RNA STRUCTURE ANALYSIS VIA THE RIGID BLOCK MODEL

by

MAURICIO ESGUERRA NEIRA

A dissertation submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Chemistry and Chemical Biology

Written under the direction of

Wilma K. Olson

and approved by

_____

_____

_____

_____

New Brunswick, New Jersey

May, 2010

ABSTRACT OF THE DISSERTATION

RNA Structure Analysis via the Rigid Block Model

by Mauricio Esguerra Neira

Dissertation Director: Wilma K. Olson

RNA structure is at the forefront of our understanding of the origin of life, and the mechanisms of life regulation and control. RNA plays a primordial role in some viruses. Our knowledge of the importance of RNA in cellular regulation is relatively new, and this knowledge, along with the detailed structural elucidation of the transcription machine, the ribosome, has propelled interest in understanding RNA to a level which starts to closely resemble that given to proteins and DNA.

In the process of progressively understanding the landscape of functionality of such a complex polymer as RNA, one practical task left to the structural chemist is to understand the details of how structure relates to large-scale polymer processes. With this in mind the fundamental problems which fuel the work described in this thesis are those of the conformations which RNA's assume in nature, and the aim to understand how RNA folds.

The RNA folding problem can be understood as a mechanical problem. Therefore efforts to determine its solution are not foreign to the use of statistical mechanical methods combined with detailed knowledge of atomic level structure. Such methodology is mainly used in this work in a long-term effort to understand the intrinsic structural features of RNA, and how they might relate to its folding.

As a thing among things, each thing is equally insignificant; as a world each one equally significant.

If I have been contemplating the stove, and then am told; but now all you know is the stove, my result does indeed sound trivial. For this represents the matter as if I had studied the stove as one among the many, many things in the world. But if I was contemplating the stove, it was my world, and everything else colorless by contrast with it ...

For it is equally possible to take the bare present image as the worthless momentary picture in the whole temporal world, and as the true world among shadows.

Ludwig Wittgenstein

As a molecule among molecules, each molecule is equally insignificant; as a world each one equally significant.

If I have been contemplating RNA, and then am told; but now all you know is RNA, my result does indeed sound trivial. For this represents the matter as if I had studied RNA as one among the many, many molecules in the world. But if I was contemplating RNA, it was my world, and everything else colorless by contrast with it ...

For it is equally possible to take the bare present image as the worthless momentary picture in the whole temporal world, and as the true world among shadows.

Anonymous Chemist

Acknowledgements

I would first like to give a special thanks to Dr. Yurong Xin, whose patience, help, and collaboration since the very beginning of my joining of the Olson lab have been fundamental for the development of this work. I would like to thank Dr. Olson's extreme patience, and room for freedom on carrying out this research. Finally I thank all colleagues at the Olson lab.

I would like to dedicate this thesis to David and Stella Case, without them these words would not exist.

Table of Contents

List of Tables

List of Figures

Chapter 4

RNA Base Pair Steps

4.1 Analysis (Albany Poster) and Django Webserver

Results shown in Albany and steps part of methods paper.

This gives us the force constant matrices per base-step which are used in the next section.

4.2 Persistence Length of RNA

A quantity commonly used to quantify the stiffness of polymers is the so-called persistence length a. To determine this quantity for DNA or RNA a variety of theoretical and experimental techniques are used. Some common experimental techniques to determine a are Electron Microscopy (EM), gel electrophoresis, sedimentation velocities, electrical birefringence Atomic Force Microscopy (AFM) , Magnetic Tweezers, and Small Angle X-Ray Scattering (SAXS). For reviews of such techniques applied to the determination of RNA persistence length, we refer the reader to Hagerman [**?**], Abels et al. [1], and Caliskan et al. [2]. We will use their results for comparison with those coming from the "realistic" model developed by Olson and collaborators [3] to describe DNA. The "realistic" model is dependent on high resolution crystallographic data. Initial studies started with small numbers of data for the deformabilities of the ten unique base-pair steps [3]. A more complete picture applied to the study of DNA sequence dependent deformability became available in 1998 [4]. The base-pair step deformability data for DNA has been constantly refined as more high resolution DNA and DNA-protein structures have been added to the Nucleic Acid Database (NDB) [5]. Although such data has been available for DNA since 1998, it had not been so for RNA, until now [6].

A detailed description of the "realistic model" along with the scheme of the C++ code developed by Czapla and Zheng to implement it, and a brief account of various definitions of persistence length and models from which a can be derived are included in Appendix D

4.3 AMBER: Persistence Length of Base-Pair Step Patterns

I guess it needs some input here in order to work on latex compilation.

References

[1] Abels, J. A., Moreno-Herrero, F., van derHeijden, T., Dekker, C., and Dekker, N. H. (2005) Single-Molecule Measurements of the Persistence Length of Double-Stranded RNA. *Biophysical Journal,* **88**, 2737–2744.

[2] Caliskan, G., Hyeon, C., Perez-Salas, U., Briber, R. M., Woodson, S. A., and Thirumalai, D. (2005) Persistence Length Changes Dramatically as RNA Folds. *Phys Rev Lett,* **95**, 268303.

[3] Olson, W. K., Babcock, M. S., Gorin, A., Liu, G., Marky, N. L., Martino, J. A., Pedersen, S. C., Srinivasan, A. R., Tobias, I., and Westcott, T. P. (1995) Flexing and Folding Double Helical DNA. *Biophysical Chemistry,* **55**, 7–29.

[4] Olson, W. K., Gorin, A. A., Lu, X.-J., Hock, L. M., and Zhurkin, V. B. (1998) DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proceedings of the National Academy of Sciences,* **95**, 11163–11168.

[5] Balasubramanian, S., Xu, F., and Olson, W. K. (2009) DNA Sequence-Directed Organization of Chromatin: Structure-Based Computational Analysis of Nucleosome-Binding Sequences. *Biophysical Journal,* **96**, 2245–2260.

[6] Olson, W. K., Esguerra, M., Xin, Y., and Lu, X.-J. (2009) New Information Content in RNA Base Pairing Deduced from Quantitative Analysis of High-Resolution Structures. *Methods,* **47**, 177–186.

Appendix D

Persistence Length

Nucleic Acids and other polymers can be understood as mechanical objects [1, 2] and therefore engineering approaches can be used for their understanding. Usually the methods followed by the engineering approach consider the polymer as a long continuous rod, and are known as continuum elastic theory. This type of approach leaves little space for taking into account the nature of the subunits which make up the polymer, that is, it is mainly applicable to homopolymers made up of identical subunits. In nucleic acids this is not necesarily the case, a more general approach should take into account the possibility of having different subunits making up the polymer. Olson and collaborators have developed a sequence dependent model, refered to as the "realistic" model [3]. Such model is harmonic and depends on determination of force-constant analogs derived from X-Ray cristallographic data taken from the Nucleic Acid Database (NDB) [4, 5]. Within the context of the "realistic" model Czapla et al. [6] have suggested a gaussian sampling methodology which allows the determination of global polymer properties like the persistence length a following a matrix approach suggested by Flory [7] and expanded upon by Olson et al. [**?**, **?**]

In what follows we summarize various definitions of persistence length and how it's computed using different models.

D.1 Persistence Lenght Definitions

There are two parallel perspectives which can be used to define the persistence length of nucleic acids. One is a more mathematical, or physical one, where it is understood as the resistence to deformation of a curve in space, or thin rod, a physical object. The other one is a stochastic one where it is understood as "a measure of the distance over which the direction of the DNA is maintained [8]" and has the classical formulation by Flory which states that the persistence length is:

"the average sum of the projections of all bonds $j \geq i$ on bond i in an indefinetely long chain. The bond i is taken to be remote from either end of the chain, i.e., $1 \ll i \ll n$". Paul J. Flory, Statistical

Mechanics of Chain Molecules. 1969

This perspective has a more chemically flavored tone, since it assumes some type of bonded connectivity between polymeric units, whether the bond is a "real" one, or a "virtual" one.

“**Bend-persistence length**:
A length scale beyond which the elastic cost of bending is totally negligible” Philip Nelson, Yearbook of Science and Technology, McGraw Hill 1999

“In a randomly shaken rod any particular point in the rod will be pointing in a random direction, but nearby points will be pointing in roughly the same direction, that is, these nearby points are persistent. Points farther away than the bend-persistence length are said to be uncorrelated.” Philip Nelson, Yearbook of Science and Technology, McGraw Hill 1999

“**twist-persistence length**:
that is, a rubber rod not only resists bending but also twisting”

"basic mechanical property that quantifies stiffness" Abels et al, Biophysical Journal, 2005, 2737-2744

"Classical elasticity tells us that a thin, straight rod that is bent into an arc has a bending energy $E = Bl/2R^2$, where B is the bending elastic constant of the rod, l is the length of the rod and R is the radius of arc. Setting $R = l$ gives us the energy of a 1 radian bend along the rod, and solving for when $E\kappa_B T$ gives us the length of rod along which a thermally excited bend of 1 radian typically occurs: $lB/\kappa_B T$. This is called the persistence length..." John F. Marko and Simona Cocco, Physics World, March 2003

$$E = \frac{Bl}{2R^2} \tag{D.1}$$

$$R = l \tag{D.2}$$

$$E \sim k_B T l \sim \frac{B}{k_B T} \tag{D.3}$$

"The persistence length a is a measure of the stiffness of a polymer chain and is related to the limiting value of the characteristic ratio at infinite chain length

$$a = \frac{\nu}{2}(C_\infty + 1) \tag{D.4}$$

"

"length at which the orientation of the sequential bonds which make up a polymer chain, stop being correlated. That is, if you have just two bonds, or a few, they will be correlated, which is the case in most molecules, but, in polymers, you have a long chain of sequential bonds. At some length, bonds will become uncorrelated, but up to that length they were correlated, this is what is meant by persistence length, and, in this context it's obvious that is an exclusive property of polymers." My understanding so far.

Biopolymers can be either rigid or flexible.

They can be classified according to whether their persistence length (a) is greater, smaller, or similar to the contour length (L) of the polymer.

Model Type	Polymer Characteristic	a to L relation
Rigid Rod	Rigid	$a \gg L$
Gaussian chain	Flexible	$a \ll L$
Worm-like chain	Semi-flexible	$a \approx L$

Notice that for $a \gg L$, there is a definition problem, since L has to be large enough to be a good approximation to the definition of persistence length, which is defined for an infinite chain length.

Worm-like-chain = Porod-Kratky = Freely Rotating Chain in limit l=0 and n=infinity

Rigid biopolymers: actin, microtubules

Flexible biopolymers:

Semi-flexible biopolymers: High force extension DNA.

If the persistence length is of the same order of the length of the polymer, then the polymer is classified as semi-flexible

Think about the "energy" based perspective of Nicolas, and the stochastic based perspective of Flory and others.

Polymer	a (nm)
α-helix	80-100
coiled-coil	150-300
Ideal DNA	51
Ideal RNA	70-80

Table D.1: Persistence lengths for some biopolymers with filament structures.

D.2 end-to-end

The end-to-end vector r is the vector which connects the ends of a polymer chain. It can be defined as the sum of the vectors connecting the monomer units in a chain. These connecting vectors are sometimes called virtual bond vectors l.

From the end-to-end vector the quantity which is usually of interest is it's magnitude.

$$r = \sum_{i=1}^{n} l_i \tag{D.5}$$

$$r^2 = r \cdot r = \sum_{i,j} l_i \cdot l_j \tag{D.6}$$

Equation D.5, can also be written:

$$r^2 = \sum_{i} l_i^2 + 2 \sum_{i \neq j} l_i \cdot l_j \tag{D.7}$$

To describe a polymer it's necessary to think about the various conformations it can adopt due to its flexibility, therefore, it is important to think of polymer related quantities in terms of the average of their possible conformations. For the end-to-end vector the average of its values is denoted as $< r >$, and the average of its norm, also called the second moment of the end-to-end distribution, is denoted by $< r^2 >$:

$$< r^2 >= \sum_{i} < l_i^2 > +2 \sum_{i<j} < l_i \cdot l_j > \tag{D.8}$$

When there is no correlation between succesive bonds we can write:

$$< l_i \cdot l_j >= 0 \tag{D.9}$$

So that equation D.8 keeps only the bond auto-correlation term:

$$< r^2 >= \sum_i < l_i^2 >= n < l^2 > \tag{D.10}$$

This equation is used to describe a so-called freely-jointed chain.

D.3 Models

Nelson in book says:

$$dE = \frac{1}{2} K_B T [A\beta^2 + Bu^2 + C\omega^2 + 2Du\omega] ds \tag{D.11}$$

A$\kappa\beta$ T = Bend stiffness B$\kappa\beta$ T = Stretch stiffness C$\kappa\beta$ T = Twist stiffness D$\kappa\beta$ T = Twist-stretch coupling

If only the bend stiffness survives then the model is called an inextensible model, also Porod-Kratky, or WLC.

D.3.1 Kuhn - Freely Jointed Chain (FJC)

D.3.2 Porod-Kratky - Worm Like Chain (WLC)

D.3.3 Olson - Realistic

The Hamiltonian for a [9]

D.4 Suggested Reads

From Equilibrium Statistics of Plischke and Bergersen they suggest to read: Des Cloiseaoux and Janik () Rubinstein and Colby (Polymer Physics)

References

[1] Marko, J. F. and Cocco, S. (2003) The Micromechanics of DNA. *Physics World,* **16**, 37–41.

[2] Nelson, P. (2004) Biological Physics: Energy, Information, Life, W. H. Freeman and Company, .

[3] Olson, W. K., Marky, N. L., Jernigan, R. L., and Zhurkin, V. B. (1993) Influence of Fluctuations on DNA Curvature. A Comparison of Flexible and Static Wedge Models of Intrinsically Bent DNA. *Journal of Molecular Biology,* **232**, 530–554.

[4] Go, M. and Go, N. (1976) Fluctuations of an Alpha-Helix. *Biopolymers,* **15**, 1119–1127.

[5] Olson, W. K., Gorin, A. A., Lu, X.-J., Hock, L. M., and Zhurkin, V. B. (1998) DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proceedings of the National Academy of Sciences,* **95**, 11163–11168.

[6] Czapla, L., Swigon, D., and Olson, W. K. (2006) Sequence-dependent effects in the cyclization of short dna. *Journal of Chemical Theory and Computation,* **2**, 685–695.

[7] Flory, P. J. (1969) Statistical Mechanics of Chain Molecules, Interscience Publishers, .

[8] Olson, W. K., Babcock, M. S., Gorin, A., Liu, G., Marky, N. L., Martino, J. A., Pedersen, S. C., Srinivasan, A. R., Tobias, I., and Westcott, T. P. (1995) Flexing and Folding Double Helical DNA. *Biophysical Chemistry,* **55**, 7–29.

[9] Czapla, L. The Statistical Mechanics of Free and Protein-Bound DNA by Monte Carlo Simulation PhD thesis Rutgers, The State University of New Jersey (2009).