

Is the Dominance of Even Carbon Atom Molecules Odd?

T. P. Radhakrishnan

School of Chemistry, University of Hyderabad, Hyderabad 500 046, India

Received May 18, 1999

The unexpected higher frequency of occurrence of even carbon atom molecules over odd carbon atom ones revealed through previous database searches is shown to be a consequence of the predominantly bimolecular nature of bond formation processes and the fact that combinations of odd and odd as well as even and even atom molecules result in even atom products whereas odd and even atom molecule combinations alone lead to odd atom products.

INTRODUCTION

Topological statistics gathered on structural files by Petitjean and Dubois¹ as well as the search of several databases by Sarma et al.² have revealed that the molecules containing carbon atoms form a smooth distribution with respect to the number of carbon atoms present in them; the distribution has a peak in the range of 12–16 carbons. Interestingly in both cases it was observed that there is a higher frequency of occurrence of molecules containing an even number of C atoms compared to their near neighbors with an odd number of C atoms; i.e., for most of the even numbers p in the ranges considered, the number of molecules with p C atoms was found to be larger than the number of molecules with $(p - 1)$ or $(p + 1)$ C atoms! This bias was particularly conspicuous at larger C numbers where the number of molecules is small. On the basis of these observations it was speculated that there may be some underlying parity preference between odd and even carbon atom molecules. The statistics presented by Petitjean and Dubois showed that similar odd/even bias exists for H atom distribution as well. They proposed a graph theoretical explanation for their observation based on the node connectivities and the fact that the number of odd-connected nodes in a graph is always even.¹ This explanation can be shown to be invalid since H atoms, as well as C atoms in many instances, possess odd connectivity and hence only the total number of these atoms (and other odd-connected atoms) need necessarily be even. Possible explanations for the unexpected and counterintuitive odd/even bias in the C atom distribution have been proposed by Stanley³ and Kaser.⁴ However, these have been shown to be inadequate by Sarma et al.⁵ The origin of the observed parity preference has not been explained satisfactorily so far.

Given that the numbers of molecules of various sizes (odd as well as even) increase and decrease together forming a relatively smooth distribution, the bias toward even C atom molecules is a consequence of the fact that the total number of even C atom molecules is larger than the total number of odd C atom molecules. Therefore, it suffices to seek the origin of the latter to explain the former observation. This parity preference appears to contradict simple statistical and intuitive considerations which would predict that the number of odd and even atom molecules should be similar; however,

it may be noted that this need be established only when the sampling size is sufficiently large. Any explanation of the origin of this disparity based on chemical arguments related to specific synthetic routes cannot be general since the sample under consideration includes a wide variety of molecules covering a wide range of sizes and odd/even bias exists even for H atom distribution.¹ Concepts such as isoprenoid generation of molecules and dimerization³ explain the odd–even disparity only in limited regions of the data and for specific systems. Therefore we have sought an explanation for this interesting observation through a nonchemical argument. Consider the following facts: (i) bond formation processes are predominantly bimolecular in nature; (ii) the addition of two odd or two even atom molecules gives rise to an even atom molecule whereas only the addition of an odd and an even atom molecule results in an odd atom molecule. We show in this paper that these simple observations are highly suggestive of a process that leads to the odd/even bias which is apparent in the large, but as shown later not sufficiently large, sample under consideration. A similar argument can be applied to the case of H atom distribution¹ as well.

RESULTS AND DISCUSSION

We define generation 1 as having e_1 even C atom molecules and o_1 odd C atom molecules. Generation 2 (e_2 , o_2) consists of generation 1 and the products of bimolecular additions among the $(e_1 + o_1)$ molecules from generation 1. Generation 3 is made up of generation 1 and the products from the $(e_2 + o_2)$ molecules and so on. Since the nature of the molecules is completely unspecified, the formation of isomers cannot be taken into account fully. The isomerism resulting from the combination of different fragments to form molecules of a given size can be considered up to an extent within a general scheme and is implicit in our generation-building protocol; in fact this is central to the development of the bias observed toward even C systems. Isomerisms such as geometric isomerism or stereoisomerism cannot be invoked without specifying the nature of the molecules and hence is not incorporated in our general approach. An example involving the lower alkanes is presented later to illustrate our generation procedure and the level of isomerism that can be included. In the generation scheme we consider,

if one starts with small enough odd and even atom precursors, any molecule can be generated using bond formations without resorting to bond-breaking processes. On the basis of the definitions above, the following arguments illustrate that $e_i > o_i$ for all $i > 1$ irrespective of the starting values, e_1 and o_1 (except in three special cases; even in those cases $e_i > o_i$ for $i > 3$).

Bimolecular additions among the molecules in generation 1 can provide up to $[e_1(e_1 + 1)/2 + o_1(o_1 + 1)/2]$ even atom molecules through even + even and odd + odd combinations and $[e_1o_1]$ odd atom molecules through even + odd combinations. Therefore

$$e_2 = \left\lfloor \frac{e_1(e_1 + 1)}{2} + \frac{o_1(o_1 + 1)}{2} \right\rfloor + e_1 \quad \text{and}$$

$$o_2 = e_1 o_1 + o_1$$

so that

$$\begin{aligned} e_2 - o_2 &= \left[\frac{e_1(e_1 + 1)}{2} + \frac{o_1(o_1 + 1)}{2} \right] - e_1 o_1 + (e_1 - o_1) \\ &= \frac{(e_1 - o_1)^2 + 2(e_1 - o_1) + (e_1 + o_1)}{2} \end{aligned} \quad (1)$$

Note that e_1 and o_1 are positive integers and $(e_1 + o_1) \geq 1$. Equation 1 shows that $(e_2 - o_2) \geq 0$ in all cases; it is zero only when $e_1 = 0, o_1 = 1$. The following may also be easily verified.

$$(e_2 - o_2) > (e_1 - o_1) \quad \text{in all cases}$$

$$(e_2 - o_2) < |e_1 - o_1| \quad \text{when } e_1 = 0, o_1 = 1 \text{ or } 2$$

$$(e_2 - o_2) = |e_1 - o_1|$$

when $e_1 = 0, o_1 = 3$ and $e_1 = 1, o_1 = 2$ or 3

$$(e_2 - o_2) > |e_1 - o_1| \quad \text{in all other cases}$$

Equation 1 may be generalized for any $i > 1$

$$(e_i - o_i) = \frac{(e_{i-1} - o_{i-1})^2 + 2(e_1 - o_1) + (e_{i-1} + o_{i-1})}{2} \quad (2)$$

Equation 2 shows that $(e_i - o_i) > (e_1 - o_1)$ for all $i > 1$. Since $(e_2 - o_2) \geq |e_1 - o_1|$ in most cases as shown above and $e_2 > e_1$ and $o_2 \geq o_1$, equation 2 shows that $(e_3 - o_3) > (e_2 - o_2)$ except for the two special cases of $e_1 = 0, o_1 = 1$ or 2; in the latter two cases, $(e_3 - o_3) = (e_2 - o_2)$. Since e_i and o_i keep increasing, the exceptions to $(e_i - o_i) > (e_{i-1} - o_{i-1})$ disappear after $i = 3$ so that we conclude

$$(e_i - o_i) > (e_{i-1} - o_{i-1})$$

for all $i > 3$ and therefore

$$(e_i - o_i) > 0$$

for all $i > 1$, except in the special cases of $i = 2$ when $e_1 = 0$ and $o_1 = 1$ and for $i = 3$ when $e_1 = 0$ and $o_1 = 1$ and 2.

The generation scheme may be illustrated using the following example involving simple acyclic alkanes. Consider generation 1 consisting of a molecule with one carbon

Table 1. (a) C_n molecules generated (up to Generation 5) Starting with a C_1 Molecule Using Bimolecular Additions^a and (b) Graphs of the C_n systems, up to Generation 4^b

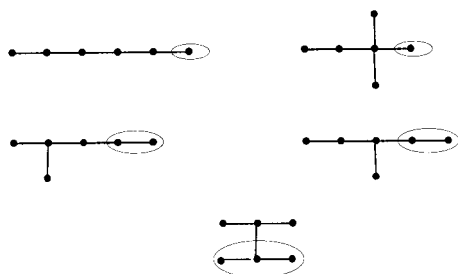
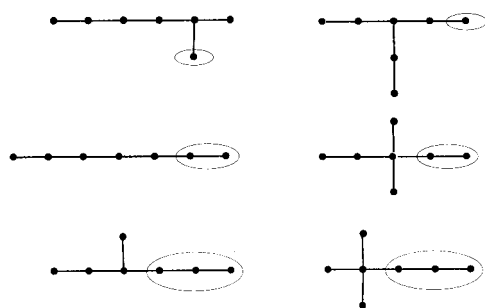
generation	even carbon molecules	odd carbon molecules
(a)		
1	—	C ₁
2	C ₂	C ₁
3	C ₂ C ₄	C ₁ C ₃
4	C ₂ C ₄ (2) C ₆ (2) C ₈ (1)	C ₁ C ₃ C ₅ (2) C ₇ (1)
5	C ₂ C ₄ (2) C ₆ (5) C ₈ (8) C ₁₀ (9) C ₁₂ (7) C ₁₄ (3) C ₁₆ (1)	C ₁ C ₃ C ₅ (3) C ₇ (6) C ₉ (8) C ₁₁ (7) C ₁₃ (4) C ₁₅ (1)
(b)		
1	-	•
2		•
3	 	•
4	 	•

^a Numbers in parentheses are numbers of isomers formed. ^b The subgraphs added in each generation are demarcated using an ellipse. The graphs in generation 4 show only one possible choice.

atom alone denoted as C_1 . Therefore $e_1 = 0$ and $o_1 = 1$. Generation 2 would consist of C_2 and C_1 so that $e_2 = 1$ and $o_2 = 1$. Table 1 lists the various species that are generated up to generation 5. The dominance of even C molecules is evident from generation 4. Table 1 depicts the corresponding molecular graphs up to generation 4; it may be noted that those shown in generation 4 are not unique. Similarly some possible cases of C_6 and C_7 systems formed in generation 5 are shown in Figure 1. The molecules in the latter generations clearly illustrate the level of isomerism that is implicit in our protocol. It may be noted that the number of isomers formed in the various cases are within the well-known alkane isomer count.⁶ Formation of more isomers may be visualized as occurring through processes such as cyclizations and rearrangements of those generated in our scheme. Graph theoretical counting procedures⁷ show that the number of isomers increase steadily with the number of carbon atoms with no odd/even preference. Therefore, the second level of isomer formation from the various even and odd molecules generated in our scheme will be evenly matched. For instance, no bias is expected between the number of isomers

Table 2. Numerically Evaluated e_i , o_i , $(e_i - o_i)$, and d_i for Six Generations Starting with (a) $e_1 = 0$, $o_1 = 1$; (b) $e_1 = 3$, $o_1 = 1$; (c) $e_1 = 2$, $o_1 = 2$, and (d) $e_1 = 0$, $o_1 = 4$

generation, i	e_i	o_i	$(e_i - o_i)$	d_i
(a)				
1	0	1	-1	—
2	1	1	0	0
3	2	2	0	0
4	6	5	1	0.090 909 1
5	36	31	5	0.074 626 9
6	1 162	1 117	45	0.019 745 5
(b)				
1	3	1	2	—
2	10	4	6	0.428 571 4
3	68	41	27	0.247 706 4
4	3 210	2 789	421	0.070 178 4
5	9 044 313	8 952 691	91 622	0.005 091 0
6	80 975 145 890 230	80 970 939 596 284	4 206 293 946	0.000 026 0
(c)				
1	2	2	0	—
2	8	6	2	0.142 857 1
3	59	50	9	0.082 568 8
4	3 047	2 952	95	0.015 836 0
5	9 002 258	8 994 746	7 512	0.000 417 4
6	80 973 061 350 044	80 973 024 136 470	37 213 574	0.000 000 2
(d)				
1	0	4	-4	—
2	10	4	6	0.428 571 4
3	65	44	21	0.192 660 6
4	3 135	2 864	271	0.045 174 2
5	9 018 360	8 978 644	39 716	0.002 206 8
6	80 973 441 582 670	80 972 643 903 844	797 678 826	0.000 004 9

C₆ molecular graphs**C₇ molecular graphs****Figure 1.** Graphs of one possible set each of C₆ and C₇ systems formed in generation 5. The subgraphs added after generation 4 in each case are demarcated using an ellipse.

that may be generated from C₄, C₆, and C₈ systems together and the number generated from C₃, C₅, and C₇ systems together in generation 4. Thus the e_i and o_i in our scheme represent one set of molecules (though not a unique set) generated at the i th level from bimolecular additions which induces the odd/even bias in the molecular count at that level.

We have examined the growth of e_i and o_i with i for different initial values e_1 and o_1 . The numerically evaluated e_i , o_i , and $(e_i - o_i)$ for $i = 1-6$ for four different sets of e_1 and o_1 are collected in Table 2, including the case illustrated in Table 1. It is seen that irrespective of whether $e_1 > o_1$, $e_1 = o_1$, or $e_1 < o_1$, the number of even carbon atom molecules is always larger than the number of odd carbon atom molecules (except in some of the early generations); the difference increases with the generation. Table 2 also provides the relative difference between odd and even atom molecules, d_i , defined as $(e_i - o_i)/(e_i + o_i)$. d_i in each case shows a steady decrease as e_i and o_i increase; the difference becomes relatively very small when the total number of molecules becomes quite large. One may conclude that it would become inconspicuous at large enough sample size as would be expected on the basis of statistical arguments.

CONCLUSION

We have shown that the dominance of even C atom molecules observed in database searches is a consequence of the dominance of bimolecular processes and the fact that addition of odd and odd as well as even and even atom molecules lead to even atom products whereas odd and even atom molecule combinations alone lead to odd atom products. The number of molecules currently known and considered in the database search² is of the order of 7 million; apparently this number is small enough and the bias in favor of even C atom molecules is clearly discernible. The trend of the relative difference we have illustrated is consistent with the relatively weaker parity preference observed at the peak of the distribution and the stronger preference observed at large molecular sizes where the numbers of molecules are small.

ACKNOWLEDGMENT

Financial support from the Department of Science and Technology (Swarnajayanti Fellowship) during the course of this work is acknowledged.

Supporting Information Available: Values of e_i (number of even C atom molecules), o_i (number of odd C atom molecules), the difference, $(e_i - o_i)$, and the relative difference $(e_i - o_i)/(e_i + o_i)$ are tabulated for starting values of $e_1 = 0, \dots, 6$ and $o_1 = 0, \dots, 6$ and $i = 1, \dots, 6$; the relative difference is provided for $i > 1$. This material is available free of charge via the Internet at <http://pubs.acs.org>.

REFERENCES AND NOTES

- (1) Petitjean, M.; Dubois, J. E. *J. Chem. Inf. Comput. Sci.* **1990**, 30, 332. Petitjean, M.; Dubois, J. E. *Collect. Czech. Chem. Commun.* **1990**, 55, 1404.
- (2) Sarma, J. A. R. P.; Nangia, A.; Desiraju, G. R.; Zass, E.; Dunitz, J. D. *Nature* **1996**, 384, 320.
- (3) Stanley, P. G. *Nature* **1997**, 385, 782.
- (4) Kaser, M. *Nature* **1997**, 385, 782.
- (5) Sarma, J. A. R. P.; Nangia, A.; Desiraju, G. R.; Zass, E.; Dunitz, J. D. *Nature* **1997**, 387, 464.
- (6) Blair, C. M.; Henze, H. R. *J. Am. Chem. Soc.* **1932**, 54, 1538.
- (7) (a) Harary, F.; Palmer, E. M.; Robinson, R. W.; Read, R. C. In *Chemical Applications of Graph Theory*; Balaban, A. T., Ed.; Academic Press: London, 1976; p 11. (b) Read, R. C. In *Chemical Applications of Graph Theory*; Balaban, A. T., Ed.; Academic Press: London, 1976; p 25. (c) Balaban, A. T. In *Chemical Applications of Graph Theory*; Balaban, A. T., Ed.; Academic Press: London, 1976; p 63.

CI990046F