Mathematical and Biological Scientists Assess the State-of-the-Art in RNA Science at an IMA Workshop *RNA in Biology, Bioengineering and Biotechnology*

Tamar Schlick[§]

Department of Chemistry and Courant
Institute of Mathematical Sciences
New York University, 251 Mercer Street
New York, New York 10012

Running title: IMA Workshop on RNA

[§]To whom correspondence should be addressed (Phone: 212-998-3116; e-mail: schlick@nyu.edu; fax: 212-995-4152)

May 21, 2009

Abstract

Highlights of the IMA workshop *RNA in Biology, Bioengineering, and Biotechnology* are summarized, including recent developments in RNA secondary structure prediction and RNA design, innovative mathematical constructs for RNA structure, bioinformatics advances in RNA structure analysis and prediction, and experimental progress in RNA folding and imaging.

In small things considered is the world revealed.

Robert Lee Hotz, Feb. 2008.

1 Introduction

Though Wall Street Journal reporter Hotz was referring to stunning recent discoveries concerning regulation of monarch butterfly migration by two light-sensing genes that serve as the monarch's timing device — to trigger northern journeys in the spring and southern returns in the fall (Zhu et al., 2008) — the expanding world of small RNAs is perhaps more aptly described in terms of such global ramifications. Indeed, the emergence of numerous classes of non-coding RNAs that critically regulate gene expression has brought to the fore a renewed interest in RNA biology. As a result, scientists are intensely investigating these new functions of RNA and rapidly exploiting RNA's newly realized powers for exciting applications in bioengineering, therapeutics, and nanotechnology.

Currently, however, most advances in RNA science are dominated by experimental approaches. Theoretical contributions by modeling and simulation have certainly contributed to these advances, but their systematic application has been hampered by realized difficulties in RNA structure classification, prediction and modeling, including the rugged energy landscape of RNAs, the associated multiple pathways, conformational dependence on divalent ions, and the dearth of solved RNA tertiary structures compared to proteins. Still, a very active and devoted community of RNA bioinformaticians has arisen as well as formed an RNA Ontology Consortium (ROC) to help define standards for RNA investigations and propel this young field.

For a week in late Fall 2007, about 200 such researchers from the US and abroad that included mathematicians, biologists, computer scientists, chemistry, bioengineers, and physicists convened at the University of Minnesota's Institute of Mathematics and its Applications for a workshop entitled "*RNA in Biology, Bioengineering and Biotechnology*". Organized by Tamar Schlick (New York University) and Eric Westhof (University of Strasbourg) with invaluable assitance by IMA staff, the workshop sought to bring a blend of both experts, with extensive hands-on experience on RNA research, and novices from allied fields who were interested in this young and vibrant field of immense scientific importance. The IMA's mission is to immerse and stimulate different scientists from the mathematical and other disciplines to explore new territories and ideas of research in novel, interdisciplinary, and collaborative ways. These developments are made possible by the intensity of the activities both inside and outside the lecture room which encourage the scientists to participate and be part of a community with shared interests and goals.

Our esteemed biopolymer subject generated exciting new themes, approaches, and problems concerning RNA structure analysis, design, and prediction, all of which were intensely discussed through-

out the workshop, from the lecture room to informal meetings during the week and several special cultural events. Indeed, the mathematicians became acquainted by current biological problems in the RNA field, including regulatory and catalytic RNAs, RNA folding, dynamics, and engineering, while biologists were exposed to new approaches for viewing and studying RNA using seemingly theoretical mathematical constructs from graph theory or topology, or Feynman diagrams from physics.

In this short report, we highlight some of the topics presented in the workshop, while revealing some interesting discussion themes and pointing to areas for future research developments. Full workshop information, including schedule, abstracts, and podcasts can be obtained via the IMA website: http://ima.umn.edu/2007-2008/W10.29-11.2.07/.

2 RNA Secondary Structure Prediction

Michael Zuker (Rensselaer Polytechnic Institute), the developer of the widely used prediction program Mfold, recently replaced by the more comprehensive UNAfold package, opened the meeting with an overview of computational methods for RNA secondary structure determination, both by comparative analysis (sequence alignment) and free energy minimization via empirically derived energy parameters based on basic physiochemical laws (Zuker and Stiegler, 1981; Zuker, 2003) (Fig. 1). Zuker emphasized that problems in prediction can occur since some high-probability base pairs are not correct while correct ones have very low probability. Moreover, in homologous RNAs, high probability base pairs in one sequence may correspond to low-probability base pairs in another. Thus, caution is warranted in using Rfam data and interpreting computational data.

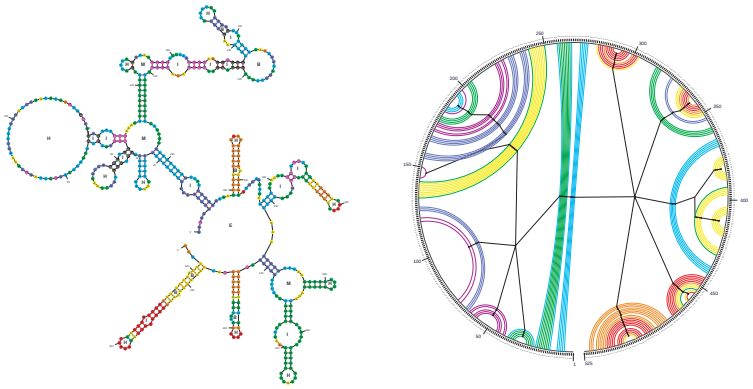

Figure 1: Representations of RNA secondary structures – traditional tree structure (left, from Bill Zimmerly's website http://www.fp.ucalgary.ca/group2introns/species.htm) and circle plot (right) of group II intron from Microscilla species. Figure is taken with permission from Zuker's presentation and made by UNAfold using mfold_util.

David Mathews (University of Rochester) described the dynamic programming method Dynalign by his group to find the lowest free-energy secondary structure conformation by simultaneously using several structural predictions with sequence alignment data (Mathews et al., 1999; Mathews and Turner, 2002). This approach can make calculations tractable for long sequences.

The subsequent discussions highlighted that for riboswitches multiple conformations will have various free energies (thus low entropy by itself is not a sufficient criterion) and that sequence alignment would be much improved with a probabilistic framework attached to the alignment, a difficult but ongoing area of research. Both talks emphasized the challenges that remain in secondary structure prediction for RNA and the unique conformational and functional properties of RNAs that can make predictions difficult.

3 RNA Design

The RNA design session consisted of fascinating lectures and discussions by Shapiro, Jaeger, Pierce, and Inoue.

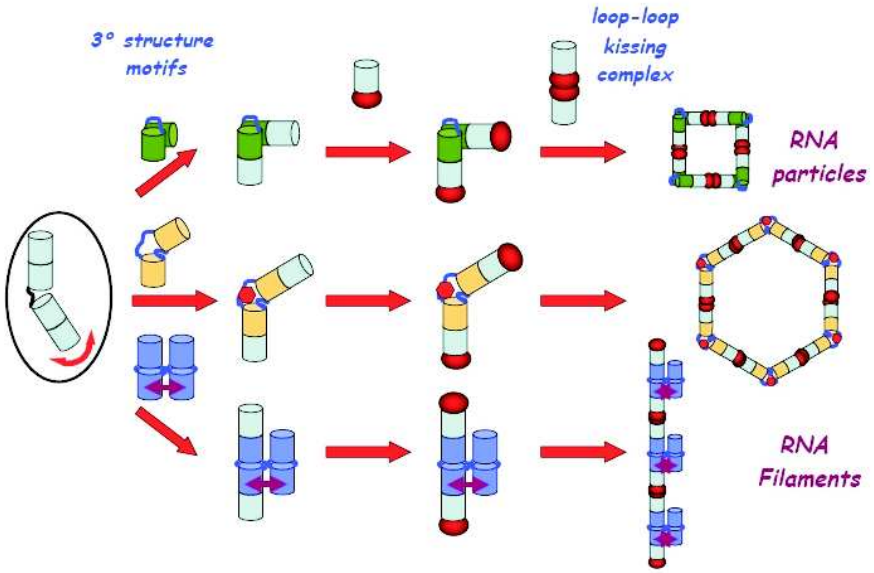

Figure 2: Schematic diagram of various RNA design to build novel nano-RNAs (RNA particles and RNA filaments) for self-assembling building blocks from RNA tertiary motifs. Figure is taken with permission from Jaeger's presentation.

Bruce Shapiro (National Cancer Institute) described a large suite of tools that allows users to design RNA-based nanoparticles with desired functionalities (Yingling and Shapiro, 2007; Bindewald et al., 2008). In particular, the new relational database RNAJunction reveals many low as well as some

high-order junctions (e.g., up to 9) in which RNAs can self assemble. In conjunction with NanoTiler, computational modeling and ingenuity can yield many design constructs.

Luc Jaeger (University of California, Santa Barbara) showed how systematic annotation and analysis of ribosomal motifs can define a 'proto-language' he coined *RNA architectonics* to build novel nano-RNAs for self-assembling building blocks (Chworos et al., 2004; Jaeger and Chworos, 2006) (Fig. 2). The versatility of RNAs can lead to many envisioned usages in medicine and technology.

Niles Pierce (California Institute of Technology) described construction of versatile new materials, such as biosensors, by programming synthetic nucleic acid systems to self-assemble using equilibrium prediction approaches for the interacting nucleic acid strands (Dirks et al., 2004; Yin et al., 2008).

Tan Inoue (Kyoto University) focused on RNA/protein constructs. He described design and construction of a self-folding RNA scaffold that can be re-wired by combining it with a ribozyme or protein element in a reaction site (Ikawa et al., 2004; Saito and Inoue, 2008). He demonstrated how in silico and in vitro tools can be combined to develop a prototype for a multi-functional RNA. For example, the designed RNA with functional protein could be built to recognize an antibody in a cancer cell and halt a biological reaction or help image cellular processes. Of course, validation of this idea is a future goal. His group has accumulated RNP motifs for such varied design objectives."*Evolution is cleverer than you are*" (Leslie Orgel), the speaker reminded us, highlighting the challenges that face RNA designers.

4 The Varied Mathematician's Toolkit

Talks by mathematical scientists highlighted the many innovative approaches from mathematics that can be applied to RNA studies.

Jes Frellsen (University of Copenhagen) described a probabilistic approach that models RNA's conformational space as continuous rather than discrete states using a Dynamic Bayesian Network for modeling backbone and base dihedral angles (Hamelryck et al., 2006). The advantage of the approach is its rapid probabilistic sampling of locally RNA-like structures, as demonstrated for five different RNA target structures compared to 1500 decoys. Besides applications to structure prediction, the model can be used to validate experimentally determined structures.

Henri Orland (Saclay) described a topological classification of RNA pseudoknots by using the concept of topological genus in matrix field theory (Vernizzi et al., 2005) (Fig. 3). In combination with exact enumeration of RNA structure (albeit simplified) and MC simulations, pseudoknots in RNA databases can be identified.

Asamoah Nkwanta (Morgan State University) presented a way to describe an RNA secondary structure as a random walk, a lattice path (Ndifon and Nkwanta, 2006). His group's goal is to use such

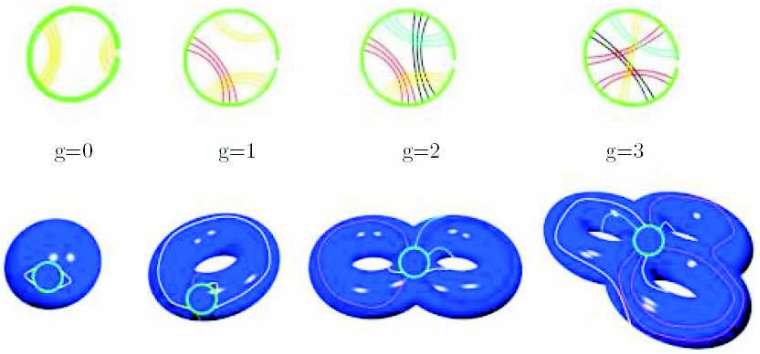

Figure 3: Topological classification of RNA pseudoknots using the concept of topological genus (g) in matrix field theory. Figure is taken with permission from Orland's presentation.

combinatorics constructs to enhance research on biochemical and chemical sensors.

Yongwu Rong (George Washington University) suggested how to use Feynman diagrams (which can code secondary structure) to describe RNA folding from transition polynomials (Rong and Luse, 2006). Namely, a partition function summed over all possible Feynman diagrams can give the probability distribution of conformational states.

Christine Heitsch (Georgia Institute of Technology) described a combinatorial graph theory approach (Heitsch et al., 2002) to model the base pairing of large RNA sequences, such as the 9400 base Hepatitis C viral genome (HCV) whose secondary structure is available from the laboratory of A. C. Palmenberg at UW Madison (http://www.virology.wisc.edu/acp/RNAFolds/RNAFolds_hep.html) (Fig. 4). The mathematical model yields local and global constraints in the secondary structure model and suggests ways to study viral genomes.

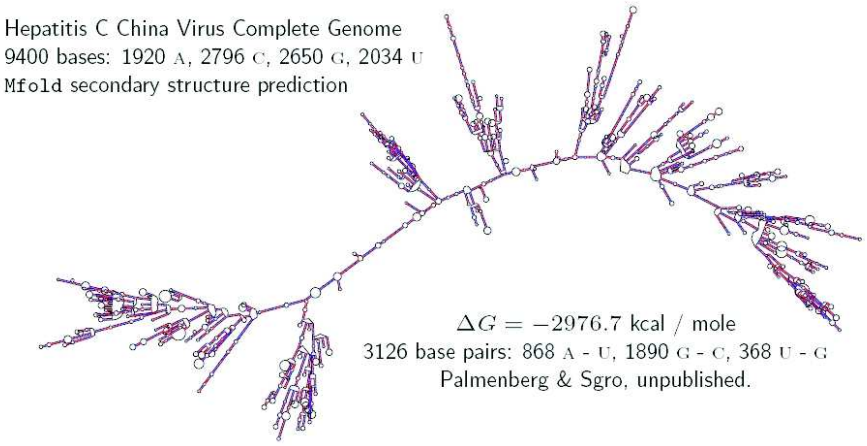

Figure 4: RNA viral genome secondary structure. Figure is taken with permission from Heitsch's presentation.

Building on her lab's expertise in protein structure prediction, **Ruth Nussinov** (Tel Aviv University and the National Cancer Institute) described an algorithm (ARTS) and database (DARTS) for multiple-alignment of RNA tertiary structures and assembly of the fold repertoire of RNAs (Dror et al., 2006).

Ivo Hofacker (University of Vienna) described prediction applications specific to RNA/RNA co-structures. Such interaction partners are important to identify because most recently discovered ncRNAs interact with other RNAs (e.g., microRNAs regulate mRNA translation) (Muckstein et al., 2006). In his group's approach (termed RNAup), efficient hybridization of two RNAs is computationally assessed by an interaction energy to describe complementarity and a probability for base pairing. The program uses available folding algorithms and structural information to define such potential interactions and local accessibility of the RNAs. An interesting possible application is to predict siRNA binding to microRNAs.

Jakob Pedersen (University of Copenhagen) described the application of phylogenetic models combined with classical algorithms such as sequence alignments (program EvoFold) to screen multiple sequences of genomes of vertebrates and Drosopholids in search for structural RNAs, namely transcribed regions with functional structure (Pedersen et al., 2006).

Jens Reeder (University of Bielefeld) described a graphical approach for motifs search (Locomotif) that employs dynamic programming algorithms and thermodynamic models of RNA secondary structure formation and executes them on a readily available bioinformatics server which includes the C-language source code: http://bibiserv.techfak.uni-bielefeld.de/locomotif (Reeder et al., 2007).

Namhee Kim and **Hin Hark Gan** (New York University) described a new approach for enhancing in vitro selection of novel RNAs (Kim et al., 2007a,b). Such selection experiments are widely performed using random sequence pools to search for RNAs with desired properties (such as catalysis or binding). However, such applications are limited because yields can be small, especially for complex RNAs. To accelerate such discoveries, the group has proposed using mathematical expressions for describing the sequence permutation or synthesis procedure (Fig. 5). Such processes can then be optimized in terms of the reactants (nucleotide ratios entered in the different experimental vials) to optimize a desired yield, expressed in terms of RNA secondary structures. The optimization is made possible by using mixing matrices for the nucleotide ratios and graph theory representations of RNA secondary structures. The optimization task then becomes a simple linear problem. This pool design procedure has been automated on a web-server RAGPOOLS (http://rubin2.biomath.nyu.edu) to aid researchers to perform in vitro selection experiments more efficiently.

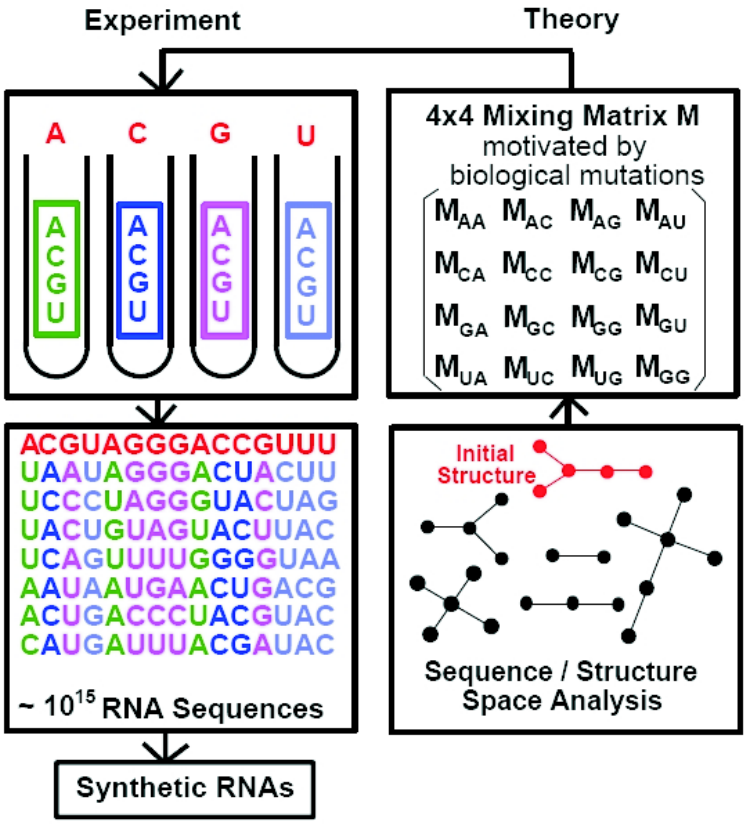

Figure 5: Schematic diagram of modeling RNA in vitro selection. RNA pools are designed by optimization of mixing matrices and analysis of pool structural distributions using tree graphs. Figure is taken with permission from Kim and Gan's presentation.

5 Experimental Advances in RNA Structure and Processes

The workshop also benefited from expositions by several leading experimentalists who presented new discoveries concerning RNA structure and processes.

Tao Pan (University of Chicago) described recent findings concerning *RNA pausing* during transcriptional folding that serves to guide the folding process (Wong et al., 2007). The phylogenetically-conserved pause sites in these ncRNAs are located between the upstream and downstream portions of the native long-range helix products and, in a such a way, constrain RNA segments to guide the correct folding. In response to the many questions that followed, Pan described at least two types of observed pausing and emphasized the many aspects of pausing yet to be discovered.

Tobin Sosnick (University of Chicago) described new advances of electron cryo-microscopy in tandem with single-particle reconstruction to visualize small RNAs such as 154 residues (≈50kD), about two times smaller than typical limits (Baird et al., 2005). He showed how divalent-ion-dependent intermediates in the folding of the catalytic domain RNase P RNA can be imaged when stably popu-

lated. He proposed that in combination with molecular modeling (with experimental constraints, such as from SAXS and chemical mapping) and all-atom simulations, cryo-EM could be a rapid alternative to RNA crystallography.

Steven Brenner (University of California, Berkeley) presented recent findings showing that all conserved members of the SR family of splice regulators have an unproductive alternative mRNA isoform targeted for decay. Because this splice pattern is conserved in mouse and associated with highly conserved regions that are common to mouse and human, this natural mode of regulation may be biologically significant (Lareau et al., 2007).

Structural biologists **Sarah Woodson** (Johns Hopkins University) and **David Lilley** (University of Dundee) illuminated the attendees on the complex RNA folding process by describing recent dynamics and folding analysis of ribozymes (Koculi et al., 2007; Wilson et al., 2007) (Fig. 6).

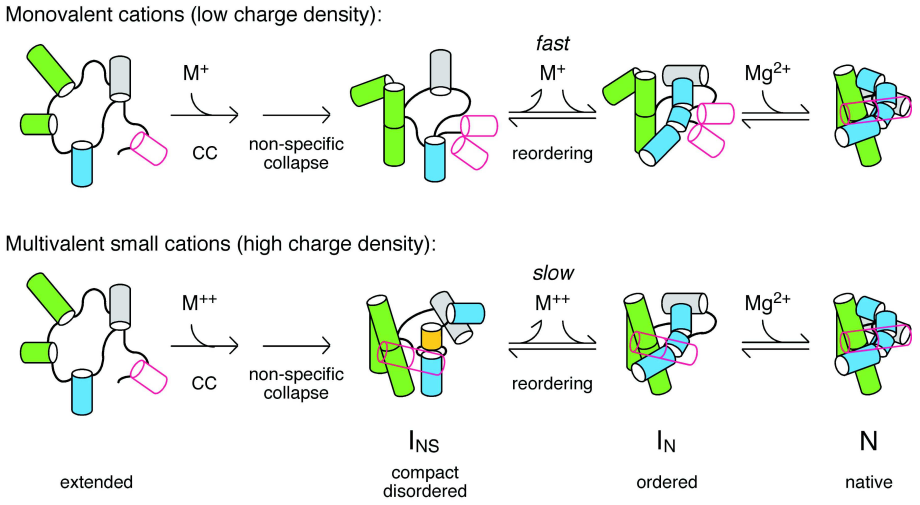

Figure 6: RNA folding and dynamics. Monovalent ions allow fast dynamics. Figure is taken with permission from Woodson's presentation.

Woodson described experiments in which varied cation type and quantity as well as ribozyme sequence permutations helped capture intermediates along the pathway and understand ion effects on the pathway and the intermediates that form (Koculi et al., 2007). For example, small multivalent counterions stabilize the RNA more than the large monovalent ions and, in low charge densities, the transition state ensemble becomes broader and accelerates native structure formation.

Lilley described structural studies using the FRET technique of two catalytic RNAs that generate site-specific cleavage by trans-esterification reactions, the hairpin and VS ribozymes (Wilson et al., 2007). Experiments for the hairpin ribozyme capture multiple cycles of cleavage and ligation and measure rates of internal reactions. For the large VS ribozyme (for which no crystal is available), low-resolution SAXS helped propose a three-dimensional structure of the ribozyme using modeling and

intuition. Mutations of the two nucleotides critical to the folding process led to more than 1000-fold impairment in catalysis. The team is working on solving the structure by crystallography.

Theoretician **Devarajan Thirumalai** (University of Maryland) described complementary computational studies that similarly sought to explore the energy landscape of folding. Already for a hairpin formation, complex kinetics already emerge with multiple pathways (Hyeon and Thirumalai, 2008).

6 RNA Structure Analysis

The workshop ended with a series of talks on recent innovative computational efforts for analyzing RNA secondary and tertiary interactions and ultimately predicting RNA structures. Many of these works are performed in the scope of the RNA Ontology Consortium (ROC, http://roc.bgsu.edu/) (Leontis et al., 2006), an effort that seeks to unify RNA structural description and present to the community standard tools for annotating, analyzing, and predicting RNA structures.

Francois Major (University of Montreal) described an approach developed and enhanced over several years to predict RNA secondary and tertiary structures using a geometric approach based on nucleotide cycle building blocks (motifs) (St-Onge et al., 2007; Parisien and Major, 2008). Combined with sequence alignments and low-resolution data, these building block definitions can be incorporated to predict RNA conformation states.

Neocles Leontis and **Craig Zirbel** from Bowling Green State University described new structural analysis studies of ribosomal RNAs that lead to definitions of RNA motifs (using symbolic notation) and RNA assembly patterns that help identify RNA motifs more broadly (Sarver et al., 2008). The suite of programs in FR3D (http://rna.bgsu.edu/FR3D/) can help annotate RNAs and perform various searches for recurrent RNA motifs (see also Fig. 7 and (Xin et al., 2008)).

Eric Westhof (University of Strasbourg) described fundamental concepts that have accumulated regarding RNA architecture and reactivity (Brion and Westof, 1997; Lescoute and Westof, 2006). He emphasized the hierarchical assembly of RNA and the versatility of A-minor motifs that are mutationally robust. He illustrated RNA's versatility by describing the assembly of the bacterial ribosomal decoding A site in which a dynamic equilibrium between a bulging or hidden adenines contributes to the docoding process during codon/anitcodon recognition. In contrast, specificity stemming from helical lengths, helical stacking, and junction assembly help define a global native fold for many RNAs. He expressed the hope that RNA three-dimensional structure analysis, sequence alignment, and motif annotation will help derive common evolutionary rules in structural RNAs that could ultimately be used in three-dimensional structure prediction.

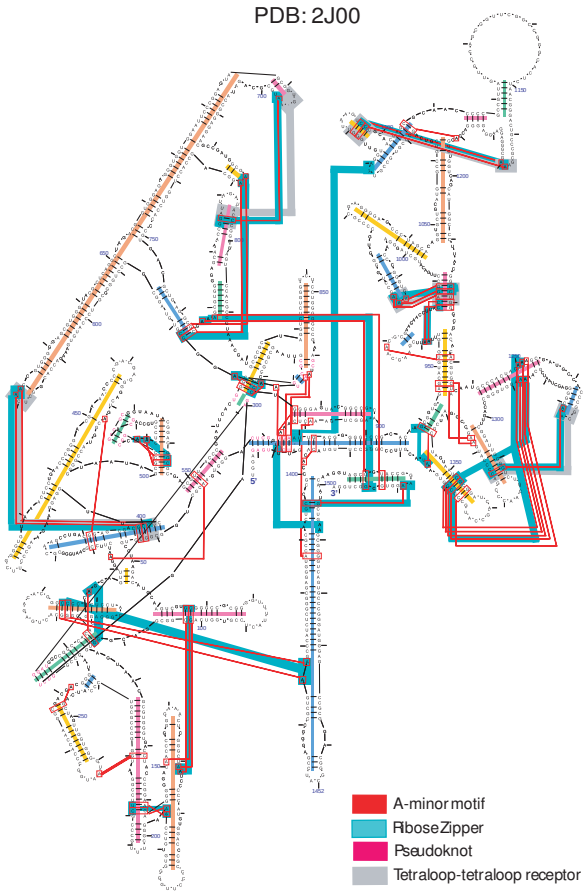

Figure 7: Annotation of tertiary motifs in 16S rRNA structure using FR3D and other programs. Coaxial stacking between helices are formed and stabilized by a variety of long-range interactions (e.g., ribose zipper and A-minor motif) that work in a cooperative way. Figure is provided by the Schlick group.

7 Concluding Remarks

Clearly, the large as well as small RNAs pose many open questions that require close collaborations between the experimentalists and the modelers. With the growing number of genome databases, computational tools for RNA analysis, prediction, and design will be critically needed. As demonstrated in the workshop, many innovative approaches and productive collaborations already are in place. Thus, there is no reason not to expect that RNA science will catch up to protein endeavors in the near future. In fact, since many functionally important RNAs are quite small, computational strategies should have a profound impact on RNA biology and chemistry, including the exciting applications to medicine and technology.

Acknowledgments

We thank Namhee Kim and Yurong Xin for transcribing the workshop discussions and Cheri Shakiban for her devoted work in helping organize the workshop. We also thank Douglas Arnold, John Baxter, Eve Marofsky, Holly Pinkerton and other devoted staff at the IMA for their assistance.

References

Baird, N., Westof, E., Qin, H., Pan, T. and Sosnick, T., (2005) Structures of a Folding Intermediate Reveals the Interplay Between Core and Peripheral Elements in RNA Folding, *J. Mol. Biol.*, 352, pp. 712–722.

Bindewald, E., Hayes, R., Yingling, Y., Kasprzak, W. and Shapiro, B., (2008) RNAJunction: a Database of RNA Junctions and Kissing Loops for Three-Dimensional Structural Analysis and Nanodesign, *Nucleic Acids Res.*, D36, pp. 392–397.

Brion, P. and Westof, E., (1997) Hierarchy and Dynamics of RNA Folding, *Ann. Rev. Biophys. Biomol. Struc.*, 26, pp. 113–137.

Chworos, A., Severcan, I., Koyfman, A., Weinkam, P., Oroudjev, E., Hansma, H. and Jaeger, L., (2004) Building Programmable Jigsaw Puzzles With RNA, *Science*, 306, pp. 2068–2072.

Dirks, R., Lin, M., Winfree, E. and Pierce, N., (2004) Paradigms for Computational Nucleic Acid Design, *Nucleic Acids Res.*, 32, pp. 1392–1403.

Dror, O., Nussinov, R. and Wolfson, H., (2006) The ARTS Web Server for Aligning RNA Tertiary Structures, *Nucleic Acids Res.*, 34, pp. W412–415.

Hamelryck, T., Kent, J. and Krogh, A., (2006) Sampling Realistic Protein Conformations Using Local Structural Bias, *PLoS Comput. Biol.*, 2, pp. 1121–1133.

Heitsch, C., Condon, A. and Hoos, H., (2002) *From RNA Secondary Structure to Coding Theory: A Combinatorial Approach*, Shapporo, Japan: Springer-Verlag.

Hyeon, C. and Thirumalai, D., (2008) Multiple Probes are Required to Explore and Control the Rugged Energy Landscape of RNA Hairpins, *J. Am. Chem. Soc.*, 130, pp. 1538–1539.

Ikawa, Y., Tsuda, K., Matsumura, S. and Inoue, T., (2004) De Novo Synthesis and Development of an RNA Enzyme, *Proc. Natl. Acad. Sci. USA*, 101, pp. 13750–13755.

Jaeger, L. and Chworos, A., (2006) The Architectonics of Programmable RNA and DNA Nanostructures, *Curr. Opin. Struct. Biol.*, 16, pp. 531–543.

Kim, N., Gan, H. and Schlick, T., (2007a) Computational Proposal of Structured RNA Pools for In Vitro Selection of RNAs, *RNA*, 13, pp. 478–492.

Kim, N., Shin, J., Elmetwaly, S., Gan, H. and Schlick, T., (2007b) RAGPOOLS: RNA-As-Graph-Pools – A Web Server for Assisting the Design of Structured RNA Pools for In Vitro Selection, *Bioinformatics*, 23, pp. 2959–2960.

Koculi, E., Hyeon, C., Thirumalai, D. and Woodson, S., (2007) Charge Density of Divalent Metal Cations Determines RNA Stability, *J. Am. Chem. Soc.*, 129, pp. 2676–2682.

Lareau, L., Inada, M., Green, R., Wengrod, J. and Brenner, S., (2007) Unproductive Splicing of SR Genes Associated With Highly Conserved and Ultraconserved DNA Elements, *Nature*, 446, pp. 926–929.

Leontis, N., Altman, R., Berman, H., Brenner, S., Brown, J., Engelke, D., Harvey, S., Holbrook, S., Jossinet, F., Lewis, S., Major, F., Mathews, D., Richardson, J., Williamson, J. and Westof, E., (2006) The RNA Ontology Consortium: An Open Invitation to the RNA Community, *RNA*, 12, pp. 533–541.

Lescoute, A. and Westof, E., (2006) The Interaction Networks of Structured RNAs, *Nucleic Acid Res.*, 34, pp. 6587–6604.

Mathews, D., Sabina, J., Zuker, M. and Turner, D., (1999) Expanded Sequence Dependence of Thermodynamic Parameters Improves Prediction of RNA Secondary Structure, *J. Mol. Biol.*, 288, pp. 911–940.

Mathews, D. and Turner, D., (2002) Dynalign: an Algorithm for Finding the Secondary Structure Common to Two RNA Sequences, *J. Mol. Biol.*, 317, pp. 191–203.

Muckstein, U., Tafer, H., Hackermuller, J., Bernhart, S., Stadler, P. and Hofacker, I., (2006) Thermodynamics of RNA-RNA Binding, *Bioinformatics*, 22, pp. 1177–1182.

Ndifon, W. and Nkwanta, A., (2006) An RNA Foldability Metric; Implications for the Design of Rapidly Foldable RNA Sequences, *Biophys. Chem*, 120, pp. 237–239.

Parisien, M. and Major, F., (2008) The MC-Fold and MC-Sym Pipeline Infers RNA Structure From Sequence Data, *Nature*, 452, pp. 51–55.

Pedersen, J., Bejerano, G., Siepel, A., Rosenbloom, K., Lindblad-Toh, K., Lander, E., Kent, J., Miller, W. and Haussler, D., (2006) Identification and Classification of Conserved RNA Secondary Structures in the Human Genome, *PLos Comput. Biol.*, 2, p. e33.

Reeder, J., Reeder, J. and Giegerich, R., (2007) Locomotif: From Graphical Motif Description to RNA Motif Search, *Bioinformatics*, 23, pp. i392–400.

Rong, Y. and Luse, K., (2006) Examples of Knots With the Same Polynomials, *J. Knot Theory and Its Ramifications*, 15, pp. 749–749.

Saito, H. and Inoue, T., (2008) RNA and RNP as New Molecular Parts in Synthetic Biology, *J. Biotechnol.*, 132, pp. 1–7.

Sarver, M., Zirbel, C., Stombaugh, J., Mokdad, A. and Leontis, N., (2008) FR3D: Finding Local and Composite Recurrent Structural Motifs in RNA 3D Structures, *J. Math. Biol.*, 56, pp. 215–252.

St-Onge, K., Thinault, P., Hamel, S. and Major, F., (2007) Modeling RNA Tertiary Structure Motifs by Graph Grammers, *Nucleic Acid Res.*, 35, pp. 1726–1736.

Vernizzi, G., Orland, H. and Zee, A., (2005) Enumeration of RNA Structures by Matrix Models, *PLoS Comput. Biol.*, 2, pp. 1121–1133.

Wilson, T., McLeod, A. and Lilley, D., (2007) A Guanine Nucleobase Important for Catalysis by the VS Ribozyme, *EMBO J.*, 26, pp. 2489–2500.

Wong, T., Sosnick, T. and Pan, T., (2007) Folding of Noncoding RNAs During Transcription Facilitated by Pausing-Induced Nonnative Structures, *Proc. Natl. Acad. Sci. USA*, 104, pp. 17995–18000.

Xin, Y., Laing, C., Leontis, N. and Schlick, T., (2008) Annotation of Tertiary Interactions in RNA Structures Reveals Variations and Correlations, *RNA*, 14, pp. 2465–2477.

Yin, P., Choi, H., Calvert, C. and Pierce, N., (2008) Programming Biomolecular Self-Assembly Pathways, *Nature*, 451, pp. 318–322.

Yingling, Y. and Shapiro, B., (2007) Computational Design of an RNA Hexagonal Nanoring and an RNA Nanotube, *Nano Lett.*, 7, pp. 2328–2334.

Zhu, H., Yuan, Q., Casselman, A., Emery-Le, M., Emery, P. and Reppert, S.M., (2008) Cryptochromes Define a Novel Circadian Clock Mechanism in Monarch Butterflies That May Underline Sun Compass Navigation, *PLoS Biol.*, 6, pp. 138–155.

Zuker, M., (2003) Mfold Web Server For Nucleic Acid Folding and Hybridization Prediction, *Nucleic Acids Res.*, 31, pp. 3406–3415.

Zuker, M. and Stiegler, P., (1981) Optimal Computer Folding of Large RNA Sequences Using Thermodynamics and Auxiliary Information, *Nucleic Acids Res.*, 9, pp. 133–148.