The size of RNA as an ideal branched polymer

Li Tai Fang, William M. Gelbart, and Avinoam Ben-Shaul

Citation: J. Chem. Phys. **135**, 155105 (2011); doi: 10.1063/1.3652763
View online: http://dx.doi.org/10.1063/1.3652763
View Table of Contents: http://jcp.aip.org/resource/1/JCPSA6/v135/i15
Published by the American Institute of Physics.

The size of RNA as an ideal branched polymer

Li Tai Fang,[1] William M. Gelbart,[2] and Avinoam Ben-Shaul[1,a)]
[1]*Institute of Chemistry and the Fritz Haber Research Center, Hebrew University of Jerusalem, Jerusalem 91904, Israel*
[2]*Department of Chemistry and Biochemistry, University of California, Los Angeles, California 90024, USA*

(Received 15 August 2011; accepted 26 September 2011; published online 20 October 2011)

Because of the branching arising from partial self-complementarity, long single-stranded (ss) RNA molecules are significantly more compact than linear arrangements (e.g., denatured states) of the same sequence of monomers. To elucidate the dependence of compactness on the nature and extent of branching, we represent ssRNA secondary structures as tree graphs which we treat as *ideal* branched polymers, and use a theorem of Kramers for evaluating their root-mean-square radius of gyration, $\hat{R}_g = \sqrt{\langle R_g^2 \rangle}$. We consider two sets of sequences—random and viral—with nucleotide sequence lengths (N) ranging from 100 to 10 000. The RNAs of icosahedral viruses are shown to be more compact (i.e., to have smaller $\hat{R}_g$) than the random RNAs. For the random sequences we find that $\hat{R}_g$ varies as $N^{1/3}$. These results are contrasted with the scaling of $\hat{R}_g$ for ideal *randomly* branched polymers ($N^{1/4}$), and with that from recent modeling of (relatively short, $N \leq 161$) RNA tertiary structures ($N^{2/5}$). © *2011 American Institute of Physics*. [doi:10.1063/1.3652763]

I. INTRODUCTION

Many theoretical studies have been concerned with the analysis and prediction of the secondary structure of single-stranded RNA (ssRNA) molecules—see, for example, Refs. 1–8. Similarly, many experimental measurements have been made of the secondary and tertiary structures of these molecules.[9–11] In addition, molecular dynamics simulation programs have been developed for coarse-grained RNA models, providing important insights into the flexibility and conformational freedom of RNA in solution.[12,13] These theoretical and experimental approaches become problematic, however, when applied to *long* RNA molecules, i.e., ones longer than about 1000 nucleotides (nt).

Here we focus on the relationship between the 2D (secondary) and 3D sizes of long ssRNA molecules. More explicitly, using a mapping of the secondary structures onto tree graphs and a theorem due to Kramers,[14,15] we formulate an ideal branched polymer model of ssRNA that enables us to estimate the root-mean-square radius of gyration ($\hat{R}_g$) corresponding to any given secondary structure of the molecule. Numerical calculations of $\hat{R}_g$ are presented for viral vs. random sequences involving lengths up to 10 000 nt. For random RNA sequences with uniform base composition, the scaling behavior of $\hat{R}_g$ with sequence length (N) is found to be $\hat{R}_g \sim N^{1/3}$, in contrast to the $\hat{R}_g \sim N^{1/4}$ result derived for ideal *randomly* branched polymers.[15–18]

II. THEORY

At room and physiological temperatures ssRNA in solution folds on itself, developing hydrogen-bonded interactions between complementary pairs of nucleotides—the "Watson-Crick" pairs G-C and A-U, and the "wobble" pair G-U. The result is a branched structure composed of rigid duplexes of successive base pairs (bp), joined by semiflexible loops of nt, as illustrated in Fig. 1(a) for a 200-nt-long sequence. The number of bp per duplex, k, varies between 2 and a maximal value $k_{\max}$ that increases logarithmically with N, e.g., $k_{\max} \approx 12$ for random sequences of length $N \approx 1000$ with uniform base composition.[19] However, the average duplex length, $\langle k \rangle$, is independent of N, and most duplexes are rather short. For random sequences of uniform base composition $\langle k \rangle \approx 4$.[2,19–21] In addition to the unpaired (i.e., single stranded) nt, the loops contain several paired (i.e., H-bonded) nt—those belonging to the first base pair of each of the duplexes emanating from the loop. It is not difficult to show that, on average, each loop is connected to $\langle d \rangle = 2 - 2/L \approx 2$ duplexes, where L is the total number of loops. Here, and in the rest of this Letter, we shall regard the exterior loop—the loop containing the two ends of the molecule—on the same footing as the other loops (as in the case of a circular RNA). Thus for both open ("linear") and circular RNAs $L = S + 1$, where S is the total number of duplexes in the structure. The long ssRNA sequences of interest here contain many loops (and hence duplexes), i.e., $L, S \gg 1$, so that we can set $L = S$.

The fraction of nt in duplexes is $f = 2\langle k \rangle S/N$, and—for both random and many viral sequences of nearly uniform nt compositions—it is found that $f \approx 0.6$, independent of N,[19] a value consistent with the 0.54 estimate of Dima *et al.*[22] from statistical-potential analyses of PDB data sets, and with current values predicted by M-fold[1] and RNA-fold[4] folding algorithms. For these molecules the average number of unpaired nt per loop is $\langle l_{ss} \rangle = (1 - f)N/L = (1 - f)N/S = 2\langle k \rangle (1 - f)/f \approx 5$, where we have set $\langle k \rangle = 4$.[2,19–21] The total number of nt per loop is thus $\langle l \rangle = \langle l_{ss} \rangle + 2\langle d \rangle = \langle l_{ss} \rangle + 4 \approx 9$. This, in turn, implies that, on average, an internal loop is comprised of 2 H-bonded links

a) Author to whom correspondence should be addressed. Electronic mail: abs@fh.huji.ac.il.

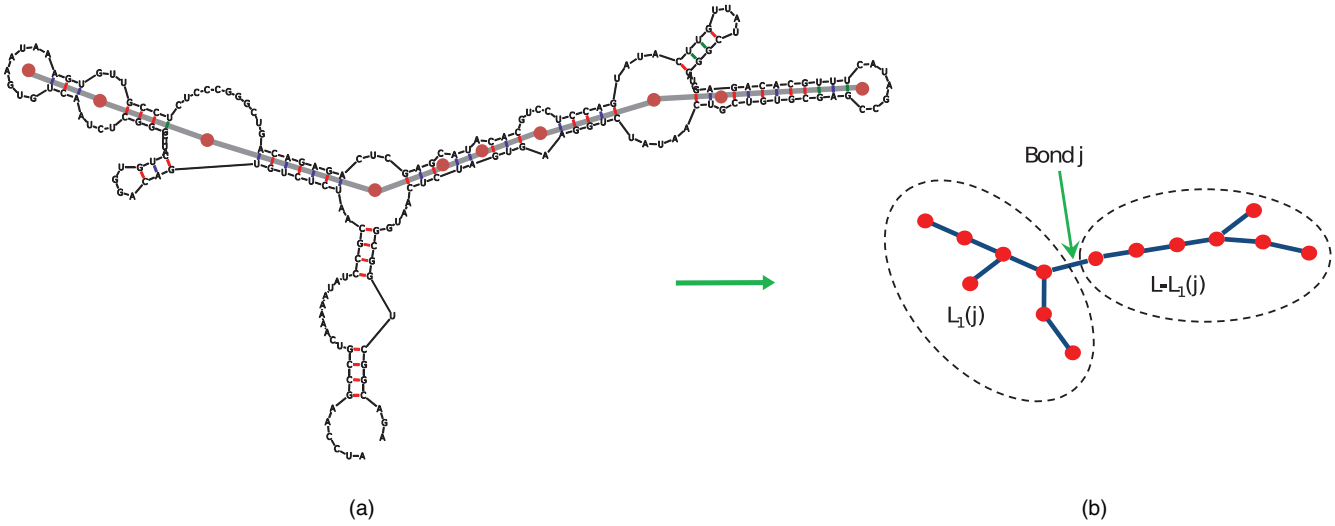

FIG. 1. Mapping of an RNA secondary structure (a) onto a branched tree graph (b). Duplexes are mapped into bonds, and loops into vertices (of order 1 [hairpin loops], 2 [bubbles and bulges], and 3 or higher [multi-loops]). The grey path connecting the two extreme hairpin loops in (a) denotes the contour of the maximum ladder distance for this secondary structure.

and (about) 7 covalent (single-strand) bonds. (In a hairpin loop there is one H-bonded link and at least 4 ss links.)

From detailed analyses[12,23,24] of the flexibility of internal loops it is known that there are significant correlations in the orientations of the duplexes emanating from the same loop. Nevertheless, we shall assume here that the loops can be regarded as flexible joints, as in the earlier work on ideal *randomly* branched polymers, and will focus on the consequences of *non*-random branching. The flexibility of the loops does not change the scaling of $\hat{R}_g$, e.g., the worm-like chain and ideal chain models of linear polymers share the same scaling law.[15] Similarly, we ignore features of secondary structure associated with "kissing hairpins" and pseudoknot structures[8] because they do not have an analogue in the classic randomly branched polymer model[15–18] with which we are comparing RNA molecules here. Also, these higher-order features do not allow the mapping to tree graphs that we use to calculate 3D sizes; further, they only effect the short-range structure of RNA and are not expected to change the mass-size scaling behavior. Finally, we neglect excluded volume effects, again because our main purpose is to compare the mass-size scaling of the classic "randomly branched" *ideal* polymers with the ideal polymers whose branching is governed by the "rules" of RNA self-complementarity. In both instances, we are treating tree graphs that behave ideally but whose distributions of branch points are different and which thus have different mass-size scaling behavior.

The mapping of a ssRNA secondary structure to a branched tree graph is illustrated schematically in Fig. 1(b). The duplexes are reduced to bonds of equal length, and the interior loops to vertices. The particular example shown here corresponds to the predicted minimum free energy (MFE) structure[4] of a short ssRNA sequence composed of 200 nt of uniform base composition and random distribution along the sequence. Recall, however, that in the present work we are mainly interested in longer sequences, involving thousands of nt. Assuming that all vertices are flexible joints, the tree graph may be regarded as an ideal branched polymer free to move in three dimensions, with no correlations between neighboring bond orientations and (as explained above) with neglect of excluded volume interactions. This picture is obviously highly approximate, because not all duplexes are of equal length, and the internal loops are not perfectly flexible. (Small bulges and bubbles connecting a pair of duplexes are quite stiff, for example, but the duplexes involved may be regarded as part of a longer duplex; this would imply an effectively larger $\langle k \rangle$ value, but will not change the scaling of $\hat{R}_g$ with N.)

Based on a theorem due to Kramers, the above ideal branched polymer model provides a direct and useful relationship between the secondary structure of ssRNA and its radius of gyration, and hence its 3D size. According to this theorem, the mean square radius of gyration of any ideal polymer (linear or branched) is given by[14,15]

$$\langle \hat{R}_g^2 \rangle = \frac{b^2}{L^2} \sum_{j=1}^{L-1} L_1(j)[L - L_1(j)], \tag{1}$$

where the angular brackets indicate the averaging over all possible spatial conformations of the polymer. The sum on the right hand side extends over all possible divisions of the polymer into two parts, one part containing $L_1(j)$ segments (in our case loops) and the other part containing $L_2(j) = L - L_1(j)$ segments, as illustrated in Fig. 1. Because there are no closed loops in the branched polymer representation of ssRNA, each bond, j, divides the polymer into two distinct sub-structures. The total number of divisions equals the number of bonds, $S \approx L$. The constant b in Eq. (1) is the linear dimension of one polymer segment. For our branched polymer RNA model this unit length may be regarded as the average distance between the centers of two neighboring loops. (Equivalently, the segment length b is the sum of the average duplex length and the average loop diameter.) A crude estimate of this quantity, based on the known bond lengths and bp separations, yields $b \sim 3$ nm (for random sequences with uniform base composition).[25,26] Better estimates of b are both difficult and unwarranted, especially in view of the fact that its numerical

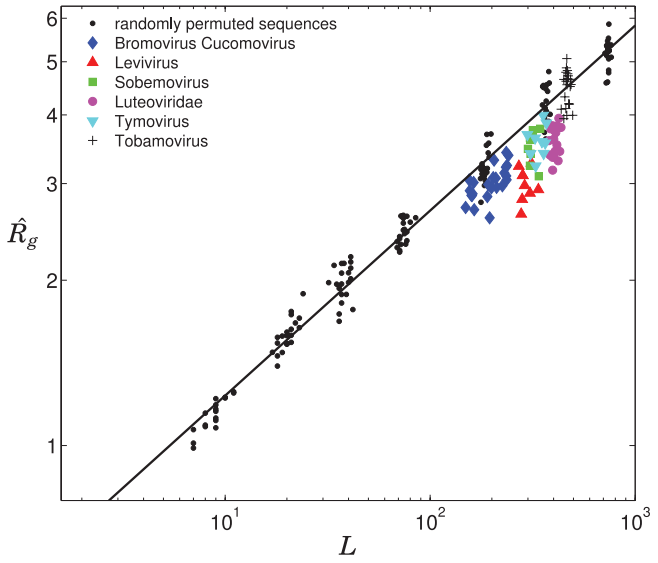

FIG. 2. A log-log plot of the radius of gyration, $\hat{R}_g$ (in units of segment length, b), as a function of the number of loops in the secondary structure, L. The black dots correspond to the minimum free energy structures of random ssRNA of uniform base composition. Each of the 7 clusters of dots includes the results obtained for 20 randomly shuffled sequences of the same length, $N = 100, 250, 500, 1000, 2500, 5000, 10000$ nucleotides. The slope of the straight line is 0.33 ± 0.01. Also shown are the radii of gyration calculated for viral RNAs belonging to 6 viral families, all of which involve icosahedral capsids[27,28] except for the tobamoviruses whose capsids are cylindrical.[29]

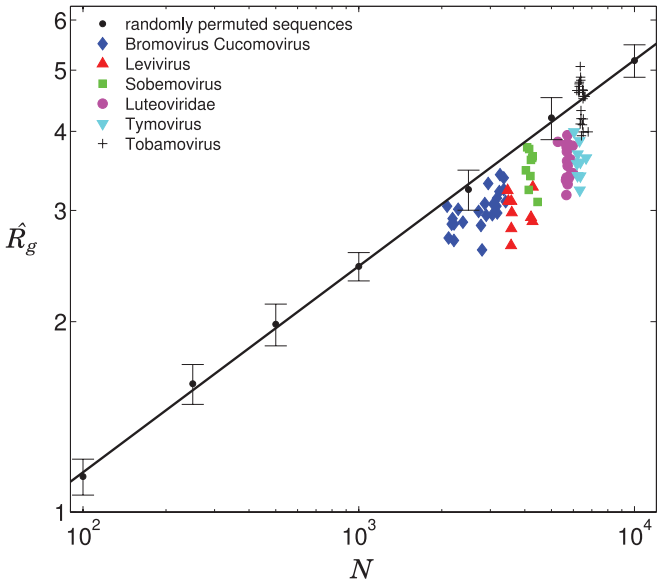

FIG. 3. A log-log plot of the radius of gyration, $\hat{R}_g$ (in units of segment length, b), as a function of RNA sequence length, N. Each black dot represents the average result obtained for 20 randomly shuffled sequences of equal base composition. Other notation as in Fig. 2. The slope of the straight line is 0.33 ± 0.01.

value does not affect the scaling law describing the increase of the (root mean square, rms) radius of gyration $\hat{R}_g \equiv \sqrt{\langle R_g^2 \rangle}$ with N.

III. RESULTS

Figure 2 shows $\hat{R}_g$ as a function of L for two sets of ssRNAs. The first set includes random sequences composed of $N = 100, 250, 500, 1000, 2500, 5000$, and 10 000 nt, all with uniform base composition, (25% of G, C, A, and U). For each value of N we have sampled 20 randomly shuffled sequences, using the same shuffling algorithm as in earlier work;[20] sampling more than 20 sequences changes average $\hat{R}_g$ values and their standard deviations by only a few percent. The MFE secondary structures for each sequence was calculated using the RNA-fold algorithm,[4] and the $\hat{R}_g$ values of these structures were calculated using the tree graph mapping of Fig. 1 and Eq. (1). (The base-pairing predictions of RNA folding algorithms are known to deteriorate with increasing length, especially for molecules with thousands of nucleotides, but RNA-fold is expected nevertheless to provide a good estimate of coarse-grained, large-scale, properties such as the dependence of $\hat{R}_g$ on sequence length, N, or the number of loops, L). The values of L corresponding to the 20 MFE structures of a given N are clustered around an average value $\bar{L}$, which increases with N. (For large N the increase is linear, because $\bar{L} = N(1 - \bar{f})/\overline{\langle l \rangle} = N\bar{f}/2\overline{\langle k \rangle}$, and as noted above $\bar{f}$ and $\overline{\langle k \rangle}$ approach constant values, independent of N; see, for example, Ref. 19 and references cited therein. In addition to the random sequences, Fig. 2 shows the similarly computed radii of gyration of RNAs corresponding to those of several viral families, whose average base compositions are nearly uniform (approximately 24% G, 22% C, 27% A, and 27% U, except for the Tymoviruses whose compositions vary in the range 15%–18% G, 32%–42% C, 17%–24% A, and 22%–29% U).

In Fig. 3 we show the radius of gyration as a function of sequence length, N, for the same random and viral RNA sequences considered in Fig. 2. From both figures it is seen clearly that the RNAs of icosahedral viruses, as indicated by their radii of gyration, are distinctly more compact than the random sequences of the same length. The Tobamoviruses, whose $\hat{R}_g$ is similar to the radius of gyration of random RNA of equal sequence length, strengthen this conclusion because they are the only ones whose capsids are cylindrical rather than icosahedral. Similar conclusions have previously been reached[20] by comparing another structural characteristic of ssRNAs, namely, their MLD, as discussed briefly below.

Further, Figs. 2 and 3 reveal a linear dependence of $\log \hat{R}_g$ on $\log L$ and $\log N$, respectively, for the random ssRNAs. Moreover, the slopes of the two lines are identical, 0.33 ± 0.01. That is,

$$\hat{R}_g \sim L^{\alpha} \sim N^{\alpha}, \quad \alpha = 0.33 \pm 0.01. \tag{2}$$

Recalling that $\bar{L} = N(1 - \bar{f})/\overline{\langle l \rangle}$ it follows that the ratio $(1 - \bar{f})/\overline{\langle l \rangle}$ should be a constant independent of N. Indeed, as already mentioned above, both $\bar{f}$ and $\overline{\langle l \rangle}$ approach asymptotically constant values. More explicitly, $\bar{f} \approx 0.6$ and $\overline{\langle l \rangle} \approx 9.3$, implying $\bar{L} \approx 0.043N$.

It is interesting to consider the relationship between $\hat{R}_g$ and another measure of RNA size—the "maximum ladder distance", ζ. Bundschuh and Hwa[7] introduced the *ladder* distance between two arbitrary base pairs, i and j, in a given secondary structure of a ssRNA molecule, as the number of bp crossed in traversing the shortest path connecting i and j. The

MLD, ζ, is the *longest* ladder distance among all such paths. Equivalently, ζ is the largest number of bp rungs crossed along the shortest route between any two hairpins loops. For instance, the MLD corresponding to the short sequence in Fig. 1(a) is the path (highlighted in grey) leading from the extreme left hairpin to the extreme right hairpin, yielding $\zeta = 38$. In a previous study, we have used the MLD as a measure of the "extendedness" of viral RNAs vs. non-viral (random and yeast) RNAs.[20] We found, consistent with the results in Figs. 2 and 3, that the RNAs of icosahedral viruses are indeed more compact than the non-viral ones.

The MLD is a property of the *secondary* structure of ssRNA, and therefore not a direct measure of the 3D size of the molecule. Nevertheless, we have conjectured[20] that such a measure could be derived by treating the MLD contour as the backbone of a *linear* polymer chain, with the loops regarded as flexible joints and the duplexes as rigid bonds, ignoring all side-branches from the MLD contour. Based on a comprehensive set of MLD calculations, involving averaging over Boltzmann ensembles of structures for each of many randomly shuffled sequences corresponding to each of many sequence lengths N, it was found that, to a very good approximation, the MLD scales with sequence length as $\langle\zeta\rangle \sim N^{2/3}$; more precisely, the exponent of N was found to be 0.67 $\pm$ 0.01.[20] Treating the MLD contour as an *ideal* linear polymer it thus follows that $\hat{R}_g \sim \langle\zeta\rangle^{1/2} \sim N^{1/3}$, identical to the scaling relationship found above for the *ideal branched polymer* model that we have used here to calculate $\hat{R}_g$.

IV. DISCUSSION

A procedure similar to the above analysis has previously been suggested to calculate the radius of gyration of an ideal *randomly* branched polymer,[18] with $\hat{R}_g$ regarded as the mean physical distance between two arbitrary endpoints of the branched polymer. The path connecting the endpoints is treated as a linear chain whose segments are the links between adjacent branching points. By an elegant construction (of a random 1D walk surrounding the entire polymer), the number of links along the path is shown to scale as $N^{1/2}$, implying the familiar result $\hat{R}_g \sim N^{1/4}$—originally derived by Zimm and Stockmayer.[16] As is well known, this model is physically unrealistic for large N, because it predicts a diverging density $\rho \sim N/\hat{R}_g^3 \sim N^{1/4}$, reflecting the neglect of excluded volume effects (see below). Such effects are also not included in our tree graph model of RNA, which predicts $\hat{R}_g \sim N^{1/3}$ (and hence $\rho \sim N/\hat{R}_g^3 \sim 1$). Alternatively, and in accordance with our treatment of ssRNA as a branched polymer whose constituent segments are loops (connected by duplexes), the prediction $\hat{R}_g \sim L^{1/3}$ implies a constant density of loops.

The scaling behavior, $\hat{R}_g \sim N^{1/3}$, found here for random-sequence RNAs using our tree-graph model, is qualitatively different from the $\hat{R}_g \sim N^{1/4}$ scaling law of the ideal *randomly* branched polymer. Since both models assume ideal behavior (i.e., no excluded volume effects) we conclude that the branching pattern of ssRNA is distinctly *non*-random, corresponding to a less compact and less branched structure than that of the *randomly* branched polymer. Indeed, the folded structures of long random RNAs involve generally a hierarchy of loops of different orders, d, where d is the number of duplexes connected to the loop in question. Numerical analyses of many sequences[30] reveal that $L(d)$, the number of loops of degree d, peaks at $d = 2$ (corresponding to "bubbles" and "bulges"), decreasing rapidly as d increases; e.g., the average total number of loops for sequences of length $N = 6000$ is $L \approx 460$, with $L(d) \approx 109, 256, 84, 10, 1$ for $d = 1, 2, 3, 4, 5$, respectively. (The maximum loop degree, $d_{\max}$, increases slowly (logarithmically) with sequence length N.) The highest order loops are naturally found somewhere in the middle of the folded structure, serving as the origin of several (namely $d \approx d_{\max}$) branches emanating outwards. Each branch is, in turn, the folded structure of the smaller RNA molecule corresponding to the sequence comprising this branch. It is of course folded as well, but, being shorter, its maximum loop degree is smaller. This pattern continues until the branches and hence the loops become too small to enable the formation of a new generation of arms.[19] Thus, rather crudely, and only on average, the folded sequence of ssRNA may be regarded as a self-similar structure consisting of several folding generations. Whether this picture can explain why $\hat{R}_g \sim N^{1/3}$ is still an open question.

Many studies have been concerned with the role of excluded volume interactions in randomly branched polymers, revealing a significant increase of the scaling exponent ν, from $\nu = 1/4$ of the ideal polymer to $\nu \approx 0.5$ for the self-interacting polymer, see, e.g., Refs. 31–34. Theoretical analyses of excluded volume interactions in ssRNA are scarce and, at present, are limited to coarse-grained molecular dynamics simulations of relatively short sequences, no longer than several hundreds of nucleotides. In a recent study of this kind by Hajdin *et al.*,[35] involving a series of short ($N = 27 - 161$) RNA sequences, a scaling exponent of $\nu \approx 0.41$ was obtained for $\hat{R}_g$. They also comment that their results are consistent with the earlier work of Hyeon *et al.*[36] (who found $\nu \approx 0.33$ from direct calculation of $\hat{R}_g$ using PDB data) provided one excludes the very short ($N \leq 24$) and the (long, but protein-complexed) 16S and 23S ribosomal RNAs from the RNA structures analyzed by these authors. Interestingly, returning to our interpretation of the MLD of ssRNA as the contour length of a *real* (i.e., self-avoiding) linear polymer, the result predicted for the radius of gyration would have been $\hat{R}_g \sim \langle\zeta\rangle^{3/5} \sim N^{2/5}$, instead of the $\hat{R}_g \sim \langle\zeta\rangle^{1/2} \sim N^{1/3}$ dependence corresponding to the ideal polymer. Fortuitously perhaps, the 2/5 exponent is practically identical to the value $\nu \approx 0.41$ obtained for short RNAs by Hajdin *et al.*[35] It must be remembered, however, that these latter conclusions pertain to structures involving a small number of (about 5–15) duplexes and loops, and considerable scatter in base composition and in radii of gyration. Accordingly, it will be important to extend these approaches to molecules as long as thousands of nt. More generally, much work remains to be done on the effects of branching on the large-scale 3D structures of long ssRNAs.

ACKNOWLEDGMENTS

We thank Dr. Aron Yoffe and Dr. Ajay Gopal for many helpful discussions. The financial support from the Israel

Science Foundation (ISF grants 695/06 and 1448/10) and the Archie and Marjorie Sherman Chair (to A.B.S.), and the US National Science Foundation (Grant No. CHE07-14411 to W.M.G.) are gratefully acknowledged.

[1]M. Zuker, Nucleic Acids Res. **31**(13), 3406 (2003).
[2]I. L. Hofacker, P. Schuster, and P. F. Stadler, Discrete Appl. Math. **88**(1–3), 207 (1998).
[3]D. H. Mathews, RNA **10**(8), 1178 (2004).
[4]I. L. Hofacker, W. Fontana, P. F. Stadler, L. S. Bonhoeffer, M. Tacker, and P. Schuster, Monatsh. Chem. **125**(2), 167 (1994).
[5]R. Nussinov and A. B. Jacobson, Proc. Natl. Acad. Sci. U.S.A. **77**(11), 6309 (1980).
[6]D. Thirumalai and C. Hyeon, Biochemistry **44**(13), 4957 (2005).
[7]R. Bundschuh and T. Hwa, Phys. Rev. E **65**(3), 031903 (2002).
[8]M. Bon and H. Orland, Physica A **389**(15), 2987 (2010).
[9]J. Flinders and T. Dieckmann, Prog. Nucl. Magn. Reson. Spectrosc. **48**(2–3), 137 (2006).
[10]J. Lipfert, J. Ouellet, D. G. Norman, S. Doniach, and D. M. J. Lilley, Structure **16**(9), 1357 (2008).
[11]K. M. Weeks, Curr. Opin. Struct. Biol. **20**(3), 295 (2010).
[12]M. A. Jonikas, R. J. Radmer, A. Laederach, R. Das, S. Pearlman, D. Herschlag, and R. B. Altman, RNA **15**(2), 189 (2009).
[13]C. Hyeon and D. Thirumalai, Proc. Natl. Acad. Sci. U.S.A. **102**(19), 6789 (2005).
[14]H. A. Kramers, J. Chem. Phys. **14**(7), 415 (1946).
[15]M. Rubinstein and R. H. Colby, *Polymer Physics*, 1st ed. (Oxford University Press, Oxford, 2003).
[16]B. H. Zimm and W. H. Stockmayer, J. Chem. Phys. **17**(12), 1301 (1949).
[17]P. G. Degennes, Biopolymers **6**(5), 715 (1968).
[18]K. A. R. Grosberg and A. Yu., *Statistical Physics of Macromolecules* (American Institute of Physics, New York, 1994).
[19]L. T. Fang, A. M. Yoffe, W. M. Gelbart, and A. Ben-Shaul, J. Phys. Chem. B **115**, 3193 (2011).
[20]A. M. Yoffe, P. Prinsen, A. Gopal, C. M. Knobler, W. M. Gelbart, and A. Ben-Shaul, Proc. Natl. Acad. Sci. U.S.A. **105**(42), 16153 (2008).
[21]A. M. Yoffe, M. Prinsen, W. M. Gelbart, and A. Ben-Shaul, Nucleic Acid Res. **39**(1), 292 (2011).
[22]R. I. Dima, C. Hyeon, and D. Thirumalai, J. Mol. Biol. **347**(1), 53 (2005).
[23]M. H. Bailor, X. Y. Sun, and H. M. Al-Hashimi, Science **327**(5962), 202 (2010).
[24]A. Lescoute and E. Westhof, RNA **12**(1), 83 (2006).
[25]D. P. Aalberts and N. O. Hodas, Nucleic Acids Res. **33**(7), 2210 (2005).
[26]P. Taylor, F. Rixon, and U. Desselberger, Virus Res. **2**(2), 175 (1985).
[27]C. M. Shepherd, I. A. Borelli, G. Lander, P. Natarajan, V. Siddavanahalli, C. Bajaj, J. E. Johnson, C. L. Brooks, and V. S. Reddy, Nucleic Acids Res. **34**, D386 (2006).
[28]C. M. Fauquet, M. A. Mayo, J. Maniloff, U. Desselberger, and L. A. Ball (Eds.), *Virus Taxonomy: VIIIth Report of the International Committee on Taxonomy of Viruses* (Elsevier Academic Press, San Diego, 2005), p. xii.
[29]P. R. Whitfeld and T. J. V. Higgins, Virology **71**(2), 471 (1976).
[30]A. M. Yoffe, "Predicting new biophysical properties of nucleic acids," Ph.D. dissertation, UCLA, 2009.
[31]S. Redner, J. Phys. A **12**(9), L239 (1979).
[32]T. C. Lubensky and J. Isaacson, Phys. Rev. A **20**(5), 2130 (1979).
[33]M. Daoud, P. Pincus, W. H. Stockmayer, and T. Witten, Macromolecules **16**(12), 1833 (1983).
[34]A. M. Gutin, A. Y. Grosberg, and E. I. Shakhnovich, Macromolecules **26**(6), 1293 (1993).
[35]C. E. Hajdin, F. Ding, N. V. Dokholyan, and K. M. Weeks, RNA **16**(7), 1340 (2010).
[36]C. Hyeon, R. I. Dima, and D. Thirumalai, J. Chem. Phys. **125**(19), 194905 (2006).