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 3

RNA Base-Pairing

3.1 Canonical and Noncanonical Base-pairs

As seen in Figure 1.2, there can be various base-pairing patterns between heterocyclic bases in nucleic acids due to a variety of possible hydrogen bonding interactions. The most prevalent hydrogen bonding pattern is known as canonical Watson-Crick, all other possible patterns are known as non-canonical base-pairs and are more common in RNA than in DNA. We used 3DNA to find all base-pairs in a non-redundant database of X-ray determined RNA structures from the PDB with resolutions less than or equal to 3.5 Å. We also constrained our search to helical regions in RNA. Such helical regions are composed of 3 consecutive base-pairs or more, and they need not be covalently bonded by the sugar-phosphate backbone between consecutive base-pairs. For more details the reader is refered to Olson et al. [1].

In the helical regions data we quantify:

Abundances (Counts) Deformabilites Helical Context

NON-REDUNDANT DATABASE AND CONSTRAIN TO HELICAL REGIONS.

We use a non-redundant dataset of RNA structures. By non-redundant we mean to say that, for the main source of RNA structural information, which is the ribosome, we used only one of the available structures per organism, that is, one for each of *Deinococcus Radiodurans*, *Haloarcula marismortui*, *Escherichi coli*, and *Thermus thermophilus*.

3.2 Clustering of Yurong's Classification

RNA Type	Counts	G	C	A	U
small helices	78	891	753	404	442
drug-RNA	36	932	862	365	433
protein-RNA	207	4001	3457	1771	1731
protein-tRNA	9	175	155	98	87
rRNA	13	3866	2949	1939	1785
tRNA	13	205	159	124	112
ribozyme	113	2434	2086	1438	1150
Total	469	12504	10421	6139	5740

Table 3.1: Classification of RNA Types in Non-Redundant Dataset at less than 3.5 Å (For Base-Pairs in Helices of 3 base-pairs or more).

References

[1] 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.