plastic pattern processing capabilities. The adaptive properties of such systems arise in large measure from transformations in the input-output behavior of cells. They cannot be simply analyzed in terms of the rewiring of given components, as in connectionist models.

The phrase "vertical information processing" places the picture in a larger perspective [10]. Macroscopic environmental influences are transduced to hormonal and nerve signals that impinge on cells and that are transduced to more mesoscopic intracellular signals. The latter trigger macromolecular conformational changes that link the mesoscopic with the truly microphysical (electronic) scale. Macromolecular actions are then amplified into macroscopic cellular outputs and finally to the more macroscopic activities of the organism. The computational power arises from the synergistic orchestration of physical-dynamical processes with characteristics unique to the various scales in this hierarchy. Conformational interactions play the linchpin role, due to their enabling importance for structural integration and dynamical coordination.

3 Algorithmic Specification

The processing modes of organisms are clearly different from those of machines. Systems so dominated by self-organizing dynamics and cross-scale interactions are not effectively programmable in the conventional sense [3]. Concomitantly, the number of interactions available for information processing, and the degree of parallelism obtainable, is very much larger. This disparity has a positive side. The power of digital computers, as compared to other modeling approaches, is that many more types of abstraction are feasible. Abstractions that address the general features outlined above can be implemented and used for experimentation, subject to the realization that not all of the principles can be captured in any single simulation and that the greater the emphasis on the processing power of the components, and the number of levels of processing, the smaller the number of components that can be represented.

Our group has built a variety of such virtual systems, ranging from systems that focus on molecular level interactions to vertically structured artificial worlds designed to study ecosystem evolution processes [4]. The term algorithmic specification is appropriate, since the goal is to use the language tools of computer science to formalize models of complex biological systems that are refractory to traditional analytic techniques [2]. Here is not the place to describe these models. But I would briefly like to mention two that illustrate how different choices of abstraction highlight different aspects of the complex reality.

4 CKSD Simulation System

The first and most recent system, to be referred to as the CKSD simulator, highlights aspects connected with conformation, kinetics, structure and dynamics (Fig. 2). Of the features discussed in the previous section the following are represented: molecular pattern recognition, self-assembly of larger molecular structures, fusion of information, distribution of pattern recognition activities, and interaction between processes at very different spatial scales. The system is rich enough for the integrative role of subcellular fine structure to play a role and to perform experiments on adaptive self-stabilization.

Conformation is represented in an abstract manner, as states of macromolecules that determine specificity and binding interactions either with other macromolecules or with small metabolites and ions (referred to as microcomponents). Extended structures can arise from these interactions, including in principle cytoskeletal and membrane structures. Dynamics refers to the changes in these structures. Kinetics refers to biochemical reaction-diffusion processes. The response of the macromolecules can be highly sensitive to milieu context, including neighboring large molecules and concentrations of small molecules and ions.

The macromolecular representations include conformational states, catalytic activity, binding interactions, and diffusional motion in three dimensional space. Microcomponents are represented by concentrations. These provide the milieu in

which macromolecules and polymacromolecular aggregates function. Dynamics refers to configurations and reconfigurations of the structural organization of a reaction system resulting from the binding interactions between macromolecules.

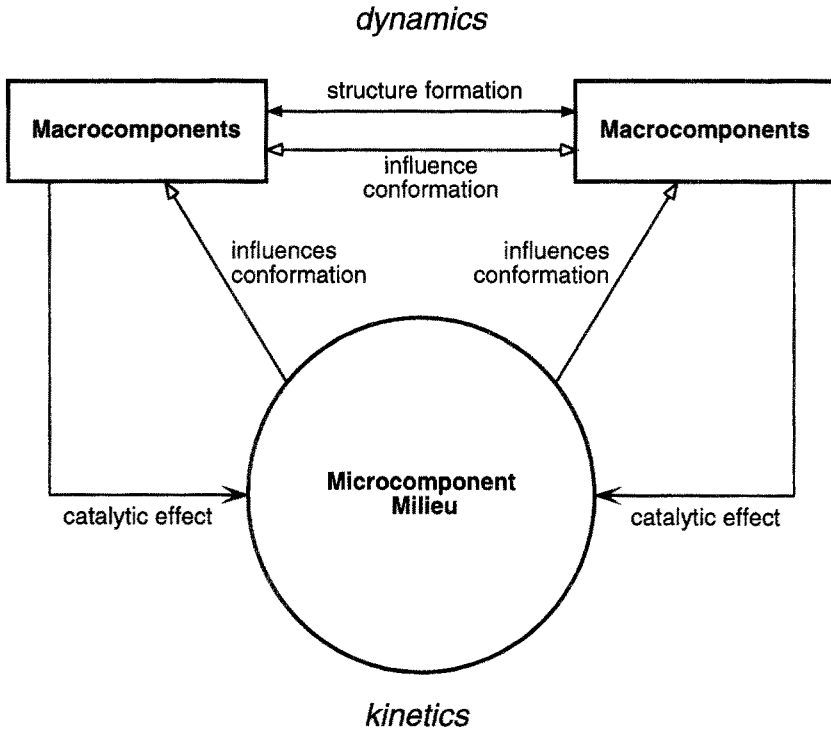


Fig. 2. Overview of the CKSD simulator.

Each macromolecule in the CKSD system is represented by a dodecahedron. Each side of the dodecahedron represents a domain (or subunit) of the molecule. The latter could be a protein, a catalytically active RNA or even a membrane region. Each domain has a finite automaton description. The different states of these automata represent different conformational states; the state transitions represent either conformational changes or chemical modifications, including ionization. Catalytic activity affecting metabolites and also attractive or repulsive interactions with other macrocomponents are assigned to the subunit states. The state transitions of the subunits can depend on the states of neighboring subunits in the same macromolecule and in adjacent macromolecules. They can also depend on the presence of surrounding microcomponents.

A simple specification language facilitates the communication of complex models to the simulation engine. Virtual measurement instruments are used to monitor the time development of a simulation. Active virtual devices are also

provided and make it possible to intervene during a running simulation, e.g., to change the local concentration of selected molecular species. The simulator is implemented in C++ and can run on any UNIX workstation with sufficient RAM.

The CKSD simulator is well suited for representing the main features connected with conformation-based computing. The functionality of macromolecules depends on the pattern of surrounding milieu features. Self-assembly of poly-macromolecular structures can occur. Adaptive self-stabilization is possible, since the diffusion and structure formation properties of both micro and macrocomponents is influenced by local milieu, including charge interactions. An adaptive feedback loop between structural changes and kinetic behavior is therefore possible.

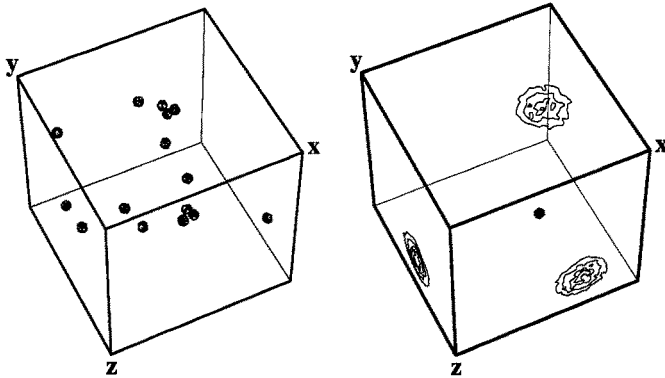


Fig. 3. Macrocomponent distribution (left) and concentration gradient of one microcomponent form the milieu (right). For clarity only one of the macrocomponents is shown on the right.

Selected outputs from a sample run are shown in Figs. 3 and 4. The macrocomponent distribution at a single point in time is illustrated in Fig. 3 (left hand side). The macrocomponents are embedded in a milieu of microcomponents. The right hand cube in Fig. 3 illustrates the concentration gradient of a single microcomponent species together with one of the macrocomponents from the left hand cube. Actually the simulation in this instance accommodated 30 microcomponent types (a parameter limited only by the computing resources available). Thus a full picture would include the distribution of all macrocomponents (15 in this instance) and an overlay of 30 microcomponent concentration gradients. The temporal development of one of these spatial concentration gradients is shown in Figure 4.

Each cube shows projections of the local concentration in the simulation space along the three axes. The lines in a projection plane connect points at which the projected concentrations have equal values. The environment of the macrocomponents is a pattern resulting from a superposition of numerous such gradients and also adjacent macromolecular states. We could choose to represent the time development of any one of the other microcomponents, the distribution of all of them at any one instant of time, and also the distribution of states of the macromolecules that respond to and act on the milieu pattern.

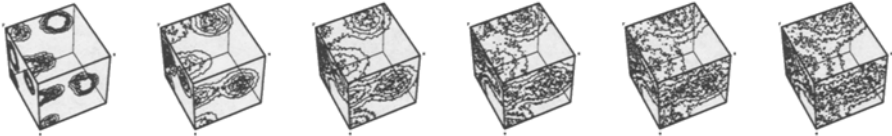


Fig. 4. Spatiotemporal development of a microcomponent gradient. The time evolution runs from left to right. The full milieu pattern would comprise a superposition of gradients for many different types of microcomponents.

Implementation of the CKSD system has only recently been completed and experiments are just beginning. The system can be used as an empirical modeling tool, drawing on experimental data to set transition tables, or as a theoretical tool to investigate general properties of complex molecular information processing systems. We anticipate that it will be possible to use the model to experiment on adaptive self-stabilization, as noted above, and also to address the significance of structure-kinetic interactions for information processing and control. Potential application areas include complex biochemical networks, cellular and metabolic engineering, and design of conformation-driven molecular computing/control modules in which the conformational pattern processing is distributed in a massively parallel way.

5 ANM Framework

The second system, called the artificial neuromolecular (ANM) brain, highlights the vertical flow of information across different levels. The ANM system can be viewed as a population of networks built up from pattern processing neurons. Signal patterns impinging on these neurons are integrated in space and time by the cytoskeletal network and associated modulating proteins. Neuron firing is triggered by readout enzymes that are suitably activated. Variation-selection evolution is used to generate a repertoire of pattern processing neurons. A memory system, basically a synaptic facilitation scheme with hierarchical controls, is

used to select cytoskeletal neurons from the repertoire and to orchestrate them into groupings adapted to perception-action tasks. The system has been used for maze navigation tasks involving recognition of a series of 64-bit patterns [10] and more recently used for Chinese character recognition. Here again what is important are the features represented: molecular readout, subcellular signal integration, structure-function plasticity based on redundancy and modulating interactions, and multicellular networks built from moldable building blocks.

What is the relationship between the ANM architecture and the CKSD simulator discussed in the previous section? The notion of percolation provides the conceptual link. From the natural systems point of view it is important to understand how information impinging on the organism filters down to the molecular level and how the molecular level processing percolates up to control the macroscopic behavior of the organism. This is manifestly important in developmental and immunological processes. The working hypothesis of the ANM framework is that it is also key to the computational capabilities of the brain. From the artificial systems point of view it is important to understand what kind of computer architectures could gainfully utilize conformational pattern recognition. We regard the ANM system as a precursor architecture into which conformation-driven pattern recognition could eventually be embedded.

The CKSD simulator can supply cellular dynamics for the ANM system from both of the above points of view. The full simulator is too computationally costly to embed in the higher level architecture for real time artificial intelligence applications. However, it would be quite feasible to use a simplified version of the simulator to evolve dynamic processes with pattern processing capabilities, and to evolve these dynamics in the context of an ANM type design.

6 From Silicon to Carbon

Recent years have seen expansive progress in the technologies required to implement carbon-based information processing devices [5, 6, 11]. Some of the pertinent technologies include: protein engineering, recombinant DNA, directed evolution, Langmuir-Blodgett films, and optomolecular electronic interfaces. Biosensors, optical interfaces to dynamic chemical reactions for image processing, DNA computing, and bacteriorhodopsin (BR) films and sensors are at the prototype stage. Optical systems that utilize BR are at the product stage.

These prototypes, like the simulation systems alluded to in the previous section, capture the natural biomolecular mode of processing in only a fragmentary way. Yet one can discern an emerging synthesis. Virtual molecular computers already have applicative value, as evidenced by the ANM system. The basic idea is complementary to the classical artificial intelligence objective of giving the machine a representation of the external world. Give it instead, or as well, a representation of the internal world of biological information processing. Some essential aspects of biological information processing are too computationally costly for real time applications and in any event would be more efficient if realized in carbon rather than silicon. Simulations that push these limits pro-

vide insight into natural biological processing and guidelines for device construction. Hybrid approaches that yield novel computational synergies are feasible. As carbon-based devices are developed they can be used to directly realize components of virtual systems, such as the ANM brain. The unification of computer science and biology will undoubtedly lead to a broader and deeper understanding of both subjects.

Acknowledgment. This research was supported by NSF Grant No. ECS-9409780.

References

1. Chen, J.C., Conrad, M.: Learning synergy in a multilevel neuronal architecture. *BioSystems* **32** (1993) 111–142
2. Conrad, M.: Algorithmic specification as a technique for computing with informal biological models. *BioSystems* **13** (1981) 303–320
3. Conrad, M.: The price of programmability. In: *The Universal Turing Machine: A Fifty Year Survey* (Herken, R., ed.), pp. 285–307, Oxford: Oxford University Press (1988) (in Germany, Hamburg: Verlag Kammerer and Unverzagt)
4. Conrad, M., Rizki, M.M.: The artificial worlds approach to emergent evolution. *BioSystems* **23** (1989) 247–260
5. Conrad, M.: Molecular computing. In: *Advances in Computers* (Yovits, M.C., ed.), vol. 31, pp. 235–324. Boston: Academic Press (1990)
6. Conrad, M.: Molecular computing: the lock-key paradigm. *Computer IEEE* **25** (1992) 11–20
7. Conrad, M.: Quantum molecular computing: the self-assembly model. *Int. J. Quantum Chem.: Quantum Biol. Symp.* **19** (1992) 125–128
8. Conrad, M., Ebeling, W.: M.V. Volkenstein, evolutionary thinking and the structure of fitness landscapes. *BioSystems* **27** (1992) 125–128
9. Conrad, M.: Emergent computation through self-assembly. *Nanobiology* **2** (1993) 5–30
10. Conrad, M.: Multiscale synergy in biological information processing. *Optical Memory and Neural Networks* **4** (1995) 89–98
11. Conrad, M., Zauner, K.-P.: Molecular computing: steps toward integration. *Oyo Buturi* (Japan Society of Applied Physics) **64** (1995) 1002–1006
12. Liberman, E.A., Minina, S.V., Mjakotina, O.L., Shklovsky-Kordy, N.E., Conrad, M.: Neuron generator potentials evoked by intracellular injection of cyclic nucleotides and mechanical distension. *Brain Research* **338** (1985) 33–44
13. Matsumoto, G., Sakai, H.: Microtubules inside the plasma membrane of squid giant axons and their possible physiological function. *J. Membr. Biol.* **50** (1979) 1–14