Computer Enumeration of Walks on Directed Graphs

K. Balasubramanian*

Department of Chemistry, Arizona State University, Tempe, Arizona 85287-1604

Received 17 May 1990; accepted 19 July 1990

A vectorized computer code is developed for the enumeration of walks through the matrix power method for directed graphs. Application of this code to several graphs is considered. It is shown that the coefficients in the generating functions for signed graphs are much smaller in magnitude. It is shown that self-avoiding walks on some graphs can be enumerated as a linear combination of walk GFs of directed paths and rooted-directed paths.

INTRODUCTION

The enumeration of walks on graphs and lattices of chemical interest has numerous applications in many areas of chemistry.[1–11] Random walk models are useful in the treatment of diffusion, conformations of flexible polymers, among other applications. Walks on graphs are also useful in chemical coding, characterization of graphs, applications of extended Hückel methods to solids through moment generating functions,[14] etc. Randić and co-workers,[2,3] as well as the present author[11] have formulated computational techniques for enumerating walks and self-returning walks for ordinary graphs. However, walks on directed graphs, weighted graphs, signed graphs, and graphs containing complex edge weights have not been explored at all. We develop here a vectorized computer code for the enumeration of walks and spectral moments of directed and unsymmetrical graphs, in general. The second section describes the computational techniques and codes while the third section comprises results and discussions.

COMPUTATION OF WALKS

A walk on a graph is defined as a sequence of edges that a walker can traverse continuously starting and ending on any vertex. The possibility of repetition of edges is permitted in a random walk. A self-returning walk is a random walk in which one starts and ends the walk in the same vertex. A self-avoiding walk is a walk in which no vertex appears more than once. A walk of length k is a walk consisting of k edges of a graph or lattice in a continuous manner. This is also known as the distance of a walk. Similarly one can define self-avoiding and self-returning walks of length k. Suppose N_k is the number of walks of length k, the function W, shown below, is known as the generating function for the walk

$$W = \sum_{k=0}^{\infty} N_k x^k,$$

where x is a dummy variable and x^k represents walks of length k. Similarly one defines the generating function for self-avoiding walk (SAW) and self-returning walk (SRW) as

$$\text{SAW} = \sum_{k=0}^{n-1} L_k x^k,$$

$$\text{SRW} = \sum_{k=0}^{\infty} M_k x^k.$$

Note that N_o in W is the number of vertices in the graph and that SAW does not have a nonvanishing term beyond x^{n-1}, since one cannot have a self-avoiding walk of length more than $n - 1$ on a graph containing n vertices. Similarly M_o is zero if the graph in question does not have loops. Hence, we omit the constant term in W, SRW, and SAW in this manuscript.

The adjacency matrix A of an ordinary graph is defined as

$$A_{ij} = \begin{cases} 0 & \text{if } i = j \\ 1 & \text{if } i \neq j \end{cases}$$

and i and j are connected by an edge

In general different powers of the adjacency matrix enumerate walks of different lengths.[2] The ijth element of A^k of $(A^k)_{ij}$ enumerates the total number of walks of length k between the vertices i and j. Consequently the trace of A^k or trA^k is the total number of self-returning walks also known as spectral moments. The sum of all the elements of A^k is N_k. In symbols,

$$N_k = \sum_{ij} (A^k)_{ij},$$

$$M_k = \sum_{i} (A^k)_{ii} = \text{tr}A^k,$$

For symmetrical graphs it can be shown that[12]

$$W = \frac{P_{\bar{G}}(1/x)}{xP_G(1/x)},$$

where $P_{\bar{G}}$ is the characteristic polynomial of $A - J$, where J is a matrix in which all entries are unities.

The author used his code to compute characteristic polynomials[13] to construct W of graphs before.[11] The coefficients M_ks in SRW (self-returning walk generating function) are simply S_ks, the spectral moments of graphs. S_k is defined as follows

$$S_k = \text{tr}A^k = \sum_{i} (A^k)_{ii}.$$

The spectral moments S_k's and the coefficients in the characteristic polynomials are related by the following expressions (see, for example, reference (9))

$$S_1 = 0, \qquad -mC_m = \sum_{k=1}^{m} S_k C_{m-k}.$$

Thus the spectral moments S_ks (same as M_k) can be recursively obtained using the above relation once the coefficients C_k's in the characteristic polynomial of the graph are determined. All analytical expressions discussed above are valid only for ordinary graphs.

A signed graph is defined as a bidirected graph for which a signed adjacency matrix is shown below:

$$A_{ij}^{(s)} = \begin{cases} 0 & \text{if } i = j \\ 1 & \text{if } i \neq j,\ i \text{ and } j \text{ are connected and } i > j. \\ -1 & \text{if } i \neq j,\ i \text{ and } j \text{ are connected and } i < j. \end{cases}$$

Note that the adjacency matrix $A_{ij}^{(s)}$ is not symmetrical. The adjacency matrix of a weighted directed graph can be defined as

$$A_{ij}^{w} = \begin{cases} 0 & \text{if } i = j. \\ w_{ij} & i \neq j, \text{ for an edge from } i \text{ to } j. \end{cases}$$

The possibility of nonzero diagonal elements could also be included in the above definition through loops.

The powers of adjacency matrices of weighted, directed, and bidirected graphs (different weights for different directions) do not enumerate the "number of walks" in a strict sense. This is because the ij matrix element A_{ij}^k is given by

$$A_{ij}^k = \sum a_{ii_1} a_{i_1 i_2} a_{i_2 i_3} \ldots a_{i_{k-1} j},$$

where the sum is over all such terms starting with the vertex i and terminating at vertex j. Note that if all matrix elements are 0 and 1 for ordinary graphs the above sum gives the number of walks from i to j of the length k. For a signed graph since some of the matrix product shown above will be negative while the others will be positive and, consequently, A_{ij}^k gives the net effect of all possible walks from i to j of length k. There are some advantages to the evaluation of such generating functions since the coefficients in general tend to be smaller in magnitude. For signed graphs the coefficients of x^k vanish if k is odd since the sum of contributions of all walks cancel out. Note that the walks of signed and directed graphs are dependent on labeling of vertices, in general.

There are also other advantages in using weighted graphs. For some trees and cyclic graphs a combination of properly weighted directed graphs could be used to enumerate the number of self-avoiding walks as we show here.

In the next section, we shall use the term walk-generating function (WGF) and self-returning walk generating function (SRWGF) to simply refer to

$$\text{WGF} = \sum_{k=1}^{n} N_k x^k,$$

$$N_k = \sum_{ij} \sum a_{ii_1} a_{i_1 i_2} \ldots a_{i_{k-1} j},$$

$$\text{SRWGF} = \sum_{k=1}^{n} M_k x^k,$$

$$M_k = \sum_{i} \sum a_{ii_1} \ldots a_{i_{k-1} i}.$$

Thus the coefficients in WGF and SRWGF do not necessarily enumerate the number of walks in the sense they are interpreted for nondirected ordinary graphs. These coefficients measure the net effect of such walks for signed and weighted graphs.

The computation of powers of matrices is efficiently accomplished through the use of vector processors. We developed a vectorized Fortran '77 code and was compiled on an IBM 3090/300 with the vector option and an optimization level = 3. The DO loops were properly organized to achieve maximum vectorization. All matrices were stored as two-dimensional arrays. The original adjacency matrix, the $A^{(k-1)}$ matrix and the A^k matrix were saved in each iteration $k = 1, 2, \ldots$ etc. The codes thus developed were tested on several graphs. For all the graphs considered here, this code took only a few seconds.

RESULTS AND DISCUSSION

Table I compares the results obtained for a series of ordinary and signed graphs containing eight vertices shown in Figure 1. Note that the spectral moments (the number of self-returning walks) differ only in

*Camille and Henry Dreyfus Teacher-Scholar.

Journal of Computational Chemistry, Vol. 12, No. 1, 106–112 (1991)

CCC 0192-8651/91/010106-07$04.00

Table I. WGF and SRWGF for both regular and signed graphs containing eight vertices.[a]

Graph (Fig. 1)	WGF		SRWGF	
	Ordinary graph	Signed graph	Ordinary graph	Signed graph
I	$14x + 26x^2 + 48x^3 + 90x^4 + 168x^5 + 316x^6 + 592x^7 + 1114x^8$	$-2x^2 + 2x^4 + 4x^6 + 10x^8$	$14x^2 + 38x^4 + 116x^6 + 374x^8$	$-14x^2 + 38x^4 - 116x^6 + 374x^8$
II	$14x + 28x^2 + 52x^3 + 140x^4 + 196x^5 + 392x^6 + 742x^7 + 1484x^8$	$-4x^2 + 8x^4 - 24x^6 + 84x^8$	$14x^2 + 42x^4 + 140x^6 + 490x^8$	$-14x^2 + 42x^4 - 140x^6 + 490x^8$
III	$14x + 30x^2 + 60x^3 + 126x^4 + 258x^5 + 538x^6 + 1110x^7 + 2308x^8$	$-6x^2 + 10x^4 - 26x^6 + 88x^8$	$14x^2 + 46x^4 + 176x^6 + 718x^8$	$-14x^2 + 46x^4 - 176x^6 + 718x^8$
IV	$14x + 30x^2 + 62x^3 + 130x^4 + 272x^5 + 570x^6 + 1194x^7 + 2502x^8$	$-6x^2 + 10x^4 - 30x^6 + 114x^8$	$14x^2 + 46x^4 + 182x^6 + 766x^8$	$-14x^2 + 46x^4 - 182x^6 + 766x^8$
V	$14x + 32x^2 + 60x^3 + 138x^4 + 264x^5 + 606x^6 + 1170x^7 + 2682x^8$	$-8x^2 + 26x^4 - 102x^6 + 426x^8$	$14x^2 + 50x^4 + 200x^6 + 842x^8$	$-14x^2 + 50x^4 - 200x^6 + 842x^8$
VI	$14x + 34x^2 + 70x^3 + 162x^4 + 342x^5 + 784x^6 + 1666x^7 + 3808x^8$	$-10x^2 + 30x^4 - 116x^6 + 512x^8$	$14x^2 + 54x^4 + 242x^6 + 1142x^8$	$-14x^2 + 54x^4 - 242x^6 + 242x^6 + 1142x^8$
VII	$16x + 32x^2 + 64x^3 + 128x^4 + 256x^5 + 512x^6 + 1024x^7 + 2048x^8$	$-8x^2 + 16x^4 - 48x^6 + 160x^8$	$16x^2 + 48x^4 + 160x^6 + 576x^8$	$-16x^2 + 48x^4 - 160x^6 + 544x^8$

[a]For all signed graphs vertices were labeled sequentially for the largest paths and then branching vertices in the order of increasing vertex numbers they are attached to.

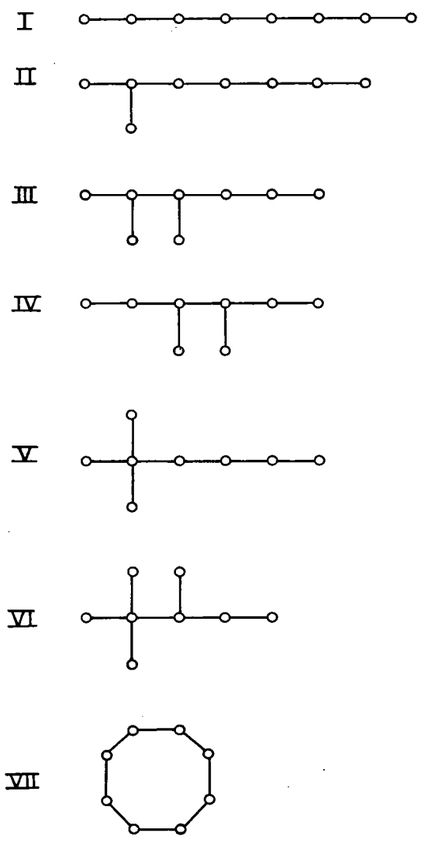

Figure 1. Graphs I–VII containing eight vertices (see Table I for walk GFs).

signs for ordinary and signed graphs in Table I. However, the actual walk generating functions differ significantly for ordinary and signed graphs. Note that a walk from a vertex i to j $(i > j)$ is considered positive while a walk from i to j $(i < j)$ is negative because of weights $+1$ and -1, for these two edges. Consequently, the coefficients of all terms with odd powers are zero as there are equal and opposite contributions for every term since $A^k_{ij} = -A^k_{ji}$ if k is odd.

Consider a directed chain of length 8 (L_8). Label the vertices 1–8 sequentially from left to right. Let the edges be directed from i to j if and only if $j > i$. This means, for example, the matrix element a_{12} is 1 but a_{21} is 0. The resulting adjacency matrix is unsymmetrical as expected. The powers of this matrix A enumerate exactly half of the self-avoiding walks (SAW) and in general this is the case for any L_n graph. For the directed L_8 graph thus obtained

$$2W = \text{SAW} = 14x + 12x^2 + 10x^3 + 8x^4 + 6x^5 + 4x^6 + 2x^7$$

The enumeration of self-avoiding walks can in general be a difficult problem. Randić et al.[3] have formulated an algorithm based on growing paths from nonequivalent vertices of graphs. This algorithm works reasonably well for graphs which do not contain several rings. For lattice graphs and complex polycyclic graphs this algorithm can experience a combinatorial explosion. There is in general no polynomial algorithm to compute self-avoiding walks. It is hoped that the present study based on matrix powers of directed graphs would lead to finding SAWs as combination of generating functions of directed graphs.

Similarly SAW generating functions of cyclic graphs (C_n) can be obtained using an unidirectional graph with edges from 1 to 2, 2 to 3, 3 to 4, etc. up to n to 1. If there is no edge from a vertex i to j then a_{ij} is set to zero. The WGF-SRWGF of such a directed C_8 gives SAW as shown below.

$$\text{SAW} = 2(\text{WGF-SRWGF}) = 16x + 16x^2 + 16x^3 + 16x^4 + 16x^5 + 16x^6 + 16x^7$$

Figure 2 shows a directed square lattice graph. As seen from Figure 2, all edges are unidirectional. This means if there is an edge from i to j the matrix element $a_{ij} = 1$ and the element $a_{ji} = 0$. First 16 coefficients in the WGF and SRWGFs are shown

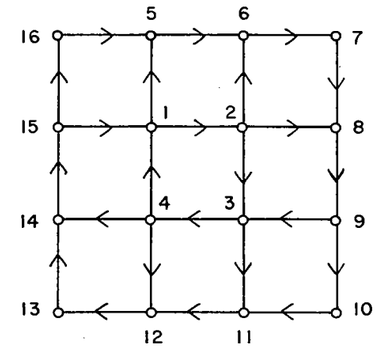

Figure 2. A directed square lattice.

Table II. Walk generating functions and spectral moments of an ordinary and signed square lattice Figure 3.

First 36 coefficients in the WGF of the signed lattice

0.0	-64.0	0.0
288.0	0.0	-1888.0
0.0	14552.0	0.0
-122424.0	0.0	1085360.0
0.0	-9939368.0	0.0
92921072.0	0.0	-880695112.0
0.0	8427265752.0	0.0
-81204846832.0	0.0	786670439280.0
0.0	-7653034773952.0	0.0
74706758257384.0	0.0	-731333896332248.0
0.0	7176356509174024.0	0.0
-70561564162709040.0	0.0	694993818269838920.0

First 36 spectral moments of the signed square lattice

0.0	-120.0	0.0
768.0	0.0	-5928.0
0.0	49776.0	0.0
-436720.0	0.0	3929784.0
0.0	-35943776.0	0.0
332651088.0	0.0	-3107174424.0
0.0	29246452648.0	0.0
-277102859472.0	0.0	2640656572104.0
0.0	-25292741018600.0	0.0
243358540192920.0	0.0	-2350980128975368.0
0.0	22793564085322064.0	0.0
-221700022093408552.0	0.0	2162499752812994880.0

First 36 Coefficients in the WGF of the ordinary square lattice

120.0	416.0	1464.0
5200.0	18568.0	66536.0
238920.0	859120.0	3091776.0
11132712.0	40098720.0	144462120.0
520511176.0	1875612440.0	6758904784.0
24357018608.0	87776778392.0	316330383696.0
1140000681800.0	4108389392200.0	14806049461384.0
53359002540304.0	192298822608736.0	693020229440328.0
2497556681651424.0	9000879365019048.0	32438038544485608.0
116902632034269848.0	421302477910682736.0	1518321569834059408.0
5471841605796314872.0	19719835270901438480.0	71067829310366984328.0
256119604528111987240.0	923023153701966669000.0	3326460491911600982832.0

First 36 Spectral moments of the ordinary square lattice

0.0	120.0	0.0
912.0	0.0	8712.0
0.0	92688.0	0.0
1049400.0	0.0	12365448.0
0.0	149738072.0	0.0
1848602576.0	0.0	23142098592.0
0.0	292655854632.0	0.0
3728313849248.0	0.0	47752512334440.0
0.0	614002413827832.0	0.0
7917124500058776.0	0.0	102293612786053392.0
0.0	1323629434618917840.0	0.0
17145199314926894472.0	0.0	222253147896045465840.0

below for this graph. We omit the constant term for convenience. This is equal to the number of vertices for WGF and zero for SRWGF.

$$\text{WGF} = 24x + 34x^2 + 50x^3 + 70x^4 + 102x^5 + 144x^6 + 212x^7 + 300x^8 + 440x^9 + 620x^{10} + 908x^{11} + 1282x^{12} + 1882x^{13} + 2658x^{14} + 3898x^{15} + 5500x^{16}$$

$$\text{SRWGF} = 4x^4 + 12x^6 + 44x^8 + 80x^{10} + 196x^{12} + 252x^{14} + 668x^{16}.$$

Next we consider a few signed square and honeycomb lattices. It may be recalled that signed graphs are bidirected graphs with a weight $+1$ for an edge i to j if $i > i$ and -1 if $j < i$. Note that because of the introduction of negative weights the coefficents in WGF enumerate the "net effect" of walks of various lengths as opposed to the total number of walks of length k. Hence, the coefficients of odd powers vanish in WGF.

Table II compares the WGF and spectral moments of a signed square lattice (Fig. 3) and the corresponding ordinary square lattice. The ordinary square lattice is a bidirectional graph with the same weight $+1.0$ for both directions. As seen from Table II, the coefficients of odd powers of the signed lattice vanish while the coefficients of even powered terms of the signed lattice are significantly smaller as these measure the net effect of taking walks of various lengths while for ordinary lattices these are the number of walks of specified length. The odd coefficients vanish for signed graphs since the matrix product terms cancel out. Note that the WGF for signed lattice is label-dependent. Interested readers can obtain the labels used for Table II from the author.

The number of "non-self-returning" walks can be evaluated using WGF-SRWFG. The coefficients of various terms in WGF-SRWGF truly enumerate non-self-returning walks only for regular lattices. Thus the coefficient of x^2 in WGF-SRWGF of the ordinary lattice in Figure 3 (Table II) is 296 while it is 56 for a signed lattice.

Table III shows the WGF and SRWGFs of a honeycomb lattice containing 54 vertices (Fig. 4). Note that since the coefficients of WGF of this lattice grow

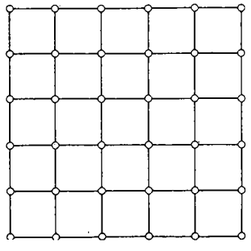

Figure 3. A square lattice containing 36 vertices. For the WGF, SRWGFs of this regular and signed lattice see Table II.

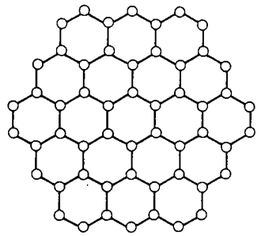

Figure 4. A honeycomb lattice containing 54 vertices. See Table III for WGF, SRWGFs of regular and signed lattices.

astronomically a quadruple precision arithmetic was invoked. Hence, the execution was slowed down since vectorization is not possible with the quadruple precision arithmetic. Yet the CPU time taken was relatively insignificant for the problem at hand. Again the general feature of the coefficients for the honeycomb lattice graph follows the square lattice graph discussed above. The labeling for this lattice can be obtained from the author.

The code developed above could be potentially useful for enumerating self-avoiding walks in a faster manner compared to any known procedure up to now. The procedure outlined by Randić et al.[3] for finding the number of self-avoiding walks is a non-polynomial algorithm and thus the CPU time required grows astronomically for larger complex graphs containing cycles. However, if matrix power procedure can be suitably adapted for a combination of weighted graphs such that the resulting generating function is the self-avoiding walk polynomial then it could be very powerful. For example, consider the tree in Figure 5. If one enumerates the largest path (L_n) starting from the vertex labeled 1 one gets L_5. The SAW of L_5 is given by the matrix powers of the appropriate directed graph discussed before

$$\text{SAW}(L_5) = 8x + 6x^2 + 4x^3 + 2x^4.$$

Now we start with the vertex labeled 6 which was not visited before and enumerate the largest possible paths. It can be seen that they are 6-3-4-5 and 6-3-2-1 for the graph in Figure 5. A restricted SAW for these two graphs can be obtained such that all walks start from the vertex labeled 6 (root). This would be the sum of sixth-row elements in the powers of the

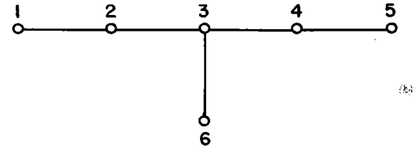

Figure 5. A branched tree containing six vertices. The self-avoiding walk GF (SAW) of this tree can be obtained as a linear combination of the SAW of a path and SAWS of rooted paths.

Table III. The walk GF, SRWGF of an ordinary honeycomb lattice graph and signed lattice graph in Figure 4.

First 49 Coefficients in the WGF of the ordinary lattice

144.0	396.0	1092.0
3048.0	8520.0	23952.0
67404.0	190248.0	537384.0
1520400.0	4303980.0	12195024.0
34566936.0	98032224.0	278091468.0
789114456.0	2239569336.0	6357221520.0
18047469516.0	51240288672.0	145490932872.0
413131015680.0	1173161674668.0	3331534906728.0
9461111559432.0	26868895603920.0	76307011425516.0
216713019869040.0	615474205858104.0	1747987354097952.0
4964429538833100.0	14099463174807672.0	40043996791048344.0
113729614430783760.0	323006083973758284.0	917378655666919104.0
2605476514877159976.0	7399055531735159744.0	21016749890964025644.0
59690501598005305992.0	169529439135407481192.0	481487702415096150096.0
1367493939412230949356.0	3883879336873015091856.0	11030777660087721846360.0
31329004736045445534432.0	88978917233724658427340.0	252713053122361419265176.0
717741839846367542003832.0		

First 53 coefficients in SRWGF of the ordinary lattice

0.0	144.0	0.0
648.0	0.0	3648.0
0.0	22968.0	0.0
154224.0	0.0	1079496.0
0.0	7778376.0	0.0
57258360.0	0.0	428438136.0
0.0	3247259328.0	0.0
24866881128.0	0.0	192030980040.0
0.0	1493227064928.0	0.0
11678269142544.0	0.0	91774395035448.0
0.0	724139076916344.0	0.0
5733298732288680.0	0.0	45523925960816232.0
0.0	362355328593353520.0	0.0
2890192973530857408.0	0.0	23092914702266609208.0
0.0	1847880279806554602720.0	0.0
1480516346318066986416.0	0.0	11874435754527365234568.0
0.0	9532393846629582072902 4.0	0.0
765807880412372719390896.0	0.0	

First 54 Coefficients in the WGF of signed lattice

0.0	-48.0	0.0
140.0	0.0	-588.0
0.0	2904.0	0.0
-15664.0	0.0	88852.0
0.0	-519652.0	0.0
3100560.0	0.0	-18761008.0
0.0	114712316.0	0.0
-707160284.0	0.0	4388510200.0
0.0	-27386424808.0	0.0
171718652564.0	0.0	-1081146490940.0
0.0	6831398518040.0	0.0
-433015371801 28.0	0.0	275235661534492.0
0.0	-1753781166456532.0	0.0
11199417483834872.0	0.0	-71657265362655496.0
0.0	459280822426293668.0	0.0
-2948284844179093460.0	0.0	18952350966817137136.0
0.0	-12198239714795430384 8.0	0.0
7859897642610690822 44.0	0.0	-50696022269821999126 92.0

First 54 coefficients in the SRWGF of the signed lattice (fig. 4)

0.0	-144.0	0.0
648.0	0.0	-3360.0
0.0	18712.0	0.0
-108624.0	0.0	646872.0
0.0	-3916448.0	0.0
23981240.0	0.0	-148039824.0
0.0	919483888.0	0.0
-5738467032.0	0.0	35953419016.0
0.0	-225993008008.0	0.0
1424460990432.0	0.0	-9000065170080.0
0.0	56983984551544.0	0.0
-361467127423128.0	0.0	2296721555875032.0
0.0	-14615032268628000.0	0.0
93128384827526752.0	0.0	-594158930982146936.0
0.0	37950341727245036 64.0	0.0
-24265045239591069824.0	0.0	155297119435234833032.0
0.0	-99478915918769205662 4.0	0.0
6377579173459523249504.0		0.0-40917825524595466717680.0

adjacency matrix of this graph. Suppose we call this rooted-self-avoiding walk (RSAW) then

$$\mathrm{RSAW}(L_4) = x + 2x^2 + 2x^3.$$

Since there are two paths (6-3-4-5 and 6-3-2-1) starting from the vertex 6, SAW of the graph in Figure 5 is given by

$$\mathrm{SAW\ (Fig.\ 5)} = \mathrm{SAW}\ (L_5) + 2\mathrm{RSAW}(L_4)$$
$$= 10x + 10x^2 + 8x^3 + 2x^4.$$

The above result can be easily verified.

Although the above procedure for complex cyclic graphs can get complicated it may be a better and efficient alternative to compute the SAW of a given graph as a linear combination of SAWs of L_ns and rooted paths. Such advancement and extension of these procedures to complex cyclic graphs could be the topic of future investigatons.

CONCLUSION

In this investigation we developed a vectorized computer code in FORTRAN 77 to enumerate the walk generating functions (WGF) and self-returning walk generating functions (SRWGF) for directed graphs, signed graphs and weighted graphs. The code was applied to several graphs and lattices. It was also shown that the self-avoiding walks on some graphs can be enumerated as a linear combination of WGFs of directed maximal paths and rooted-directed maximal paths.

References

1. M. Randić, *J. Chem. Inf. Comput. Sci.*, **18,** 101 (1978).
2. M. Randić, *J. Comp. Chem.*, **1,** 386 (1980).
3. M. Randić, G.M. Brissey, R.B. Spencer, and C.L. Wilkins, *Computer Chem.*, **3,** 5 (1979).
4. M. Randić, and C.L. Wilkins, *J. Chem. Inf. Comput. Sci.*, **20,** 36 (1980).
5. W. Brastow and A. Schinitzel, *J. Stort. Phys.*, **4,** 103 (1972).
6. F.T. Wall and D.J. Klein, *Proc. Nat. Acad. Sci.*, **76,** 1529 (1979).
7. M.F. Sykes and M.E. Fisher, *Adv. Phys.*, **9,** 315 (1960).
8. G.K. Baker, *J. Math. Phys.*, **7,** 2238 (1966).
9. R.A. Marcus, *J. Chem. Phys.*, **43,** 2643 (1965).
10. E.W. Montroll, in *Applied Combinatorial Mathematics,* E.F. Beckenbach, Ed., Wiley, New York, pp. 96–143.
11. K. Balasubramanian, *Computers and Chemistry,* **9,** 43 (1985).
12. D.M. Cvetković, M. Doob, and H. Sachs, *Spectra of Graphs,* Academic, New York, 1986.
13. K. Balasubramanian, *J. Comp. Chem.*, **5,** 387 (1984).
14. J.K. Burdett, in *Graph Theory and Topology in Chemistry,* R.B. King and D.H. Rouvray, Eds., Elsevier, New York, 1987, pp. 302–324.

Ab Initio Study of Ascorbic Acid Conformations

Mohammad A. Al-Laham and G.A. Petersson*

Department of Chemistry, Wesleyan University, Hall-Atwater Laboratory, Middletown, Connecticut 06457

Paul Haake

Department of Molecular Biology and Biochemistry and Department of Chemistry, Wesleyan University, Hall-Atwater Laboratory, Middletown, Connecticut 06457

Received 13 March 1990; accepted 6 July 1990

The STO-3G optimized structures of nine different staggered conformers of ascorbic acid are presented. The largest energy difference between the nine local minima is 5.1 kcal/mol. Comparison of the relative energies of the fully optimized structures of ascorbic acid conformers with those of nonoptimized conformers shows that full optimization is essential to obtain meaningful results. However, optimization of the ring structure is almost independent of optimization of the side-chain structure. One of the STO-3G optimized gas phase conformers is very close to the X-ray structure of the crystal.

INTRODUCTION

Ascorbic acid or Vitamin C (Fig. 1) has been a molecule of great biochemical interest since it was isolated and characterized[1] in 1928. The crystal structure was determined in 1968 by Hvoslef using X-ray crystallography[2a] and neutron diffraction.[2b] Little theoretical work has been done on ascorbic acid, mainly because of its size. The most comprehensive work so far is an STO-3G minimal basis set study without geometry optimization by Carlson, Cable and Pedersen.[3] Other work includes semiempirical[4] calculations on free radicals derived from ascorbic acid and α-hydroxytetronic acid to help determine the structure and UV spectrum of ascorbic acid, and to study the π-electrons in the ring. *Ab initio* SCF calculations[5] on α-hydroxytetronic acid have been used as a model for the electronic structure of ascorbic acid. The object of this work is to study the importance of fully optimizing the structure of a large molecule such as ascorbic acid, by comparing the fully optimized results from this work with the previous nonoptimized results of Carlson, Cable and Pedersen.[3] In this article, we consider the separate optimization of the α-hydroxytetronic acid (ring) and the 1,2-diol sidechain, and we compare the results with those obtained when we fully optimized the molecule as one unit, in order to study the separability of calculations on large systems.

Accurate *ab initio* calculations of conformational geometries and energies require at least a double-zeta plus polarization basis set and inclusion of correlation effects on at least the second-order Møller–Plesset level of theory. A complete optimization of all ascorbic acid conformers at the MP2/6-31G* level of theory is not practical at this time. On the other hand, SCF calculations using the STO-3G minimal basis set give qualitatively correct results through accounting for both the electrostatic interactions and the Pauli repulsive interactions between electrons (the RMS error in the STO-3G rotational barriers for ethane, methanol, and acetaldehyde is 0.52 kcal/mole).[6] This level of calculation also allows us to compare our optimized results with the nonoptimized results of Carlson, Cable, and Pedersen.[3]

The local minima on the ascorbic acid potential energy surface are related by rotations about the two dihedral angles, τ_1 (C3-C4-C5-C6 in Figure 1) and τ_2 (C4-C5-C6-O6 in Figure 1). Carlson, Cable, and Pedersen[3] investigated the nine staggered conformations obtained by rotations of 0°, 120°, and 240° about τ_1 and τ_2 with respect to the crystal structure (τ_1 = 187.8° and τ_2 = 295.3°). We shall use the notation $(\Delta\tau_1,\Delta\tau_2)$ to indicate the conformation with $\tau_1 \approx 187.8° + \Delta\tau_1$ and $\tau_2 \approx 295.3° + \Delta\tau_2$. For example, the (0,0) conformer is the STO-3G optimized conformer corresponding to the crystal structure. We have not performed a similar systematic investigation of the possible conformations of the hydroxyl groups because of the large number of conformers (324) that would need to be considered. In order to be consistent with the earlier nonoptimized study with which we wished to compare our results, we instead limited our search to finding the global minimum for the hydroxyl groups for each of the nine conformations of the side-chain. We have optimized all side-chain variables in all conformations, while

*To whom all correspondence should be addressed.

Journal of Computational Chemistry, Vol. 12, No. 1, 113-118 (1991)

CCC 0192-8651/91/010113-06$04.00