

A cellular automata model of the soluble state

Lemont B. Kier^a and Chao-Kun Cheng^b

^a *Department of Medicinal Chemistry, Virginia Commonwealth University, Richmond, Virginia 23298, USA*

^b *Department of Mathematical Sciences, Virginia Commonwealth University, Richmond, Virginia 23298, USA*

Received 6 March 1996; revised 14 January 1997

Cellular automata models of solubilities in a solvent (water) have been dynamically synthesized. Rules relating water-water, water-solute, and solute-solute relationships have been systematically varied in order to assess their influence on the emergent property of solubility. The results reveal the prominent influence of rules governing the probabilities of solute-water joining and breaking. This influence manifests itself in significant changes in the emergent properties of relative solubility and solubility changes with water “temperature”. The study demonstrates the validity and potential value of cellular automata to model solution phenomena.

1. Introduction

The liquid state of matter lying between the order of the solid and the randomness of the vapor, has been characterized as a chaotic state. The dynamic events postulated taking place produce a substance with attributes unique among liquids. A static definition of water structure is not meaningful; understanding being derived from dynamic simulations of the interactions between components of the liquid. These simulations have led to a general model with its origins in the work of Zachariasen [17] and Bernal and Fowler [1]. Subsequent work has led to the emergence of a theory of liquid water as a random network of hydrogen bonded molecules, rapidly breaking and exchanging partners [4,12–14]. Rice and Skeats [13] have defined this model as having i) a narrow range of separation of nearest neighbors, a condition very close to that of ice, ii) an irregular connectivity within the network with various structures of multiple waters randomly interwoven throughout the network, iii) a significant distribution of angles between nearest neighbor bonds and iv) very few unbonded or small clusters of water molecules. Other characteristics have been described by investigators studying different aspects of the problem [15].

The goal of understanding water may be approached by using one of two philosophical concepts. In the classical Newtonian concept, we seek understanding through the reduction of the whole into its parts or ingredients with subsequent study of these parts and their limited interactions. These analyses have been viewed by some as constituting the only way to achieve fundamental knowledge about a system. In contrast, there has arisen in recent years an alternative approach to the understanding of nature. This approach employing both reduction and synthesis, seeks to model a complex system by combining separate elements to form a whole which has behavior and attributes that have emerged unpredictably from the parts. Kier and Testa have recently written on this topic as it relates to drug research [11]. Nowhere is the reality of emergent behavior more evident than in the properties of a solution.

2. Aqueous solubility

One of the most important aspects of water is its ability to function as a solvent. Indeed in the biological realm, aqueous solubility is an essential characteristic of most biomolecules. Certain intermolecular forces have been ascribed between solute molecules and solute–water pairs to account for the observation of solubility. The classical scenario is i) the removal of a molecule from the solid phase, ii) the creation of a cavity in the solvent to accept the solute molecule and finally iii) the acceptance of the solute molecule into the cavity. The estimation of solubility from molecular structure is actively pursued. It is a complicated problem because of the impact of the electronic structure of mantle atoms, the topology, the size and the similarity to the solvent. The traditional approach has been to employ a reductionist analysis of molecules to the level of fragments and atoms to explain the measured property. This is the basis of quantitative structure–activity relationships, much in use today to evaluate phenomena including solubility. Such an approach omits the consideration of water as a dynamic entity. Physical phenomena such as solubility arise as emergent properties, unpredictable from an exclusively reductionist approach.

A second consideration that has guided our earlier work on water and solution phenomena [6–10], is that a route to some understanding of solubility may begin with relationships among the ingredients in the form of rules of engagement. Thus we do not attempt to define and estimate forces between molecular features, but we write down stochastic rules that govern whether two objects engage or repel; whether they form aggregates or whether they break apart. Such a non-traditional approach is a long way from the classical idea of ascribing to molecular fragments all of the forces influencing a phenomenon; it epitomizes the phenomenon with rules. The outcomes or emergent properties from the rules may be unforeseen and thus the dynamic process of synthesis must be used. If we can obtain a profile of the extent of some phenomenon like solubility based on rules, then further understanding of the phenomenon may be at hand. This is the philosophy we have adopted in our cellular automata models of water [6–10]. This is the approach we employ here to study aqueous solubility.

3. Cellular automata

Cellular automata is a dynamic method for the modeling of the emergent behavior of a complex system [2,3,16]. It is composed of a lattice or grid of sites called cells in one, two or three dimensions. The intrinsic quality or status of any cell, i , is the state of that cell. Each cell i , is in a local situation or map called a neighborhood. This neighborhood of tessellated cells influences the changes that cell i may experience. These changes or movement parameters are encoded into a set of rules. With the successive response of each cell i to the movement parameters, an iteration of the entire ensemble of cells occurs producing a configuration of the system and a unit of "time". The rules of a cellular automata are local, hence, there is no action at a distance and the laws are uniform, that is the rules are the same for every cell. The discreteness in time, space and rules results in an error-free specification of the dynamics. The functioning of a computer is by the same process.

4. The rules

Our model is made up of a domain composed of cells on the surface of a torus. Each cell, i , has four joined neighbors, j , and the cells two places beyond i , labeled k in figure 1. The cells, i and j , constitute the von Neumann neighborhood while the cells, j and k , constitute the extended von Neumann neighborhood. The state of each cell reflects whether it is empty or occupied with a specific molecule. The configuration of the entire system after any number of iterations is defined by the set of state values. After a discrete number of iterations, discrete 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 model is thus stochastic and not deterministic.

Two parameters are selected for a simulation to govern the probabilities for movement of the contents of the cells in the grid. The breaking probability, $P_B(XY)$ is the probability for molecule, X , in cell i , figure 1 to detach from the molecule of type Y , at some j cells when there is exactly one occupied j cell. The value for $P_B(XY)$, lies in the closed unit interval. When X and Y are the same molecular types, it is denoted by $P_B(X)$ or $P_B(Y)$ as symbolized in our previous study [7].

The second parameter, $J(X)$, describes the movement of the molecule, X , in cell i toward or away from the molecule Y , in cell k in the extended von Neumann neighborhood when the intermediate j cell is vacant. It represents the ratio of the probability that a molecule of type X at cell i will move toward a k cell occupied by a molecule of type Y while the intermediate j cell is vacant, and the probability that a molecule of type X at cell i will move toward a vacant k cell while the intermediate j cell is vacant. J is a positive real number. When $J(XY) = 1$, the molecule in cell i has the same probability of movement toward an occupied cell k as when cell k is empty. When $J(XY) > 1$, it indicates that the molecule in cell i has a greater

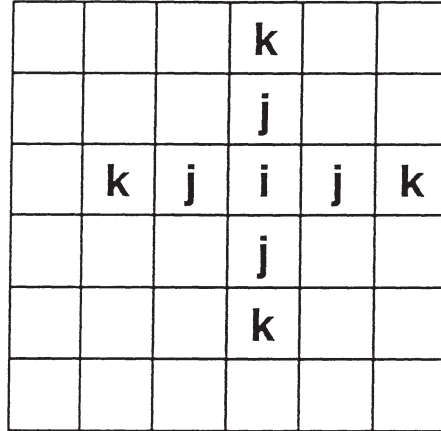


Figure 1. Cells i and j represent the von Neumann neighborhood while cell i , j and k represent the extended von Neumann neighborhood.

probability of movement toward an occupied cell k than when cell k is empty. When $J(XY) < 1$, it indicates that molecule i has a lesser probability of movement toward an occupied cell k than when cell k is empty. When X and Y are the same molecular types, it is denoted by $J(X)$ or $J(Y)$.

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

$$p_m = \begin{cases} \text{sum of } p_m(d) \text{ over all possible } d, & \text{if the sum } < 1, \\ 1, & \text{otherwise,} \end{cases} \quad (1)$$

where d is a direction in which the molecule in cell 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. So there are $4 - n$ many directions that i can move. We define:

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

where $J(d) = 1$ if the k cell in the direction d is vacant, otherwise $J(d) =$ the J value with respect to the molecule types in cell i and cell k in the direction d , and where P_B is the product of the breaking probabilities contributed from the occupied cells.

The choice of $p_m(d)$ and p_m allows for a simple computation of moving probabilities while limiting the influence from molecules in k cells to a portion of the moving probability, thus the influence from one direction does not overshadow the influence from other directions. Furthermore, when all J values used in the system are equal to 1, $p_m = (P_B)$; 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.

5. Initial conditions

The size of the grid for our 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 an unspecified solute molecule. 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), the solute (L) and a tessellated pairing of the two (WL) 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.

6. Attributes recorded

At the end of each dynamic synthesis two attributes were recorded as average values over the last 100 iterations of 2000 iteration runs. The first of these is the fraction of isolated solute molecules not joined to other L molecules, labeled $f_0(L)$. The values range over a unit interval. In this study we interpret this attribute to model the state of a molecule most likely to be in solution. No doubt different solutes may exhibit behavior characteristic of being in solution while existing in states of binding as dimers and higher x -mers. Since we don't define what our solute is but are pursuing a general model, we can accept this approximation as valuable to our understanding of the rule-based influences. The second attribute recorded is the average number of L molecules joined to other L molecules, averaged over the last 100 iterations. We use the symbol $T(L)$ to denote this attribute. Its value ranges between 0 and 4.0. We assume that high values describe a clustered state of the solute hence a relatively low solubility.

7. Influence of solute-related rules on solubility

In this first series of simulations we are seeking information about the influence of various solute rules on the attributes interpreted as solubility. The rules governing the water state and transitions are held constant at: $P_B(W) = 0.25$, $J(W) = 1.0$. The sets of rules governing the trajectories of the solute are shown in table 1. For each transition function, a high and a low value of the rule values were employed. For the breaking probabilities, $P_B(WL)$ and $P_B(L)$, the rules reflect high and low probability

Table 1
Solubility attributes from dynamics of various rules.

Rule set	$P_B(WL)$	$P_B(L)$	$J(WL)$	$J(L)$	$f_0(L)$	$T(L)$
1	0.15	0.85	1.5	0.5	1.00	0.00
2	0.15	0.85	1.5	1.5	1.00	0.00
3	0.15	0.85	0.5	0.5	0.99	0.01
4	0.15	0.85	0.5	1.5	0.98	0.02
5	0.15	0.15	1.5	0.5	0.92	0.08
6	0.15	0.15	1.5	1.5	0.92	0.08
7	0.15	0.15	0.5	0.5	0.87	0.14
8	0.15	0.15	0.5	1.5	0.83	0.18
9	0.85	0.85	1.5	0.5	0.82	0.18
10	0.85	0.85	1.5	1.5	0.81	0.20
11	0.85	0.85	0.5	1.5	0.73	0.30
12	0.85	0.85	0.5	0.5	0.63	0.46
13	0.85	0.15	1.5	0.5	0.56	0.54
14	0.85	0.15	1.5	1.5	0.49	0.66
15	0.85	0.15	0.5	0.5	0.30	1.01
16	0.85	0.15	0.5	1.5	0.17	1.48

values of 0.85 and 0.15. For the joining parameters, $J(WL)$ and $J(L)$, the high and low parameter values chosen were 1.5 and 0.5. This variety is designed to cover a large part of the parameter space chosen for this study. The $f_0(L)$ and $T(L)$ values synthesized from each rule set are shown in table 1.

8. Results

The fraction of solute molecules not bound to other solute molecules, $f_0(L)$ in table 1 is shown to be most strongly dependent on the value of $P_B(WL)$. The upper half of the ranked list of $f_0(L)$ values correspond without exception to the low numerical values of $P_B(WL)$. The lower the probability of breaking WL bonds, the higher the fraction of single solute molecules. A change in the $P_B(WL)$ value from 0.85 to 0.15 can increase the $f_0(L)$ fraction by as much as 0.66, with an average increase of 0.38.

Changes in $P_B(L)$ from 0.15 to 0.85 produce the next largest effect on $f_0(L)$ among the rule sets in table 1. In both cases the higher $P_B(L)$ values give rise to higher $f_0(L)$ values. The average increment in $f_0(L)$ created by the $P_B(L)$ value changes, shown in table 1, is about 0.24. The influence on $f_0(L)$ due to changing $J(WL)$, in table 1 is modest. The average value of these increments is about 0.13. Finally, the change in $J(L)$ values produces virtually no change in $f_0(L)$. The relative influence of these parameters can be summarized as: $P_B(WL) > P_B(L) > J(WL) > J(L)$ by ratios of 3:2:1:0.

In table 1 are also recorded the average tessellation value $T(L)$, characterizing

Table 2
Influence of various parameters on the $P_B(W)$ - $f_0(L)$ relationship.

Rule set	$P_B(WL) = 0.9$		$P_B(W) = 0.25$		$P_B(W) = 0.75$	
	$P_B(L)$	$J(L)$	$J(WL)$	$f_0(L)$	$f_0(L)$	$\Delta f_0(L)$
1	0.2	4.00	0.25	0.13	0.27	0.14
2	0.2	0.25	0.25	0.30	0.77	0.47
3	0.7	4.00	0.25	0.45	0.71	0.26
4	0.2	4.00	4.00	0.46	0.79	0.33
5	0.7	0.25	0.25	0.68	0.92	0.24
6	0.2	0.25	4.00	0.71	0.82	0.11
7	0.7	4.00	4.00	0.81	0.85	0.04
8	0.7	0.25	4.00	0.86	0.94	0.08

the degree of bonding. These values range from zero to four. As expected, they are inversely related to the $f_0(L)$ values in the same table. The correlation is very high.

9. Influence of water rules on solubility

This study was designed to evaluate the influence of the rules governing the breaking and joining of water clusters upon the fraction of unbound solute molecules. In an earlier study, [6] we found that the $P_B(W)$ parameter was in close correspondence to the liquid water temperature based on several criteria. We also found that increasing the $P_B(W)$ value produced an increase in the f_0 value corresponding to a model of an increased solubility with a higher water “temperature”. We wish to extend that study to evaluate the “temperature-solubility” effect influenced by other parameters.

In this study, we held $P_B(WL)$ constant at 0.90 and the $J(W)$ parameter at 1.0. We ran the dynamics for 5000 iterations beginning with a random distribution of 100 solute molecules and 2000 water molecules in a 55×55 grid. Two values of $P_B(W)$ were chosen for selective evaluation, 0.25 and 0.75. The attributes are reported as averages over the last 100 iterations of 5000 iteration runs. The three variable transition functions, $P_B(L)$, $J(L)$ and $J(WL)$ are each represented in the parameter sets by two rule values, one high and the other low. The values of $f_0(L)$ for each rule set and the two $P_B(W)$ values are shown in table 2.

10. Results

The principal observation from table 2 is the increase in the fraction of unbound solute molecules, $f_0(L)$, when the water breaking probability is increased. This increase ranges from a very modest 0.04 to a large value of 0.47. There is no simple correlation between the incremental change in $f_0(L)$ due to the increase in $P_B(W)$ and the value of $P_B(L)$ itself. There is a clear relationship between $\Delta f_0(L)$ and the value of $J(WL)$. With only one exception, the low values of $J(WL)$ contribute to the high values of $\Delta f_0(L)$.

11. The precision of multiple runs

The question arises concerning the relative reproducibility of multiple dynamic models using stochastic parameters. To develop some insight into this problem we have selected a test set of parameters representing values central in the parameter space shown in table 1. These parameter values are:

$$\begin{aligned} P_B(W) &= 0.25, & J(W) &= 1, \\ P_B(WL) &= 0.5, & J(WL) &= 1, \\ P_B(L) &= 0.5, & J(L) &= 1. \end{aligned}$$

A dynamic simulation was run using these parameters. At 4000 iterations, average values of each of three attributes were computed over successive 100 iteration segments until the iteration count reached 6000. Thus 20 sets of average values were collected in this run. In addition, standard deviations of each of these averages were computed. This process was then repeated 20 times, collecting average values and standard deviations in each run. An average of these 20 sets of averages was then computed for each of the three attributes along with the standard deviations. These results are:

$$\begin{aligned} \text{average } f_0(L) &= 0.830 \pm 0.012, \\ \text{average } T(L) &= 0.182 \pm 0.014, \\ \text{average } n_{\text{HB}}(W) &= 2.923 \pm 0.005, \end{aligned}$$

where $n_{\text{HB}}(W)$ is the average count of bonded neighbors to a water molecule, interpreted earlier [6] as the average hydrogen bonding state. This analysis reveals a high degree of reproducibility for these three attributes emerging from dynamics based upon stochastic parameters.

12. Discussion

A series of cellular automata have been synthesized using stochastic rules to influence the changing configurations of solute molecules in a solvent (here interpreted as water). The purpose of the study is to evaluate the influence of the rules on two ‘‘solubility’’ attributes. From a systematic variation of the rules, we can conclude that, under the conditions of our simulations, the water–solute breaking parameter, $P_B(WL)$, is the most influential on the extent of single solute molecule, $f_0(L)$, formation. This implies that a very influential factor in solubility is the relationship between solute and solvent. This is not a new idea certainly, but it is noteworthy that it arises as a spontaneous event, or an emergent property from this dynamic synthesis. Furthermore, this influence appears to be greater than the solute–solute cluster breaking parameter, $P_B(L)$, although this later rule is a close second in average influence. This finding is characteristic of a model in which the solute is formed in situ from a reaction. We have obtained similar results by creating a model of the dissolution of a solid block of solute molecules [8].



Figure 2. A dynamic simulation of 100 solutes and 2000 water molecules in a 55×55 grid. The figure enlarges a section of the total grid space. The $P_B(WL)$ value is 0.10. The black cells are water, the gray are solute and the white are cavities.

The finding that low values of the water–solute breaking parameter, $P_B(WL)$, produce a high fraction of unbound or single solute molecules, $f_0(L)$, is in agreement with our earlier study [7]. In that series of simulations we were able to draw a correspondence between low $P_B(WL)$ values and high “polarity” of molecules so described. In those cases, the solute molecules were found to reside mostly within the patches formed of water molecules in the random network. The single or unbound configurations of the solute molecules is described by high $f_0(L)$ values. This is interpreted as “being soluble”. In contrast, the high $P_B(WL)$ values studied earlier [7, 9] led to configurations of solute molecules outside of water patches, within the cavities between them. This was interpreted as epitomizing a non-polar molecule. We related this configuration to the conventional concept of the hydrophobic effect. In the present study we see that high $P_B(WL)$ values lead to configurations interpreted as having lower “solubilities”. Two figures, 2 and 3, show configurations synthesized from low and high values of $P_B(WL)$, respectively. It is apparent that there is a significant difference in the $f_0(L)$ values. Figure 2, from $P_B(WL) = 0.1$, shows numerous single solute molecules among the water patches with a high $f_0(L)$ value. We relate these to “polar molecules”. In contrast, in figure 3, the solute molecules synthesized from



Figure 3. The same conditions as in figure 2, but with $P_B(WL) = 0.80$.

$P_B(WL) = 0.8$ are concentrated within the cavities. These exist as single, dimeric and higher x -mer aggregates, resulting in a lower $f_0(L)$ value. The solute is less soluble than the study shown in figure 2. We can say that these are a “non-polar molecules”. Overall, our results reveal that the polarity of a molecule, describable by the $P_B(WL)$ rules, is influential on the solubility.

Changes in the water–water breaking probability, $P_B(W)$, clearly influence the $f_0(L)$ values. As shown in earlier studies [7,9], this rule closely parallels the temperature. At high values of $P_B(W)$ the $f_0(L)$ values are higher; a clear parallel to the general observation of higher solubility with a higher temperature. The extent of the $f_0(L)$ increase with $P_B(W)$ is noticeably dependent upon the water–solute joining parameter $J(WL)$. The conclusion here is that a lower probability of a solute molecule moving to bind with a water molecule leads to an increase in the solubility when the water “temperature” is increased. This is interpretable from table 2, where low values of $J(WL)$ have initially lower values of $f_0(L)$ hence there is a potentially greater range of $f_0(L)$ values possible. It is also quite apparent in this study that an attribute of the solute, the $f_0(L)$, changes when a rule governing only water, $P_B(W)$ changes. This is a vivid example of an emergent property arising in a complex system from the non-linear, hence unpredictable interactions of two ingredients.

These studies describe interpretations of the rules used in our cellular automata

dynamics of solution phenomena. This is a new approach to the study of these systems at the molecular system level. We are using rules to define the non-linear contribution of molecules to emergent properties in a complex system. This is the same philosophy employed in molecular dynamics [5], however, the simplicity, visualizability and adaptability of complex systems using cellular automata are strong factors in its favor as a method of dynamic modeling.

Acknowledgements

The authors appreciate helpful discussions with Jurgen Venitz and Glen Kellogg. The program employed for the CA calculations is the steadily evolving program DINGHAO. The A.D. Williams Foundation is gratefully acknowledged for a grant to obtain computer facilities.

References

- [1] J.D. Bernal and R.H. Fowler, A theory of water and ionic solution, *J. Chem. Phys.* 1 (1933) 515.
- [2] G.B. Ermentrout and L. Edelstein-Keshet, Cellular automata approaches to biological modeling, *J. Theor. Biol.* 160 (1993) 97.
- [3] D. Farmer, T. Toffoli and S. Wolfram, eds., *Cellular Automata*, *Physica D* 10 (1984).
- [4] A. Geiger, F.H. Stillinger and A. Rahman, Aspects of the percolation process for hydrogen bond networks in water, *J. Chem. Phys.* 70 (1979) 4186.
- [5] J.M. Haile, *Molecular Dynamics Simulation* (Wiley, New York, 1992).
- [6] L.B. Kier and C.-K. Cheng, A cellular automata model of water, *J. Chem. Inf. and Comp. Sci.* 34 (1984) 647.
- [7] L.B. Kier and C.-K. Cheng, A cellular automata model of an aqueous solution, *J. Chem. Inf. and Comp. Sci.* 34 (1994) 1334.
- [8] L.B. Kier and C.-K. Cheng, A cellular automata model of dissolution, *Pharm. Res.* 12 (1995) 1521.
- [9] L.B. Kier, C.-K. Cheng, B. Testa and P.-A. Carrupt, A cellular automata model of the hydrophobic effect, *Pharm. Res.* 12 (1995) 615.
- [10] L.B. Kier, C.-K. Cheng, B. Testa and P.-A. Carrupt, A cellular automata model of micelle formation, *Pharm. Res.* 13 (1996) 1419.
- [11] L.B. Kier and B. Testa, Complexity and emergence in drug research, *Advances in Drug Research* 26 (1995) 1.
- [12] M. Mezei and D.L. Beveridge, Theoretical studies of hydrogen bonding in liquid water and dilute aqueous solutions, *J. Chem. Phys.* 74 (1981) 622.
- [13] S.A. Rice and M.G. Skeats, A random network model for water, *J. Phys. Chem.* 85 (1981) 1108.
- [14] R.J. Speedy, Self-replicating structures in water, *J. Phys. Chem.* 88 (1984) 3364.
- [15] H.E. Stanley and J. Teixeira, Interpretation of unusual behavior of H₂O and D₂O at low temperatures: tests of a percolation model, *J. Chem. Phys.* 73 (1980) 3404.
- [16] S. Wolfram, Statistical mechanics of cellular automata, *Rev. Mod. Physics* 55 (1984) 601.
- [17] W.H. Zachariasen, The atomic arrangement in glass, *J. Am. Chem. Soc.* 54 (1932) 3841.