To Appear In: Bonchev, D. and Rouvray, D. (eds.) (2004, in press). Complexity in Chemistry, Biology and Ecology. Kluwer Academic Press, New York, N.Y.

Cellular Automata Models of Complex Biochemical Systems

Lemont B. Kier and Tarynn M. Witten

Center for the Study of Biological Complexity
Virginia Commonwealth University, Richmond, Va 23284

1. Reality, systems, and models

1.1 Introduction

The role of a scientist is to study nature and to attempt to unlock her secrets. In order to pursue this goal, a certain process is usually followed, normally starting with observations. The scientist observes some part of the natural world and attempts to find patterns in the behaviors observed. These patterns, when they are found in what may be a quite complicated set of events, are then called the laws of behavior for the particular part of nature that has been studied. However, the process does not stop at this point. Scientists are not content merely to observe nature and catalog patterns, they seek explanations for the patterns. The possible explanations, that scientists propose, take the form of hypotheses and theories in the form of models. The models serve as representations about how things work behind the scenes of appearance. One way to describe the modeling process is to express it as a pictorial algorithm or flow diagram shown in Figure 1.

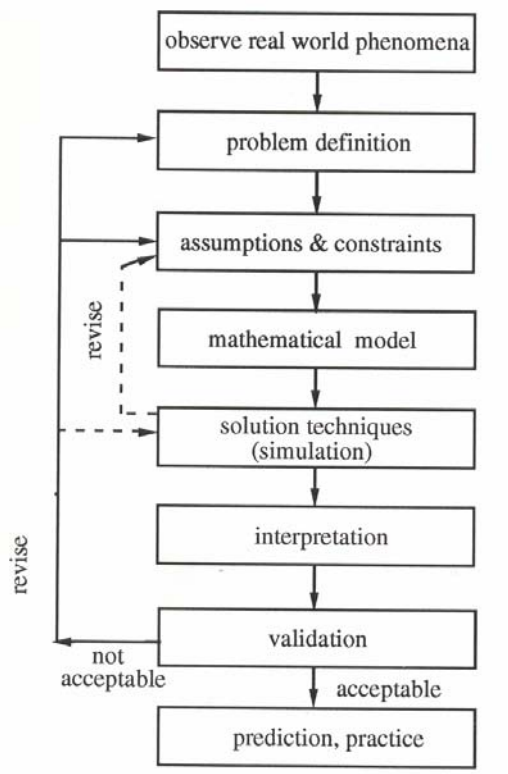

Figure 1: A flow diagram of the modeling process.

This chapter is about one such type of model and how it can be used to understand some of the more complex patterns of chemistry and biology. Let us begin by attempting to understand some essential principles behind modeling/simulation. We will then examine how, in certain scenarios, the models/simulations "*take on a life of their own*," that is, they move from being *complicated* to being *complex*.

1.2 The "what" of modeling and simulation

A model is an observer/scientist's attempt to represent, using a set of rules that they have deduced from observation and scientific deduction, the behavior of a system of interest. Consequently, a model is an abstraction of the whole system. By definition, due to its reduction, the model has access to fewer states than the original system. Hence, scientific interest lies in just a part of the whole system. Clearly, scientific logic dictates that the output of the model system should be consistent with the original system but only for a restricted set of inputs.

Many models in science take the form of mathematical relationships, equations connecting some property or set of properties with other parameters of the system. Some of these relationships are quite simple, *e.g.*, Newton's second law of motion, which says that force equals mass times acceleration, F = ma. Newton's gravitational law for the attractive force F between two masses m_1 and m_2 also takes a rather simple form,

$$F = Gm_1m_2/r^2,$$

where r^2 is the square of the distance separating the masses and G is a constant that correctly dimensionalizes the units between the two sides of the equation. However, many mathematical relationships are much more complicated, and rely on the techniques of calculus, differential and

partial differential equations, and abstract algebra to describe the rates of change of the quantities involved. Such an example would be the basic equation of quantum theory, the Schrödinger equation, which takes a more formidable form:

$$-\frac{\eta^2}{2m}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)\psi = E\psi$$

In chemical kinetics one finds linked sets of differential equations expressing the rates of change of the interacting species [1,2,3,4]. Overall, mathematical models have been exceedingly successful in depicting the broad outlines of an enormously diverse variety of phenomena in nature. Some scientists have even commented in surprise at how well mathematics works in describing nature. So successful have these mathematical models been that their use has spread from the hard sciences to areas as diverse as biology [5], medicine [6], economics [7] and the analysis of athletic performance [8]. In fact, many of the social and psychological sciences now use mathematical models to describe the behaviors of social systems [9, 10], the spread of information in society, and the dynamics of psychological interaction [11]

In other cases models take a more pictorial form. In the early atomic models an atom was first pictured by J. J. Thomson as a "plum pudding", with negative electrons (the "plums") embedded in a spread-out positive charge (the pudding), and then later by Ernest Rutherford and Niels Bohr as a planetary system with a tiny positive core surrounded by circling electrons, a model called the "nuclear atom". Today, within quantum theory, the nuclear atom picture has been further transformed into one with a positive nucleus surrounded by a cloud of electron probabilities. In biology, the double-helix model of the structure of the genetic material DNA proposed by James Watson and Francis Crick led to an explosion of studies in the field of molecular genetics. Charles Darwin's model of evolution by means of natural selection pictures species, composed of a collection of individuals with a variety of different traits, interacting with their environments. Individuals with some traits are better suited to survive and reproduce, thereby passing on these traits to their offspring. Over time new traits are introduced through mutations, environments gradually (or sometimes rapidly) change, and new forms develop from the old ones. The modern model of the human brain envisions regions devoted to different functions such as sight, motor movements, and higher thought processes. In geology, the tectonic plate model of the Earth pictures expansive continental plates moving gradually over the planet's surface, generating earthquakes as they meet and slide over one another. And in psychology, the Freudian approach pictures human behavior as resulting from the actions of invisible components of the mind termed the id, the ego, and the superego. Thus, the modeling process could be pictorially represented as in Figure 2.

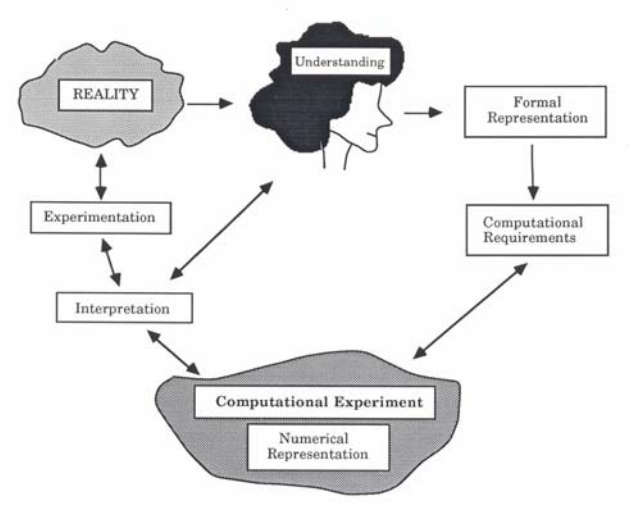

Figure 2: The modeling process including mental components.

The key feature of successful models is that they produce results consistent with the experimental observations. Successful models capture the essential features of the systems of interest. While it is not always true, scientists also hope that the newly created model will go beyond this simple reproduction to predict new features of the systems that may have previously escaped notice. In this latter case the predictions provide an important means for testing the validity of the models. There are many different philosophical approaches to defining the art of model-building and its components. Consequently, at this point it is helpful to dissect models into their most significant parts, so that we can start from a common basis.

1.2.1. The System

Studies in chemistry, or any realm of science, commonly consist of a series of directed examinations of parts of nature's realm called systems. A *system* is an identifiable fragment of the world that is recognizable and that has attributes that one can identify in terms of form and/or function. We can give examples at any level of size and complexity, and in essentially any context. Indeed, a dog is a system at a pet show; whereas the human heart is a system to the cardiologist; a tumor cell is a system to the cancer specialist; a star or planet or galaxy is a system to an astronomer, a molecule, or a collection of molecules, is a system to a chemist; and a macromolecule in a cell is a system to a molecular biologist. A system is, then, whatever we choose to focus our attention upon for study and examination.

1.2.2. States of the System

A system is composed of parts that can be recognized and identified. As time goes by, a system under study may acquire different attributes as a result of changes among its parts, and over time its appearance or function may change. Moreover, as time goes by, our own technological capabilities may expand, thereby allowing us to identify parts of the system we would not have been able to previously identify. Each of the different stages through which the system passes in its evolution is called a state of the system. A dog grows old over time, passing through stages recognized in general terms as puppy, dog, and old dog. A heart may change its pattern of contractions, going from normal to tachycardia to ventricular fibrillation, each of which we

categorize as different states of functioning. A solution of ethyl acetate in water may slowly decompose to mixtures of ethyl acetate, acetic acid and ethanol, through a sequence of states characterized by their different compositions. Water may start as a solid (ice), become a cool liquid, then a warmer liquid, and finally appear as a vapor at higher temperature, passing through these different stages as it is melted, heated, and vaporized. For the purposes of our discussion, we will refer to each set of conditions as a *state* of the system under observation/study. It is the various states of the system upon which we focus attention when we study any system.

1.2.3. Observables

Our studies require us to analyze and describe the changes that occur in the systems we are interested in, as they evolve with time. To accomplish this analysis properly we need to record specific features that characterize what is occurring during the evolutionary process. The features assigned for this purpose are referred to as the *observables* of the study system. For example, we distinguish the puppy from the old dog by changes in its physical appearance and its behavior. The changes in a heart's rhythm are recorded on special charts monitoring electrical signals. The changes occurring in a solution of ethyl acetate in water can be characterized by changes in the solution's acidity, by spectroscopic readings or by detection of the odor of acetic acid. To be as precise as possible in a scientific investigation it is necessary to assign numerical values to the characteristics that distinguish one state of a system from another.

The state of a system is studied through detection and recording of its observables. To record an observable, we must "*probe*" the system with a "*measurement*" instrument of some type. This requires an "*interaction*," which we will discuss in the next section, as well as the existence of a device that is capable of recognizing the particular system observable as well as reporting back to the scientist/observer the value of that observable. It is extremely important to understand that observables of a system are intimately tied to the technological capability of the scientist. Thus, gene sequencing, common in today's technological arsenal, was not a probe available to scientists fifty years ago. Thus, if our observable was the gene sequence of an organism, the probe did not exist to provide the requisite information. Hence, the observable, while of scientific interest, was not accessible.

1.2.4. Interactions

There are two types of *interaction* with which we must deal. The first is the interaction of the scientist/observer with the system under study. The second is the interaction of the parts of the system with each other (both known and unknown).

The scientist interacts with the system in two ways; through setting the actual experiment up to be observed and through measurement probing of the system. Each of these interactions can obviously affect how the scientist sees the system and thereby subsequently affect the resultant measurements and through this, affect the scientific observations, conclusions and the modeling effort. For example, removing a wolf from the pack may help you to understand the isolated wolf's health status, but does not tell us anything about how the social hierarchy of the wolf pack affects the wolf's health. Thus, if you do not know that that social hierarchy and dynamics is important, the single wolf experiment removed an interaction necessary for the study of the wolf and consequently affects the conclusions available to the scientist, thereby affecting the accuracy of the conclusions and any subsequent model-building exercise.

The second interaction, the scientist-system interaction, is more subtle. However, it must be mentioned. The act of actually inserting a probe into a system clearly affects the system. Thus, the scientist is forced to ask the question of whether or not the measurements being obtained are actually those of the isolated system of interest or of the system-probe complex. This question, while seemingly philosophical, has important ramifications in quantum-level measurements and in abstract theoretical biology [12,13].

The parts of a complex system naturally interact with one another, and the fascinating evolutionary dynamics of complex systems depends crucially upon the nature of these interactions. The interactions supply the driving forces for the changes that we observe in systems. In addition, we can change the behavior of a system by introducing new elements or ingredients. Intrusions of this kind produce new interactions, which in turn alter the system. By carefully choosing the added factors and interactions we, as scientists, can develop new patterns of observables that may be revealing. Interaction with your dog might include exercising to increase his running stamina, which in turn will lead to a new, improved set of health-related (state) indicators. Electrical stimulation of a fibrillating heart can introduce interactions that lead to the conversion of the heart from the fibrillating state to a normal, healthy state of performance. Heating the ethyl acetate solution will eventually accelerate the hydrolysis reaction and distill away the resulting ethanol, leaving a solution of acetic acid. The interactions introduced and the accompanying changes in a system's observables produce information about the nature of the system and its behavior under different conditions. With enough observables we may be able to piece together a reasonable description, a model, of how the system operates.

1.3 Back to models

From a carefully selected list of experiments with a system we can evoke certain conclusions. The mosaic of information leads us to piece together a description of the system, what is going on inside it, the relationships among its states, and how these states change under different circumstances. In the case of our dog, the exercise tests may lead us to theorize that the dog is in good or poor health. With the heart, the electrical impulses that we record can reveal a pattern of changes (observables) that we theorize to belong to a healthy (or diseased) heart. By subjecting the solution of hydrolyzing chemicals to fractional distillation and chemical analysis we may theorize that we originally had a system of water and ethyl acetate.

We can arrive at our theories in two main ways. In the first, as illustrated above, we subject a system to experimental perturbations, tests, and intrusions, thereby leading to patterns of observables from which we may concoct a theory of the system's structure and function. An alternative approach, made possible by the dramatic advances that have occurred in the area of computer hardware in recent times, is to construct a computer model of the system and then to carry out simulations of its behavior under different conditions. The computer "experiments" can lead to observables that may be interpreted as though they were derived from interactions.

1.3.1. Simulations

It is important to recognize the different concepts conveyed by the terms "model" and "simulation", even though these terms are sometimes used interchangeably. As noted above, a model is a general construct in which the parts of a system and the interactions between these parts

are identified. The model is necessarily simpler than the original system, although it may itself take on a rather complicated form. It consists of ingredients and proposals for their interactions.

Simulations are active imitations of real things, and there are generally two different types of simulations, with different aims. In one approach a simulation is merely designed to match a certain behavior, often in a very limited context. Thus a mechanical bird whistle may simulate a sound resembling that of a bird, and does so through a very different mechanism than the real source of the sound. Such a simulation reveals little or nothing about the features of the original system, and is not intended to do so. Only the outcome, to some extent, matches reality. A hologram may look like a real object, but it is constructed from interfering light waves.

A second type of simulation is more ambitious. It attempts to mimic at least some of the key features of the system under study, with the intent of gaining insight into how the system operates. In the context of our modeling exercise, a simulation of this sort means letting our model "run." It refers to the act of letting the parts of our model interact and seeing what happens. The results are sometimes very surprising and informative.

1.3.2. Why are modeling and simulation important?

Beginning in the late 1800's, mathematicians began to realize that biology and ecology were sources of intriguing mathematical problems. The very complexity that made life difficult for experimental biologists intrigued mathematicians and led to the development of the field of Mathematical Biology. More recently, as computers became more cost-effective, simulation modeling became more widely used for incorporating the necessary biological complexity into the original, often over-simplified mathematical models.

The experimentalists felt that the theoretical analyses were deficient in a variety of areas. The models were far too simple to be useful in clinical or practical biological application. They lacked crucial biological and medical realism. Mathematical modelers balked at the demands for increased levels of biological complexity. The addition of the required biological reality, desired by the life scientists, often lead to alterations in the formulation of the mathematical models, alterations that made the models intractable to formal mathematical analysis.

With the advent of the new high performance computer technologies and the deluge of 'omic'-data, biological and biomedical reality is finally within the grasp of the bio-modeler. Mathematical complexity is no longer as serious an issue, as new mathematical tools and techniques have grown at nearly the same speed as the development of computational technology [14].

The role of high performance computing and modeling in the sciences has been documented by numerous federal publications (NSF [15,16], for example). Many of the grand challenge bio-computational problems of the 1990's still remain so. Some of these problems, such as multi-scale simulations, and such grand challenge computational problems as linking heart and kidney simulations, which are now beginning to become addressable, were only pipedreams during the 1990's [17-19]. More recently, such models and their simulations are being termed "*in silico*" modeling and simulation.

Modeling and simulation provide the scientist with two very useful tools. The first of these is validation of the theoretical understanding and its model implementation. The second of these tools is that, the more complete the model, the more it provides an experimental laboratory for further research on the very system being modeled. Thus, "*in silico*" models can both validate current viewpoints/perspectives of the dynamical evolution of a system and can provide an environment in which the scientist can explore potential new theories and their consequences. It is this second aspect of models and their simulations that is of particular interest to us. Let us take a moment to address modeling in chemistry and molecular biology.

1.4. Models in chemistry and molecular biology

Chemistry and molecular biology, like other sciences, progresses through the use of models. They are the means by which we attempt to understand nature. In this chapter we are primarily concerned with models of complex systems, those whose behaviors result from the many interactions of a large number of ingredients. In this context two powerful approaches have been developed in recent years for chemical investigations: molecular dynamics and Monte Carlo calculations [20-25]. Both techniques have been made possible by the development of extremely powerful, modern, high-speed computers.

Both of these approaches rely, in most cases, on classical ideas that picture the atoms and molecules in the system interacting *via* ordinary electrical and steric forces. These interactions between the ingredients are expressed in terms of force fields, *i.e.*, sets of mathematical equations that describe the attractions and repulsions between the atomic charges, the forces needed to stretch or compress the chemical bonds, repulsions between the atoms due to their excluded volumes, etc. A variety of different force fields have been developed by different workers to represent the forces present in chemical systems, and although these differ in their details, they tend generally to include the same aspects of the molecular interactions. Some are directed more specifically at the forces important for, say, protein structure, while others focus more on features important in liquids. With time more and more sophisticated force fields are continually being introduced to include additional aspects of the inter-atomic interactions, *e.g.*, polarizations of the atomic charge clouds and more subtle influences associated with quantum chemical effects. Naturally, inclusion of these additional features requires greater computational effort, so that a compromise between sophistication and practicality is required.

The molecular dynamics approach has been called a brute-force solution of Newton's equations of motion [20]. One normally starts a simulation using some assumed configuration of the system components, for example an X-ray diffraction structure obtained for a protein in crystalline form or some arrangement of liquid molecules enclosed in a box. In the protein case one might next introduce solvent molecules to surround the protein. One then allows the system, protein-in-solvent or liquid sample, to evolve in time as governed by the interactions of the force field. As this happens one observes the different configurations of the species that appear and disappear. Periodic boundary conditions are usually applied such that molecules leaving the box on the right side appear on the left side; those leaving at the top appear at the bottom, and so forth. The system's evolution occurs *via* time steps (iterations) that are normally taken to be very short, e.g., 0.5 –2.0 femtoseconds (fs, 10^{-15} seconds), so that Newton's second law of motion

$$F = ma = m(dv/dt),$$

can be assumed to hold in a nearly linear form. The evolution of the system is followed over a very great number of time steps, often more than a million, and averages for the features of interest of the system are determined over this time frame. Because the calculation of the large number of interactions present in such a system is very computationally demanding, the simulations take far longer than the actual time scale of the molecular events. Indeed, at present most research-level simulations of this type cover at best only a few tens of nanoseconds of real time. (Note that 10^6 steps of 1 fs duration equal one nanosecond, 10^{-9} s.) Whereas such a timeframe is sufficient to examine many phenomena of chemical and biochemical interest, other phenomena, which occur over longer time scales, are not as conveniently studied using this approach.

The Monte Carlo method for molecular simulations takes a rather different approach from that of the molecular dynamics method [24-26]. Rather than watching the system evolve under the influence of the force field, as done in molecular dynamics, a very large number of possible configurations of the system are sampled by moving the ingredients by random amounts in each step. New configurations are evaluated according to their energies, so that those lowering the energy of the system are accepted whereas those raising the system energy are conditionally "weighted", or proportionately accepted, according to their potential energies. The weighting is normally taken to have the form of the Boltzmann distribution, *i.e.*, to be proportional to $e^{-\Delta V/kT}$, where ΔV is the potential energy change, k is Boltzmann's constant, and T is the absolute temperature. From statistical analysis of a large, weighted sample (ensemble) of such configurations one can ascertain many of the important thermodynamic and structural features of the system. A typical sample size employed for this purpose might encompass between one and ten million configurations.

Both the molecular dynamics and the Monte Carlo approaches have great strengths and often lead to quite similar results for the properties of the systems investigated. However, these methods depend on rather elaborate models of the molecular interactions. As a result, as noted above, both methods are highly computationally demanding, and research-level calculations are normally run on supercomputers, clusters, or other large systems. In the next chapter we shall introduce an alternative approach that greatly simplifies the view of the molecular system, and that, in turn, significantly reduces the computational demand, so that ordinary personal computers suffice for calculations and elongated time frames can be investigated. The elaborate force fields are replaced by simple, heuristic rules. This simplified approach employs another alternative modeling approach using *cellular automata*. However, before we begin this discussion, we must first address the general subject of complexity.

2. General principles of complexity

2.1. Defining complexity: complicated vs. complex

Up to this point we have been discussing "*complicated*" systems where, by complicated, we mean that they may be organized in very intricate ways, but they exhibit no properties that are not already programmed into the system. We may summarize this by saying that complicated systems are no more than the sum of their parts. Moreover, should we be able to isolate all of the parts and provide all possible inputs, we would, in theory, know everything that there is to know about the system. Complicated systems also have the property that one key defect can bring the entire system to its knees. Thus, in order to make sure that such a problem does not occur, the system must have built-in redundancy. Redundancy is necessary because complicated systems do not adapt. A familiar

example here is household plumbing. This is a relatively complicated system (to me) where there are many cut-off valves that can be used to deal with a leak. The leak does not solve itself.

There are systems, however, where the "whole is greater than the sum of the parts." Systems of this type have the property that decomposing the system and analyzing the pieces does not necessarily give clues as to the behavior of the whole system." We call such systems, "*complex*" systems. These are systems that display properties called "*emergence*," "*adaptation*," and "*self-organization*." Systems that fall under the rubric of complex systems include molecules, metabolic networks, signaling pathways, ecosystems, the world-wide-web, and even the propagation of HIV infections. Ideas about complex systems are making their way into many fields including the social sciences and anthropology, political science and finance, ecology and biology, and medicine. Let us consider a couple of familiar examples. Consider a quantity of hydrogen and oxygen molecules. Gas laws are obeyed, the system can be defined by the nature of both gases. We would call this a simple system.

$$nH_2 + mO_2 \rightarrow kH_2O$$

If we now ignite the system producing a reaction leading to water, we now have a complex system where a knowledge of H_2 and O_2 no longer tell us anything about the behavior and properties of H_2O. The properties of water have emerged from the proto-system of the two gases. The two gases have experience the dissolvence of their properties in this process, an event that will be described later. A second familiar example is the array of amino acids, twenty in nature, that are available for polymerization. When this process occurs, a large, new molecule, a protein, appears. The behavior and properties of the protein are not discernable from a simple list of the amino acids.

$$X \text{ Amino Acids} \rightarrow \text{Protein}$$

The spatial structure and functions of this protein are non-liner functions of the kinds and numbers of the amino acids and their order of linkage in the protein. The amino acids have surrendered their individual properties and functions, blending these into the whole, the protein molecule.
In order to understand the distinction between complex and complicated systems, we need to make some definitions.

2.2. Defining complexity: agents, hierarchy, self-organization, emergence, and dissolvence

2.2.1. Agents

The concept of an agent emerges out of the world of computer simulation. Agent-based models are computer-driven tools to study the intricate dynamics of complex adaptive systems. We use agent-based models because they offer unique advantages to studying complex systems. One of the most powerful of these advantages is the ability of such modeling systems to be used to study complex social systems, complex biological and biomedical systems, molecules, and even complex financial systems that we could not model using mathematical equations or which may be intractable mathematically. Agent-based approaches allow us to examine, not only the final outcome of a simulation, but the whole history of the system as the interactions proceed. Moreover, agent-based models allow us to examine the effects of different "*rules*" on a system.

An *agent* is the lowest level of the model or simulation. For example, if the environment is a checkerboard, then the agents are the checkers; if it is a chess board, then the agents are the chess pieces. Thus, agents act within their environment. However, this is an extremely broad definition of an agent. Let us look at the structure of agents in a closer fashion.

The agent exists within the environment; the agent interacts with the environment/other agents by performing actions on/within it, the environment/other agents may or may not respond to those actions with changes in state. Agents are assumed to have a repertoire of possible actions available to them [27]. These actions can change the state of other agents or the environment. Hence, if we were to look at a basic model of agents interacting within their environment, it would look as follows. The environment and the agents start in an initial configuration/set of states. The agent begins by choosing an action to perform on other agents or on the environment. As a result of this action, the environment/other agents can respond with a number of possible states. What is important to understand is that the outcome of the response is not predictable. On the basis of the response received, the agent again chooses an action to perform. This process is repeated over and over again. Because the interaction is history dependent, there is a non-determinism within the system.

Agents do not act without some sort of rule-base. We build agents to carry out tasks for us. Depending upon what is being modeled, the rule-base can be simple interaction rules or it can be more sophisticated rules about achievement, maintenance, utility, or other performance rules. Thus, provide rules about how the agents function in the environment. Checkers have rules about jumping, kinging, and movement; similarly, chess pieces have rules. While checkers consists of one type of agent (*homogeneous*), "the checker," with a simple set of rules, chess is a "*multi-agent*" (*heterogeneous*) game having different agents with different interaction rules. Closer to home we recognize the rules, called valence, that proscribe the bonding patterns of atoms to form molecules. Additionally, agent interaction rules can be static (unchanging over the lifetime of the simulation) or dynamic. They can operate on multiple temporal and spatial scales (local and global). They can be direct, indirect, or even hierarchical. And, one can even assume a generalized form of inter-agent communication by allowing the agents to see the changes caused by other agents and to alter their operational rules in response to those observations. In the upcoming section on cellular automata, we will illustrate some of these concepts in more detail.

2.2.2. Hierarchy

The concept of *hierarchy* is intuitive 28,29]. If we look at complex systems, we see that they are made up of what we might call "layers" of structure. The human body contains numerous examples of hierarchical structures. The excretory system contains numerous organs. Those organs contain numerous cells, those cells contain numerous subcellular metabolic and signaling systems, and those systems contain numerous atoms and molecules. At each level of "organization," there are rules, functions, and dynamics that are being carried out. Similar arguments can be made about the cardiovascular system, the nervous system, the digestive system, etc. Even the brain can be subdivided in a hierarchical fashion. The brain is formed from the cerebrum, the cerebellum, and the brain stem. The cerebrum is divided into two hemispheres. Each hemisphere is divided into four lobes. Each lobe is further divided into smaller functional regions [30]. And again, in each area of the brain, and at all levels, the "brain" system is engaging in various "hierarchically related" behaviors. Other examples of hierarchies include ecosystems [31] and social systems [9].

What makes hierarchies interesting is not just that they exist, but also how they are structured and how the levels of the hierarchies are interconnected. Moreover, one can ask how hierarchies evolved into the particular forms that they currently display. Additionally, one can ask questions concerning how the functions of the network evolved as they did. These questions lead us to ask about relational aspects of a hierarchy/network, self-organizational aspects of the hierarchy/network, and emergent properties of such systems. Some of these questions will be addressed, in greater

detail, in the chapter by Don Mikulecky in this volume. However, what is important to understand is that hierarchies have both temporal and spatial organization and that these organizational forms create the pathways for patterns of behavior and, as we shall see in a moment, these patterns of behavior and structural forms are often not predictable; they are self-organizational and emergent.

Some of the groundbreaking, original work in this field was done by Nicolas Rashevsky [32] and Robert Rosen [33-36] (who was Rashevsky's student). Their work involved examining biological systems from the standpoint of agents (although at that time they were not called that) and the relationships between the agents. Rosen's work was extended by Witten [37] (who was Rosen's student) in an effort to study the dynamic complexity of senescence in biological systems. A simple argument is as follows. It is certainly clear that every biological organism or system *O* is characterized by a collection *P* of relevant biological properties P_i which allows the observer to recognize our biological organism as a specific organism. That is, these properties allow us to distinguish between organisms. This collection of properties P_i is the set of biological properties representing the organism *O*. It can be the very "coarse" set of sensitivity (S) to stimuli, movement (M), ingestion (I), and digestion (D). Or, at a slightly less coarse (more detailed) level, we might consider the set of all hormones (H) in an organism, and the set of that organism's metabolic responses (R). It is also clear that many of our biological properties P_i will be related to each other in some way. For example, ingestion (I) must come before digestion (D). We may denote this ordering through the use of arrows as follows, I ->D. For those readers who are familiar with business management, an organizational chart is a perfect example of such an interrelated collection of properties. If we were to say that two properties were related, but not indicate how, we would write I-D. We call such figures *graphs*. When the edges of the graph are directed, we term such graphs *directed graphs*. The elements at the intersection of two or more edges are called the nodes or vertices of the graph. Hence, biological hierarchies may, in some cases, be represented by directed graphs: or what we might call *dependency networks* [37]. In summary, we have seen that abstracting away the fine grain aspects of biological systems allows us to represent their complexity (hierarchical structure) by either directed or undirected graphs.

Why is such an approach important? First, it allows us to represent basic processes of the system without being involved in the details at microscopic levels. Such a representation is useful, particularly if the mathematical modeling becomes intractable to analysis. More important is the fact that using hierarchical representation of a system allows us to understand *relationships* between components in a way that is not amenable to traditional mathematical modeling and simulation techniques. It is not so much about what the boxes in the network actually do as about how they are interconnected and how that set of connections creates the possibility for various dynamical behaviors. This approach is currently undergoing a great resurgence with the new studies of genomic [38], metabolic [39,40], and other networks [41]. In fact, this approach is being used to study ecological systems such as food webs [42], human sexual contacts and linguistics [43], and even e-mail networks and telephone networks [44]. Biochemical systems can also be studied with these "*topological*" approaches. Seminal work in this area has been done by Bonchev and his collaborators [45,46]. These structural or topological approaches are generalizable across systems spanning vast orders of hierarchical magnitude. One could say that one of the major characteristics of complex systems is that there are common behaviors, a number of levels or scales that dynamically interact and have many components. A marvelous example of such a complex system can be found in dealing with hierarchical modularity in the bacterial E. coli [47]. In this paper, the authors

demonstrate that the metabolic networks of 43 distinct organisms are organized into many small, highly connected, topologic modules that combine in a hierarchical manner into large, less cohesive units. It follows then that, within the biochemical pathways of metabolic networks, there is a large degree of hierarchical structure [48].

As a consequence of such topological organizational properties, networks generate properties that cannot be simply inferred from the behavior of the components (*emergence*). They develop unpredictable temporal behavior (*chaos*). And they develop the ability to organize themselves in ways that were not obvious from the component pieces (*self-organization*). Let us briefly address these three properties and then illustrate them by focusing on cellular automata models of biochemical systems.

2.2.3. Self-organization and emergence.

Self-organization and emergence are two properties of complex systems that are very much intertwined. Like hierarchy, their meaning seems intuitive. And yet, there is far more to self-organization and emergence than can be easily reviewed, much less covered in this brief chapter. Let us begin with the concept of self-organization. Stuart Kauffman, one of complexity theory's greatest proponents spoke of self-organization in the following fashion, "*Self-organization* is matter's incessant attempts to organize itself into ever more complex structures, even in the face of the incessant forces of dissolution described by the second law of thermodynamics [49]. By means of a simple example, suppose we have the following set of letters, L, S, A, T, E. Moreover, suppose that each of them was written on a card that had magnets placed on its four edges. Obviously, just sitting there, the letters have no intrinsic value other than representing certain sounds; they exist as a collection of objects. If, however, we put them into a shoebox, shake the box, and let them magnetize to each other, we might obtain any number of letter combinations; LTS, ATE, TLSA, STLAE, STALE, etc. Again, intrinsically, these organized structures have no meaning. At the lowest hierarchy level, they represent new organized structures that occurred because we shook them around in a shoebox. Suppose, however, that we now give them context. That is, at a higher hierarchical level, that of language, these strings of letters acquire a new property; that of meaning. Meaning, through the self-organization process of being shaken around in the shoebox, becomes an emergent property of the system. It could not have been inferred from the simple lower-level collection of letters. Suddenly, some of the organized structures, like ATE and STALE, lose their sense as strings of valueless symbols and acquire this new property of being a word with meaning. Similarly, one can make the same argument by putting the magnetized words into the shoebox and creating strings of words. Some of the word strings will acquire meaning and will be called sentences. Perhaps, along with exogenous factors, a language evolves. Language, the emergent property of the interaction of the letters and the environment/culture, could never have been predicted from the properties of the magnetic letters themselves. Nor could it have been predicted from the strings of letters. Thus language itself could be viewed as an emergent property.

Well, at this point, you are wondering what this has to do with anything chemical. Let's take a look at some examples from the biological and chemical world. First we consider a very simple example. Atoms, the lowest hierarchical unit, have certain properties. These properties allow them to combine, in various ways, to form molecular structures. When this happens, the properties of the atoms are lost, in part, to the overall molecular properties. Another simple example is the laser. A solid state laser consists of a rod of material in which specific atoms are embedded. Each atom may be excited by energy from outside, leading it to the emission of light pulses. Mirrors at the end of the

rod serve to select these pulses. If the pulses run axially down the rod, then they will be reflected several times and stay longer in the laser. Pulses that do not emit axially leave the laser. If the laser is pumped with low power, the rod will illuminate, but it will look more like a lamp. However, at a certain pumping power, the atoms will oscillate in phase and a single pulse of light will be emitted. Thus, the laser is an example of how macroscopic order emerges from self-organization. What is interesting about this particular example is that the order is not near equilibrium for the system. In fact, the laser beam is dissipative and far from thermal equilibrium.

Other examples of self-organization occur in the kinetics of the autocatalytic formation of sugars from formaldehyde (formol reaction) [50]. Radical new self-organizational behaviors have been discovered in numerous biochemical systems; enzyme reactions [51], glycolysis [52,53], and the Bray-Liebafsky reaction of iodate and hydrogen peroxide [54] to name just a few. One of the most striking of these reactions and certainly one of the most famous is the Belousov-Zhabotinsky reaction [55]. What happens is that under certain non-equilibrium conditions, this system behaves with all sorts of unexpected and unpredictable behaviors. In terms of its reactants, the BZ-system is not an unusual one. A typical preparation consists of cerium sulfate, malonic acid, and potassium bromate, all dissolved in sulfuric acid. It is easy to follow the pattern formation because an excess of cerium Ce^{4+} ions gives a pale yellow color to the solution, where as if there is an excess of Ce^{3+} ions, the solution is colorless. Depending upon the initial mixing conditions, ionic potential traces show sustained oscillations, damped oscillations, and chaotic oscillations. When viewed spatially, the reaction displaces spiral waves, some of which have multi-arm spirals.

Cellular and subcellular biochemical signaling pathways are also extremely complex. They allow the cell to receive, process, and to respond to information. Frequently, components of different pathways interact and these interactions result in signaling networks. Under various conditions, such networks exhibit emergent properties that are; they exhibit properties that could not have been inferred from the behaviors of the parts. Such properties include integration of signals across multiple time-scales, generation of distinct outputs depending upon input strength and duration, and self-sustaining feedback loops. Moreover, the feedback can result in bi-stable behavior with discrete steady-state activities not available to any of the component pieces. One of the consequences of emergent properties is that it raises the possibility that information for learned behavior of biological systems may be stored within intracellular biochemical reactions that comprise signaling pathways [56].

2.2.4. Emergence

A corollary to emergence is the loss of identity, properties, and attributes (called property space) of the agents as they progressively self-organize to form complex systems at a higher hierarchical level. Testa and Kier have addressed this issue where they have referred this reciprocal event as *dissolvence* [57,58]. It is the reduction in the number of probable states of agents as they engage each other in the synergy with fellow agents. This is a partial loss, they do not disappear but are dissolved into the higher system. As hydrogen and oxygen are consumed in a reaction to form water, these atoms loose their identity as gases with free movement and become joined with each other to change state and to become ensnared in a fixed relationship. To quote H. G. Wells and J. S. Huxley [59]: *He escapes from his ego by this merger and acquires an impersonal immortality in the association; his identity* ***dissolving*** *into greater identity.*

2.2.5. The next step

In the preceding discussion, we have seen how complex behaviors can emerge from the combination of simple systems. Moreover, we have seen how these behaviors are not predictable from the pieces that compose the system. This raises the following question. How does one model such behaviors? In the next section, we will address one modeling approach to handing systems that might display self-organization and emergence phenomena, that of cellular automata.

3. Modeling emergence in complex biosystems

3.1. Cellular Automata

Cellular automata were first proposed by the mathematician Stanislaw Ulam and the mathematical physicist John von Neumann a half century ago [60-62] although related ideas were put forth earlier by the German engineer Konrad Zuse [63]. Von Neumann's interest was in the construction of self-reproducing automata. His idea was to construct a series of mechanical devices or automata that would gather and integrate the ingredients that could reproduce themselves. A suggestion by Ulam [61] led him to consider grids with moving ingredients, operating with rules. The first such system proposed by van Neumann was made up of square cells in a matrix, each with a state, operating with a set of rules in a two-dimensional grid. With the development of modern digital computers it became increasingly clear that these fairly abstract ideas could in fact be usefully applied to the examination of real physical systems [64-67] As described by Wolfram [68] cellular automata have five fundamental defining characteristics:

- They consist of a discrete lattice of cells
- They evolve in discrete time steps
- Each site takes on a finite number of possible values
- The value of each site evolves according to the same rules
- The rules for the evolution of a site depend only on a local neighborhood of sites around it

As we shall see, the fourth characteristic can include probabilistic as well as deterministic rules. An important feature sometimes observed in the evolution of these computational systems is the development of unanticipated patterns of ordered dynamical behavior, or emergent properties to be used to drive further experimental inquiry.

Cellular automata is one of several approaches to the modeling of complex dynamic systems. It is a model because it is used as an abstraction of a system in which a portion has been selected for study. The principle features are the modeling of state changes and /or the movement of parts of a system. Such a simple model would be expected to have a very wide range of applications in nature. Indeed, there are many studies reported in the literature such as music, arts traffic, cities and so forth.

The results of a CA model are new sets of states of the ingredients called the configuration of the system. This configuration arises from many changes and encounters among the ingredients of the CA. These changes may occur over a very long period of "time" in the model. The ability of a computer to carry out many changes simulating a long time is a huge advantage of the machine. Before the computer, it would have taken unfathomable amounts of individual calculations over a vast amount of real time. The CA then, is a platform on which many changes can take place, data collected and reported.

3.2. The general structure

The simulation of a dynamic system using cellular automata requires several parts that make up the process. The cell is the basic model of each ingredient, molecule or whatever constitutes the

system. These cells may have several shapes as part of the matrix or grid of cells. The grid containing these cells may have boundaries or be part of a topological object that eliminates boundaries. The cells may have rules that apply to all of the edges or there may be different rules for each edge. This latter plan may impart more detail to the model, as needed for a more detailed study. This is a grid with ingredients, A an B occupying the cells shown in Figure 3.

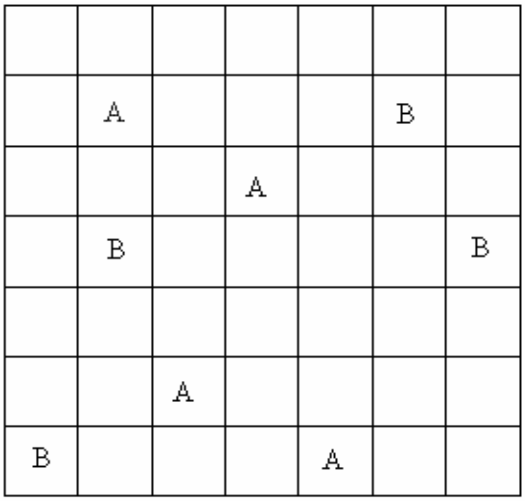

Figure 3: A cellular automata grid with occupied cells containing ingredients A and B.

3.2.1. The cells

Cellular automata have been designed for one, two or three-dimensional models. The most commonly used is the two-dimensional grid. The cells may be triangles, squares, hexagons or other shapes in the two-dimensional grid. The square cell has been the one most widely used over the past 40 years. Each cell in the grid is endowed with a primary state, i.e., whether it is empty or occupied with a particle, object, molecule or whatever the system requires to study the dynamics see, Figure 1. Information is contained in the state description that encodes the differences among cell occupants in a study.

3.2.2. The cell shape

The choice of the cell shape is based on the objective of the study. In the case of studies of water and solution phenomena, the square cell is appropriate since the water molecule is quadravalent to hydrogen bonding to other water molecules or solutes. A water molecule donates two hydrogens and two lone pair electrons in forming the tetrahedral structure that characterizes the liquid state. The four faces of a square cell thus correspond to the bonding opportunities of a water molecule.

3.2.3. The grid boundary cells

The moving cell may encounter an edge or boundary during its movements. The boundary cell may be treated as any other occupied cell, following rules that permit joining or breaking. A common practice is to assume that the grid is simulating a small segment of a very large dynamic system. In this model, the boundaries should not come into play in the results. The grid is then considered to be the surface of a torus shown in Figure 4.

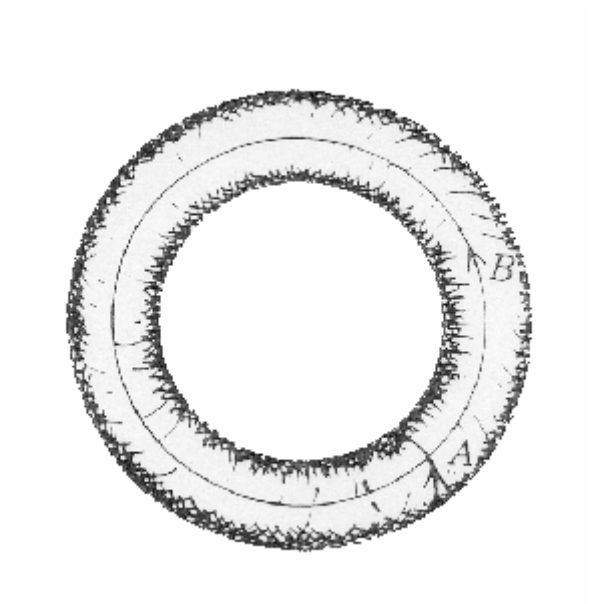

Figure 4: A two-dimensional grid mapped onto the surface of a torus.

In this case the planer projection of this surface would reveal the movement of a cell off the edge and reappearing at the opposite edge onto the grid as shown in Figure 5.

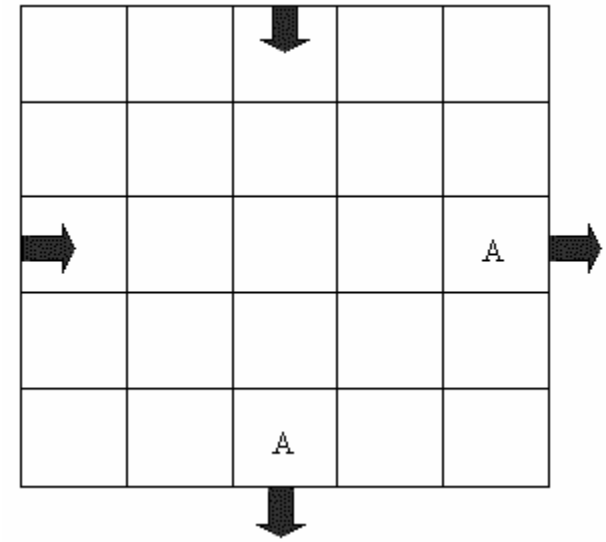

Figure 5: Paths of movement of ingredients on a torus, projected on a two-dimensional grid.

In some cases it is necessary to establish a vertical relationship among occupants. This establishes a gravity effect. For these studies the grid is chosen to be the surface of a cylinder with a boundary condition at the top and bottom which is an impenetrable boundary.

3.2.4. Variegated Cell Types

Until recently, all of the cellular automata models assumed that each edge of a cell had the same state and movement rules. Recent work in our laboratory has employed a variegated cell where each edge may have its own state and set of movement rules, Figure 6.

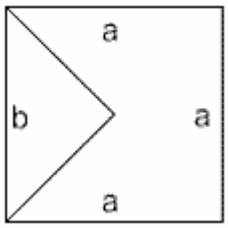

Figure 6: A variegated cell with two different sets of face states.

The cell shown in Figure 7 has three faces, a, with the same state and movement parameters while the other face, b, has a different state and movement parameters.

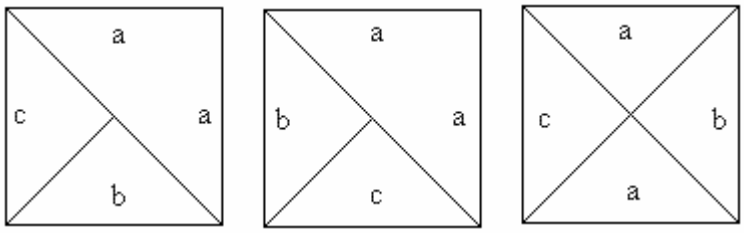

Figure 7: Variegated cells with three different face states.

The three cells shown here have two faces, a, with identical states and movement parameters while the other two faces are different from a and from each other. Note that the faces, a, can be adjacent on the cell or they may be opposite.

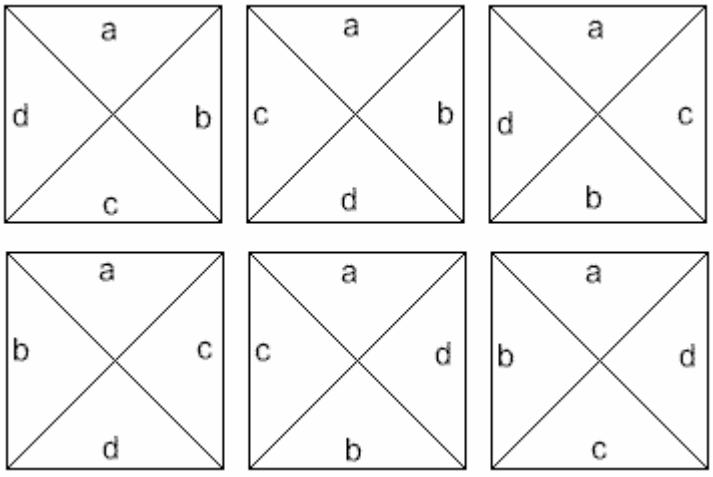

Figure 8: Variegated cells with four different face states.

Finally Figure 8 shows six arrangements of cell faces wherein all of the faces have different states and parameters. Note that mirror images or chirality are present among these cells. These variegated cells can be used for studies in which there are attempts to model different features within the same molecule.

3.3. Cell movement

The dynamic character of cellular automata is developed by the simulation of movement of the cells. This may be a simultaneous process or each cell, in turn, may execute a movement. Each cell computes its movement based on rules derived from the states of other nearby cells. These nearby cells constitute a neighborhood. The rules may be deterministic or they may be stochastic, the latter process driven by probabilities of certain events occurring.

.

3.3.1. Neighborhoods

Cell movement is governed by rules called transition functions. The rules involve the immediate environment of the cell called the neighborhood. The most common neighborhood used in two-dimensional cellular automata is called the von Neumann neighborhood, Figure 9, after the pioneer of the method.

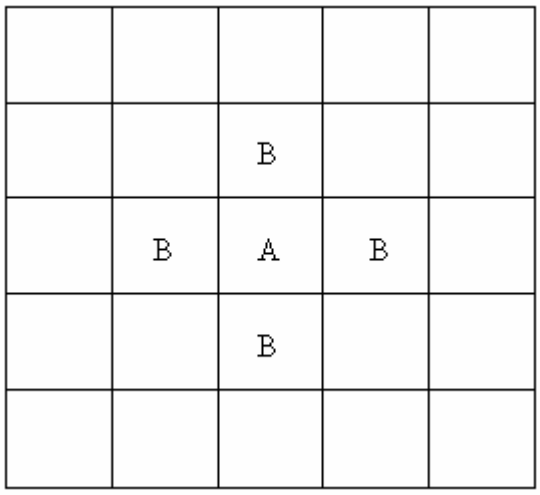

Figure 9: The von Neumann neighborhood. One cell, A, is in the center of four cells, B, adjoining the four faces of A.

.

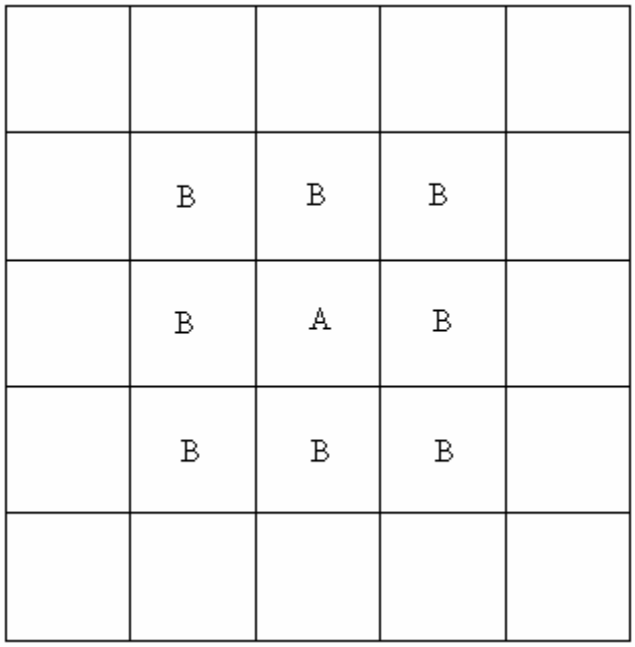

Figure 10: The Moore neighborhood.

Another common neighborhood is the Moore neighborhood, Figure 10, where cell, A, is completely surrounded by cells, B.

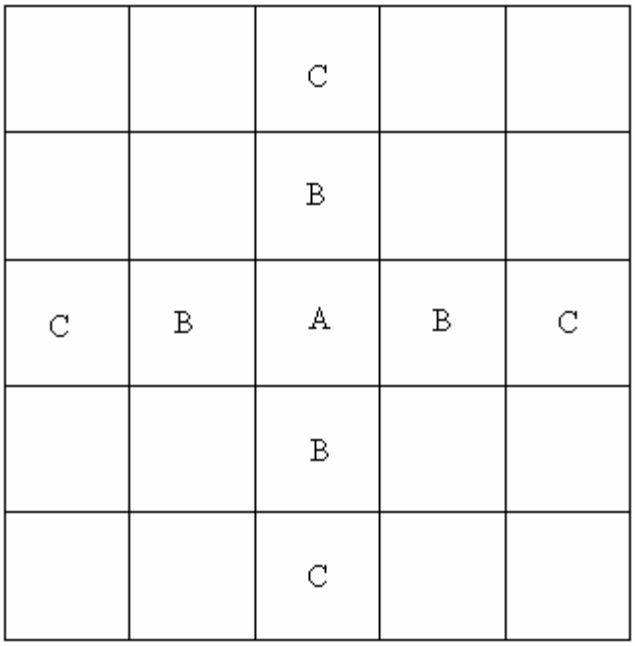

Figure 11: The extended von Neumann neighborhood.

Other neighborhoods include the extended von Neumann neighborhood shown in Figure 11, where the C cells beyond, B, are identified and allowed to participate in movements of the occupant of the A cell.

3.3.2. Synchronous/asynchronous movement

When we speak of movement of a cell or the movement of cell occupants, we are speaking of the simulation of a movement from one cell to another. Thus a molecule or some object is postulated to move across space, appearing in a new location at time t+1. In the cellular automata models the actual situation is the exchange of state between two adjacent cells. If we are modeling the movement of a molecule from place A to the adjacent place B then we must exchange the states of cells A and B. Initially, at time t, cell A has a state corresponding to an occupant molecule, while adjacent cell B is devoid of a molecule, i.e., it is empty. At time t+1, the states of the cells A and B have exchanged. Cell A is empty and cell B has the state of the occupant molecule. This exchange gives the illusion, and the practical consequences of a movement of an ingredient from cell A to cell B. We speak of the movement of cells or of the movement of cell occupants; either way we are describing the process of simulating a movement as stated above. This effect is shown in Figure 12 for synchronous movement.

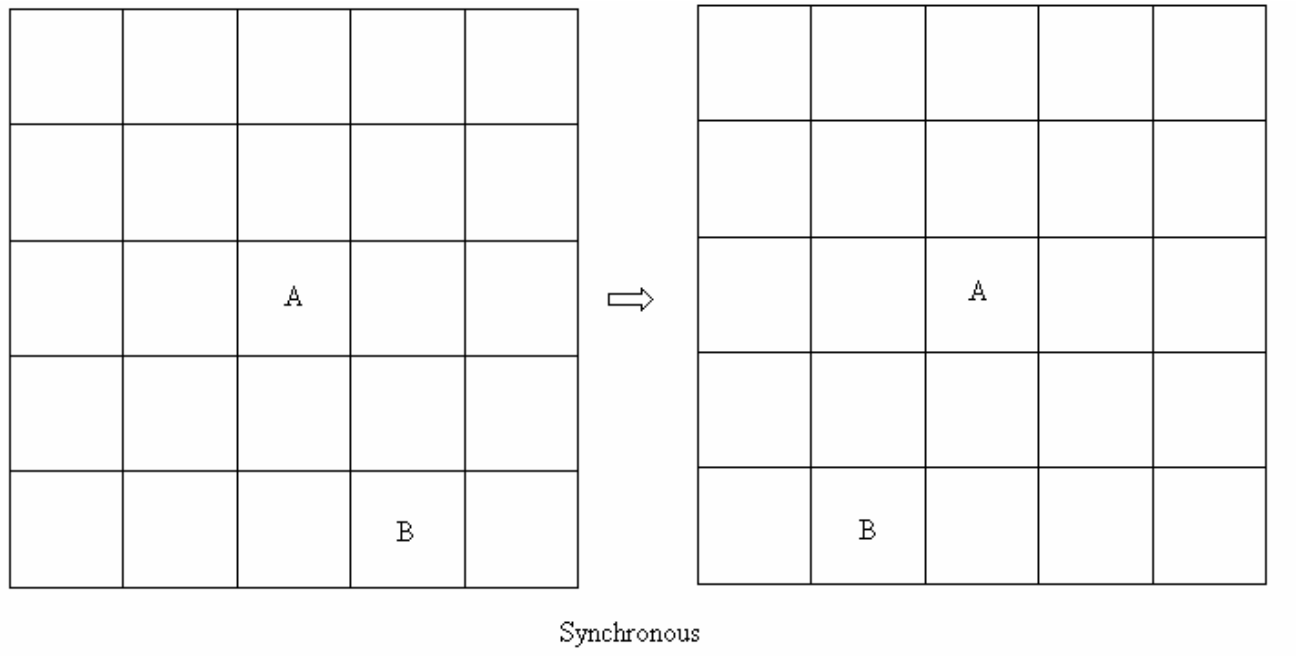

Figure 12: Synchronous movement of all ingredients n the grid.

Asynchronous movement is shown in Figure 13.

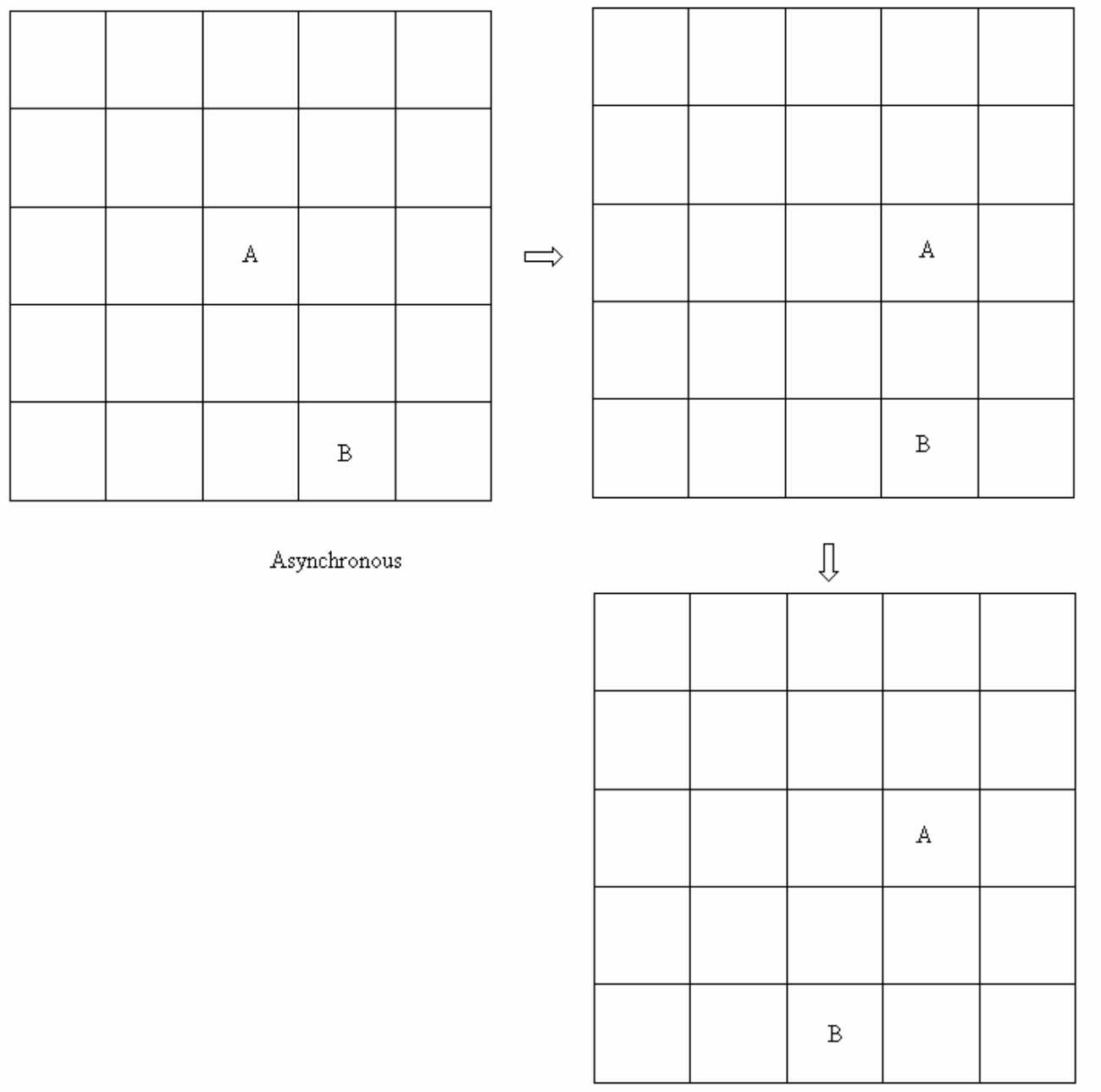

Figure 13: Asynchronous movement of two ingredients in the grid.

The movement of all cell occupants in the grid may occur simultaneously (synchronous) or it may occur sequentially (asynchronous). When all cells in the grid have computed their state and have executed their movement (or not) it is one iteration, a unit of time in the cellular automata model. In asynchronous movement each cell is identified in the program and is selected randomly for the choice of movement or not. The question of which type of movement to use depends upon the system being modeled and the information sought from the model but it should reflect reality. If the system being studied is a slow process then synchronous motion may be best represent the process. In contrast, if the system is very fast, like proton hopping among water molecules where

the cellular automata is using a few thousand cells, then an asynchronous model is desirable. A synchronous execution of the movement rules leads to possible competition for a cell from more than one occupant. A resolution scheme must thus be in place to resolve the competition, otherwise this may interfere with the validity of the model.

3.3.3. Deterministic/probabilistic movement rules

The rules governing cell movement may be deterministic or probabilistic. Deterministic cellular automata use a fixed set of rules, the values of which are constant and uniformly applied to the cells of the same type. In probabilistic cellular automata, the movement of i is based on a probability-chosen rule where a certain probability to move or not to move is established for each type of i cell at its turn. Its state, (empty or occupied) is determined, then its attribute as an occupant is determined. The probability of movement is next determined by a random number selection between two predefined limits. As an example the random choice limits are 0 to 1000. A choice of numbers between 0 and 200 are designated as a "move" rule while a choice in the remaining number set, 201 to 1000, is a "no-move" rule; the case representing a probabilistic rule of 20% movement. Each cell then chooses a random number and behaves according to the rule corresponding to that numerical value.

3.4. Movement (transition) rules

The movement of cells is based upon rules governing the events inherent in cellular automata dynamics. These are rules that describe the probabilities of two adjacent cells separating, two cells joining at a face, two cells displacing each other in a gravity simulation or a cell with different designated edges rotating in the grid. These events are the essence of the cellular automata dynamics and produce configurations that may possibly mirror physical events.

3.4.1. The free movement probability

The first rule is the movement probability, P_m. This rule involves the probability that an occupant in a cell, A, will move to an unoccupied adjacent cell. An example is cell A that may move (in its turn) to any unoccupied cell. As a matter of course this movement probability, P_m, is usually set at 1.0, which means that this event always happens (a rule).

3.4.2. Joining parameter

A joining trajectory parameter, J(AB), describes the movement of a molecule at, A, to join with a molecule at, B, or at C when an intermediate cell is vacant, shown in Figure 14.

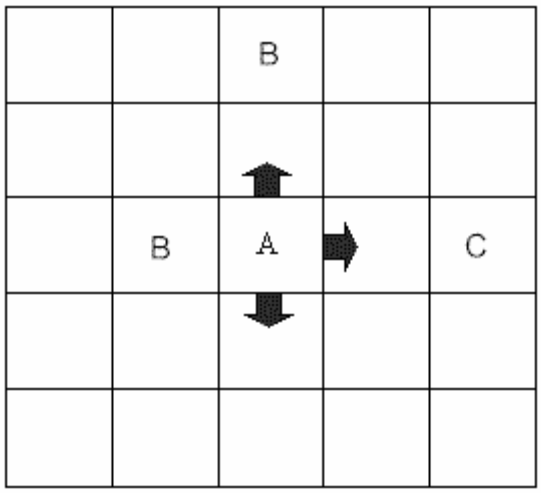

Figure 14: Results of a trajectory parameter operating on cell A ingredient.

This rule is computed after the rule to move or not to move is computed as described above. J is a non-negative real number. When J = 1, the molecule A has the same probability of movement toward or away from C as for the case when the C cell is empty. When J > 1, molecule A has a greater probability of movement toward an occupied cell B than when cell B is empty.

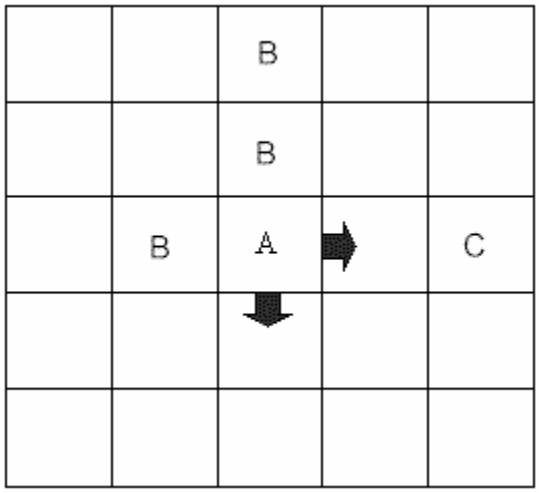

Figure 15: An arrangement on a grid where a movement away from grid C is favored when 0<J<1.

When 0<J < 1, the molecule A in Figure 15 has a lower probability of such movement. When J = 0, molecule A can not make any movement.

3.4.3. Breaking rules

Just as two cells can join together, so their joined state can be broken. This is a movement governed by the trajectory rule called the breaking probability, P_B. To comprehend this it is

convenient to refer to the occupant of a cell as a molecule. The $P_B(AB)$ rule is the probability for a molecule, A, bonded to molecule, B, to break away from, B, shown in Figure 16.

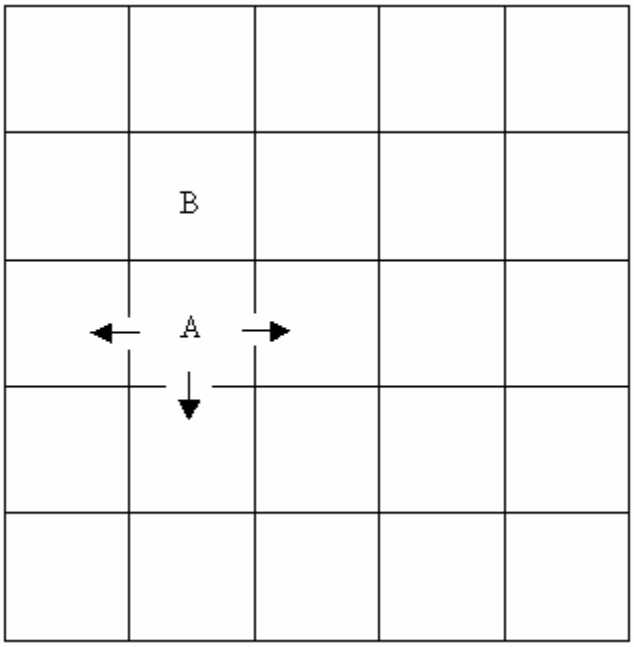

Figure 16: The breaking of A away from B governed by the P_B rule assigned.

The value for $P_B(AB)$ lies between 0 and 1. If molecule A is bonded to two molecules, B and C, as shown in Figure 17,

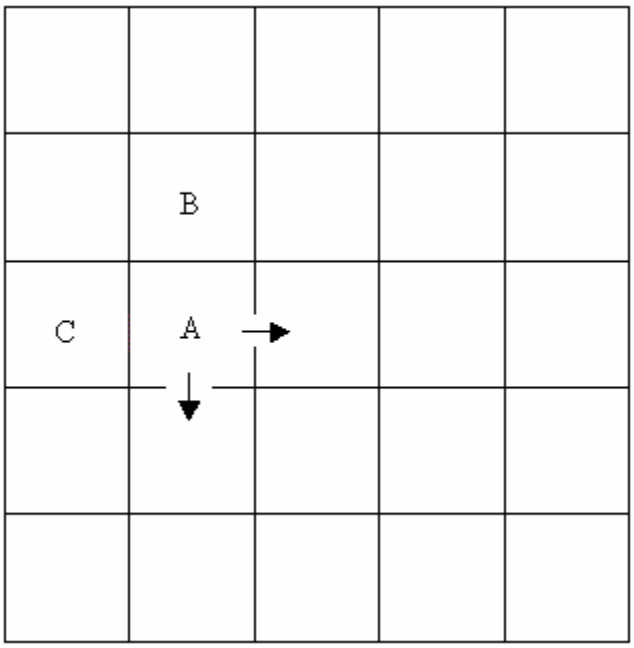

Figure 17: A grid arrangement where A is bonded to two other ingredients.

the simultaneous probability of a breaking away event from both B and C is $P_B(AB)*P_B(AC)$. If molecule A is bound to three other molecules (B, C, and D), shown in Figure 18

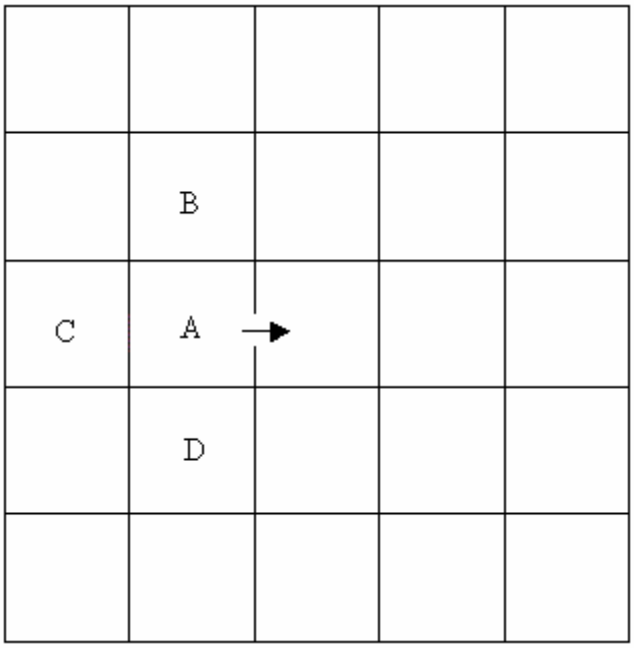

Figure 18: The grid arrangement where ingredient A is bonded to three other ingredients.

the simultaneous breaking probability of molecule A is $P_B(AB)*P_B(AC)*(P_B(AD)$. Of course if molecule, A, is surrounded by four molecules, Figure 19, it cannot move.

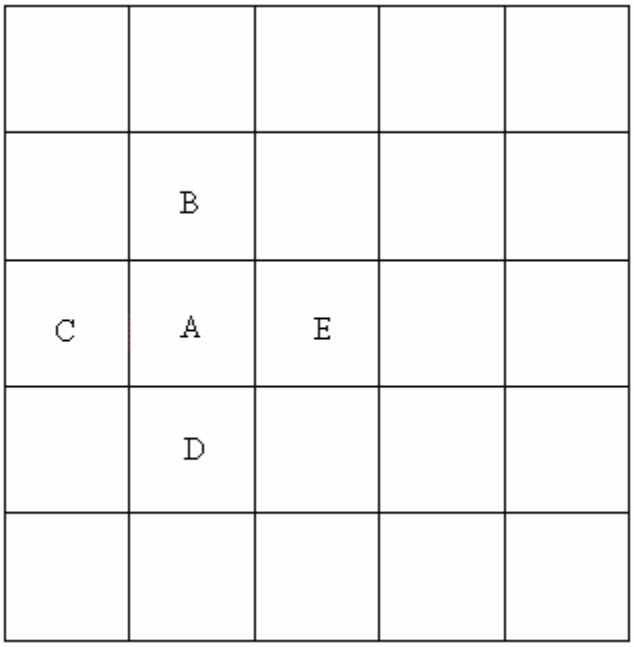

Figure 19: Arrangement of ingredient A in which all four faces are bound to other ingredients.

3.4.4. Relative gravity

The simulation of a gravity effect may be introduced into the cellular automata paradigm to model separating phenomena like the de-mixing of immiscible liquids or the flow of solutions in a chromatographic separation. To accomplish this a boundary condition is imposed at the upper and lower edges of the grid to simulate a vertical verses a horizontal relationship. The differential effect of gravity is simulated by introducing two new rules governing the preferences of two cells of different composition to exchange positions when they are in a vertically joined state. When molecule A is on top of a second molecule B, then two new rules are actuated. The first rule $G_D(AB)$ is the probability that molecule, A will exchange places with molecule B, assuming a position below B. The other gravity term is $G_D(BA)$ which expresses the probability that molecule B will occupy a position beneath molecule A. These rules are illustrated in Figure 20.

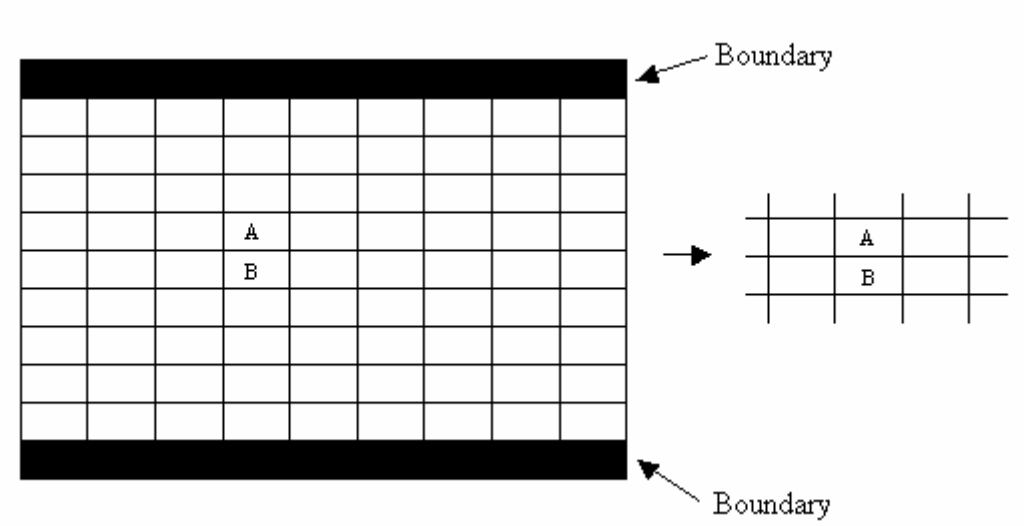

Figure 20: Sequence of movements reflecting the relative gravity rule.

In the absence of any strong evidence to support the model that the two gravity rules are complementary to each other in general cases, the treatment described above reflects the situations in which the gravity effects of A and B are two separate random events. Based on an assumption of complimentarity, the equality $G_D(AB) = 1 - G_D(BA)$ may be employed in the gravity simulation. Once again the rules are probabilities of an event occurring. The choice of actuating this event made by each cell, in turn, is based on a random number selection from a set of numbers used for that particular event.

3.4.5. Absolute gravity

The absolute gravity measure of a molecule, A, denoted by absG(A), is a non-negative number. It is the adjustment needed in the computation to determine its movement if, A, is to have a bias to move down (or up).

3.4.6. Cell rotation

In cases where a variegated cell is used it is necessary to insure that there exists a uniform representation of all possible rotational states of that cell in the grid. To accomplish this, the variegated cells are rotated randomly, (90^0), every iteration after beginning the run. This process is illustrated for four iterations in Figure 21.

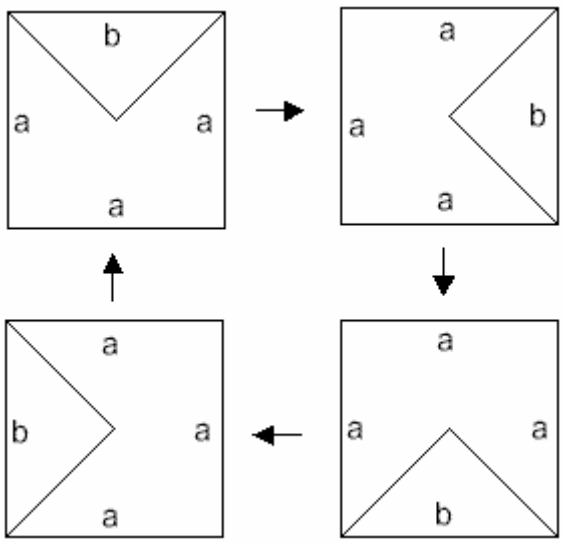

Figure 21: Cell rotations in four iterations.

3.5. Collection of data

A cellular automata simulation of a dynamic system provides two classes of information. The first, a visual display may be very informative of the character of a system as it develops from initial conditions. This can be a dynamic portrayal of a process that opens the door to greater understanding of systems. The second source of information is in the counts of cells in different configurations such as those in isolation or those that are joined to another cell. This is called the configuration of the system and is a rich source of information from which understanding of a process and the prediction of unforeseen events may be derived.

3.5.1. Number of runs

It is customary to collect data from several runs, averaging the counts over those runs. The number of iterations performed depends upon the system under study. The data collection may be over several iterations following the achievement of a stable or equilibrium condition. This stability is reckoned as a series of values that exhibit a relatively constant average value over a number of iterations. In other words there is no trend observed toward a higher or lower average value.

3.5.2. Types of data collected

From typical simulations used in the study of aqueous systems, several attributes are customarily recorded and used in comparative studies with properties. These attributes used singly or in sets are useful for analyses of different phenomena. Examples of the use and significance of these attributes will be described in later examples. The designations are:

f_0 - fraction of cells not bound to other cells
f_1 - fraction of cells bound to only one other cell
f_2 - fraction of cells bound to two other cells
f_3 - fraction of cells bound to three other cells
f_4 - fraction of cells bound to four other cells

In addition, the average distance of cell movement, the average cluster size and other attributes may be recorded.

4. Examples of cellular automata models

4.1. Introduction

Over the past decade we have focused our attention on the use of cellular automata dynamics to model some of the systems of interest to the chemist and biologist. The early work in our laboratory has been directed toward the study of water and solution phenomena. This has resulted in a number of studies modeling water at different temperatures leading to a structural profile of degrees of bonding related to temperature. Such a profile is a structural surrogate for temperature in the correlations with properties of water. Properties normally related to temperature may now be related to structure. Studies include cellular automata models of water as a solvent [69], dissolution of a solute [70], solution phenomena [71], the hydrophobic effect [72], oil and water de-mixing [73], solute partitioning between two immiscible solvents [74], micelle formation [75], diffusion in water [76], membrane permeability [77], acid dissociation [78] , and dynamic percolation [79]. These studies have been summarized in reviews [80-82]. We will discuss some more recent cellular automata models carried out at the Center.

4.2. Water structure

Evidence shows that bulk water contains a significant amount of free space referred to as cavities or voids. It is obvious that water could not permit the diffusion of solutes through it if there was no space between water molecules. In ice, this is not the case since water molecules are bound to approximately four other water molecules. The choice of how many water molecules should be represented on a CA grid of a certain size was explored by Kier and Cheng [80]. Two approaches were taken. On the basis of estimates of the volume of a water molecule and the number of water molecules in a mole, an estimate of about 69% occupancy of a grid was deduced.

The second approach was to conduct CA runs with varying water concentrations. The attributes of the CA configuration were interpreted and compared with experimental values. After a sufficient number of runs, the average number of cells joining each cell was recorded. This attribute was judged to be a model of the average number of hydrogen bonds per molecule of water. A good correspondence of this value was found for a water concentration of about 69% of the grid cells. Another attribute from these experiments, the number of free, unbound water molecules was recorded. This small percentage of the total number of waters was compared with the number of free waters from experiment. The best correspondence was found for a CA system containing about 69% water molecules in the grid. From this information, a system modeling water was adopted using 69% water in the grid.

4.2.1. Water movement rules

Three rules must be chosen to impart a “water character” to the occupied cells that we designate. The first of these is the movement probability P_m. There is no practical reason to believe that anything other than $P_m = 1.0$ for water has any real significance. This choice characterizes water as a freely moving molecule whenever it is possible. The other two rules governing the joining and breaking of water molecules are critical to their behavior and to the emergent attributes of the CA dynamics. Recall that the joining rule, J, encodes the probabilities of water molecules to join others to form a bond (a hydrogen bond in the case of water). The breaking probability, P_B, describes the tendency of bound waters to break apart. The selection of these rules is essential if the model is to have any validity.

The linkage between rules J and P_B can be made relative to a range of values of one of them. As described earlier, the P_B value ranges from zero to one, therefore the J value may be chosen as a function of P_B.

$$\text{Log } J = -1.5\ P_B + 0.6$$

The wisdom of this choice can be tested by comparing the attributes from the dynamics with physical properties.

4.2.2. The attributes recorded

A CA run leads to a configuration that is constant in an average sense. Several attributes of this configuration may be recorded and used for further study. The configuration may be analyzed for the numbers of molecules with no neighbors, one neighbor, and so on up to four neighbors. The fraction of each state are represented as shown in the f_x vs "Temperature" plot above. The distribution of these fractional values becomes a profile of the state of a molecule. It is observed that these states change with different J and P_B rules and that these states have some correspondence to the temperature of the system. The f_x attributes computed for different values of P_B and the corresponding J rules are plotted in Figure 22.

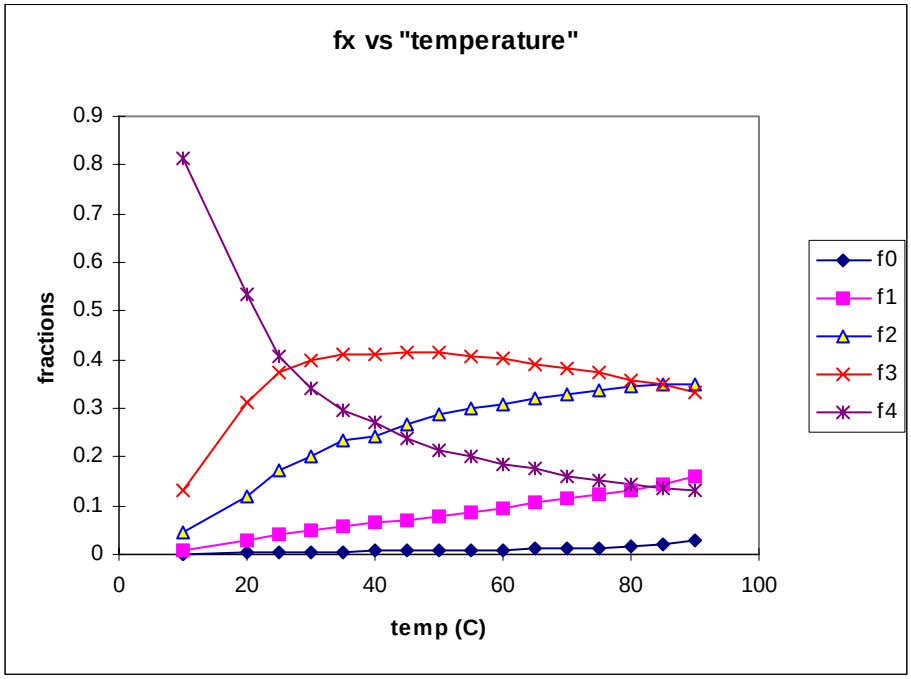

Figure 22: Fractions of cell binding states as a function of modeled temperature.

If we assign a "temperature" of the water to a P_B value according to the relationship:

$$T\ (^0C) = 100\ P_B$$

then sets of f_x values can be related to temperatures of water. From this relationship, selected physical properties at different temperatures may be related to f_x values at those temperatures. An analysis of several properties demonstrate the relationship with selected f_x values and the general validity of our CA model of water [78,79]. Some of these analyses are shown in Table 1.

Table 1
Water Properties Related To
Cellular Automata Attributes

Property	Equation	r^2 Correlation
Vapor Pressure	Log P_v (mm Hg) = 13.77 (f_0 + f_1) + 0.795	0.987
Dielectric Constant	ε = -224 f_1 + 86.9	0.989
Viscosity	η (centipoise) = 3.165 f_4 - 0.187	0.989
Ionization	-Log K_w = -20.94 f_H + 16.43	0.999
Surface Tension	γ (dynes/cm) = 16.07 N_{HB} + 22.35	0.970
Compressibility	κ (x10^6/Bar) = -53.82 f_3 + 66.66	0.953

The conclusion that selected values of the movement rules produce a meaningful profile of f_x values, makes it possible to proceed with some confidence that this CA model of water has validity.

4.3. Cellular automata models of molecular bond interactions

One emergent property arising from an ensemble of agents in a complex system is the extent of molecular interactions, that is the pattern of behavior as free moving agents join and break with their neighbors. In the condensed phase, molecules like water move about, bind with each other in response to input of energy usually in the form of heat. A vigorous pattern of repulsive interactions leads to the transition of the liquid to the gaseous state. The extent of these interactions may be modeled by recording the encounters of molecules with a thermometer bulb. A displacement of mercury in the thermometer is taken to be a model of a certain number of interactions among the liquid molecules. At the phase transition the recorded temperature is the boiling point. This emergent property is a consequence of the structure of the molecules exhibiting this behavior. What is it about the structure difference between two different molecules that gives rise to two different boiling points? It is the topological structure of the molecule, permitting more or fewer binding interactions among molecules. The greater the number of binding interactions, the greater the tendency of the molecules to remain in contact with a neighbor. This translates into a lower propensity of the molecules to be liberated from the bulk liquid and to remain as an ingredient in the liquid. The fewer the intermolecular bond interactions, the higher the recorded temperature. If we could reckon the relative extent of these binding states, we may have a structural model of the water at the temperature of the boiling point.

A novel approach to modeling intermolecular interactions was proposed by Kier [81] in which encounters of molecules were dissected into the encounters of individual bonds, an approach called *disjecta membra* Each bond type was simulated by an occupied cell on a cellular automata grid. Each occupied cell has a particular state value, derived from the descriptors, C_{ij} in Table 2.

Table 2
Alkane Bond Types and

Connectivity Descriptors, C_{ij}

Bond Type, $\delta_i - \delta_j$ [(a)]	$C_{ij} = (\delta_i \delta_j)^{-0.5}$
1,1	1.000
1,2	0.707
1,3	0.577
1,4	0.500
2,2	0.500
2,3	0.408
2,4	0.354
3,3	0.333
3,4	0.289
4,4	0.250

(a) The symbols δ_i and δ_j are the counts of the number of carbon atoms bonded to atoms i and j. Atoms i and j are bonded to each other.

The rules for joining and breaking of each type of CA cell with another CA cell were derived from the bond encounter possibilities using the product of the bond types in the encounter, $(C_{ij})(C_{kl})$, between two molecules. To scale these possibilities to the trajectory J rules for our cellular automata model we have adopted the relationship:

$$J \text{ for } (C_{ij})(C_{kl}) = 4(C_{ij})(C_{kl})$$

The breaking probabilities, $P_B(C_{ij})(C_{kl})$, were derived from the relationship used in studies of water:

$$\text{Log } J\ (C_{ij})(C_{kl}) = -1.5\ P_B\ (C_{ij})(C_{kl}) + 0.6$$

Each carbon-carbon bond in an alkane was represented by a particular state of a cell. One hundred molecules were modeled in a grid of 3025 cells. For example, the *disjecta membra* model of 3-methyl pentane uses 200 cells with a state corresponding to bond type (1,2), 100 cells with a state corresponding to bond type (1,3), and 200 cells with a state corresponding to bond type (2,3). Each cell moved randomly during one iteration, joining another cell, breaking from another cell or moving freely in unoccupied grid space. The dynamics were run for 990 iterations and then during the next 10 iterations the count of the number of joined cells was recorded. This process was repeated for 25 runs and the count of joined cells was averaged from this data. The count of joined cells was called the beta, β, value. Thirty eight alkanes including all of the pentanes, hexanes, heptanes and octanes and three cycloalkanes were modeled and the beta values recorded

.

4.3.1. Results

The interpretation of the β value is a relationship with a physical property that is highly dependent upon intermolecular binding interactions. One such property is the boiling point. The β values, expressed in a quadratic equation, relates to the boiling point with the statistics shown:

$$\text{B.p. } (^0\text{C}) = 0.584\ \beta - 0.0004\ \beta^2 - 71.517$$
$$r^2 = 0.991,\ s = 2.996,\ n = 38,\ F = 2027$$

4.3.2. Discussion

The count of cell encounters averaged over time (iterations), encoded by the beta value is very closely related to the boiling point of the 38 alkanes. The standard deviation of only 3.00 degrees is better than any one-variable quadratic analysis we have found reported. The cellular automata dynamics improves the modeling of bond encounters compared to the static count of all possible bimolecular encounters. The quality of the relationships revealed here supports the cellular automata dynamic model using the *disject membra* simulation of the important topological features of these molecules.

We propose that treating the bonds of a molecule as *disjecta membra* is a model with a limited objective and a limited relationship to reality. It is an example of the analysis of a complex system using reduction to isolate relevant parts followed by synthesis using cellular automata dynamics to create a model that reveals some information about emergent properties and the role that the ingredients contribute to the whole. In this case, we considered just the bonds in alkanes, endowed with numerical values reflecting their accessibility to other bonds in other molecules. Our dynamic model in a limited way, simulates the conditions of a molecule in its milieu.

4.4. Diffusion in water

The diffusion of solutes through water is a means of transport of both vital and noxious compounds within the body. We see this phenomenon everywhere, across a synapse, through a nephron, within cells, along blood vessels; where water goes, so go the solutes via diffusion. Water is uniquely structured to facilitate this passage by virtue of its reactivity in bulk. A water molecule will form up to four hydrogen bonds with neighbors. This is a dynamic pattern, with hydrogen bonds forming and breaking at a rate of 10^{-14} seconds. Water molecules constitute about 2/3 of the space they occupy in bulk, thus large volumes of space are available for solute passage as the architecture of the water changes. It is clear that these spaces, voids or chreodes are the passages through which all diffusion occurs through water.

The term, chreodes was used by Kier [82] in a recent article describing a theory of the facilitated and preferred passage of ligands across the landscape of a protein to a receptor or enzyme active site. The chreode is a temporarily connected series of evanescent cavities in the water enshrouding the field of amino acid side chains on the surface of a protein. The varying hydropathic states of these side chains produces an influence on the nearby bulk water ranging from the hydrophobic structuring of water to electrostriction, This varying pattern creates evanescent, favored pathways, chreodes, through the water whereby a molecule experiences a facilitated diffusion from anywhere on the protein surface, to the receptor.

The influences on the molecule that govern its diffusion characteristics are its size, its hydropathic state, and the temperature (structure) of the water. Some experimental evidence and modeling has demonstrated these influences, leading to the conclusion that more hydrophobic molecules diffuse faster than hydrophilic molecules of the same size [73]. The studies of these influences are scarce because of the difficulties in conducting diffusion measurements with varying parameters among the ingredients. The modeling results mirror reality as far as the hydropathic state influence on the rate of diffusion. A major advantage of these kinds of models is that it is possible to manipulate one variable while holding others constant, thus creating a profile of a system and its behavior under several sets of conditions.

In this study we examined by modeling, the influence of water temperature and solute hydropathic state on the diffusion of the solute through water. In addition, we modeled the water

and the cavities within it to attempt to explain the observed behavior. The modeling is accomplished using asynchronous, probabilistic cellular automata, as we have described above.

4.4.1 Temperature and Hydropathic State Influences on Diffusion Rate

In order to establish the relationship between the extent of diffusion and the water temperature, we first recorded the diffusion every 100 iterations up to 1000 iterations for several modeled temperatures of water, labeled W. The hydropathic states of the solute molecules, labeled S, were held constant at an intermediate value. This was accomplished by using the parameters for solute-water interaction as $P_B(W,S) = 0.5$ and $J(W,S) = 0.7$. This modeled a solute with a mid-level hydropathic state. The solute-solute interaction parameters were held constant at $P_B(S,S) = 0.5$ and $J(S,S) = 0.7$. The diffusion was shown to be linear with time (iterations), characteristic of a random walk where the distance traveled is known to be linear with a function of time. These results produced a confidence in the model and led us to the use of a common iteration time, 1000 iterations, to compare various hydropathic states with their influence on diffusion.

The extent of diffusion at 1000 iterations was then recorded for various combinations of water temperature and solute hydropathic state. These results are shown in Table 3.

Table 3
Extent of Diffusion at 1000 Iterations

Hydropathic State	Temperature						
$P_B(W,S)$	0.10	0.20	0.30	0.40	0.50	0.60	0.70
0.10	0.50	0.54	0.29	0.20	0.14	0.19	0.21
0.20	1.24	1.75	1.25	1.20	1.15	1.08	1.20
0.30	1.74	3.40	3.74	4.02	3.44	3.13	3.25
0.40	2.05	3.60	5.24	4.78	5.61	5.52	4.34
0.50	2.68	3.90	4.18	6.22	6.96	7.11	7.02
0.60	2.36	4.14	6.42	7.85	7.56	8.95	8.45
0.70	2.11	3.92	7.08	8.02	9.36	9.49	9.27

The diffusion was measured as the average count of cells traversed by the solute from the center of the grid after 1000 iterations. This average was computed from 40 runs. The distance was counted as cell faces, not diagonal distances since the solutes only move in rectangular patterns.

4.4.2. Results

From this table we see that diffusion is modest at all temperatures for relatively hydrophilic solutes. As the hydrophobic character increases, the diffusion rate increases for all temperatures. The response to these parameters is, however, not linear. As can be seen, the diffusion of mid-level hydrophobic solutes goes through a maximum at mid-temperatures. When the solute hydropathic state is non-polar the diffusion increases increased temperature. With intermediate hydropathic states, modeled with $P_B(W,S) = 0.3$, 0.4, and 0.5, the diffusion rate is maximal at intermediate temperatures modeled with $P_B(W) = 0.4$, 0.5, and 0.6, falling to lower rates at higher temperatures.

We surmise that this non-linear behavior must be due to at least two intersecting changes in the complex system formed by the solute and water at different temperatures and hydropathic states.

To explore these possibilities we have modeled the water architecture at different states relevant to conditions described above.

4.4.3. The architecture of water

Another effect emerging from temperature changes of water is the distribution of the cavities among the molecules. A study of the average cluster size and distribution of the clusters may shed some light on its behavior in the presence of solutes of differing hydropathic states. We have run cellular automata simulations of water at various temperatures, recording the average size of the cavities. These simulations treated the cavity spaces as discrete entities and so we could make the measurements of their average size and distribution. Table 4 reveals the architecture of water cavities at various temperatures.

Table 4
Average Cavity Structure

$P_B(W)$ (WaterTemp.)	Ave. Cavity Cluster Size (cells)
0.20	7.2
0.30	4.2
0.40	3.0
0.50	2.4
0.60	2.1
0.70	1.8
0.80	1.6

4.4.4. Discussion

The rate of diffusion declines at a certain temperature because the increase in the number of hydrogens available for hydrogen bonding increases. This produces a greater extent of solute hydration, which produces a larger effective size hence, resistance to diffusion through the cavities. As the temperature rises from cold to about mid-temperature, the average size of the water cavities decreases and openings between cavities occur. This permits diffusion of a solute to occur more readily between groups of cavities. As the temperature rises above the mid level, the cavities become smaller in size but more connected. As a result, there is more cavity surface area exposed to solutes. If the solute is hydrophilic, it will form more hydrogen bonds with this increased polar surface of the cavity. These hydrogen bonds produce an effectively larger, hydrated solute molecules, reducing the diffusion rate. These two intersecting effects produced a non-linear rate of diffusion for solutes in the mid-range hydropathic state.

4.5. Chreode theory of diffusion in water

A major focus of attention in drug-receptor studies is the structural influence on the encounter of these two molecules. Variation in the structure of the drug is often related, through models, to the binding affinity with a receptor or the catalytic rate with an enzyme. This leads to progress in the design of new, active biomolecules and some understanding of the processes underway. Less attention has been paid to the journey of the drug to the receptor.

4.5.1. The role of diffusion

An early view described the approach of a drug through bulk water to a receptor. This three-dimensional random walk is postulated by Welch [83] and others to be too slow to permit

coordination of the heterogeneous metabolic processes in living systems. The current view invokes some period of residence of the drug in the vicinity of the protein surface, on the way to an encounter with a receptor. The nature of this two-dimensional passage is not generally agreed upon and is the subject of several models. The central idea is that the approach of a drug to a receptor is a process assumed to be diffusion, acting as a limiting condition to the rate of the reaction, [84-86].

It is well-known that water is an essential ingredient in the reactions of biological phenomena, indeed the complex system of drug-receptor-water is described by Kier and Testa [87] as a triad that is at the core of biological transformations. Any model of two-dimensional diffusion of a drug across a membrane or protein surface must include the participation of water. Water is an evanescent substance, constantly changing its architecture by making and breaking hydrogen bonds. Diffusion through water is a passage through the voids between clusters of hydrogen-bonded water molecules. The presence of solute molecules has a varying influence on water molecules in their vicinity depending on the interactions possible between the different species. Polar solutes form hydrates, collections of water molecules in intimate contact with the solute molecule. In contrast, non-polar solute molecules are not effectively bound to water molecules, allowing the local organization of water to occur, driven by the preference of water to bind to water rather than water binding to solute. This phenomenon is called the hydrophobic effect.

Evidence may be interpreted as describing a facilitation of reaction rates and receptor activation due to the rapid diffusion of molecules to target sites in essentially a two-dimensional domain. This diffusion most likely occurs across the surface of a protein, involving structural features not part of the receptor. The process of guidance of the drug molecules to the receptor would be expected to result in several effects. These include some period of retention on the protein surface, a minimum extent of binding delaying the passage, and some influence facilitating the movement of the drug to the receptor. These considerations and the demonstrated role of water led us to consider a model involving the immediate layers of water enshrouding the protein.

4.5.2. Drug molecule diffusion and the hydrophobic effect

We have proposed that drug molecules encounter the surface of a protein molecule and are captured within the first few layers of water on the surface. They are then guided to the receptor over a series of cavity paths in the water created by the hydrophobic effect responding to the hydropathic state of each amino acid side chain. These paths are preferences reminiscent of the chreodes envisioned by Waddington [88] in his description of an epigenetic landscape. He coined the word chreode from the Greek words for “necessary” and “route” or path. He defined it as a representation of a temporal succession of states of a system. That system is characterized by a property that a dynamic system will tend to respond to perturbations by returning to the chreode. There are two characteristics encoded in this concept. The first is the presence of a degree of progress as one proceeds from the initial to the final state of the system. In some periods of the traverse through the chreode, there is a great deal of progress; in other periods, less so. The second characteristic is the relative strength of the tendency of the ingredient to return to the trajectory created by a chreode. Because this definition is close to the phenomenon Kier adopted this term to characterize the system.

The hydrophobic effect arises from the greater attraction and binding of water to itself, rather than water binding to another molecule (a solute or a stationary molecular fragment). The relative

hydropathic state of the other molecule influences the degree of aggregation of the water molecules in the vicinity. On the surface of a membrane or protein, Welch [89] referred to as the microviscosity. To reveal this effect, we have previously calculated the relative hydrophobicity of a solute in water in a cellular automata simulation [90]. The hydropathic state is encoded in the rule, P_B(WS), where a high probability reflects a hydrophobic solute and a low value reflects a hydrophilic solute. This earlier study illustrated the ability of the hydropathic state of a solute to organize the water in its vicinity. From this, we infer that the amino acid side chains, with variable hydropathic states, may organize the water and the cavities into some pattern which could function as our postulated chreode.

4.5.3. The hydropathic state and ligand diffusion

Our previous studies have shown that the diffusion of a solute through a solution is influenced by the hydropathic state of other solutes [91]. The diffusion of a solute was demonstrated to be faster if the another solute is hydrophobic. In our model of a chreode on a hydrodynamic landscape, it is necessary to consider the influence on diffusion of multiple stationary ingredients representing the side chains on a protein surface.

4.5.4. Creation of a model chreode

With the information gleaned from the gate studies and hydropathic influences, we next explored the possibility of a guided or directed trajectory for the diffusant; in essence a model of a chreode on the hydrodynamic landscape. For this study, we used a cellular automata grid located on the surface of a torus with a central region representing a target for a ligand, shown in the figure.

A random distribution of the five cell types representing the five groups of amino acid side chains based on hydropathic state, Figure 23, were introduced onto a cellular automata grid. All of these cells, (A,B,C,D,E) were separated from each other by 3 cell spaces, not explicitly shown but present in the calculation. The system contains water in the same proportion as used in the previous studies. The water is allowed to move freely and to interact with the ligand and stationary cells representing amino acid side chains of various hydropathic states. The diffusant was started at a position 38 cells from the target and endowed with a neutral hydropathic state in this study. The count of iterations necessary for the ligand, S, to traverse the grid and touch the central target was averaged over 200 runs. These values and those in the following study had a standard error of the mean of 6%. The number of iterations necessary in this study to traverse the distance to the center was calculated to be 12,429.

S C C D D B E B C D

D B E B C C A C D C

B D A D A C D B B D

A C B A D A A C E A

B D C E B E C B A C

E C A D ♥ A B E B B

C A C A C E A C D C

C A B D A E C B A D

A D A B B A B A C B

C A C A D B A E A C

Figure 23: A grid representing the protein landscape of a receptor, ♥. The side chains represented by the letters are randomly scattered over the grid. The water moves freely among them. The ligand, S, is allowed to diffuse among the side chains until it reaches the receptor.

A second model was a created in which the same number of A,B,C,D,and E structures were randomly scattered, Figure 24. Each cell was separated from another by a 3-cell space. Eighteen of the 100 cells were organized to form a chreode pattern shown in underlined italics. In this pattern the order of increasing hydrophobic character is assigned in the order A,B,C,D,E. The dynamics

were run as before and the time for the S cell to traverse the spaces to the central target was averaged over 200 runs. In this study the average time to diffuse to the center was calculated to be 8,609 iterations. This grid models the possible diffusion of a ligand across a surface with specifically positioned stationary cells coordinating their hydropathic states to facilitate diffusion toward the center of the grid. This models a chreode as we have defined it in this study. It was found that the rate of diffusion of a ligand is faster in the ordered stationary cell model as compared to a random distribution of these same amino acid side chain cells on the grid. The diffusion observed in the second study is a consequence of the existence of a pattern of cells and their hydropathic states, a possibility existing on the surface of a protein.

S C C D A B E B C D

D B E B B C A C D C

B D A D C C D B B D

A C B A D A A C E A

B D C E E E C B A C

B C D E ♥ E D C B A

C A C A E E A C D C

C A B D D E C B A D

A D A B C A B A C B

C A C A B B A E A C

Figure 2: A grid representing the protein landscape of a receptor, ♥. The side chains represented by the letters are randomly scattered over the grid. The water moves freely among them. A number of side chains are arranged in an order of increasing hydophobicity around the receptor. The ligand, S, is allowed to diffuse among the side chains until it reaches the receptor.

4.5.6. Discussion

The diffusion of ligands across protein surfaces has been considered as a route for these molecules to reach the active sites. This model is offered to explain the rapid response of receptors and the fast rates of enzyme catalysis. Two general mechanisms have been proposed in the past to define the facilitation of the diffusion. In the first case, the forces of interaction between ligands and the side chains are suggested to be electrostatic. If this were true the attraction between ligand and certain side chains might be strong enough to ensnare the molecule in regions of the protein surface, retarding diffusion. These side chains would, in essence, function like an active site or like a binding site rather than a diffusion-promoting feature. In contrast, van der Waals forces have been proposed between ligands and side chains, serving to facilitate the diffusion to the active site. These forces require a very close approach of the ligand and side chain, presenting a deterrent to rapid diffusion because of steric entanglement.

A theory proposed in this study invokes the participation of the layers of water molecules immediately adjacent to the protein surface. Each amino acid side chain intruding into the bulk water exercises an influence on the water architecture. This takes the form of a hydrophobic effect from hydrophobic side chains and side chain hydration with hydrophilic side chains. The cavities between clusters of hydrogen-bonded water molecules are in a dynamic state, joining with other cavities, breaking away, reforming, all in a manner resembling a dynamic peristaltic pump. These cavities form the proposed chreodes facilitating the passage of diffusants across the protein surface.

Several consequences of the theory may be derived from a consideration of known phenomena. The measurement of the affinity of a molecule may reflect a more complex system than just a drug-receptor encounter. It may be that the measurement is including some residence of the molecules in chreodes. Another observation that may have a connection with the chreodes is the phenomenon of lag, where some time must pass before an effect is observed in a pharmacological test system. The lag may reflect the time needed for the drug molecule to displace the transmitter from the chreodes leading to the active site. The sequel to that observation is persistence, where the measured effect continues for some time after the washout of the ligand from the test system. The velocity of enzyme reactions may, in part, be explained by the facilitated trajectory of the ligand through chreodes to the active site and the velocity of departure of the product through chreodes. The chreodes may exhibit some selectivity of their occupants and may possibly reject some molecules that are detrimental to fitness of the system. Finally, the mechanism of non-specific anesthetic agents may depend on their ability to interfere with the chreodes carrying the normal transmitter to the receptor.

4.5.7. A theory of volatile anesthetic action

A theory was proposed by Kier [92] for the clinical actions and behavior of the volatile anesthetic agents. Evidence supports the non-specific and ubiquitous effects of these drugs in which many ligand-receptor systems may be involved. The drugs interfere with the diffusion of ligands to receptors across the hydrodynamic landscape on the protein surface. The hydropathic states of the amino acid side chains surrounding an active site, influence the water to form chreodes. These chreodes are altered by the presence of the volatile anesthetic drugs, leading to a reduction in the diffusion rate of numerous ligands to their active sites. This effect is sufficient to decrease the responses of numerous receptors, leading to responses characteristic of the anesthetic states.

4.6. Modeling Biochemical Networks

Dynamic evolutionary networks have recently been recognized as a universal approach to complex systems, ranging from quantum gravity to biological cells and organisms, ecosystems, social groups, and market economy. The network approach is a non-reductionist approach enabling analysis of the systems as a whole, which makes it an ideal tool for systems biology. Network topology is generally used in characterizing networks, focusing on their connectivity, neighborhood and distance relationships. Network complexity has also been recently quantitatively characterized [93]. This abundance of cellular networks data, produced by microarrays, 2D-gel chromatography, mass-spectra, and other techniques, brings about another dimension of the network approach, allowing the tracing of the continuous changes of network species and their interactions. The large size of the metabolic, protein, and gene regulatory networks makes impractical many of the traditional methods for dynamic modeling. It is the purpose of this study to outline the potential of cellular automata as a basic method for the dynamic modeling of networks for biological and medical applications [94].

4.6.1. The MAPK Cascade Signaling Pathway

We have begun the analysis of a protein network [94], selecting the mitogen-activated protein kinase (MAPK) cascade as an example of importance, studied recently by numerical solving of the reaction rate equations [95]. This is a signaling pathway, relaying signals from the plasma membrane to targets in the cytoplasm and nucleus. The cascade is shown schematically in Figure 1.

The first reaction of the MAPK cascade implies the detailed reaction mechanism shown here:

$$\mathbf{A + E1 \rightarrow AE1 \rightarrow BE1 \rightarrow B + E1}$$

where *A* stands for MAPKKK, and *B* for MAPKKK*. A product of a reaction may be a molecule, cell, etc., but it may also be a catalyst or enzyme. In the MAPK cascade, this is the case with the activated MAPKKK and the MAPKK-PP, which are catalysts for the forward reactions in the second and third cascade level, respectively. The three substrates in the cascade have some prescribed initial concentration.

It is assumed that the enzymes have considerably smaller concentration than the substrates. The enzymes are modeled as being selective, each for a specific encounter (complex) and a specific reaction outcome. Finally, the model requires that each protein (but not necessarily each enzyme) is free to move about and to encounter other proteins and enzymes. The selectivity of the enzymes determines the conversion of one protein to another in a discrete way, thus the network display encodes these encounters and outcomes

The question asked in an analysis of a network such as Figure 1 is, what is the consequence of the change in a concentration of a protein or an enzyme in the system. The change in the substrates concentrations changes directs the reversible reactions toward the desirable outcome. The change in an enzyme concentration changes the rate with which the reaction reaches the equilibrium state. However, in non-equilibrium reactions, which frequently occur in biochemistry,

different enzyme concentrations produce a different steady-state, thus influencing the degree of conversion of substrates into products. One approach to these questions and to describe the patterns of behavior of the dynamic network is to model the system using cellular automata (CA). We will follow the general method used by Kier [96,97] in setting up a CA model of enzyme activity.

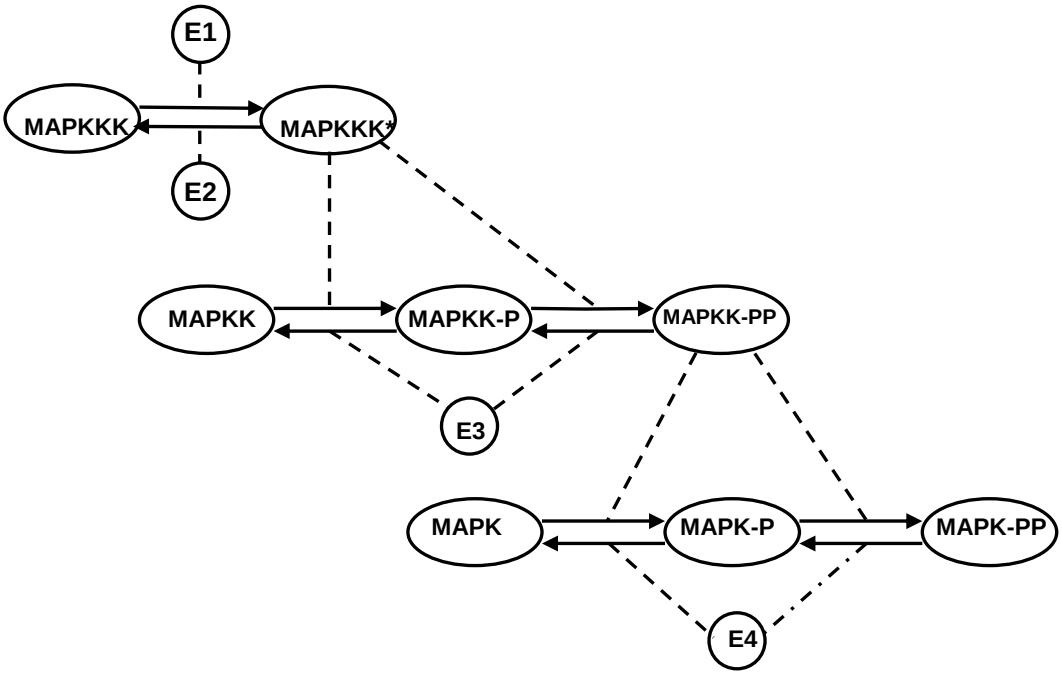

Figure 1. The MAPK signaling cascade. The substrates and products are represented by oval contours, reactions by arrows, and the catalysts action by dashed lines. E3 and E4 are MAPKK-protease and MAPK-protease, respectively. P stands for phosphate, PP for diphosphate.

4.6.2. The CA Modeling Design

Each molecule involved in the MAPK pathway is represented by a number of cells in the CA grid. The numbers chosen reflect the relative concentration of that protein. Each of the cells representing all other proteins moves about freely in the grid. They may encounter each other but this has no consequence. The only encounters that have a consequence are those between a specific protein (substrate) and a specific enzyme, as shown in the network. When such an encounter occurs, there is modeled a complex (enzyme-substrate). This complex has an assigned probability of converging to a new complex (enzyme-product). Following this there is a probability assigned for the separation of these two species.

Our studies of the MAPK cascade were performed using a CA grid of 100 by 100 cells. Each model was obtained as the average of 50 runs, each of which included 5000 iterations, a number sufficiently large to enable reproducing the steady-state of reaction. The grid used had no boundary conditions, thus movement past an edge puts the substance at the opposite face. A network to be studied is represented by groups of CA cells, each group representing one of the network species. The number of cells in each group reflects the relative concentrations of each network ingredient. We have systematically altered the initial concentrations of several proteins (MAPKKK, MAPKK, and MAPK) and the competencies of several enzymes (MAPKK- and MAPK-proteases, and the hypothetical enzymes E1 and E2 that affect the forward and reverse reactions of activation and deactivation of MAPKKK). The basic variable was the concentration of MAPKKK, which was varied within a 25-fold range from 20 to 500 cells. The concentrations of MAPKK and MAPK were kept constant (500 or 250 cells) in most of the models. The four enzymes, denoted by E1, E2, E3, and E4, were represented in the CA grid by 50 cells each. In one series of models, we kept the transition probabilities of three of the enzymes the same, (P = 0.1), and varied the probability of the fourth enzyme within the 0 to 1 range. In another series, *all* enzyme probabilities were kept constant, whereas the concentrations of substrates were varied. The last series varied both substrate concentrations and enzyme propensities. Recorded were the

variations in the concentrations of the three substrates MAPKKK, MAPKK, and MAPK, and those of the products MAPKKK*, MAPKK-P, MAPKK-PP, MAPK-P, and MAPK-PP.

4.6.3. Modeling Enzymes Activity

Upgrading or downgrading enzymes activity is one of the typical ways the cell reacts to stress and interactions with pathogens. We studied systematically the variations of one of the four enzymes E1 to E4 at constant concentrations of the substrates MAPKKK, MAPKK and MAPK, and constant propensity of the other three enzymes. We illustrate this type of pathway modeling in Fig. 2 with the variation of the MAPKK-protease (E3 enzyme in Fig. 1), which reverses the two-step reaction of MAPKK phosphorilation. It is shown that the concentration of the MAPKK- and MAPK-diphosphates (marked as *E* and *H* respectively) passes through a maximum near relatively low enzyme transition probability ($P \approx 0.02$). At the point of its maximum, the concentration of MAPK-PP reaches over 80% of its maximum, whereas that of MAPKK-PP is slightly over 50%.This shows the potential for a strong influence on the concentrations of the two diphosphates in the MAPK cascade by inhibiting the MAPKK-protease. In contrast, the level of steady-state concentrations of the two monophosphates (marked by D and G in Fig. 2) is not sensitive to the activity of the enzyme modeled, except for the extreme case of very strong inhibition ($P \rightarrow 0.001$).

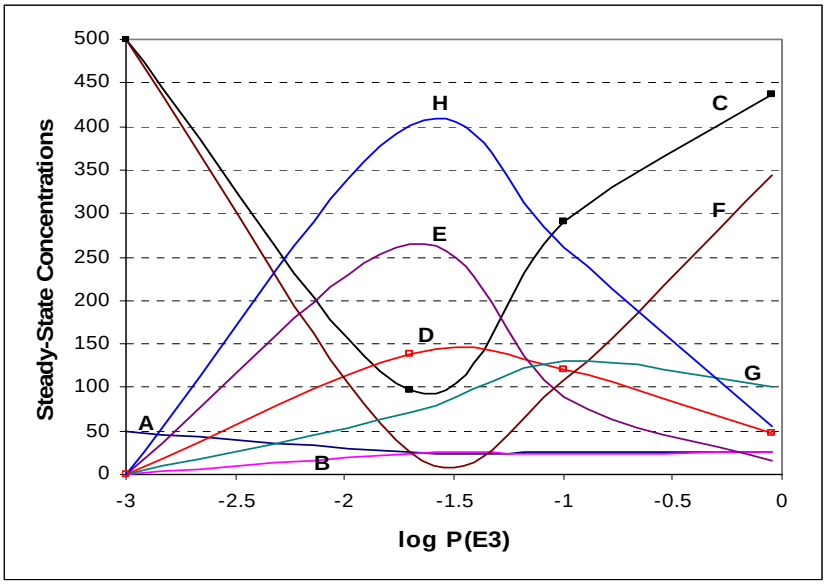

Figure 2. Influence of the MAPKK-protease propensity P(E3) on the steady-state concentrations of the MAPK cascade species (A = MAPKKK, B = MAPKKK*, C = MAPKK, D = MAPKK-P, E = MAPKK-PP, F = MAPK, F = MAPK-P, H = MAPK-PP). Enzyme propensities P(E1) = P(E2) = P(E4) = 0.1, substrate initial concentrations $[A_o] = 50$, $[C_o] = [F_o] = 500$.

4.6.4. Varying the Concentration of MAPKKK at Constant Enzyme Propensities

Another line of analysis is to study how the variations in the initial concentration of MAPKKK, the major downstream effector, affects the MAPK-cascade when keeping the enzyme activity constant. Fig. 3a shows the dynamics of the concentrations of all substrates and products for a 25-fold range of $[MAPKKK_o]$ from 20 to 500 cells, constant enzyme propensities P(E) = 0.1, and constant initial concentrations of MAPKK and MAPK equal to 500 cells. In the semilogarithmic plot of Fig. 3a, the MAPK-PP and MAPKK-PP ascending curves are sigmoids, as are the descending curves of the respective substrates MAPK and MAPKK, respectively. The sigmoidal pattern is also manifested in Fig. 3b with the MAPK curve in predicted stimulus/response semilogarithmic plot, the input stimulus in which is expressed in multiples of the EC_{50}, the concentration of MAPKKK that produces 50% maximal response. This behavior is typical for alosteric enzymes disobeying Michaelis-Menten kinetics, and thus evidences for a cooperative effect of cascade enzymes. Our finding confirms the result of

Huang and Ferell [6], obtained by numerical solving of the differential rate equations, assuming the concentration of enzyme E1 as a basic variable.

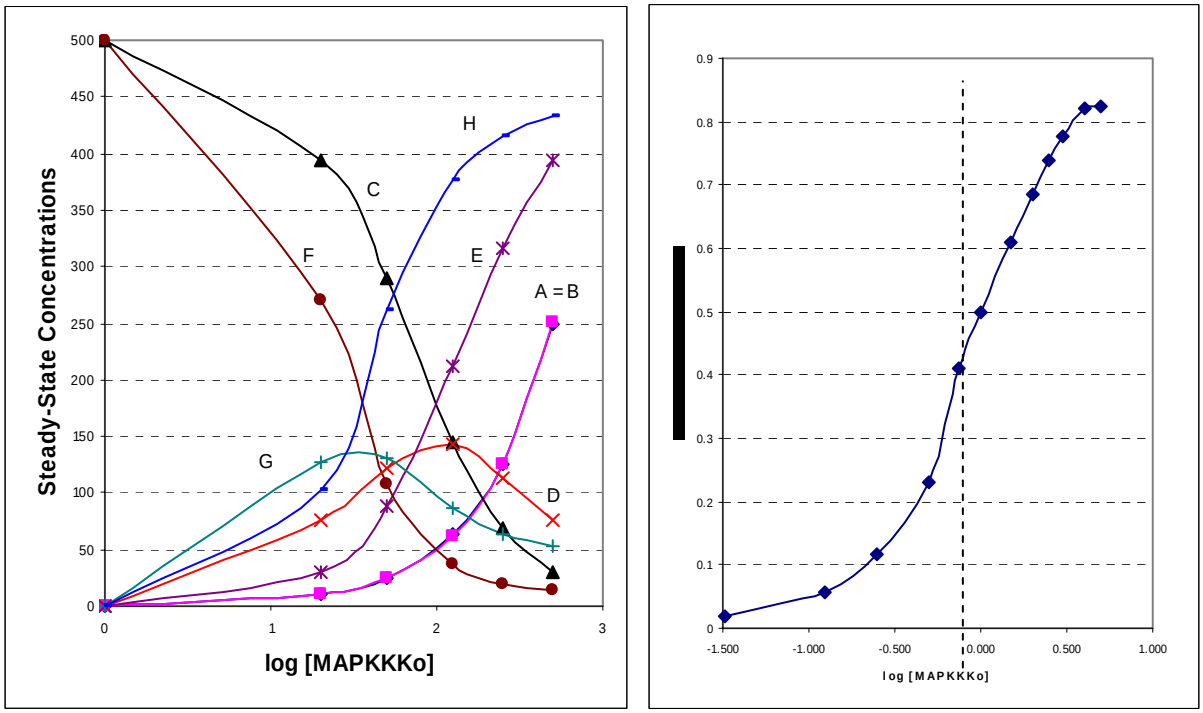

Figure 3. a) Semi-logarithmic plot of the steady-state concentrations of substrates and products dependence on the initial concentration of MAPKKK in MAPK cascade (A = MAPKKK, B = MAPKKK*, C = MAPKK, D = MAPKK-P, E = MAPKK-PP, F = MAPK, F = MAPK-P, H = MAPK-PP). P(E1) = P(E2) = P(E3) = P(E4) = 0.1, $[C_o] = [F_o] = 500$; b) Predicted relative stimulus/response curve for MAPK-PP with input stimulus (the MAPKKK intial concentration) expressed in multiples of EC_{50}.

4.6.5. Simultaneous Variation of MAPKKK Concentration and Enzymes Competence

The dynamics of the MAPK cascade signaling pathway can be analyzed in more details at the simultaneous variation of substrate concentrations and enzyme propensities. We varied the concentration of MAPKKK within a 25-fold range: 20, 50, 125, 250, 500 cells, keeping constant the concentrations of MAPKK and MAPK at level 500 cells. The variable enzyme competence was studied within the 900-fold range from 0.001 to 0.9, keeping the other enzymes' competence at the 0.1 level. This type of CA modeling is better illustrated on contour plots showing the concentration profiles of products and substrates.

Varying $[MAPKKK_o]$ and enzyme E1 propensity did not provide surprising results because both variables increase the yield of reaction products. Yet, the CA models showed a pattern of dominance of

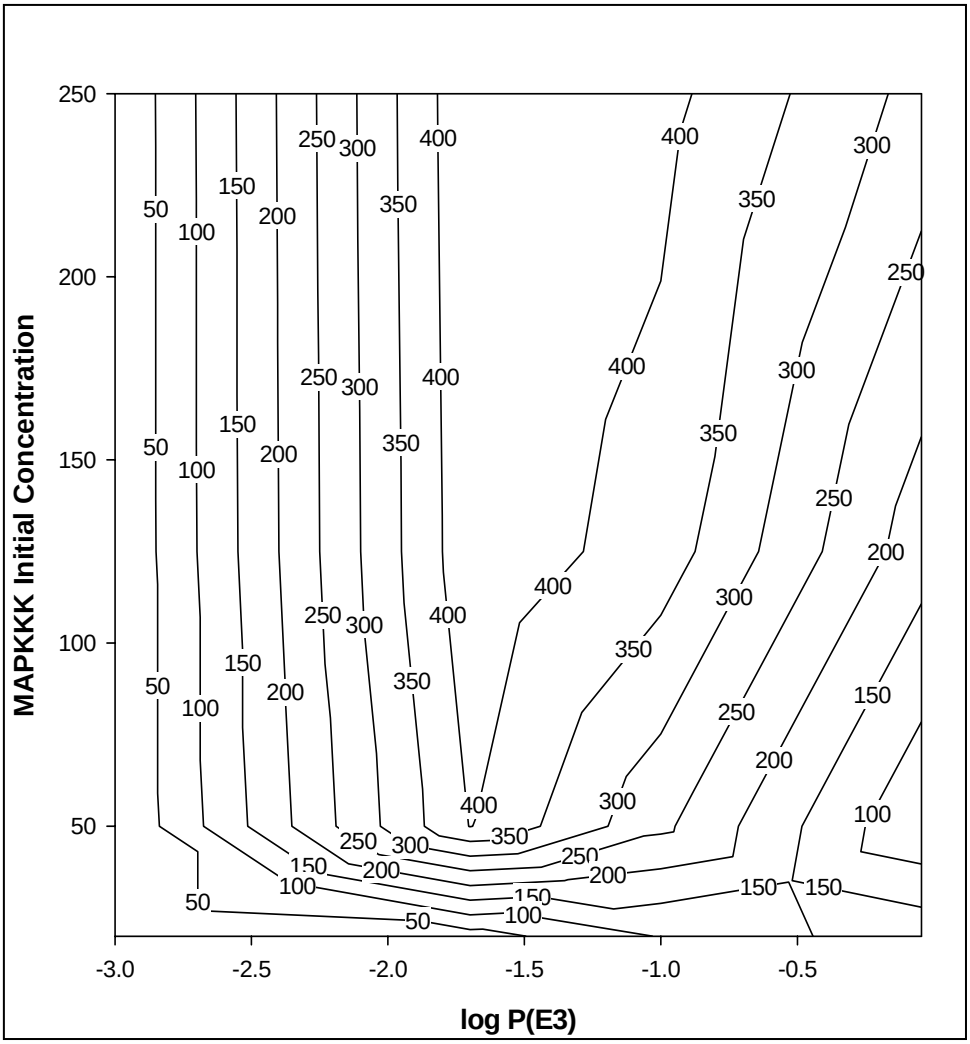

Figure 4. Contour plot of the MAPK-PP steady-state concentration at variable MAPKKK initial concentration and variable MAPKK-protease propensity. P(E1) = P(E2) = P(E4) = 0.1, $[MAPKK_o] = [MAPK_o] = 500$.

the concentration of substrate A over the competence of the enzyme E_1. Thus, at $[MAPKKK_o] = 20$, the 45-fold increase in enzyme E1 activity from 0.02 to 0.90 results in only about 4.5 fold increase in the MAPK $\rightarrow$ MAPK-PP conversion ratio. In contrast, even at relatively low activity of E1 (P = 0.02), the increase of $[MAPKKK_o]$ from 20 to 250 (12.5-fold) increases the production of MAPK-PP 7-fold (from 50 to 350).

The variation of enzymes E2, E3, and E4 propensity produces contour plots with interesting features. This is illustrated for enzyme E3 in Fig. 4, the contour lines in which show levels of constant steady-state concentration of MAPK-PP. The enzyme effect on the yield of MAPK-PP passes through a ridge at P(E3) = 0.02 to 0.1. The decrease in the MAPK-PP production is expected when E3 becomes more active, because this enzyme reduces the concentration of MAPKK-PP, which catalyzes the MAPK phosphorilation reaction. However, the increase in the MAPK-PP concentration within the P(E3) = 0.001 to 0.02 range of enzyme activity, for the broad range of [MAPKKK] $\geq$ 25 cells, is a trend that hardly could be anticipated. One can assess from the contour plots analyses that suppressing the activity of E2, E3, and E4 enzymes to a level obtained at probability P = 0.01-0.02 would enable reaching 80% of the maximum MAPK-PP concentration at relatively low

concentrations of MAPKKK. A further maximization of the MAPK-PP production can result from a more favorable combination of the four enzyme propensities, a high one for enzyme 1 and low ones for enzymes E2, E3, and E4, as follows from the patterns described above. Conversely, a substantial inhibition of these enzymes (e. g., at $P << 0.02$) would minimize the MAPK-PP steady-state concentration.

Concluding our analysis, we would like to emphasize again that the CA modeling of the MAPK cascade pathway reported here was not aimed at exact reproducing of previous work [6]. Rather, it was aimed to demonstrate the potential of the cellular automata method to model basic patterns in pathways of interest and to indicate the ways to control the pathways dynamics by selective enzyme inhibiting and concentration variations. Work is in progress to extend the methodology to large networks.

5.1 General summary

The linkage between complex, dynamic systems and cellular automata is made quite clear in this monograph. The dynamic portrayal of many phenomena have been shown to mirror reality in several important ways among the studies described. We aver that cellular automata belongs on the pantheon of methods of probing, modeling and even predicting events associated with complex systems, emerging phenomena and hierarchical patterns.

References

1. Gutfreund, H. (1995). Kinetics for the Life Sciences: Receptors, Transmitters, and Catalysis. Cambridge, England: Cambridge University Press.

2. Savageau, M.A. (1976). Biochemical Systems Analysis. Reading, MA: Addison-Wesley Publishing Co.

3. Beveridge, D.L. and Jorgensen, W.L. (1986). Computer Simulation of Chemical and Biomolecular Systems. Vol. 482. New York, NY: Annals of the New York Academy of Sciences

4. Segel, L.A. (ed.). (1991). Biological Kinetics. Cambridge, England: Cambridge University Press.

5. Levin, S.A. (ed.). (1994). Frontiers in Mathematical Biology. Berlin, Germany: Springer-Verlag, Inc.

6. Witten, T.M., Computational Medicine (in) Encyclopedia of Computer Science – 4th ed., Anthony Ralston, Edwin D. Reilly, David Hemmendinger (eds.) (Macmillan Reference, Ltd, London and Grove's Dictionaries, New York, 2000).

7. Anderson, P.W., Arrow, K.J., and Pines, D. (eds.). (1988). The Economy as an Evolving Complex System. Santa Fe Institute. Studies in the Sciences of Complexity. Redwood City, CA: Addison-Wesley Publishing Co, Inc.

8. Townend, M.S. (1984). Mathematics in Sport. Chichester, England: Ellis Horwood Press.

9. Renfrew, C. and Cooke, K.L. (1979). Transformations: Mathematical Approaches to Culture change. New

York, NY: Academic Press, Inc.

10. Mikhailov, A.S. and Calenbuhr, V. (2002). From Cells to Societies: Models of Complex Coherent Interaction. Berlin, Germany: Springer Verlag, Inc.

11. Sulis, W. and Combs, A. (1996). Nonlinear Dynamics in Human Behavior. Singapore: World Scientific Press.

12. Witten, T. M., A note on the structure of system state spaces and its implications on the existence of non—repeatable experiments, Bull. Math. Biol., 42 (1980) 267--272.

13. Witten, T.M., Some thoughts on quantum non-demolition experiments in biological systems, Bull. Math. Biol., 44 (1982) 572--584.

14. Witten, T.M., Computational Medicine (in) Encyclopedia of Computer Science – 4th ed., Anthony Ralston, Edwin D. Reilly, David Hemmendinger (eds.) (Macmillan Reference, Ltd, London and Grove's Dictionaries, New York, 2000).

15. National Science Foundation. (1992). Grand Challenges: High Performance Computing and Communications – The FY 1992 U.S. Research and Development Program.

16. National Science Foundation. (1993). Grand Challenges: High Performance Computing and Communications – The FY 1993 U.S. Research and Development Program

17. Witten, T.M. (1990). Physiology in a supercomputer: Peering inside living systems, Supercomputing Review, 3#12: 34--40

18. Witten, T.M. (1991). Supercomputing in biology and Medicine: An overview, Supercomputer, VIII#2: 37-53.

19. Witten, T.M. (1994). The role of high performance computing in medicine and public health, Future Generation Computer Systems, 10: 223--232.

20. J. M. Haile, Molecular Dynamics Simulation: Elementary Methods, Wiley: New York, 1992.

21. A. R. Leach, Molecular Modelling: Principles and Applications, Longman: Harlow, England, 1996.

22. M. P. Allen and D. J. Tildesley, Computer Simulation in Chemical Physics, Kluwer Academic Publ.: Boston, 1998.

23. D. J. Tildesley, "The Molecular Dynamics Method", in Ibid, pp. 23-47.

24. M. P. Allen, "Introduction to Monte Carlo Simulations" in Observation, Prediction and Simulation of Phase Transitions in Complex Fluids, M. Baus, L. F. Rull, and J.-P. Ryckaert, Eds, Kluwer academic Publ.: Boston, 1995, pp. 339-356.

25. W. L. Jorgensen, "Monte Carlo Simulations for Liquids", in Encyclopedia of Computational Chemistry, P. V. Ragué Schleyer, Ed., Wiley: New York, 1998, pp. 1754-1763.

26. W. L. Jorgensen and J. Tirado-Rives, J. Phys. Chem. 1996, 100, 14508-14513.

27. Belew, R.K. and Mitchell, M. (eds). (1996). Adaptive Individuals in Evolving Populations: Models and Algorithms. Santa Fe Institute Studies in the Sciences of Complexity. Reading, MA: Addison-Wesley Publishing Co.

28. Allen, T.F.H. and Starr, T.B. (1982). Hierarchy: Perspectives for Ecological Complexity. Chicago, IL: The University of Chicago Press.

29. Salthe, S.N. (1985). Evolving Hierarchical Systems: Their Structure and Representation. New York, NY: Columbia University Press.

30. Bar-Yam, Y. (1997). Dynamics of Complex Systems. Studies in Nonlinearity. Reading, MA: Perseus Books.

31. MacDonald, N. (1983). Trees and Networks in Biological Systems. Chichester, England: John Wiley and Sons, Inc.

32. Rashevsky, N. (1972). Organismic Sets, London, England: Clowes and Sons.

33. R. Rosen, A relational theory of biological systems. Bull Math. Biol., 20 (1958) 245-260.

34. R. Rosen, The representation of biological systems from the standpoint of categories. Bull Math. Biol., 20 (1958) 317-341.

35. R. Rosen, A relational theory of biological systems II. Bull. Math. Biol., 21 (1959) 109-128.

36. R. Rosen, Some comments on re-establishability. Bull Math. Biol., 27 (1965) 11-14.

37. Witten, T.M. (1984). A return to time, cells, systems, and aging: II. Relational and reliability theoretic approaches to the study of senescence in living systems. Mech. Aging and Dev., 27: 323-340.

38. Davidson, E.H., Rast, J.P., Oliveri, P. et al. (2002). A genomic regulatory network for development. Science, 295: 1669-1678.

39. Satorras-Pastor, R. Smith, E., and Sole, R.V. (2003). Evolving protein interaction networks through gene duplication. J. theor. Biol., 222: 199-210.

40. Ravasz. E., Somera, A.L., Mongru, D.A., Oltvai, Z.N. and Barabasi, A.-L. (2002). Science, 297: 1551-1555.

41. Barabasi, A.L., Albert ,R (1999). Emergence of Scaling in Random Networks. Science 286: 509-512

42. Montoya, J.M. and Sole, R.V. (2002). Small world patterns in food webs. J. theor. Biol., 214: 405-412.

43. Albert, R. and Barabasi, A-L. Statistical mechanics of complex networks. Rev. Mod. Physics, 74: 47-97.

44. Eckmann, J-P., Moses, E., and Sergi, D. (2003). Dialog in e-mail traffic. E-print archives. http://xxx.aps.org/arXiv:cond-mat/0304433.

45. Bonchev, D. Overall Connectivity and Molecular Complexity. In: Topological Indices and Related Descriptors (in) J. Devillers and A.T. Balaban, Eds., Gordon and Breach, Reading, U.K., 1999, p. 361-401.

46. Bonchev, D. The Wiener Number. Some Applications and New Developments. In: Topology in Chemistry. Discrete Mathematics of Molecules, D. H. Rouvray and R. B. King, Eds. Horwood, Chichester, U. K., 2002, p. 58-88

47. Ravesz, E. Somera, A.L., Mongru, D.A., Oltvai, Z.N. and Barabasi, A.-L. (2002). Hierarchical organization of modularity in metabolic networks. Science, 297: 1551-1555.

48. Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D. and Allon. U. (2002). Network motifs: Simple building blocks of complex networks. Science, 298: 824-827.

49. Waldrop, M.M. (1992). Complexity: The Emerging Science at the Edge of Order and Chaos. New York, NY: Simon and Schuster.

50. Decker. P. (1979). Spatial, chiral, and temporal self-organization through bifurcation in "bioids," open systems capable of a generalized Darwinian evolution. (in) Bifurcation Theory and Applications in Scientific Disciplines. Vol 316. (eds.) Gurel, O. and Rossler, O.E. New York, NY: Annals of the New York Academy of Science.

51. Olsen, L.F. and Degn, H. (1977). Chaos in an enzyme reaction. Nature, 267: 177.

52. Selkov, E.E. (1967). Self-oscillations in glycolysis 1. A simple kinetic model. European J. Biochem. 4: 79.

53. Hess, B. and Plesser, T. (1979). Temporal and spatial order in biochemical systems. (in) Bifurcation Theory and Applications in Scientific
Disciplines. Vol 316. (eds.) Gurel, O. and Rossler, O.E. New York, NY: Annals of the New York Academy of Science.

54. Sharma, K.R. and Noyes, R.M. (1976). Oscillations in chemical reactions. 13. A detailed molecular mechanism for the Bray-Liebafsky reaction of iodate and hydrogen peroxide. J. Amer. Chem. Soc., 98: 4345.

55. Nicolis, G. and Prigogine, I. (1989). Exploring Complexity: An Introduction. New York, NY: W.H. Freeman and Co.

56. Bhalla, U.S. and Iyengar, R. (1999). Emergent properties of networks of biological signaling pathways. Science, 283: 381-387.

57. Kier, L. B. and Testa, B. (1997) A systems approach to molecular structure. intermolecular recognition and emergence-dissolvence in medicinal research. Med. Res. Rev. 17: 303.

58. Testa, B. and Kier, L, B. (1997) The concept of emergence-dissolvence in drug reaearch

Adv. Drug Res. 30: 1

59. Wells, H. G., Huxley, J. S., and Wells, G. P. (1934) The Science of Life: The Literary Guild, New York p. 1475

60. S. M. Ulam, Proc. Int. Congr. Math. (held in 1950) 1952, 2, 264 Random Processes and Transformations

61. S. M. Ulam, Adventures of a Mathematician; Charles Scribner's Sons, New York, 1976.

62. J. von Neumann, Theory of Self-Replicating Automata; A. Burks,, Ed.; Univ. of Illinois Press: Urbana, 1966.

63. K. Zuse, Int. J. Theoret. Phys. 1982, 21, 589.

64. T. Toffoli and N. Margolus, Cellular Automata Machines; The MIT Press: Cambridge, MA, 1987.

65. M. Schroeder, Fractals, Chaos, Power Laws; W. H. Freeman, New York, 1991, p. 371.

66. G. Y. Vichniac, Physica D 1984, 10, 96-116. Simulating Physics with Cellular Automata

67. T. Toffoli, Physica D 1984, 10, 117-127. Cellular Automata as an Alternative (Rather than an Approximation of) Differential Equations in Modeling Physics

68. S. Wolfram, Physica D 1984, 10, vii-xii. Preface

69. L.B. Kier and C.K. Cheng, *J. Chem. Inf. & Comp. Sci.* 34, 1334 (1994)

70. L. B. Kier and C.-K.Cheng, *Pharm. Res.* 12, 1521 (1995)

71. L. B. Kier and C.-K. Cheng, *J. Math. Chem.,* 21, 71 (1997)

72. L. B. Kier, C. K. Cheng, B. Testa and P.A. Carrupt, *Pharm.Res.* 12, 615 (1995)

73. C.-K.Cheng and L. B. Kier, *J. Chem. Inf. & Comput. Sci.* 35, 1054 (1995)

74. L.B. Kier and C.-K. Cheng. in Lipophilicity in Drug Research, (Pliska, Testa and Waterbeemd, eds.) VCH Publ. (1996)

75. L.B.Kier, C.-K. Cheng, B. Testa and P.-A. Carrupt, *Pharm. Res.* 13, 1419 (1996)

76. L. B. Kier, C.-K. Cheng, B. Testa and P.-A. Carrupt. *J.Pharm. Sci.,* 86, 774 (1997)

77. L. B. Kier and C.-K. Cheng, *J. Theor. Biol.,* 186, 75 (1997)

78. L. B. Kier, C.-K.Cheng, M. Tute, and P. G. Seybold, *J. Chem. Inf. Comp. Sci.,* 38, 271 (1998)

79. L. B. Kier, C.-K. Cheng and B. Testa, *J. Chem. Inf. Comp. Sci.*,39, 326 (1999)

80. L. B. Kier, C.-K. Cheng and B. Testa, *Fut. Gen.Comput. Sys.*,16, 273 (1999)

81. L. B. Kier, C.-K. Cheng, and P. Seybold, *SAR and QSAR in Environ. Res.,* 11, 79 (2000)

82. L. B. Kier, C.-K. Cheng and P. Seybold, Rev. in Comp. Chem., 17 , 205 (2001).

83. L. B. Kier and C.-K. Cheng. *J. Chem. Inf. & Comp. Sci.,* 34, 647 (1994)

84. L. B. Kier, *J. Chem. Inf & Comp. Sci.* 40, 1285 (2000)

85. L. B. Kier, C.-K. Cheng and B. Testa, J. Theoret. Biol., 214, 415-426 (2002).

86. G. R. Welch, *Biosystems* **38**, 147 (1996).

87. G. Adam and M. Delbruck, in *Structural Chemistry and Molecular Biology*, (A. Rich and N. Davidson, eds.) Freeman, San Fransisco (1968).

88. M. Eigen, in *Quantum Statistical Mechanics in the Natural Sciences,* (S. L. Mintz, and S. M. Widemeyer, eds.) Plenum, New York (1974)

89. G. R. Welch, *J. theoret Biol.*, **68**, 267 (1977)

90. L. B. Kier and B. Testa, *Complexity*, 1, 32 (1996)

91. C. H. Waddington, *The Strategy of the Genes* , George Allen & Urwin Ltd., London (1957).

92. L. B. Kier, AANA Journal 71, 422 (2003).

93. D. Bonchev, 'Complexity of Protein-Protein Interaction Networks, Complexes and Pathways', in *Handbook of Proteomics Methods*, M. Conn, ed., Humana, New York, pp. 451-462.(2003)

94. L. B. Kier, B. Bonchev and G. Buck,

95. C.-Y. Huang and J. E. Ferell, 1996, *Proc. Natl. Acad. Sci. USA* **93**, pp. 10078 (1996)

96. L. B. Kier, L. B.,C.-K. Cheng and B. Testa *J. Molec. Graphics* **14**,. 227 (1996).

97. L. B. Kier and C.-K.Cheng, *J. Molec.* Graphics **18**, 29 (2000).

1.

References

1. Gutfreund, H. (1995). Kinetics for the Life Sciences: Receptors, Transmitters, and Catalysis. Cambridge, England: Cambridge University Press.

2. Savageau, M.A. (1976). Biochemical Systems Analysis. Reading, MA: Addison-Wesley Publishing Co.

3. Beveridge, D.L. and Jorgensen, W.L. (1986). Computer Simulation of Chemical and Biomolecular Systems. Vol. 482. New York, NY: Annals of the New York Academy of Sciences

4. Segel, L.A. (ed.). (1991). Biological Kinetics. Cambridge, England: Cambridge University Press.

5. Levin, S.A. (ed.). (1994). Frontiers in Mathematical Biology. Berlin, Germany: Springer-Verlag, Inc.

6. Witten, T.M., Computational Medicine (in) Encyclopedia of Computer Science – 4th ed., Anthony Ralston, Edwin D. Reilly, David Hemmendinger (eds.) (Macmillan Reference, Ltd, London and Grove's Dictionaries, New York, 2000).

7. Anderson, P.W., Arrow, K.J., and Pines, D. (eds.). (1988). The Economy as an Evolving Complex System. Santa Fe Institute. Studies in the Sciences of Complexity. Redwood City, CA: Addison-Wesley Publishing Co, Inc.

8. Townend, M.S. (1984). Mathematics in Sport. Chichester, England: Ellis Horwood Press.

9. Renfrew, C. and Cooke, K.L. (1979). Transformations: Mathematical Approaches to Culture change. New York, NY: Academic Press, Inc.

10. Mikhailov, A.S. and Calenbuhr, V. (2002). From Cells to Societies: Models of Complex Coherent Interaction. Berlin, Germany: Springer Verlag, Inc.

11. Sulis, W. and Combs, A. (1996). Nonlinear Dynamics in Human Behavior. Singapore: World Scientific Press.

12. Witten, T. M., A note on the structure of system state spaces and its implications on the existence of non—repeatable experiments, Bull. Math. Biol., 42 (1980) 267--272.

13. Witten, T.M., Some thoughts on quantum non-demolition experiments in biological systems, Bull. Math. Biol., 44 (1982) 572--584.

14. Witten, T.M., Computational Medicine (in) Encyclopedia of Computer Science – 4th ed., Anthony Ralston, Edwin D. Reilly, David Hemmendinger (eds.) (Macmillan Reference, Ltd, London and Grove's Dictionaries, New York, 2000).

15. National Science Foundation. (1992). Grand Challenges: High Performance Computing and Communications – The FY 1992 U.S. Research and Development Program.

16. National Science Foundation. (1993). Grand Challenges: High Performance Computing and Communications – The FY 1993 U.S. Research and Development Program

17. Witten, T.M. (1990). Physiology in a supercomputer: Peering inside living systems, Supercomputing Review, 3#12: 34--40

18. Witten, T.M. (1991). Supercomputing in biology and Medicine: An overview, Supercomputer, VIII#2: 37-53.

19. Witten, T.M. (1994). The role of high performance computing in medicine and public health, Future Generation Computer Systems, 10: 223--232.

20. J. M. Haile, Molecular Dynamics Simulation: Elementary Methods, Wiley: New York, 1992.

21. A. R. Leach, Molecular Modelling: Principles and Applications, Longman: Harlow, England, 1996.

22. M. P. Allen and D. J. Tildesley, Computer Simulation in Chemical Physics, Kluwer Academic Publ.: Boston, 1998.

23. D. J. Tildesley, "The Molecular Dynamics Method", in Ibid, pp. 23-47.

24. M. P. Allen, "Introduction to Monte Carlo Simulations" in Observation, Prediction and Simulation of Phase Transitions in Complex Fluids, M. Baus, L. F. Rull, and J.-P. Ryckaert, Eds, Kluwer academic Publ.: Boston, 1995, pp. 339-356.

25. W. L. Jorgensen, "Monte Carlo Simulations for Liquids", in Encyclopedia of Computational Chemistry, P. V. Ragué Schleyer, Ed., Wiley: New York, 1998, pp. 1754-1763.

26. W. L. Jorgensen and J. Tirado-Rives, J. Phys. Chem. 1996, 100, 14508-14513.

27. Belew, R.K. and Mitchell, M. (eds). (1996). Adaptive Individuals in Evolving Populations: Models and Algorithms. Santa Fe Institute Studies in the Sciences of Complexity. Reading, MA: Addison-Wesley Publishing Co.

28. Allen, T.F.H. and Starr, T.B. (1982). Hierarchy: Perspectives for Ecological Complexity. Chicago, IL: The University of Chicago Press.

29. Salthe, S.N. (1985). Evolving Hierarchical Systems: Their Structure and Representation. New York, NY: Columbia University Press.

30. Bar-Yam, Y. (1997). Dynamics of Complex Systems. Studies in Nonlinearity. Reading, MA: Perseus Books.

31. MacDonald, N. (1983). Trees and Networks in Biological Systems. Chichester, England: John Wiley and Sons, Inc.

32. Rashevsky, N. (1972). Organismic Sets, London, England: Clowes and Sons.

33. R. Rosen, A relational theory of biological systems. Bull Math. Biol., 20 (1958) 245-260.

34. R. Rosen, The representation of biological systems from the standpoint of categories. Bull Math. Biol., 20 (1958) 317-341.

35. R. Rosen, A relational theory of biological systems II. Bull. Math. Biol., 21 (1959) 109-128.

36. R. Rosen, Some comments on re-establishability. Bull Math. Biol., 27 (1965) 11-14.

37. Witten, T.M. (1984). A return to time, cells, systems, and aging: II. Relational and reliability theoretic approaches to the study of senescence in living systems. Mech. Aging and Dev., 27: 323-340.

38. Davidson, E.H., Rast, J.P., Oliveri, P. et al. (2002). A genomic regulatory network for development. Science, 295: 1669-1678.

39. Satorras-Pastor, R. Smith, E., and Sole, R.V. (2003). Evolving protein interaction networks through gene duplication. J. theor. Biol., 222: 199-210.

40. Ravasz. E., Somera, A.L., Mongru, D.A., Oltvai, Z.N. and Barabasi, A.-L. (2002). Science, 297: 1551-1555.

41. Barabasi, A.L., Albert ,R (1999). Emergence of Scaling in Random Networks. Science 286: 509-512

42. Montoya, J.M. and Sole, R.V. (2002). Small world patterns in food webs. J. theor. Biol., 214: 405-412.

43. Albert, R. and Barabasi, A-L. Statistical mechanics of complex networks. Rev. Mod. Physics, 74: 47-97.

44. Eckmann, J-P., Moses, E., and Sergi, D. (2003). Dialog in e-mail traffic. E-print archives. http://xxx.aps.org/arXiv:cond-mat/0304433.

45. Bonchev, D. Overall Connectivity and Molecular Complexity. In: Topological Indices and Related Descriptors (in) J. Devillers and A.T. Balaban, Eds., Gordon and Breach, Reading, U.K., 1999, p. 361-401.

46. Bonchev, D. The Wiener Number. Some Applications and New Developments. In: Topology in Chemistry. Discrete Mathematics of Molecules, D. H. Rouvray and R. B. King, Eds. Horwood, Chichester, U. K., 2002, p. 58-88

47. Ravesz, E. Somera, A.L., Mongru, D.A., Oltvai, Z.N. and Barabasi, A.-L. (2002). Hierarchical organization of modularity in metabolic networks. Science, 297: 1551-1555.

48. Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D. and Allon. U. (2002). Network motifs: Simple building blocks of complex networks. Science, 298: 824-827.

49. Waldrop, M.M. (1992). Complexity: The Emerging Science at the Edge of Order and Chaos. New York, NY: Simon and Schuster.

50. Decker. P. (1979). Spatial, chiral, and temporal self-organization through bifurcation in “bioids,” open systems capable of a generalized Darwinian evolution. (in) Bifurcation Theory and Applications in Scientific Disciplines. Vol 316. (eds.) Gurel, O. and Rossler, O.E. New York, NY: Annals of the New York Academy of Science.

51. Olsen, L.F. and Degn, H. (1977). Chaos in an enzyme reaction. Nature, 267: 177.

52. Selkov, E.E. (1967). Self-oscillations in glycolysis 1. A simple kinetic model. European J. Biochem. 4: 79.

53. Hess, B. and Plesser, T. (1979). Temporal and spatial order in biochemical systems. (in) Bifurcation Theory and Applications in Scientific
Disciplines. Vol 316. (eds.) Gurel, O. and Rossler, O.E. New York, NY: Annals of the New York Academy of Science.

54. Sharma, K.R. and Noyes, R.M. (1976). Oscillations in chemical reactions. 13. A detailed molecular mechanism for the Bray-Liebafsky reaction of iodate and hydrogen peroxide. J. Amer. Chem. Soc., 98: 4345.

55. Nicolis, G. and Prigogine, I. (1989). Exploring Complexity: An Introduction. New York, NY: W.H. Freeman and Co.

56. Bhalla, U.S. and Iyengar, R. (1999). Emergent properties of networks of biological signaling pathways. Science, 283: 381-387.

57. Kier, L. B. and Testa, B. (1997) A systems approach to molecular structure. intermolecular recognition and emergence-dissolvence in medicinal research. Med. Res. Rev. 17: 303.

58. Testa, B. and Kier, L, B. (1997) The concept of emergence-dissolvence in drug reaearch Adv. Drug Res. 30: 1

59. Wells, H. G., Huxley, J. S., and Wells, G. P. (1934) The Science of Life: The Literary Guild, New York p. 1475

60. S. M. Ulam, Proc. Int. Congr. Math. (held in 1950) 1952, 2, 264 Random Processes and Transformations

61. S. M. Ulam, Adventures of a Mathematician; Charles Scribner's Sons, New York, 1976.

62. J. von Neumann, Theory of Self-Replicating Automata; A. Burks,, Ed.; Univ. of Illinois Press: Urbana, 1966.

63. K. Zuse, Int. J. Theoret. Phys. 1982, 21, 589.

64. T. Toffoli and N. Margolus, Cellular Automata Machines; The MIT Press: Cambridge, MA, 1987.

65. M. Schroeder, Fractals, Chaos, Power Laws; W. H. Freeman, New York, 1991, p. 371.

66. G. Y. Vichniac, Physica D 1984, 10, 96-116. Simulating Physics with Cellular Automata

67. T. Toffoli, Physica D 1984, 10, 117-127. Cellular Automata as an Alternative (Rather than an Approximation of) Differential Equations in Modeling Physics

68. S. Wolfram, Physica D 1984, 10, vii-xii. Preface

69. L.B. Kier and C.K. Cheng, *J. Chem. Inf. & Comp. Sci.* 34, 1334 (1994)

70. L. B. Kier and C.-K.Cheng, *Pharm. Res.* 12, 1521 (1995)

71. L. B. Kier and C.-K. Cheng, *J. Math. Chem.*, 21, 71 (1997)

72. L. B. Kier, C. K. Cheng, B. Testa and P.A. Carrupt, *Pharm.Res.* 12, 615 (1995)

73. C.-K.Cheng and L. B. Kier, *J. Chem. Inf. & Comput. Sci.* 35, 1054 (1995)

74. L.B. Kier and C.-K. Cheng. in Lipophilicity in Drug Research, (Pliska, Testa and Waterbeemd, eds.) VCH Publ. (1996)

75. L.B.Kier, C.-K. Cheng, B. Testa and P.-A. Carrupt, *Pharm. Res.* 13, 1419 (1996)

76. L. B. Kier, C.-K. Cheng, B. Testa and P.-A. Carrupt. *J.Pharm. Sci.*, 86, 774 (1997)

77. L. B. Kier and C.-K. Cheng, *J. Theor. Biol.*, 186, 75 (1997)

78. L. B. Kier, C.-K.Cheng, M. Tute, and P. G. Seybold, *J. Chem. Inf. Comp. Sci.*, 38, 271 (1998)

79. L. B. Kier, C.-K. Cheng and B. Testa, *J. Chem. Inf. Comp. Sci.*,39, 326 (1999)

80. L. B. Kier, C.-K. Cheng and B. Testa, *Fut. Gen.Comput. Sys.*,16, 273 (1999)

81. L. B. Kier, C.-K. Cheng, and P. Seybold, *SAR and QSAR in Environ. Res.*, 11, 79 (2000)

82. L. B. Kier, C.-K. Cheng and P. Seybold, Rev. in Comp. Chem., 17 , 205 (2001).

83. L. B. Kier and C.-K. Cheng. *J. Chem. Inf. & Comp. Sci.*, 34, 647 (1994)

84. L. B. Kier, *J. Chem. Inf & Comp. Sci.* 40, 1285 (2000)

85. L. B. Kier, C.-K. Cheng and B. Testa, J. Theoret. Biol., 214, 415-426 (2002).

86. G. R. Welch, *Biosystems* **38**, 147 (1996).

87. G. Adam and M. Delbruck, in *Structural Chemistry and Molecular Biology*, (A. Rich and N. Davidson, eds.) Freeman, San Fransisco (1968).

88. M. Eigen, in *Quantum Statistical Mechanics in the Natural Sciences*, (S. L. Mintz, and S. M. Widemeyer, eds.) Plenum, New York (1974)

89. G. R. Welch, *J. theoret Biol.*, **68**, 267 (1977)

90. L. B. Kier and B. Testa, *Complexity*, 1, 32 (1996)

91. C. H. Waddington, *The Strategy of the Genes* , George Allen & Urwin Ltd., London (1957).

92. L. B. Kier, AANA Journal 71, 422 (2003).

93. D. Bonchev, 'Complexity of Protein-Protein Interaction Networks, Complexes and Pathways', in *Handbook of Proteomics Methods*, M. Conn, ed., Humana, New York, pp. 451-462.(2003)

94. L. B. Kier, B. Bonchev and G. Buck,

95. C.-Y. Huang and J. E. Ferell, 1996, *Proc. Natl. Acad. Sci. USA* **93**, pp. 10078 (1996)

96. L. B. Kier, L. B.,C.-K. Cheng and B. Testa *J. Molec. Graphics* **14**,. 227 (1996).

97. L. B. Kier and C.-K.Cheng, *J. Molec.* Graphics **18**, 29 (2000).

2.

76. L. B. Kier, C.-K. Cheng, B. Testa and P.-A. Carrupt. *J.Pharm. Sci.,* 86, 774 (1997)

77. L. B. Kier and C.-K. Cheng, *J. Theor. Biol.,* 186, 75 (1997)

78. C.-K.Cheng, M. Tute, and P. G. Seybold, *J. Chem. Inf. Comp. Sci.,* 38, 271 (1998)

79. L. B. Kier, C.-K. Cheng and B. Testa, *J. Chem. Inf. Comp. Sci.,*39, 326 (1999)

80. L. B. Kier, C.-K. Cheng and B. Testa, *Fut. Gen.Comput. Sys.,*16, 273 (1999)

81. L. B. Kier, C.-K. Cheng, and P. Seybold, *SAR and QSAR in Environ. Res.,* 11, 79 (2000)

82. L. B. Kier, C.-K. Cheng and P. Seybold, Rev. in Comp. Chem., 17 , 205 (2001).

References

1. S. Wolfram. *A New Kind of Science*, 2002, Wolfram Media, Champaign, IL.
2. Neuforth, A., Seybold, P. G., Kier, L. B., and Cheng C.–K.: 'Cellular Automata Models of Kinetically and Thermodynamically Controlled Reactions', 2000, *Int. J. Chem. Kinet.*, **32**, pp. 529-534.
3. Bonchev, D.: 'Complexity of Protein-Protein Interaction Networks, Complexes and Pathways', in *Handbook of Proteomics Methods*, M. Conn, ed., 2003, Humana, New York, pp. 451-462.
4. Kier, L. B., Cheng, C.–K. and Testa,. B.: 'A Cellular Automata Model of Enzyme Kinetics', 1996, *J. Molec. Graphics* **14**, pp. 227-231.
5. Kier, L. B. and Cheng, C.–K.: 'A Cellular Automata Model of an Anticipatory System', 2000, *J. Molec.* Graphics **18**, pp. 29-32.
6. Huang, C.–Y. F. and Ferell, J. E.: 'Ultrasensitivity in the Mitogen-activated Protein Kinase Cascade', 1996, *Proc. Natl. Acad. Sci. USA* **93**, pp. 10078-10083.