RAPID COMMUNICATIONS

PHYSICAL REVIEW A VOLUME 44, NUMBER 12 15 DECEMBER 1991

Excluded-volume effects on the stacking of RNA base pairs

Ariel Fernández
Department of Chemistry, University of Miami, Coral Gables, Florida 33124
and Department of Biochemistry and Molecular Biology, The Medical School, P.O. Box 016129, Miami, Florida 33101
(Received 5 June 1991)

We validate the implementation of a path-integral formalism to compute the thermodynamics of stacking of RNA base pairs, restricting the analysis to intramolecular interactions. We estimate the effective contact volume, an empirical parameter that determines the excluded-volume effects. This is done by comparing the analytical result with the energetics of duplex growth that takes place after a loop closure or nucleation event. As an illustrative example, the derivation is specialized to the polymer $(CG)_n$, $n\rightarrow\infty$ (C denotes cytosine, G denotes guanosine) in the limit of infinitely long tails.

PACS number(s): 87.15.Da, 36.20.Ey

The path-integral approach to derive the configurational statistics of polymers was introduced as a theoretical tool by Edwards [1]. Its usefulness in the realm of polymer folding is contingent upon a proper evaluation of the empirical parameters used to construct the Hamiltonian. This is precisely the obstacle that has hindered progress in the calculation of measurable thermodynamic quantities from first principles. In this work we shall illustrate a possible theoretical strategy to estimate the empirical parameters that determine the effect of intramolecular interactions. Our derivation is based on the assumption that the standard model Hamiltonian for an elastic chain with excluded-volume interactions and subject to configurational constraints [2] is adequate to account for the reduction of configurational entropy due to the growth of intrachain duplexes.

The statistical mechanics of a Gaussian chain, rooted in random flight statistics, have proven to be unsatisfactory to account for the thermodynamics of duplex growth which follows after the nucleation event of loop closure [3]. This inadequancy becomes most apparent in RNA folding, since water is a relatively good solvent for RNA. Therefore, a treatment that assumes conditions close to a θ regime, a regime where solvation makes excluded-volume effects negligible, would be unrealistic. Furthermore, excluded-volume effects should be far more conspicuous in RNA than in proteins, since the charge-to-mass ratio for RNA is far higher. It should be emphasized, however, that to the best of the author's knowledge, these effects have not been estimated and a direct connection to the actual energetics of RNA folding has never been drawn. In his work we shall show that progress in this direction could be achieved once it is recognized that excluded-volume interactions play a crucial role in determining the architecture of RNA structure. Thus, we shall evaluate the effective contact volume Ω, the crucial quantity to model excluded volume interactions. This parameter, also known as "tolerance volume" [4,5], refers to the volume surrounding a given residue within which another nonadjacent residue must be to qualify as being in contact with the first one.

For the sake of illustrating the point and for clarity of exposition, our results will be specialized to the case of the RNA polymer $(CG)_n$, $n\rightarrow\infty$ (C denotes cytosine, G denotes guanosine), where intrachain interactions between nonadjacent residues lead to the formation of Watson-Crick C-G base pairs. The treatment presented in this work could be applied to any long polymer, provided the thermodynamic parameters are made available for each particular stacking alternative which would arise. Thus, the choice of a periodic polymer is fully justified since it narrows down the number of stacking possibilities. On the other hand, the choice of the periodic polymer $(AU)_n$ (A denotes adenosine; U denotes uracil) might be rather risky, since the relative uncertainty in the thermodynamic parameters is larger in this case when compared with the G-C analog [6]. This is so since A-U base pairing is almost an order of magnitude weaker than G-C pairing and the contribution from stacking of the A-U hydrogen-bond doublets is more difficult to estimate than the stacking of the G-C triplets [3].

We shall assume that a loop located far away from the ends of the chain is formed as a result of a long-range contact denoted (i,j). The numbers i and j are specific values of a discrete contour variable giving the location of a residue along the chain. If $N=2n$ denotes the chain length, the contour variable takes any integer value in the interval $[0,N]$. The loop closure event will be regarded as a nucleation event. Subsequent events preserving Watson-Crick complementarity and leading to duplex growth are the result of progressive stacking of base pairs. Without loss of generality we may assume the first such event to be the formation of the contact $(i-1,j+1)$. Thus, the events that follow the formation of the contact (i,j) minimize the further decrease in configurational entropy for they are adjacent to the contact which has initiated the formation of the duplex. In other words, given that contact (i,j) has already occurred, the minimal additional loss of entropy will occur if the subsequent contact is a neighboring contact. Since the enthalpy change due to Watson-Crick pairing is negligible in water, we may conclude that stacking is the dominant force driving duplex growth and that the latter is essentially an entropically driven process. The hydrogen bonding in a single base pair is marginally advantageous, from the perspective of enthalpic contributions, with respect to hydrogen bonding

between the individual bases and water molecules [3,6]. Nevertheless, the stacking of H-bond triplets in G-C base pairs is far more effective (more easily organized) than the stacking of doublets in A-U pairs [3,6].

Given this scenario, the crucial quantity accessible from statistical mechanics is the conditional equilibrium constant $K(i-1,j+1|i,j)$, that is, the equilibrium constant for a process is defined as follows: Given an existing base pair (i,j), a new base pair $(i-1,j+1)$ adjacent to it is formed. This quantity may be computed by noting that

$$K(i-1,j+1|i,j)=K(i-1,j+1;i,j)/K(i,j)\,, \quad (1)$$

where $K(i-1,j+1;i,j)$ is the equilibrium constant for the occurrence of both contacts (i,j) and $(i-1,j+1)$ when the "reactant" is a random coil; and $K(i,j)$ is the equilibrium constant for the formation of the contact (i,j) from a random coil. Thus, we may write

$$K(i,j)=Q(i,j)/Q\,, \quad (2)$$

$$K(i-1,j+1;i,j)=Q(i-1,j+1;i,j)/Q\,, \quad (3)$$

where Q is the partition function for a chain of length N, $Q(i,j)$ is the partition function for the chain subject to the constraint that there exists the contact (i,j) and $Q(i-1,j+1;i,j)$ is the partition function subject to the constraints imposed by the existence of the two contacts (i,j) and $(i-1,j+1)$. The evaluation of $Q(i,j)$ could be made perturbatively, making use of the diagrammatic methods associated to Feynman-like propagators [2,4]. However, since we wish to explore the limit $N\to\infty$, we must resort to the well-established renormalization approach [5] for it would be incorrect to regard the quantity $v_0N^{1/2}$ (v_0 is the strength of excluded-volume interactions, $v_0\approx 0.25$ [4]) as a small quantity. Our case of interest is the limit of long tails which could be transcribed in the following conditions: (a) $i,N-j\to\infty$ as $N\to\infty$ and (b) $j-i\gg 1$ but remains finite as $N\to\infty$. The renormalization result in this case is well known [5]:

$$K(i,j)=\Omega[3/2\pi(j-i)]^{2.42}\,. \quad (4)$$

The exponent in Eq. (4) is different from that resulting from the Jacobson-Stockmayer treatment, which yields the value 1.50 [4]. This discrepancy is expected, since the latter treatment is rooted in random flight statistics and, therefore, it does not incorporate excluded-volume effects.

The equilibrium constant $K(i-1,j+1;i,j)$ may be evaluated making use of the perturbative rules developed by Chan and Dill [4]. The general result to $O(v_0)$ for configurations subject to two constraints determined by two contacts not necessarily adjacent is

$$K(i-1,j+1;i,j)=\Omega^2(3/2\pi)^3[l_2(l_1+l_3)]^{-3/2}\times[1+v_0/(2\pi)^{3/2}h]\,, \quad (5)$$

where the parameters contained in Eq. (5) are defined in Fig. 1 and h is given by

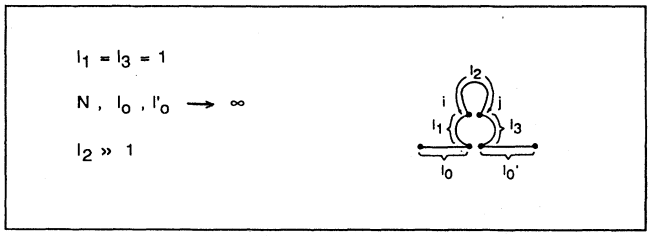

FIG. 1. Parameters describing the location of a generic given contact (i,j) and the relative location of a predicted contact. The limit case of infinitely long tails and adjacent contacts may be specified in terms of the parameters as shown in the figure.

$$h=4\left\{\pi\left[\frac{l_2(l_1+l_3)}{l_1+l_2+l_3}\right]^{1/2}+\sqrt{N}-(l_0+l_0')^{1/2}+(l_1+l_3)^{1/2}\left[\cos^{-1}\left[\frac{l_1+l_3}{l_1+l_2+4l_0}\right]^{1/2}+\cos^{-1}\left[\frac{l_1+l_3}{l_1+l_2+4l_0'}\right]^{1/2}\right]+(l_2)^{1/2}\left[2\sin^{-1}\left[\frac{l_2(l_1+l_3)}{l_2(l_1+l_3)+4l_1l_3}\right]^{1/2}-\sin^{-1}\left[\frac{l_2(l_1+l_3)}{(l_2+4l_0)(l_1+l_3)+4l_1l_3}\right]^{1/2}-\sin^{-1}\left[\frac{l_2(l_1+l_3)}{(l_2+4l_0')(l_1+l_3)+4l_1l_3}\right]^{1/2}\right]\right\}. \quad (6)$$

Specializing this result to the particular situation of adjacent base pairs, as indicated in the limit case specified in Fig. 1, we obtain

$$h=(4\sqrt{l_2}+\sqrt{128})\pi\,. \quad (7)$$

Combining Eqs. (1), (4), (5), and (7) we may now proceed to estimate the parameter Ω. Equation (1) is linked to the free-energy change $\Delta G^0_{\text{stack}}$, corresponding to the stacking of a G-C base pair adjacent to the existing one. This link may be materialized via the relation

$$\Delta G^0_{\text{stack}}\approx -RT\ln K(i-1,j+1|i,j)\,. \quad (8)$$

This relation is independent of the specific location of the loop in the chain provided the limit conditions defined in Fig. 1 are satisfied.

The free energy of a single duplex growth event has been measured [3,6] at room temperature and standard conditions, and it is approximately equal to -4.3 kcal. Thus, making use of Eq. (8), we may estimate the effective contact volume at $\Omega\approx 8488$ Å^3.

From this calculation, it might seem that the effective contact distance, of the order of 13 Å, is too large when compared with any meaningful bond distance. However, the results do not defy intuition, for it must be remem-

bered that the polymer string is solvated by the polar water molecules through its highly hydrophilic backbone of charged phosphate residues. Therefore, the "effective thickness" of the string is far larger than that of an *in vacuo* string.

The estimation presented in this work might pave the way for future attempts at using path-integral treatments to predict biologically meaningful interactions in RNA. In particular, work is in progress in our laboratory to predict the occurrence of RNA folding motifs with catalytic potential.

It is essential to emphasize that the relatively simple calculation presented in this work became only possible after our awareness that base pair stacking is essentially an excluded volume effect, when casted in the proper statistical mechanical framework.

A stacking effect might be detectable among the bases of noncovalently connected residues [3]. However, such stacking is not colinearly organized and, resulting mainly from solvophobic effects, would be far weaker. This form of stacking is entrophically driven, but since the residues are not covalently linked, the entropic contribution is far more modest when compared to the polymer situation.

The author would like to acknowledge support from the Camille and Henry Dreyfus Foundation.

[1] S. F. Edwards, Proc. Roy. Soc. London **85**, 613 (1965).
[2] A. Fernández, Chem. Phys. Lett. **154**, 396 (1989).
[3] C. Cantor and P. R. Schimmel, *Biophysical Chemistry* (Freeman, San Francisco, 1980).
[4] H. S. Chan and K. A. Dill, J. Chem. Phys. **90**, 492 (1989).
[5] J. L. Martin, M. F. Sykes, and F. T. Hioe, J. Chem. Phys. **46**, 3478 (1967).
[6] D. H. Turner, N. Sugimoto, and S. M. Freier, Ann. Rev. Biophys. Biophys. Chem. **17**, 167 (1988).