A Cellular Automata Model of Bond Interactions Among Molecules

Lemont B. Kier[†]

Department of Medicinal Chemistry, Virginia Commonwealth University, P.O. Box 980540,
Richmond, Virginia 23298

The ten types of alkane bonds have been modeled as isolated fragments using cellular automata dynamics. The rules governing the states and the trajectories of the cells simulating the bonds are derived from the bimolecular interaction accessibilites recently described. The sum of cell encounters at unit iteration (time) becomes a parameter associated with the relationship of a molecule with its neighbors. This value is found to correlate very closely with the boiling points of alkanes, pentanes through octanes. The results reinforce the concept that the molecular connectivity indices are describing the interaction possibilities among alkanes. The study introduces a new way of simulating intermolecular bond encounter dynamics among many molecules.

INTRODUCTION

In a recent study we have proposed an interpretation of molecular connectivity indices in terms of the collective accessibilities of the bonds to bimolecular encounters.[1] It was shown that the chi index, $^1\chi$, is a summation of the numerical bond terms that express their accessibility A_{ij} to encounter other bonds of another identical molecule. These bond accessibilities are derived from the δ values of the atoms comprising each bond. The δ value is the count of σ electrons associated with a carbon atom that are not bonding with a hydrogen atom. The Randic formalism, created to quantify alkane branching, used the product of the δ values for each such bonding pair of carbon atoms.[2] A number encoding each bond was thus created:

$$C_{ij} = (\delta_i \delta_j)^{-0.5} \tag{1}$$

Our derivation of the C_{ij} value led to the interpretation that this value encodes the collective influences, electronic and steric, on the ability of that bond to encounter another bond in its environment.[1] This attribute was called the bond accessibility, A_{ij}, and calculated from

$$A_{ij} = [(1/\delta_i)(1/\delta_j)]^{0.5} \tag{2}$$

The two values are, of course, the same, but the A_{ij} term is expressed as the reciprocal because this has significance in its derivation.[1] The bond accessibility, A_{ij}, is a descriptor expressing numerically the relative possibility of that bond encountering another bond (described by its A_{kl} value) in a second like molecule. The product of these two accessibilities, $A_{ij}A_{kl}$, is the contribution of that two-bond encounter to the sum of all possible two-bond encounters, P_T, of the two molecules, m and n.

$$P_T = \sum_m (A_{ij})_m \sum_m (A_{kl})_n \tag{3}$$

The contribution of each molecule to P_T is therefore P_m and P_n, and so

$$P_T = P_m P_n \tag{4}$$

It follows that

$$P_m = \sum_m (A_{ij})_m = {}^1\chi(m) \tag{5}$$

The conclusion is that the molecular connectivity index is the contribution of that molecule to the bimolecular interactions arising from all possible encounters of bonds among two like molecules. This is a concept in which the structure of a molecule is expressed in terms of attributes manifesting function. This brings molecular connectivity, as a structural paradigm, closer to a functional role of molecules, manifested by measurable physical properties. It presents a more specific idea of what the molecular connectivity indices "mean" rather than just expressions of "structure" devoid of any significant interpretation in the context of molecular encounters.

BOND ENCOUNTERS

It must be made clear that this interpretation is based on bond−bond encounters and so is a more valid model for alkanes. More complex molecules have intermolecular encounters based on these and other molecular features such as π and lone pair electrons. Still, the molecular connectivity indices have been expanded to encode some of this information which makes them of more general value in studying structure−activity relationships.[3,4] The E-state indices have been developed to encode atom contributions to intermolecular noncovalent interactions.[5] With this caveat in mind, we have examined the bimolecular encounter model applicable for alkanes.

We have shown the exact relationship between $^1\chi$ and the sum of bond encounter possibilities among two molecules. As a model, the bimolecular encounter is a limiting case. In the liquid state the alkanes engage in a host of intermolecular

† Telephone: (804) 828-6451. FAX: (804) 828-7625. E-mail: kier@hsc.vcu.edu.

Table 1. Bond Types in Csp^3 Molecules

bond type (δ_i,δ_j)	code	$\delta_i\,\delta_j$	$C_{ij} = A_{ij}$
1,1	unique	1	1.000
1,2	a	2	0.707
1,3	c	3	0.577
1,4	f	4	0.500
2,2	b	4	0.500
2,3	d	6	0.408
2,4	g	8	0.354
3,3	e	9	0.333
3,4	h	12	0.289
4,4	i	16	0.250

encounters among many molecules. By modeling the bond encounters among a large number of molecules, we might simulate the conditions of many molecules encountering each other and transcend the information resident in the bimolecular $^1\chi$ index. One such approach regards the bonds as scattered parts or fragments, a system that could be called *disjecta membra*. Each bond of a molecule is modeled as a disconnected entity, endowed with a state that corresponds to one of the types of bonds found in Table 1. These bonds move through a system containing other bonds in other states from other molecules of the same structure. This model can be simulated with cellular automata dynamics. That is the approach that we have pursued in this study.

CELLULAR AUTOMATA

The Model. Cellular automata are dynamical computational systems that are discrete in space, time, and state and whose behavior is specified completely by rules governing local relationships. They are an attempt to simplify the often numerically intractable dynamic simulations into a set of simple rules that mirror intuition and that are easy to compute. As an approach to the modeling of emergent properties of complex systems, it is of great benefit in being visually informative of the progress of dynamic events. From the early development by von Neumann,[6] a variety of applications ranging from gas phenomena to biological applications have been reported.[7-9] Our recent work has utilized cellular automata to model aqueous systems including water structure,[10] aqueous solutions,[11] oil–water partitioning,[12] the hydrophobic effect,[13] aqueous diffusion,[14] and acid dissociation,[15] all recently reviewed.[16]

Our model is composed of a grid of spaces called cells on the surface of a torus to remove boundary conditions. Each cell i has four tessellated neighbors, j, and four extended neighbors, k, in what is called an extended von Neumann neighborhood, Figure 1. Each cell has a state governing whether it is empty or is occupied by a cell with a nonzero state. The contents of a cell move, join with another occupied cell, or break from a tessellated relationship according to probabilistic rules. These rules are established at the beginning of each simulation. The rules are applied one after another to each cell at random, the complete application of the rules to all cells constituting one iteration. The rules are applied uniformly to each cell type and are local; thus there is no action at a distance. Our cellular automata model is thus kinematic, asynchronous, and stochastic. The initial conditions are random; hence they do not determine the ultimate state of the cells, called the configuration. The same
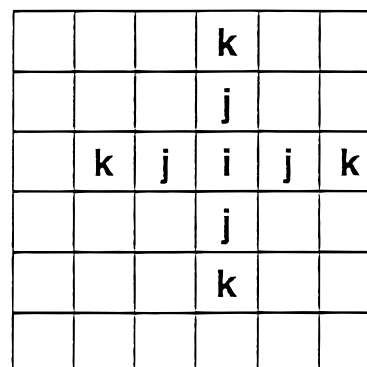
Figure 1. The von Neumann neighborhood focused on cell i.

Table 2. Intermolecular Encounter Possibilities

	bond types							
	a	c	b/f	d	g	e	h	i
bond types	1,2	1,3	2,2/1,4	2,3	2,4	3,3	3,4	4,4
a 1,2	0.50	0.41	0.35	0.29	0.25	0.24	0.20	0.18
c 1,3		0.33	0.29	0.24	0.20	0.19	0.17	0.14
b/f 2,2/1,4			0.25	0.20	0.18	0.17	0.15	0.13
d 2,3				0.17	0.14	0.13	0.12	0.10
g 2,4					0.13	0.12	0.10	0.09
e 3,3						0.11	0.10	0.08
h 3,4							0.08	0.07
i 4,4								0.06

initial conditions do not yield the same set of configurations after a certain number of iterations except in some average sense. The configurations achieved after many iterations reach a collective organization that possesses a relative constancy in appearance and in reportable counts of attributes. What we observe and record from the cellular automata simulations are emergent attributes of a complex system.

The Rules. Two parameters were adopted in our model to govern the probabilities of a cell encountering other cells in the grid. The breaking probability, P_B, is the probability for a cell at i to break away from another at j when there is exactly one occupied j cell (see Figure 1). The value for P_B lies in the closed unit interval. The second parameter, J, describes the movement of a cell at i toward or away from the cell at k in the extended von Neumann neighborhood, Figure 1, when the intermediate j cell is vacant. J is a positive real number. When $J = 1$, it indicates that the particle, i, has the same probability of movement toward or away as for the case when k is empty. When $J > 1$, it indicates that i has a greater probability of movement toward an occupied cell k than when k is empty. When $J < 1$, it indicates that i has a lower probability of such movement. Our studies on water[16] have demonstrated a relationship between P_B and J expressed as

$$\text{Log } J(\text{W}) = -1.50 P_B(\text{W}) + 0.60 \qquad (6)$$

which we adopt in this study.

A Model of Molecular Bonds. In this study we simulated each bond type as a cell with a particular state, derived from the classifications in Table 1. The rules for joining and breaking of each type of cell were derived from the encounter possibilities, $(A_{ij})(A_{kl})$, among two molecules as shown in Table 2. To scale these possibilities to the trajectory rules for our cellular automata model we have adopted the

BOND INTERACTIONS AMONG MOLECULES

J. Chem. Inf. Comput. Sci., Vol. 40, No. 5, 2000 **1287**

Table 3. Breaking and Joining Rules for All Bond Types

interacting pair	P_B	J
aa	0.20	2.00
ab	0.30	1.40
ac	0.26	1.64
ad	0.36	1.16
ae	0.41	0.96
af	0.30	1.40
ag	0.40	1.00
ah	0.47	0.80
ai	0.50	0.72
bb	0.40	1.00
bc	0.36	1.16
bd	0.47	0.80
be	0.51	0.68
bf	0.40	1.00
bg	0.50	0.72
bh	0.55	0.60
bi	0.59	0.52
cc	0.32	1.33
cd	0.41	0.96
ce	0.48	0.76
cf	0.36	1.16
cg	0.47	0.80
ch	0.51	0.68
ci	0.57	0.56
dd	0.51	0.68
de	0.59	0.52
df	0.47	0.80
dg	0.57	0.56
dh	0.61	0.48
di	0.67	0.40
ee	0.64	0.44
ef	0.51	0.68
eg	0.61	0.48
eh	0.67	0.40
ei	0.73	0.32
ff	0.40	1.00
fg	0.50	0.72
fh	0.55	0.60
fi	0.58	0.52
gg	0.59	0.52
gh	0.67	0.40
gi	0.70	0.36
hh	0.73	0.32
hi	0.77	0.28
ii	0.81	0.24

relationship

$$J \text{ for } (A_{ij})(A_{kl}) = 4(A_{ij})(A_{kl}) \qquad (7)$$

The breaking probabilities, $P_B(A_{ij})(A_{kl})$, were derived from eq 6. A complete set of joining and breaking rules for the nine bond types is shown in Table 3 using the letter codes from Table 1 to identify the bond encounter pairs. Each bond in a molecule was represented by a cell. One hundred molecules were modeled in a grid of 3025 cells. 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 elapsed time ensured that an equilibrium had been reached. This process was repeated for 25 runs, and the count of joined cells was averaged from these 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 β values are recorded in Table 4.

Table 4. Boiling Points of Alkanes vs β Values

alkane	β	$^1\chi$	bp	bp(calcd)[a]	diff[a]
pentane	219.00	2.41	36.10	37.17	−1.07
2-Me-butane	202.00	2.27	27.90	30.10	−2.20
2,2-di-Me-propane	160.00	2.00	9.50	11.66	−2.16
cyclobutane	160.00	2.00	13.00	11.66	1.34
hexane	303.00	2.91	68.70	68.69	0.01
2-Me-pentane	281.00	2.77	60.10	60.98	−0.88
3-Me-pentane	291.00	2.81	63.50	64.53	−1.03
2,3-di-Me-butane	258.00	2.64	57.90	52.50	5.40
2,2-di-Me-butane	251.00	2.56	49.70	49.84	−0.14
cyclopentane	243.00	2.50	49.30	46.75	2.55
2,4-di-Me-pentane	355.00	3.13	80.50	85.37	−4.87
2,2,3-tri-Me-butane	330.00	2.94	81.00	77.62	3.38
heptane	404.00	3.41	99.30	99.11	0.19
2-Me-hexane	372.00	3.27	90.00	90.36	−0.36
3-Me-hexane	386.00	3.31	93.40	94.29	−0.89
2,3-di-Me-pentane	367.00	3.18	89.70	88.91	0.79
2,2-di-Me-pentane	346.00	3.06	79.20	82.64	−3.44
3,3-di-Me-pentane	356.00	3.12	86.00	85.67	0.33
3-Et-pentane	396.00	3.35	93.40	97.00	−3.60
cyclohexane	332.00	3.00	80.70	78.26	2.44
octane	506.00	3.91	125.70	121.57	4.13
3-Me-heptane	489.00	3.81	119.00	118.41	0.59
3-Et-hexane	496.00	3.85	118.60	119.74	−1.14
3-Me-3-Et-pentane	480.00	3.68	118.30	116.64	1.66
3,4-di-Me-hexane	483.00	3.72	117.70	117.23	0.47
2-Me-heptane	482.00	3.77	117.70	117.04	0.66
4-Me-heptane	493.00	3.81	117.70	119.17	−1.47
2-Me, 3-Et pentane	483.00	3.72	115.70	117.23	−1.53
2,3-di-Me-hexane	467.00	3.68	115.60	113.97	1.63
2,3,3-tri-Me-pentane	444.00	3.50	114.80	108.91	5.89
2,3,4-tri-Me-pentane	445.00	3.55	113.50	109.14	4.36
3,3-di-Me-hexane	461.00	3.62	112.00	112.69	−0.69
2,2,3-tri-Me-pentane	444.00	3.48	109.90	108.91	0.99
2,4-di-Me-hexane	468.00	3.66	109.50	114.18	−4.68
2,5-di-Me-hexane	461.00	3.63	109.10	112.69	−3.59
2,2-di-Me-hexane	450.00	3.56	106.90	110.27	−3.37
2,2,3,3-tetra-Me-butane	409.00	3.25	106.30	100.41	5.89
2,2,4-tri-Me-pentane	427.00	3.42	99.30	104.90	−5.60

[a] Derived from eq 8.

RESULTS

The results of the calculation of the β values for the 38 alkanes are shown in column 2 of Table 4. The significance of this value lies in a possible relationship with a physical property that is highly dependent upon intermolecular interactions. One such property is the boiling point, entered in column 4 of Table 4. For comparison, the molecular connectivity $^1\chi$ values for the alkanes are recorded in column 3 of the same table. The β values, expressed in a quadratic equation, relate to the boiling point with the statistics shown in eq 8. In comparison the $^1\chi$ indices for the same molecules

$$\text{bp (°C)} = 0.584\beta - 0.0004\beta^2 - 71.517 \qquad (8)$$

$$r^2 = 0.991, \quad s = 2.996, n = 38, \quad F = 2027$$

produced a quadratic equation:

$$\text{bp (°C)} = 115.50\,^1\chi - 9.33(^1\chi)^2 - 184.18 \qquad (9)$$

$$r^2 = 0.984, \quad s = 4.158, \quad n = 38, \quad F = 1044$$

DISCUSSION

It must be made clear 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 revealing some information about emergent properties and the role of the ingredients contributing to the whole. Each bond type is modeled as a cell with a set of rules independent of the context of that bond in a molecule. No doubt this context is important in defining the interactions of each bond type. This deficiency may possibly be addressed by the use of cells representing higher order molecular connectivity fragments, a refinement under study. In the current study, we considered just the bonds in alkanes, endowed with numerical values reflecting their accessibility to other bonds in other molecules. Our model is an alternative to the static model of the bimolecular encounter that we have offered to explain the molecular connectivity indices.[1] It is a dynamic model that, in a limited way, simulates the conditions of a molecule in its milieu. Objectives of the study were to reinforce the central idea in our bimolecular interpretation, to confirm the reasonableness of our relative bond accessibility values, and to extend the concept of the importance of bonds and their attributes in relating structure to properties.[17] We have contributed to these objectives in this study.

The average count of the cell encounters in unit time (iterations), encoded in the β value, is very closely related to the boiling point of the 38 alkanes. The standard deviation of only 3.00 deg is better than any one-variable analysis we have found reported. It is better than the correlation with the molecular connectivity index, even though both $^1\chi$ and β are derived from identical accessibility values. It appears that the cellular automata dynamics improves the simulation of bond encounters compared to the static count of all possible bimolecular encounters. The quality of the relationships revealed here support the model that we have developed centering on the accessibility of bonds in alkanes to intermolecular encounters. The molecular connectivity indices are certainly of value in estimating these accessibilities.

The use of cellular automata to model the dynamics experienced by fragments of a molecule as though they were isolated entities is a new idea that may have some merit in further studies on ensembles of molecules. These interesting results suggest some novel possibilities.

REFERENCES AND NOTES

(1) Kier, L.; Hall, L. Intermolecular Accessibility: The Meaning of Molecular Connectivity. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 792−795.
(2) Randic, M. On Characterization of Molecular Branching. *J. Am. Chem. Soc.* **1975**, *97*, 6606−6615.
(3) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*; Academic Press: New York, 1976.
(4) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure Activity Analysis*; John Wiley & Sons: New York, 1986.
(5) Kier, L. B.; Hall, L. H., *Molecular Structure Description: The Electrotopological State*; Academic Press: San Diego, CA, 1999.
(6) Von Neumann, J. In *Theory of Self-Reproducing Automata*; Burks, A., Ed.; University of Illinois: Urbana, IL, 1966.
(7) Wolfram, S. *Theory and Applications of Cellular Automata*; World Scientific Publishing Co.: Singapore, 1986.
(8) Ermentrout, F. B.; Edelstein-Keshet, L. Cellular Automata Approaches to Biological Modeling. *J. Theor. Biol.* **1993**, *160,* 97−133.
(9) Chopard, B.; Droz, M. *Cellular Automata Modeling of Physical Systems;* Cambridge University Press: Cambridge, England, 1998.
(10) Kier, L. B.; Cheng, C.-K.; A Cellular Automata Model of Water. *J. Chem. Inf. Comput. Sci.* **1994** *34,* 647−652.
(11) Kier, L. B.; Cheng, C.-K. A Cellular Automata Model of an Aqueous Solution, *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1334−1337.
(12) Cheng, C.-K.; Kier, L. B. A Cellular Automata Model of Oil-Water Partitioning. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 1054−1059.
(13) Kier, L. B.; Cheng, C.-K.; Testa, B.; Carrupt, P.-A. A Cellular Automata Model of the Hydrophobic Effect. *Pharm. Res.* **1995**, *12,* 615−620.
(14) Kier, L. B.; Cheng, C.-K.; Testa, B.; Carrupt, P.-A. A Cellular Automata Model of Diffusion in Aqueous Systems. *J. Pharm. Sci.* **1997**, *86*, 774−778.
(15) Kier, L. B.; Cheng, C.-K.; Tute, M.; Seybold, P. A Cellular Automata Model of Acid Dissociation. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 271−275.
(16) Kier, L.; Cheng, C.-K.; Testa, B. Cellular Automata Models of Biochemical Phenomena. *Fut. Gen. Comput. Sys.* **1999**, *16*, 273−289.
(17) Randic, M. On Characterization of Chemical Structure. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 672−687.

CI000326R