Diffusion coefficients of two-dimensional viral DNA walks

Tai-Hsin Hsu and Su-Long Nyeo
Department of Physics, National Cheng Kung University, Tainan, Taiwan 701, Republic of China
(Received 19 November 2002; published 15 May 2003)

DNA sequences are represented as two-dimensional walkers based on groups of mapping rules for the nucleotides in the DNA sequences. Digital sequences from irrational and random numbers in base 4 are generated and their diffusion properties are then compared with those of 21 nucleotide sequences of animal and plant viruses. By defining the diffusion coefficient as a function of the number of steps taken in a walk, we show that the coefficients for the viral DNA sequences generally have maximum values considerably larger than those for the random-number sequences of same lengths. Moreover, using the walker diagrams generated by different mapping groups, we can study the dominance of any of the nucleotide pairs (AG or CT), (AC or GT), or (AT or CG) in a DNA sequence. Other possible studies of this approach are mentioned.

DOI: 10.1103/PhysRevE.67.051911 PACS number(s): 87.10.+e, 87.15.−v, 05.40.−a, 02.50.−r

I. INTRODUCTION

Many statistical methods have been employed to analyze DNA sequences for their information content. The most common ones include the power spectral density [1–5], the correlation function [2,3,6,7], the random-walker representation [8–11], and the mutual information function [12,13]. Each of them may only serve a specified purpose of study. For example, in the analysis of the power spectral density of a particular nucleotide, one considers a mapping of assigning that nucleotide to ''1'' and others to ''0'' to obtain a sequence of binary digits, and analyzes the periodicities of the nucleotide. In the one-dimensional DNA walk as introduced by Peng *et al.* [8], the pyrimidines and the purines in a DNA sequence are assigned for the walker to take a step up $[u(i)=+1]$ if a pyrimidine is at position i and a step down $[u(i)=-1]$ otherwise. Then the long-range correlations in nucleotide sequences may be studied.

Therefore, it is useful to consider a general approach for the analysis of nucleotide sequences. The two-dimensional walks may provide a general approach for DNA studies, since the four types of nucleotides in DNA sequences are treated equally. In Sec. II, we shall introduce the mapping rules for the two-dimensional walks. In Sec. III, we shall give the diffusion properties of the character sequences of irrational and random numbers, and compare them with those of the 21 nucleotide sequences of animal and plant viruses. Other possible studies based on the two-dimensional walks will be mentioned.

II. THE TWO-DIMENSIONAL WALKS

We first define the two-dimensional walks according to Ref. [14], in which the global fractal dimension of human DNA sequences was studied. Obviously, this approach is a generalized version of the one-dimensional walk. We shall first look at the diffusion properties of the digital sequences generated from the irrational and random numbers defined in base 4. For example, the number π in base 4 has the numerical value $\pi_4=3.021\,003\,331\,222\ldots$. For simplicity, only the digits after the decimal point will be taken as the digital sequence. Similarly, the digital sequences $\{u(i)\}$ from the irrational numbers $\sqrt{2}$ and $\sqrt{3}$ can be generated, and a random-number sequence can be obtained from the random-number generator of the software LABVIEW (Table I). LABVIEW generates a string of random numbers $\{x(i)\}$ between 0 and 1 and the string is mapped onto a digital sequence of the digits {0,1,2,3} depending on the values of $x(i)$ in the intervals: {[0.00,0.25),[0.25,0.50),[0.50,0.75), [0.75,1.00)}, which are mapped onto the digits {0,1,2,3}.

For irrational and random numbers, each string $\{u(i)\}$ is mapped onto a character sequence $\{n(i)\}$, which is finally mapped onto two binary strings $\{u_x(i)\}$ and $\{u_y(i)\}$ via a mapping rule. In mapping from $\{u(i)\}$ to $\{n(i)\}$, a digital sequence, represented by a sequence of the four digits $\{u(i)\}\in\{0,1,2,3\}$, is mapped onto the four types of nucleotides or bases: adenine (A), cytosine (C), guanine (G), and thymine (T) to form a character sequence $\{n(i)\}$. For example, we may assign $u(i)=0$ to be the nucleotide A, $u(i)=1$ to be the nucleotide C, $u(i)=2$ to be the nucleotide G, and $u(i)=3$ to be the nucleotide T. In this way, sequences of the four characters are generated and they resemble the nucleotide sequences of organisms.

Next, we define groups of mapping rules for the two-dimensional walks (Tables II–IV). These mapping rules correspond to translations from the sequence $\{n(i)\}$ to $\{u_x(i)\}$ and $\{u_y(i)\}$. There are three independent groups of mapping rules, such that in each group, there are four mapping rules that give rise to walker diagrams with reflection symmetries about the x or y axis. For instance, the mapping rule **a** of group 1 [hereafter referred to as mapping rule 1(**a**)] is defined as

TABLE I. Irrational and random numbers in base 4 and their corresponding digital sequences.

Number in base 4	Digital sequence $u(i)$
π_4-3	02100333122220202011 . . .
$\sqrt{2}_4-1$	12220021321212133303 . . .
$\sqrt{3}_4-1$	23231213223220112010 . . .
Random-number sequence from LABVIEW	22331013112303210333 . . .

TABLE II. Group 1 mapping rules.

	a		**b**		**c**		**d**	
$n(i)$	$u_x(i)$	$u_y(i)$	$u_x(i)$	$u_y(i)$	$u_x(i)$	$u_y(i)$	$u_x(i)$	$u_y(i)$
A	−1	0	−1	0	1	0	1	0
C	0	−1	0	1	0	−1	0	1
G	0	1	0	−1	0	1	0	−1
T	1	0	1	0	−1	0	−1	0

$$\mathrm{A}=(-1,0),\quad \mathrm{T}=(1,0),\quad \mathrm{C}=(0,-1),\quad \mathrm{G}=(0,1),$$

which means that the alphabets T and A denote walks on the x axis, while the alphabets G and C denote walks on the y axis. Obviously, from Table II, the mapping rule **b** generates a walker diagram that is a reflection about the x axis of the diagram with the mapping rule **a**. The mapping rule **c** generates a walker diagram that is a reflection about the y axis of the diagram with the mapping rule **a**; and the mapping rule **d** generates a walker diagram that is a reflection about the y axis of the diagram with the mapping rule **b**. However, the walker diagrams with different groups of mapping rules are different [Figs. 1(a)–1(c)]. Clearly, the one-dimensional walk defined in Ref. [8] can be obtained by projecting the two-dimensional walk onto the line $y=x$ or $y=-x$. For example, a one-dimensional walk can be obtained by projecting a two-dimensional walk with the mapping rule 1(**c**) onto the line $y=x$.

To study the diffusion properties of a two-dimensional walk, we define the position of the walker. After j steps, the position is $\mathbf{r}(j)=(x(j),y(j))$, where the components of the position read

$$x(j)\equiv\sum_{i=1}^{j} u_x(i),\quad y(j)\equiv\sum_{i=1}^{j} u_y(i).$$

For the purpose of comparing different sequences, we define the two-dimensional diffusion coefficient for a walk by [15]

$$D(n)=\frac{\langle r^2\rangle}{4n}, \tag{1}$$

where the average $\langle r^2\rangle$ is defined as the average of all the r^2 terms starting from the first step to the nth step along the sequence

$$\langle r^2\rangle=\frac{1}{n}\sum_{j=1}^{n} r^2(j)=\frac{1}{n}\sum_{j=1}^{n}[x^2(j)+y^2(j)], \tag{2}$$

TABLE III. Group 2 mapping rules.

	a		**b**		**c**		**d**	
$n(i)$	$u_x(i)$	$u_y(i)$	$u_x(i)$	$u_y(i)$	$u_x(i)$	$u_y(i)$	$u_x(i)$	$u_y(i)$
A	−1	0	−1	0	1	0	1	0
C	1	0	1	0	−1	0	−1	0
G	0	−1	0	1	0	−1	0	1
T	0	1	0	−1	0	1	0	−1

TABLE IV. Group 3 mapping rules.

	a		**b**		**c**		**d**	
$n(i)$	$u_x(i)$	$u_y(i)$	$u_x(i)$	$u_y(i)$	$u_x(i)$	$u_y(i)$	$u_x(i)$	$u_y(i)$
A	−1	0	−1	0	1	0	1	0
C	0	−1	0	1	0	−1	0	1
G	1	0	1	0	−1	0	−1	0
T	0	1	0	−1	0	1	0	−1

with n denoting the number of steps such that $1\leqslant j\leqslant n\leqslant \ell$ for a total number of steps or nucleotides ℓ in the DNA sequence. Thus, the diffusion coefficient is a function of the number of steps n. Also, we note that since the walker diagrams with the mapping rules of a group are reflections of one another, the diffusion coefficients for the diagrams in the group are the same.

TABLE V. The animal (*A*) and plant (*P*) viruses and their accession numbers [16].

Organism	Type	Accession number	Total bases (bp)
Reston Ebola virus	*A*	NC_004161	18891
Zaire Ebola virus	*A*	NC_002549	18959
Foot-and-mouth disease virus *C*	*A*	NC_002554	8115
Foot-and-mouth disease virus *O*	*A*	NC_004004	8134
Human herpes virus 1	*A*	NC_001806	152261
Human herpes virus 2	*A*	NC_001798	154746
Human immunodeficiency virus type 1	*A*	NC_001802	9181
Human immunodeficiency virus type 2	*A*	NC_001722	10359
Mumps virus	*A*	NC_002200	15384
Beet yellows virus	*P*	NC_001598	15480
Cauliflower mosaic virus	*P*	NC_001497	8024
Cucumber mosaic virus RNA 1	*P*	NC_002034	3357
Cucumber mosaic virus RNA 2	*P*	NC_002035	3050
Cucumber mosaic virus RNA 3	*P*	NC_001440	2216
Cucumber mosaic virus satellite RNA	*P*	NC_002602	336
Satellite tobacco mosaic virus	*P*	NC_003796	1058
Tobacco mild green mosaic virus	*P*	NC_001556	6355
Tobacco mosaic virus	*P*	NC_001367	6395
Tobacco necrosis satellite virus	*P*	NC_001557	1239
Tobacco necrosis virus *A*	*P*	NC_001777	3684
Tobacco necrosis virus *D*	*P*	NC_003487	3762

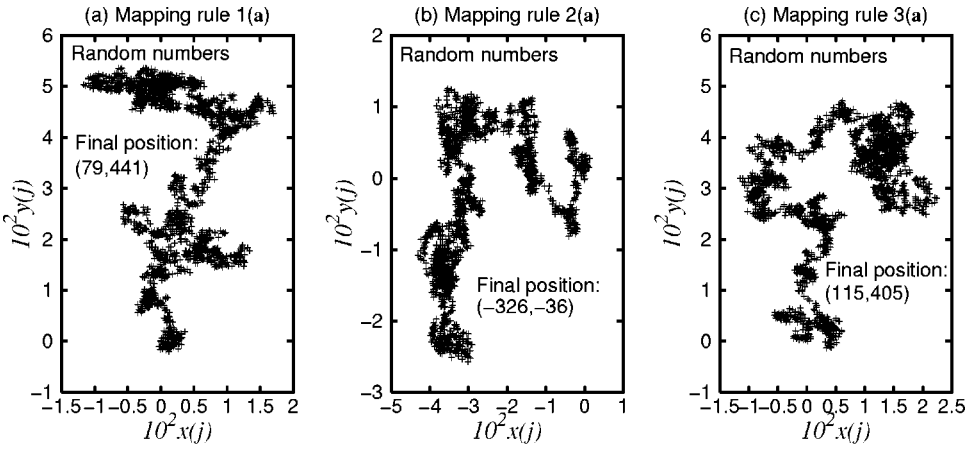

FIG. 1. Two-dimensional walker diagrams for the LABVIEW random numbers based on the mapping rules: (a) 1(**a**), (b) 2(**a**), and (c) 3(**a**).

For the nucleotide sequences, we shall consider several animal and plant viruses (Ref. [16]), with the accession numbers listed in Table V. We shall compare the viral data with π, $\sqrt{2}, \sqrt{3}$, and the random numbers from LABVIEW.

In our study, we find it useful to analyze not only the nucleotide sequences with their nucleotides in the originally listed form, which we shall refer to as the normal sequences, but also sequences with their nucleotides in the reversed order with respect to the normal sequences.

We observe that the complementary structure of a double-stranded DNA can be reflected only by the mapping rules of group 1. Consider, for example, the sequence $5'-\mathrm{AAATGGCCCC}-3'$, and apply the mapping rule 1(**a**) to it, we have

$$u_x: -,-1,-1,-1,1,0,0,0,0,0,0,0,-;$$

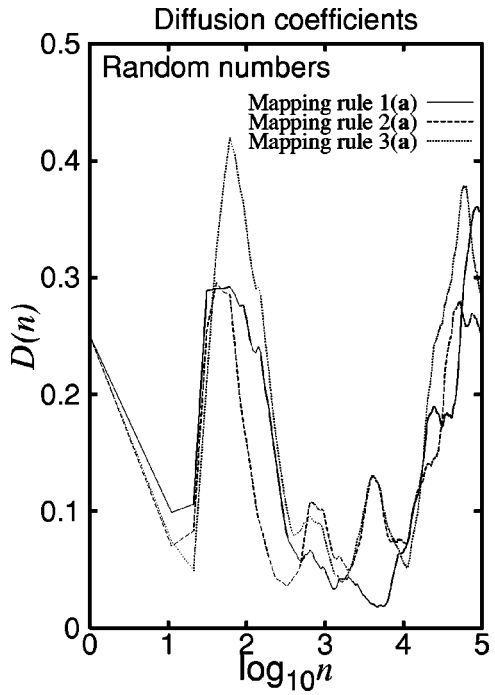

FIG. 2. Diffusion coefficients for the LABVIEW random numbers based on the mapping rules 1(**a**), 2(**a**), and 3(**a**).

$$u_y: -,0,0,0,0,1,1,-1,-1,-1,-1,-.$$

The reversed sequence of $5'-\mathrm{AAATGGCCCC}-3'$ is $3'-\mathrm{CCCCGGTAAA}-5'$. Use of the mapping rule 1(**d**) for the reversed sequence results

$$u_x: -,0,0,0,0,0,0,-1,1,1,1,-;$$

$$u_y: -,1,1,1,1,-1,-1,0,0,0,0,-.$$

The complementary sequence of the normal sequence $5'-\mathrm{AAATGGCCCC}-3'$ is $5'-\mathrm{GGGGCCATTT}-3'$, and the mapping rule 1(**a**) for $5'-\mathrm{GGGGCCATTT}-3'$ gives

$$u_x: -,0,0,0,0,0,0,-1,1,1,1,-;$$

$$u_y: -,1,1,1,1,-1,-1,0,0,0,0,-.$$

Thus, we see that the walker diagram for the reversed sequence using mapping rule 1(**d**) is the same as for the normal sequence on the complementary sequence using mapping rule 1(**a**). A similar analysis shows that the mapping rules 1(**b**) and 1(**c**) enjoy the same property.

Further, since the mapping rules within a group give the same diffusion coefficient, the group 1 mapping rules lead to

TABLE VI. The maximum values of the diffusion coefficients for the irrational and random numbers in the three mapping groups.

	Normal sequence			
Number	Group 1	Group 2	Group 3	Total length
π_4-3	0.60	0.59	0.25	100000
$\sqrt{2}_4-1$	0.32	0.29	0.28	100000
$\sqrt{3}_4-1$	0.26	0.39	0.33	100000
Random-number sequence from LABVIEW	0.36	0.31	0.42	100000

TABLE VII. The maximum values of the diffusion coefficients for the animal and plant viruses in the normal and reversed sequences in the three groups of mappings.

	Normal sequence			Reversed sequence		
Organism	Group 1	Group 2	Group 3	Group 1	Group 2	Group 3
Reston Ebola virus	2.98	22.95	24.63	1.60	33.14	34.02
Zaire Ebola virus	5.46	20.50	23.27	3.39	32.23	34.37
Foot-and-mouth disease virus *C*	1.81	2.60	4.33	1.47	2.08	3.35
Foot-and-mouth disease virus *O*	2.09	4.63	6.56	2.03	3.60	4.95
Human herpes virus 1	7.35	785.83	784.74	22.24	892.32	892.14
Human herpes virus 2	2.43	1012.82	1012.03	8.73	1109.98	1109.77
Human immunodeficiency virus type 1	20.12	28.61	14.44	15.10	23.62	11.35
Human immunodeficiency virus type 2	20.77	19.96	8.09	15.32	15.99	7.07
Mumps virus	3.25	12.40	14.17	1.78	18.90	20.42
Beet yellows virus	3.31	3.71	4.85	2.15	5.82	7.85
Cauliflower mosaic virus	13.86	22.65	24.54	11.53	15.16	17.33
Cucumber mosaic virus RNA1	0.47	0.77	1.12	0.56	0.83	1.38
Cucumber mosaic virus RNA2	0.67	1.59	1.59	0.36	1.15	1.35
Cucumber mosaic virus RNA3	0.84	0.79	1.07	0.49	0.47	0.33
Cucumber mosaic virus satellite RNA	0.70	0.31	0.75	0.39	0.39	0.39
Satellite tobacco mosaic virus	0.46	0.39	0.60	0.28	0.27	0.43
Tobacco mild green mosaic virus	1.96	10.14	8.94	2.38	10.37	9.47
Tobacco mosaic virus	1.23	5.47	4.83	2.05	6.70	5.86
Tobacco necrosis satellite virus	1.16	0.88	0.92	0.47	0.47	0.47
Tobacco necrosis virus *A*	0.93	0.88	0.62	0.84	0.64	0.72
Tobacco necrosis virus *D*	0.57	0.95	0.60	0.46	0.92	0.70

the same diffusion coefficient for the reversed sequence of a normal sequence and for the complementary sequence of the normal sequence.

III. DIFFUSION ANALYSIS AND CONCLUSION

First, we observe that from the two-dimensional walker diagrams, it is possible to classify the nucleotide sequences of viruses into three types according to the diffusion coefficients. In particular, the nucleotide sequences of viruses generally have larger maximum values of the diffusion coefficients than those of the irrational or random numbers. For long sequences, this approach can be used to indicate the abundance of certain nucleotides in the sequences; but for short sequences, say less than 5000, it is difficult to distinguish a nucleotide sequence from a random-number sequence.

In Fig. 2, the diffusion coefficients for the LABVIEW random-number sequence based on the mapping rule **a** of the three groups are plotted. We note that the maximum values of the coefficients are less than 0.5 (Table VI). For the sequences of the irrational numbers considered, the coefficients have maximum values of about 0.6. On the other hand, the maximum values of the coefficients for the nucleotide sequences can generally be quite large (Table VII), depending on the lengths of the sequences, and range from less than 1 to over 1000. Moreover, the maximum values depend also on the mapping rules.

From our analysis, we can divide the walker diagrams of the sequences into three types. A walker diagram of type I is the one that has a diffusion coefficient behaving like a monotonously increasing function with no obvious peaks between the initial and final positions, and the maximum value appears at the final position of a walk. A walker diagram of type II has a maximum value of the coefficient that occurs between the initial and final positions. While that of type III has a maximum value of the coefficient at the final position and there is at least one peak between the initial and final positions. From the 21 DNA data, we see no typical increase in diffusion coefficient of a viral DNA walk. But in general, when a two-dimensional walker follows on the average a straight path, its diffusion coefficient diagram will be of type I. The diffusion types of the sequences are listed in Tables VIII and IX.

TABLE VIII. The diffusion types for the irrational and random numbers in the three groups of mappings.

	Diffusion type			
Number	Group 1	Group 2	Group 3	Total length
π_4-3	II	II	II	100000
$\sqrt{2}_4-1$	II	II	II	100000
$\sqrt{3}_4-1$	II	II	II	100000
Random-number sequence from LABVIEW	II	II	II	10000

TABLE IX. The diffusion types for the animal and plant viruses in the normal and reversed sequences in the three groups of mapping rules.

	Normal sequence			Reverse sequence		
Organism	Group 1	Group 2	Group 3	Group 1	Group 2	Group 3
Reston Ebola virus	II	I	I	III	I	I
Zaire Ebola virus	III	I	I	III	I	I
Foot-and-mouth disease virus *C*	III	III	III	III	III	III
Foot-and-mouth disease virus *O*	III	III	I	III	III	III
Human herpes virus 1	II	I	I	II	I	I
Human herpes virus 2	II	I	I	II	I	I
Human immunodeficiency virus type 1	I	I	I	I	I	III
Human immunodeficiency virus type 2	III	III	III	I	I	I
Mumps virus	III	III	I	III	I	I
Beet yellows virus	II	III	III	III	I	I
Cauliflower mosaic virus	I	I	I	I	I	I
Cucumber mosaic virus RNA1	II	III	III	III	III	III
Cucumber mosaic virus RNA2	II	III	III	II	III	III
Cucumber mosaic virus RNA3	III	III	II	III	III	III
Cucumber mosaic virus satellite RNA	II	II	II	II	II	II
Satellite tobacco mosaic virus	II	III	II	II	III	III
Tobacco mosaic green mosaic virus	III	III	III	III	III	III
Tobacco mosaic virus	III	III	III	III	III	III
Tobacco necrosis satellite virus	II	II	II	II	II	II
Tobacco necrosis virus *A*	III	II	III	III	II	II
Tobacco necrosis virus *D*	III	II	III	III	III	III

TABLE X. The fractions of A, C, G, and T in the DNA sequences of animal and plant viruses.

	Nucleotide			
Organism	A	C	G	T
Reston Ebola virus	0.3143	0.2080	0.1983	0.2794
Zaire Ebola virus	0.3197	0.2128	0.1979	0.2696
Foot-and-mouth disease virus *C*	0.2476	0.2849	0.2559	0.2116
Foot-and-mouth disease virus *O*	0.2454	0.2908	0.2620	0.2019
Human herpes virus 1	0.1592	0.3380	0.3449	0.1580
Human herpes virus 2	0.1487	0.3504	0.3535	0.1475
Human immunodeficiency virus type 1	0.3564	0.1788	0.2423	0.2224
Human immunodeficiency virus type 2	0.3384	0.2058	0.2508	0.2049
Mumps virus	0.3074	0.2077	0.2042	0.2677
Beet yellows virus	0.2514	0.2226	0.2377	0.2883
Cauliflower mosaic virus	0.3671	0.2055	0.1942	0.2332
Cucumber mosaic virus RNA1	0.2514	0.2216	0.2431	0.2839
Cucumber mosaic virus RNA2	0.2525	0.2249	0.2315	0.2911
Cucumber mosaic virus RNA3	0.2333	0.2369	0.2383	0.2915
Cucumber mosaic virus satellite RNA	0.1964	0.2411	0.2857	0.2768
Satellite tobacco mosaic virus	0.2599	0.2146	0.2439	0.2817
Tobacco mild green mosaic virus	0.3042	0.1750	0.2343	0.2865
Tobacco mosaic virus	0.2912	0.1912	0.2416	0.2760
Tobacco necrosis satellite virus	0.2793	0.2276	0.2462	0.2470
Tobacco necrosis virus *A*	0.2815	0.2443	0.2446	0.2296
Tobacco necrosis virus *D*	0.2677	0.2196	0.2525	0.2600

We note that the walker diagrams of the reversed sequences are just the rotated diagrams of the normal sequences, with rotation angle 180°. The final positions of the walker diagrams of both sequences are identical, but their diffusion coefficients are not (cf. Table VII). The reason is that the final position of a walker diagram depends only on the numbers of the four nucleotides, which dictate the directions of the walking steps, and not on the order of the nucleotides in the sequence. For instance, consider the mapping rule 1(**a**) for a sequence of $\{n_A,n_T,n_C,n_G\}$ of the nucleotides {A, T, C, G}. The final position is then given by $(x(\ell),y(\ell))=(n_T-n_A,n_G-n_C)$, where $\ell=n_A+n_T+n_C+n_G$. However, the diffusion coefficient defined by Eq. (1) depends on how the nucleotides are ordered.

Moreover, the two-dimensional diffusion walks provide a very useful approach for the study of the abundance of nucleotides in DNA sequences. If the diffusion coefficient is large, then the two-dimensional walker diagram is most likely a straight line, and there are several possible situations for each mapping group. (1) Group 1 mapping rules: the DNA sequence has more fraction of either (AG or CT) or (AC or GT). (2) Group 2 mapping rules: the DNA sequence has more fraction of either (AG or CT) or (AT or CG). (3) Group 3 mapping rules: the DNA sequence has more fraction of either (AC or GT) or (AT or CG).

Thus, when the diffusion coefficients are large with group 2 and group 3 mapping rules but not with group 1 mapping rules, we may conclude that there are more AT or CG in the DNA sequence. When the diffusion coefficients are large with group 3 and group 1 mapping rules but not with group 2 mapping rules, there are more AC or GT in the DNA sequence. Finally, when the diffusion coefficients are large with group 1 and group 2 mapping rules but not with group 3 mapping rules, there are more AG or CT in the sequence. Hence, we may say that there is a large fractional difference between pyrimidines and purines if the diffusion coefficients are large with group 1 and group 2 mapping rules but not with group 3 mapping rules. This is clearly seen by comparing Tables VII and X. To see how the maximum diffusion coefficient depends, for example, on the CG content (in %), we plot the logarithm of the maximum diffusion coefficients based on group 2 and group 3 mapping rules for the viruses with their CG contents (Fig. 3). The maximum diffusion coefficients are seen to obey an approximate exponential decaying law with CG contents up to 50% and an approximate exponential increasing law for larger CG contents. There appears a reflection symmetry in the laws about the vertical line at CG content of 50%. The groups 2 and 3 mapping rules give about the same maximum diffusion coefficients. We note that to produce a straight-line walker diagram with groups 2 and 3 mapping rules, we require a high CG content in the sequence. In addition, how the nucleotides are ordered in the sequence is a crucial factor. For example, a repetition of several nucleotides of one type in a sequence would give a directed path along one of the axes and break the straight-line path.

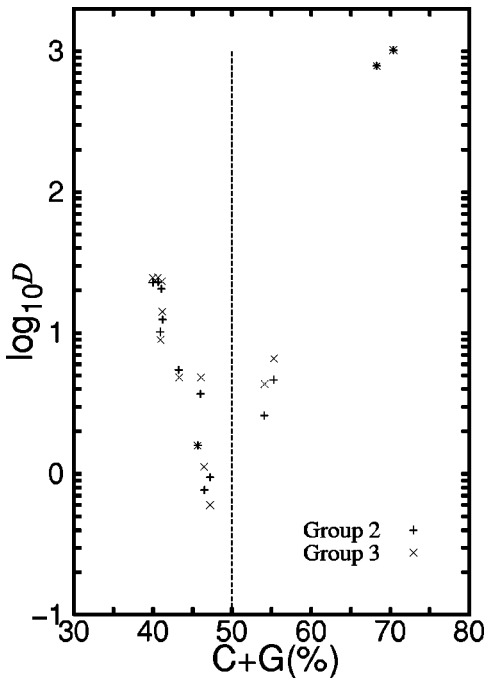

FIG. 3. Plot of the logarithm of maximum diffusion coefficients with the CG contents based on group 2 and group 3 mapping rules for the viruses.

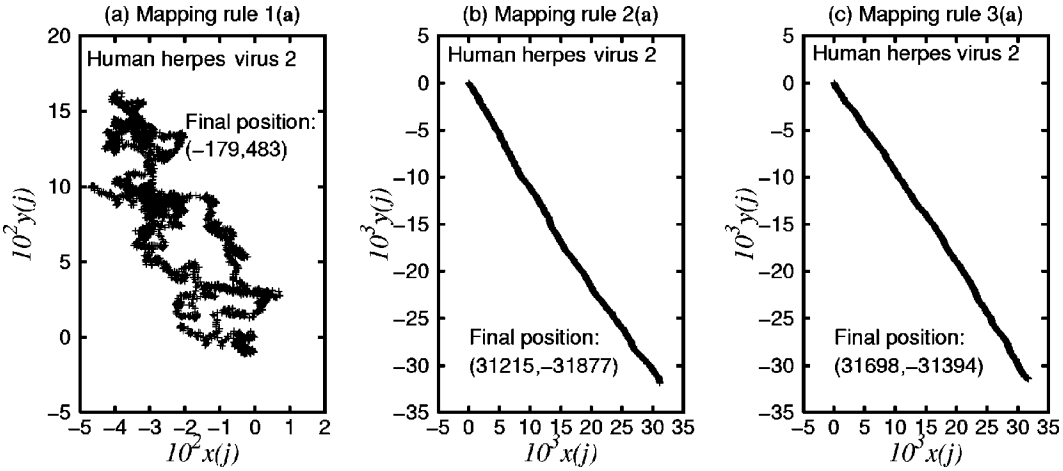

FIG. 4. Two-dimensional walker diagrams for the human herpes virus 2 based on the mapping rules: (a) 1(**a**), (b) 2(**a**), and (c) 3(**a**).

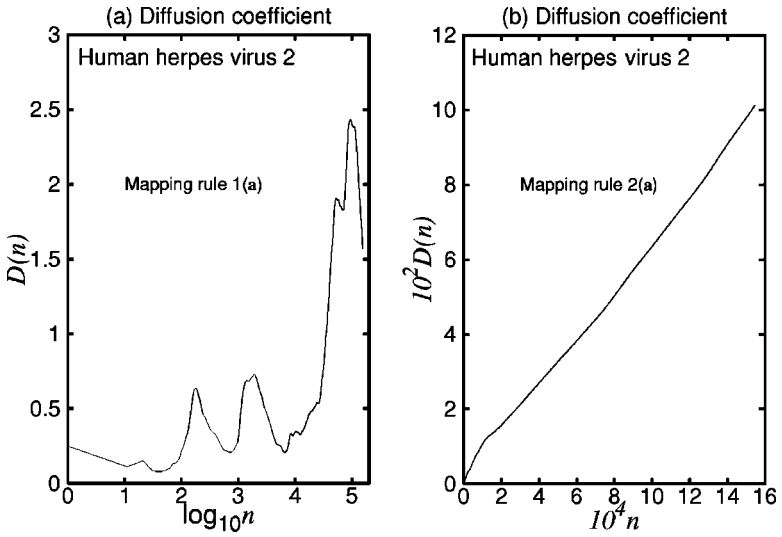

FIG. 5. The diffusion coefficients for the human herpes virus 2 based on the mapping rules: (a) 1(**a**), and (b) 2(**a**). Note that the coefficient in (b) given in the linear scales shows a linear dependence on n and no peaks appear between the initial and final positions of the walker.

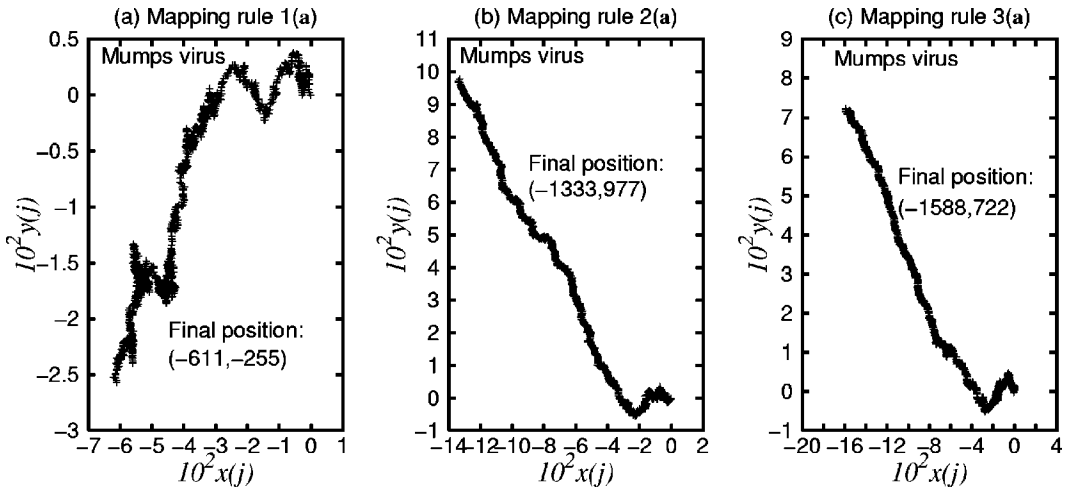

FIG. 6. Two-dimensional walker diagrams for the mumps virus based on the mapping rules: (a) 1(**a**), (b) 2(**a**), and (c) 3(**a**).

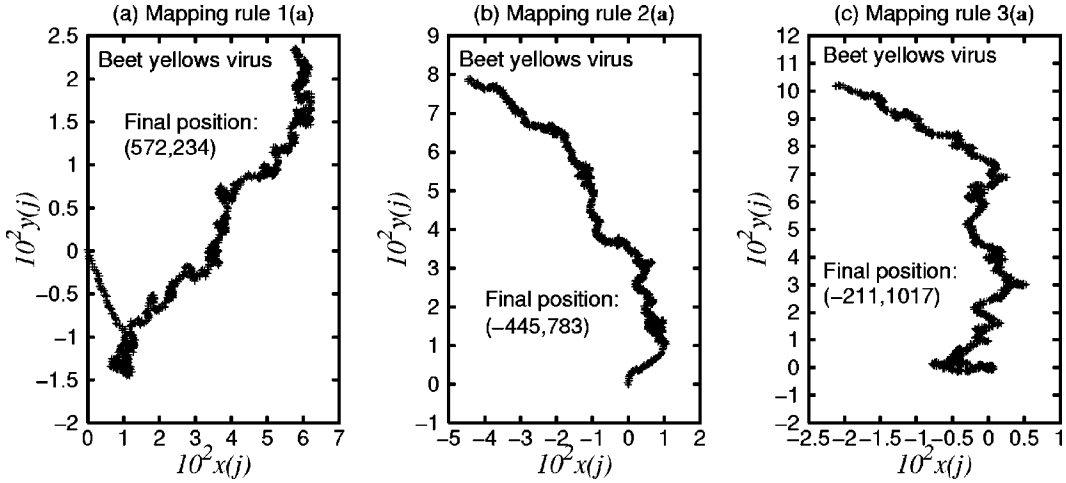

FIG. 7. Two-dimensional walker diagrams for the beet yellows virus based on the mapping rules: (a) 1(**a**), (b) 2(**a**), and (c) 3(**a**).

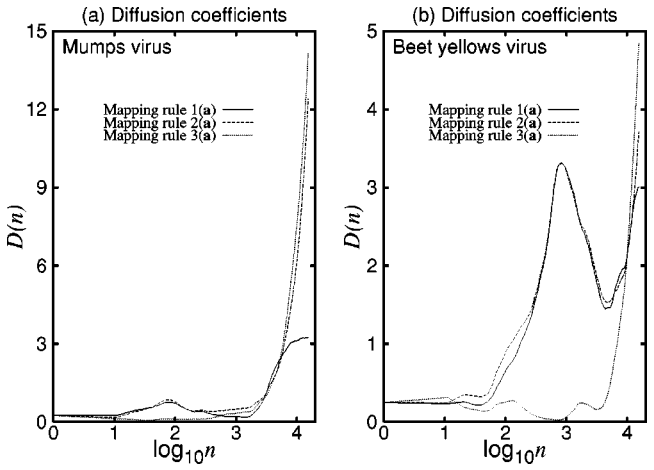

FIG. 8. Diffusion coefficients for (a) the mumps virus and (b) the beet yellows virus based on the mapping rules 1(**a**), 2(**a**), and 3(**a**).

The smallest maximum diffusion coefficient appears to be associated with the mapping group 1, with the exception of the tobacco necrosis virus *A*, human immunodeficiency virus types 1 and 2, which have the smallest maximum diffusion coefficients with the mapping group 3. Other exceptions are the Cucumber mosaic virus RNA3, Cucumber mosaic virus satellite RNA, satellite tobacco mosaic virus, and tobacco necrosis satellite virus whose smallest maximum diffusion coefficients are given by the mapping group 2.

As an example, consider the human herpes virus 2, it has large maximum values of diffusion coefficients with groups 2 and 3. Thus, the DNA sequence is rich in either AT or CG. A more detailed analysis shows that the sequence has large fractions of C and G, which make the trajectory of its two-dimensional walks to be very close to an oblique line (Fig. 4). In this case, the diffusion coefficients with mapping groups 2 and 3 are monotonously increasing functions but not with mapping group 1 (Fig. 5). Specifically, Fig. 5(b) shows that the human herpes virus 2 with the mapping rule 2(**a**) has a linear dependence on *n*. From Table X, C and G are notably more than A and T.

We next note that although the nucleotide sequences of the mumps and beet yellows viruses are about of the same length, the fractions of A, C, G, and T in their sequences are different. Consider the fractions of the nucleotides C and G in the sequences of these viruses in Table X. In the beet yellows virus, the fractions of C and G are 0.2226 and 0.2377, respectively, while in the mumps virus, they are 0.2077 and 0.2042, respectively. The relative difference between the nucleotides A and C and that between A and G in the beet yellows virus are smaller than those in the mumps virus. Also, the two-dimensional trajectories based on the mapping groups 2 and 3 for the mumps virus behave more linearly than those of the beet yellows virus [Figs. 6(a)–6(c) and 7(a)–7(c)]. Consequently, the maximum values of the diffusion coefficients for the mumps virus are larger than those of the beet yellows virus (Table VII). The diffusion coefficients based on the mapping rule (**a**) of the three groups are plotted in Figs. 8(a) and 8(b), with the corresponding types given in Table IX. From Figs. 6–8 and Table IX, we see no simple rule for determining the diffusion types from the walker diagrams.

Finally, we should mention that there are many studies that can be made with the two-dimensional walks. For instance, the DNA sequences of bacteria and human DNA sequences may be considered. Also a study of the possible connection of the two-dimensional mapping rules with the three-dimensional ones [17–19] should be of some interest. The scaling properties of the walks in two and three dimensions, which may be treated as polymers, can be analyzed. Of course, the implications of such studies remain to be seen.

ACKNOWLEDGMENT

This research was supported by the National Science Council of the Republic of China under the Contract No. NSC 91-2112-M006-012.

[1] S.V. Buldyrev, A.L. Goldberger, S. Havlin, R.N. Mantegna, M.E. Matsa, C.-K. Peng, M. Simons, and H.E. Stanley, Phys. Rev. E **51**, 5084 (1995).
[2] M. de Sousa Vieira, Phys. Rev. E **60**, 5932 (1999).
[3] W. Li, Comput. Chem. (Oxford) **21**, 257 (1997).
[4] W. Li and K. Kaneko, Europhys. Lett. **17**, 655 (1992).
[5] R.F. Voss, Phys. Rev. Lett. **68**, 3805 (1992).
[6] D. Holste, O. Weiss, I. Große, and H. Herzel, J. Mol. Evol. **51**, 353 (2000).
[7] P. Bernaola-Galván, P. Carpena, R. Román-Roldán, and J.L. Oliver, Gene **300**, 105 (2002).
[8] C.-K. Peng, S.V. Buldyrev, A.L. Goldberger, S. Havlin, M. Simons, and H.E. Stanley, Phys. Rev. E **47**, 3730 (1993).
[9] S.V. Buldyrev, A.L. Goldberger, S. Havlin, C.-K. Peng, M. Simons, and H.E. Stanley, Phys. Rev. E **47**, 4514 (1993).
[10] S.V. Buldyrev, A.L. Goldberger, S. Havlin, C.-K. Peng, H.E. Stanley, M.H.R. Stanley, and M. Simons, Biophys. J. **65**, 2673 (1993).
[11] C.-K. Peng, S.V. Buldyrev, A.L. Goldberger, S. Havlin, F. Sciortino, M. Simons, and H.E. Stanley, Nature (London) **356**, 168 (1992).
[12] H. Herzel and I. Große, Phys. Rev. E **55**, 800 (1997).
[13] H. Herzel, E.N. Trifonov, O. Weiss, and I. Große, Physica A **249**, 449 (1998).
[14] C.L. Berthelsen, J.A. Glazier, and M.H. Skolnick, Phys. Rev. A **45**, 8902 (1992).
[15] H.C. Berg, *Random Walks in Biology*, expanded ed. (Princeton University Press, Princeton, 1993).
[16] Viruses are accessible at www.ncbi.nlm.nih.gov
[17] E. Hamori and J. Ruskin, J. Biol. Chem. **258**, 1318 (1983).
[18] R. Zhang and C.-T. Zhang, J. Biomol. Struct. Dyn. **11**, 767 (1994).
[19] C.-T. Zhang, J. Theor. Biol. **187**, 297 (1997).