Pharmaceutical Research, Vol. 12, No. 4, 1995 *Report*

A Cellular Automata Model of the Hydrophobic Effect

Lemont B. Kier,[1] Chao-Kun Cheng,[1] Bernard Testa,[2] and Pierre-Alain Carrupt[2]

Received August 22, 1994; accepted November 22, 1994

Dynamic simulations of solute molecules in water are made using cellular automata. By varying the parameter governing the breaking probability of water-solute tesselated pairs, sets of configurations were modeled of solutions ranging from polar to non-polar solutes. The emergent behavior of the non-polar solute models leads us to believe that this is a possible model of the hydrophobic effect.

KEY WORDS: cellular automata; water; hydrophobic effect; complexity; emergent properties; dynamic simulation.

INTRODUCTION

Background of the Hydrophobic Effect

One of the most interesting and widely referred to solution phenomenon is the behavior of non-polar solutes in water. In an early review, Kauzmann organized the observations and understanding of the subject known at that time and developed a model based on the solute-water interaction known as hydrophobicity (1). This centered on the observation of the entropy loss when alkanes or other non-polar solutes are dissolved in water. It was suggested that among other effects, hydrophobicity was instrumental in the maintenance of the tertiary structure of proteins. The model proposed was that of the influence of the non-polar solute on nearby water molecules, causing them to become more organized or more "structured." This in turn promotes the association of the non-polar molecules into more complex systems which can, in some cases, achieve ordered forms such as micelles. The term "hydrophobic effect" has been used to describe the collective effects of the water-nonpolar solute phenomena.

Nemethy and Sheraga have extended our understanding of this phenomenon by developing a model of the system based upon the congregation of water molecules under the influence of non-polar solutes (2). These collections of water molecules were thought to be more stable relative to bulk water, principally due to a higher proportion of four-hydrogen bonded water species. Frank and Quist built a different model based on the idea that apolar molecules may occupy interstitial sites in the water clusters (3). It invoked the concept that water is a mixture of discreet species, some of which will accommodate the occupancy of a solute molecule in an interstitial space. The negative entropy observed is proposed to arise mainly from the change in function of water molecules from that of solvent to that of providing a residence within the clusters. Pierotti has proposed a model based on the scaled particle theory (4). In this model there are two processes determining the solubility; the formation of a cavity of sufficient size to hold a solute molecule, and the interaction phenomenon as a function of the nature of the solute. The quantitative aspect of the model depends upon the hard sphere diameters and the assumption of uniform distribution of water around the solute.

Dynamic Models of Water

Molecular dynamics has been used to simulate water and some of its properties. Rahman and Stillinger have used 216 molecules treated as rigid asymmetric rotors and a pair potential function in which the molecules have four tetrahedrally placed charges (5). The simulation reveals a network of strained hydrogen bonds with no significant differences in density from one area to another, thus discreet clusters are not postulated. There are no unbound water molecules evident but free OH groups are identified that persist longer than the molecular vibration period. Thus the simulation portrays the presence of different OH environments. The dynamics lead to the expectation that molecules constituting the cavities must be oriented with all hydrogen bonds directed away from the center of the cavity. This may account for the lowering of entropy when the cavity is occupied with an apolar solute. Recent reviews have updated the understanding of the hydrophobic effect in the light of these dynamic simulations (6,7,8).

Water as a Complex System

Molecular dynamic simulations represent a departure from the classical Newtonian doctrine of the reductionistic route to the understanding of nature. This is the belief that the whole is the sum of its parts and that understanding of the whole can be achieved by discovering the parts, defining them, and then attempting to recreate the whole again by some additive scheme. This concept has suffered serious deterioration in recent years with the discovery of the limitations in its ability to provide much understanding of a complex system. There has arisen a new philosophy, sometimes called post-Newtonian, in which systems are regarded as being complex when there is an extensive set of interactions occurring among the ingredients. Properties arise in a complex system that are called emergent, that is they cannot be derived from a linear sum of the properties of the ingredients. The route to understanding a complex system must be based upon reduction (analysis) to identify the essential ingredients followed by some non-linear, dynamic synthesis of a representation of the complex system. The emergent properties of the complex system may thus be modeled and some understanding achieved. This process can be viewed in a general way as:

---

[1] Departments of Medicinal Chemistry and Mathematical Sciences, Virginia Commonwealth University, Richmond, Virginia.

[2] Institut de Chimie Thérapeutique, Université de Lausanne, Lausanne, Switzerland.

0724-8741/95/0400-0615$07.50/0 © 1995 Plenum Publishing Corporation

$$\begin{matrix}\text{parts} & \xrightarrow{\text{synthesis}} & \text{complex system} \\ \text{(reduced interactions)} & \xleftarrow{\text{analysis}} & \text{(emergent properties)}\end{matrix}$$

Water as a bulk liquid may be viewed as a complex system with its properties emerging from non-linear interactions of the ingredient molecules. The use of quantum mechanical methods to analyze the complex system is a Newtonian approach with the limitations inherent in that concept. In such an analysis, a water molecule is isolated from its interacting neighbors hence an essential feature in the generation of emergent properties is lost if the process stops there. The dynamic methods attempt the reassembly through synthesis of the complex system and its emergent properties. The analysis and simulation processes used in molecular dynamics thus appears to be a viable approach to the understanding of water and the hydrophobic effect of its solutions.

CELLULAR AUTOMATA

General Principles of Cellular Automata

The use of dynamic simulations to carry out the synthesis part of the study of complex systems seems to be an easily supported procedure. It is worthwhile, at the same time, to examine several possibilities of conducting these dynamic studies beyond the use of molecular dynamics. One such method has its origins in the work of John von Neumann over half a century ago (9). In attempting to address the question of a self-reproducing machine he devised a method that is called cellular automata. In its simplest form, it is a collection of computing objects that take the form of cells in a grid or matrix of different sizes and dimensionalities. Each cell has an intrinsic quality or state characteristic of its contents. Each cell computes its relationship to its neighboring, tesselated cells according to rules chosen by the investigator. These rules take the place of physical forces and energies. The occupants of each cell, in turn, may move according to the rules chosen to describe the trajectory. After the computations of all of the cells according to the rules, a new configuration of the system is reached and one iteration has been accomplished. The configuration of the system progressively evolves with many successive iterations. Because the dynamics are discreet in space and time, as in an operating computer, selected attributes of the cellular automata can be monitored, revealing emergent properties of the system not existing in the individual cells.

Previous Studies with Cellular Automata and Water

In a previous study, we recognized the potential value of cellular automata to dynamically model liquid water (10). In that study, we used a stochastic or probabilistic set of rules governing the trajectory. The cells respond to the rules one after another to complete an iteration, thus the method is asynchronous. Each cell may change from one state to another simulating movement if the probabilities from random number generation so dictate. As a consequence, the cellular automata we use is kinematic. The rules are uniform for all cells and they are local, that is, there is no action at a distance. There is no relationship to the initial conditions which are usually set up as a random scatter of cells of different states. This model was used in a subsequent study of solutions (11).

The Rules

The model is made up of a domain composed of tesselated cells on the surface of a torus. Each cell, *i*, has four tesselated neighbors, *j*, and the cells two places beyond *i*, labeled k in Figure 1. The cells, *j*, constitute the von Neumann neighborhood for cell *i*, and the cells, *j* and *k*, constitute the extended von Neumann neighborhood for cell *i*. The state of each cell reflects whether it is empty or occupied with a specified molecule. The configuration of the entire system after any number of iterations is defined by the set of state values. After a discreet number of iterations, discreet changes in each cell occur according to rules that constitute the state and transition functions. These rules are based on probabilities hence the initial state does not uniquely specify the dynamics of any configuration. Only the parameters, not the initial states need be varied. The simulation is thus stochastic and not deterministic.

Two parameters are selected for a simulation to control the probabilities for movement of the contents of the cells in the grid. The breaking probability, $P_B(XY)$, used in an earlier study (11) is the probability for a molecule, *X*, in cell *i*, to break away from a molecule, *Y*, in cell *j* when there is exactly one occupied *j* cell, see Figure 1d. The value for $P_B(XY)$ lies in the closed unit interval. The second parameter, *J*, describes the movement of the molecule in cell *i* toward or away from the molecule in cell *k* in the extended von Neumann neighborhood when the intermediate cell *j* is vacant. It represents the ratio of the probability that a molecule at cell *i* will move toward an occupied *k* cell while the intermediate *j* cell is vacant and the probability that a molecule at cell *i* will move toward a vacant *k* cell while the intermediate *j* cell is vacant. *J* is a positive real number. When $J = 1$, the molecule in cell *i* has the same probability of movement toward an occupied cell *k* as when cell *k* is empty. When $J > 1$, it indicates that the molecule in cell *i* has a greater probability of movement toward an occupied cell *k* than when cell *k* is empty. When $J < 1$, it indicates that molecule *i* has a lesser probability of movement toward an occupied cell *k* than when *k* is empty.

In our model, the moving probability, p_m for a molecule in a cell is defined as follows:

$$p_m = \text{sum of } p_m(d) \text{ over all possible } d \text{ if the sum } <1, \text{ otherwise it is } 1. \tag{1}$$

where *d* is the direction in which *i* can move. The formula for computing the $p_m(d)$ in a certain direction *d* is as follows. Let *n* be the number of occupied *j* cells for a given *i* cell. There are 4-*n* many directions that *i* can move. We define:

$$\begin{aligned} p_m(d) &= 1/(1 + ((4 - n)/Q - 1/J(d)) \\ &\quad \text{if } n < 4 \text{ and } P_B > 0, \\ &= 0 \text{ otherwise} \end{aligned} \tag{2}$$

where $J(d) = 1$ if the *k* cell in the direction *d* is vacant, otherwise $J(d) = \mathrm{J}$. *Q* is the product of P_B values from occupied *j* cells.

The choice of $p_m(d)$ and p_m allows for a simple compu-

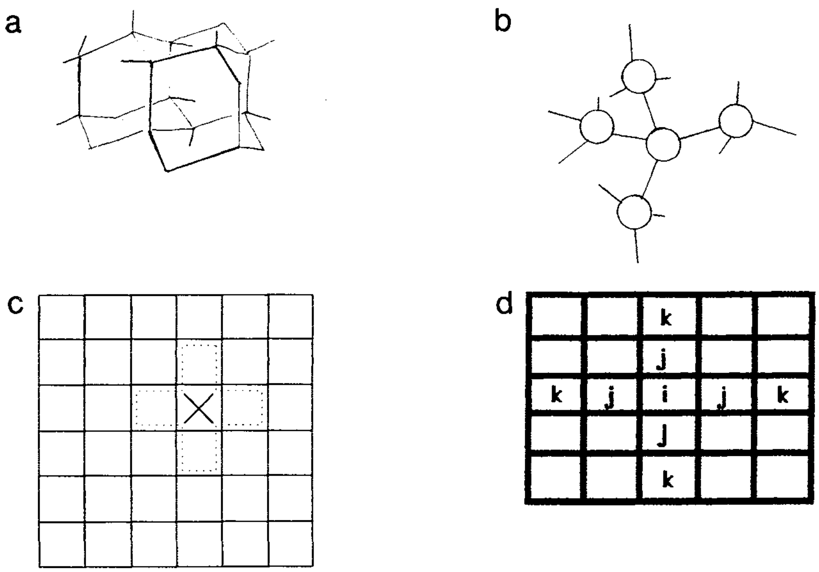

Fig. 1. (a) Ice lattic model of water. This is the proposed model of the trajectories taken by water molecules. (b) A tetrahedral fragment of the water molecule trajectory model. (c) A trace of Figure 1(b) coinciding with a von Neumann neighborhood. (d) An extended von Neumann neighborhood.

tation of moving probabilities and it also limits the influence from molecules in *k* cell to a portion of the moving probability, thus the influence from one direction does not overshadow the influence from other directions. Furthermore, when $J = 1$, $p_m = Q$; this agrees with the intuitively reasonable assumption that if the movement of a molecule is not influenced by molecules not bound to it, then its moving probability is the joint probability of probabilities for it to break away from molecules bound to it.

The Dimensionality Relationship in Our Models

In our model the trajectory of a liquid water molecule is assumed to approximate the paths of the hexagonal ice lattice, Figure 1a. Each vertex in that figure denotes a water molecule while each edge denotes a bonding relationship. This three dimensional network can be dissected into a contiguous series of tetrahedral fragments of five vertices each, Figure 1b. Some or all of these vertices in each fragment may be representative of a water molecule. The trace of each of these tetrahedral fragments can be mapped onto a two dimensional grid, Figure 1c. We equate this mapping with a cellular automata von Neumann neighborhood.

The cellular automata transition functions operate on the central cell, i, in each von Neumann neighborhood, Figure 1c. The rules in our studies composing these transition functions operate on the states of cells *i, j,* and *k* in the extended von Neumann neighborhood, Figure 1d. In our model the rules are executed for each cell in the cellular automata grid asynchronously, and at random over the grid. As a consequence, the new configuration for each cell, *i,* and its neighborhood is derived independent of all other cells remote from cells *i, j,* and *k,* in Figure 1d.

The configuration of the system, achieved after all cells respond in random order to the rules, constitutes one iteration. This configuration is a composite of the collective configurations achieved in all of the von Neumann neighborhoods. Each of these neighborhoods is a two dimensional mapping of a tetrahedral fragment of the original three dimensional model. Our model is a representation of the topology of a three dimensional system on the basis of it being a summation of discrete, orthogonal events occurring within that system. We can lay claim to the belief that the dimensionality of our model is fractional between two and three. Other studies using this approximation have been reported (12,13).

Initial Conditions

The size of the grid for this synthesis is 55 × 55 cells on a torus, thus there are no boundary conditions. As in our earlier studies, we have designated about 69% of the cells to have an occupied state. The detailed composition is 2000 cells designated as being occupied with water and 100 cells occupied with solute molecules. The initial configuration of the system is a random distribution of the cells of any state throughout the 3025 total cells in the grid. The transition functions or moving rules for water (W) and the solute (L) are designated thus:

P_B(W) . . . probability of a water molecule breaking from a water cluster

P_B(L) . . . probability of a solute molecule breaking from a solute cluster

P_B(WL) . . . probability of a water molecule breaking from a solute molecule (or vice versa) in a mixed cluster

J(W) . . . Joining parameter for two water molecules

J(L) . . . Joining parameter for two solute molecules

J(WL) . . . Joining parameter for a water and a solute molecule

Calculated Attributes

From each dynamic synthesis, after some number of iterations we can capture a set of attributes for any iteration. After allowing for a stabilization in these attribute values to occur, we may describe a configuration with the following examples:

1. Fractions of water or solute (relative to the total number in the simulation) bound to none, one, two, three, or four other molecules of the same type: f_0, f_1, f_2, f_3, f_4.
2. Average bonded (tesselated) state of water, n_{HB} (interpreted as average hydrogen bonding) or of solute molecules.

RESULTS

Influence of P_B(WL) on Calculated Water Attributes

These simulations were designed to study the emergent behavior of a solution due to the water-solute breaking parameter. In particular, we are interested in the possible changing relationship between water and solute when P_B(WL) is varied through a closed unit interval. The set of parameters for this study are: $P_B(W) = 0.25$, $P_B(L) = 0.10$, $J(W) = 1$, $J(WL) = 1$ and $P_B(WL) = 0.10, 0.20, \ldots 0.90$. Stability in the calculated values was achieved after about 5000 iterations but 15000 iterations were run for each value of P_B(WL). The values of the n_{HB} and f_4 were recorded for each value of P_B(WL) and are shown in Table I.

The results shown in the table demonstrate that the parameter P_B(WL), reflecting the relationship between the solute, L and water, W, has a significant influence on the structure of the water. Low values of this parameter describe a solute molecule with appreciable similarity to water in respect to the rules governing the breaking of W − L tesselations. The attributes, shown in the table reveal that this relationship leads to a disruption to some extent of the water structure. The average tesselation value, n_{HB}, and the fraction of four-bonded water molecules, f_4, are both lower when P_B(WL) values are low. In contrast, for higher values of the P_B(WL) parameter, reflecting a greater dissimilarity between the solute and water, the calculated attributes are higher. The finding of higher n_{HB} values (interpreted as greater hydrogen bonding among water molecules) for configurations with higher P_B(WL) values (simulating apolar molecules), agrees with molecular dynamic simulation evidence (14–16). This reflects what has been interpreted as a tendency to "increase" the water structure. The higher values of P_B(WL) may be viewed as being characteristic of a nonpolar solute, L.

Table I. Attributes of Water for Several P_B(WL) Values

P_B(WL)	f_4(W)	Ave. n_{HB}
0.1	0.28	2.90
0.2	0.29	2.91
0.3	0.29	2.92
0.4	0.30	2.93
0.5	0.32	2.96
0.6	0.31	2.97
0.7	0.32	2.97
0.8	0.34	3.00
0.9	0.33	3.00

Graphic Simulation of P_B(WL) Parameter Influence

The calculated attributes just described are quite informative and create a definite impression of the influence of this parameter on the structure of the water. Such data is really incomplete as far as an understanding of the events occurring in this dynamic synthesis. We must view the changing landscape of water and solute molecules as they respond to the rules and form the emergent behavior of this system. Accordingly we describe in words and pictures, the changes occurring as P_B(WL) changes from low to high values.

$P_B(WL) = 0.10$

Most of the solute molecules are dispersed as single entities within the patches of water molecules. The cavities among the water molecules are mostly free of any solute molecules. Figure 2a shows an isolated configuration of this simulation.

$P_B(WL) = 0.25$

A few of the L molecules are within the water patches, while about half are at the boundary between patches and the cavities. Figure 2b shows one typical configuration.

$P_B(WL) = 0.40$

No L molecules are within the water patches in this simulation. Most are within the cavities or are shared with the boundary of the patches of water. Figure 2c shows this configuration.

$P_B(WL) = 0.80$

All of the L molecules are within the cavity areas. See Figure 2d for this configuration.

DISCUSSION

We have created a model of an aqueous solution, a complex system, using the dynamic synthesis method known as cellular automata. It is a method based entirely on initial states and upon rules selected for the transition process. The goal is to find states and rules that produce emergent behavior that can be likened to some aspects of physical reality. This has been achieved to a limited extent in our previous papers (10,11). In this study, we hope to extend this correspondence by examining the model of emergent behavior involving water and a solute. In particular, we want to see

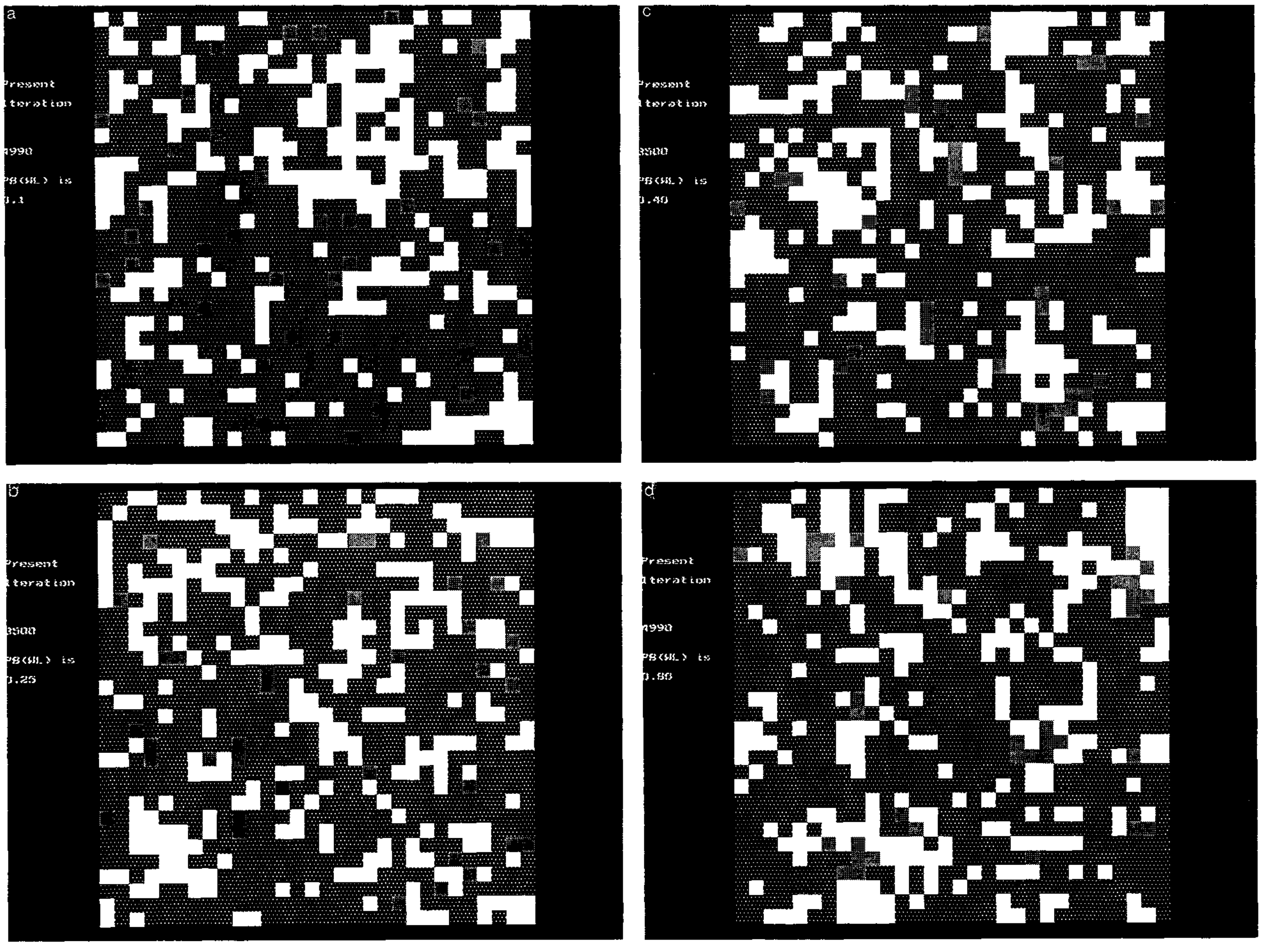

Fig. 2. Dynamic simulations of a solute in a water setting P_B = a) 0.10, b) 0.25, c) 0.40, and d) 0.80. The water cells are black, the solutes are gray and the cavity areas are white. The area shown in each case is enlarged from the original printout to more clearly illustrate the configurations described in the text.

if self-organization occurs in the dynamic synthesis using rules governing water and solute transitions.

When the breaking probability of water, W and a solute, L, P_B(WL), is low, the model is one of a close similarity between these species. When it is high, the model is one in which there is a modest to nil similarity. Interpreted broadly, we can say that a low P_B(WL) value results in a transition function typical of a polar solute. In contrast, a high P_B(WL) value is characteristic of a non-polar state. The behavior of non-polar solutes in water is what has interested biological scientists for many years. This is the set of conditions giving rise to the emergent behavior and self-organization called the hydrophobic effect.

It must be remembered that the results obtained in a stochastic cellular automata model are emergent attributes, not predictable from some simple addition of isolated ingredient attributes. Everything we see and that we record in tabular form are the attributes of a complex system, occurring dynamically as a result of interactions and transactions among the ingredients. The initial states and the probabilities selected establish trends in these emergent properties which are of interest to us in modeling reality.

The results of the dynamic synthesis in this report produce a model of the hydrophobic effect, self-organizing from initial conditions arising from a high P_B(WL) value. The simulation reveals that the solute molecules reside primarily as single entities within the cavity areas when the P_B(WL) values are high. The patches of four-bonded water molecules within the water extended network, encoded as the fraction f_4, are intact, that is they are not interspersed with solute molecules. This is reflected in the attributes shown in the table for the higher P_B(WL) values, and is in agreement with molecular dynamic simulations (14–16). The simulation shows that the solute molecules are largely segregated from the water, occupying positions within the cavities. These configurations can be described as a maintenance of the original structure of the water by accommodating the apolar molecule in its original hydrogen bond network. This is not the "structuration" frequently referred to as a consequence of the hydrophobic effect. This simulation then is compatible in several aspects with the recent evidence and models proposed for the phenomena called the hydrophobic effect (17). This simulation does not reveal any model interpretable as hydrophobic interaction, a subject that is being pursued in our laboratory.

At the other end of the P_B(WL) probability interval, we see that the simulations with low parameter values results in configurations with dramatically different emergent properties from that described above. With low P_B(WL) values the solute molecules (now characterized as polar) are widely interspersed among the patches of water in the extended network. This has the effect of reducing the structure of these patches as is demonstrated by the attribute values in the table and the configuration shown in Figure 2d. This is a simulation identifiable with the disruption of water structure produced by polar solutes. It recalls the use of urea in aqueous solutions of proteins to produce water structure changes influencing protein denaturation.

The value of this method and the models produced remain to be fully evaluated. We offer here an opening round of study and statement of progress in the hope that others will be stimulated to consider cellular automata synthesis as a productive route to the understanding of emergent behavior of complex systems of interest to drug scientists.

ACKNOWLEDGMENTS

Part of this work was performed while Professor Kier was a visiting scientist at the Université de Lausanne. Thanks is expressed to the "Fondation du 450e anniversaire de l'Université de Lausanne" for a generous award to Professor Kier in support of this research.

The computations were performed using a continuously upgraded program we call DING-HAO. This study was run on an IBM 486/66 where 1000 iterations took about 5 minutes.

REFERENCES

1. W. Kauzmann, *Adv. Protein Chem.* 14, 1-63 (1959).
2. G. Nemethy and H. A. Sheraga, *J. Chem. Phys.* 36, 3401-3408 (1962).
3. H. S. Frank and A. S. Quist, *J. Chem. Phys.* 34, 604-609 (1961)
4. R. A. Pierotti, *J. Phys. Chem.* 69, 281-287 (1965).
5. A. Rahman and F. H. Stillinger, *J. Chem. Phys.* 55, 3336-3344 (1971).
6. L. R. Pratt, *Ann. Rev. Phys. Chem.* 36, 433-449 (1985).
7. N. Muller, *Acc. Chem. Res.* 23, 23-28 (1990).
8. I. Ohmine and H. Tanaka, *Chem. Rev.* 93, 1545-2566 (1993).
9. J. von Neumann in *Theory of Self-Reproducing Automata.* (A. Burks, ed.). Univ. of Illinois Press, Urbana (1966).
10. L. B. Kier and C.-K. Cheng, *J. Chem. Inf. Comput. Sci.* 34, 647-652 (1994).
11. L. B. Kier and C.-K. Cheng, *J. Chem. Inf. Comput. Sci.* 34, (1994).
12. R.L. Blumberg, G. Shlifer and H.E. Stanley, *J. Phys. A : Math. Gen.* 13, L-147 (1980).
13. M.H. Brodsky, *Bull.Am.Phys. Soc.* 25, 260 (1980).
14. S. Okazaki, K. Nakanishi, N. Touhara, N. Watanabi, and Y. Adashi, *J. Chem. Phys.* 74, 5863-5868 (1981.
15. A. Geiger, A. Rahman, and F. Stillinger, *J. Chem. Phys.* 70, 263-271 (1979).
16. W. L. Jorgensen, J. Gao, and C. Ravimohan, *J. Chem. Phys.* 89, 3470-3478 (1985).
17. W. Blokzijl and J. B. F. N. Engberts, *Angew. Chem. Int. Ed. Engl.* 32, 1545-1579 (1993).