Published in final edited form as:

Methods. 2016 July 01; 103: 1–3. doi:10.1016/j.ymeth.2016.06.006.

Advances in RNA structure determination

Bruce A. Shapiro and

RNA Structure and Design Section, National Cancer Institute, Frederick, MD 21702, USA

Stuart F.J. Le Grice

RT Biochemistry Section, National Cancer Institute, Frederick, MD 21702, USA

The recent years have witnessed a revolution in the field of RNA structure and function. Until recently the main contribution of RNA in cellular and disease functions was considered to be a role defined by the central dogma, namely DNA codes for mRNAs, which in turn encode for proteins, a notion facilitated by non-coding ribosomal RNA and tRNA. It was also assumed at the time that less than 2% of DNA in the human genome was used to encode genes, the remainder being considered “junk”. Subsequent research has unequivocally determined that RNA mediates a plethora of functions vital to cellular activity as well as clinically-significant diseases. In turn, it was discovered that the amount of DNA that encodes functional RNAs also increased significantly. It was demonstrated, for example, that RNA can act as an enzyme, which led to the discovery of ribozymes; it can be spliced, leading to many variations in gene expression; RNA can act to silence or mediate protein expression through the RNA interference pathway. This latter discovery opened the door to the discovery of many different types of non-coding RNAs that control gene expression in a multitude of ways. The importance of RNA is further illustrated by the fact that several diseases are caused by RNA viruses, e.g. Ebola, acquired immunodeficiency syndrome, dengue fever, Zika, influenza and the common cold, just to name a few. As our understanding of the importance of RNA in cellular functions and disease progression increases, a new challenge looms on the horizon, namely the epitranscriptome. Although a few RNA modifications have been studied extensively (5’mRNA capping, alternative splicing, and polyadenylation), the functional roles of many other base modifications (greater than 100 of which have been reported) remain unknown, and are likely to influence RNA stability, trafficking, localization, enzymatic/sensing/regulatory activity and patterns of interactions with other molecules. Research in these areas will undoubtedly see RNA emerge as both a therapeutic target and a therapeutic modality. This journal issue describes several of the computational and experimental methodologies that are used to determine RNA structure and function and to use this knowledge for therapeutic purposes.

While highly-efficient methods have been reported for large scale RNA production via in vitro transcription, the availability of labeled (or otherwise-modified) nucleotides and their position-selective introduction into the biomolecule remains a formidable challenge for NMR studies. This limitation has recently been overcome by Liu et al. [1] through their development of position-selective labeling of RNA, or PLOR, which permits introduction of isotopic labels, fluorophores and other chemical modifications in a cost effective manner. A complementary approach of Longhini et al. [2] describes an efficient strategy for synthesis and purification of atom-specific nucleotides for their incorporation into RNA. Such

improvements should permit a detailed analysis of RNA structures considerably larger than those reported to date. As an alternative to NMR, advances in small angle X-ray scattering (SAXS) and atomic force microscopy (AFM), and in particular their application to structurally-complex RNAs are reviewed in this issue. By exploiting a fragment assembly search or through "natural" hierarchical moves that follow from the modular architecture of RNAs, Bhandari et al. [3] demonstrate that SAXS can provide useful outlines of the global shape, which in turn can be leveraged by computational methods to determine 3D topological folds of RNAs. AFM on the other hand, while providing nanometer spatial resolution, has now emerged as a powerful tool to study the dynamics of single- and double-stranded RNAs, as well as RNA based assemblies which are becoming increasingly important as novel unique building blocks in for RNA nanotechnology [4].

For analysis of larger RNAs, such as those of viral genomes, chemical probing remains the method of choice, where the advent of selective 2′ hydroxyl acylation analyzed by primer extension (SHAPE) has both simplified and increased the throughput of RNA analysis by automated capillary electrophoresis. Several SHAPE variations have now been documented, of which the SHAPE-seq strategy of Watters et al. that combines chemical probing with next-generation sequencing is described here [5]. Of the more "traditional" chemical probing strategies, Hulscher et al. describe the application of time-resolved X-ray-dependent hydroxyl radical and dimethyl sulfate (DMS) footprinting to study the kinetics of RNA folding both *in vitro* and *in vivo* [6]. Although such chemical probing techniques can rapidly provide secondary structural prediction, vital information can be gained from complementing them with other biochemical analyses. One such approach, outlined by Stephenson et al. [7], describes the use of single molecule Foerster resonance energy transfer (smFRET). Despite the lower resolution it offers, smFRET can be applied to much larger molecules than NMR, at a single molecule level and in close to physiological conditions of solvent and temperature. Regardless of the approach used and the resolution that structural probing provides, post-transcriptional modification (i.e. the epitranscriptome) illustrates yet another level of structural complexity imposed on RNA. Disorders associated with RNA modification include cancer, obesity, diabetes, hepatitis, neurological disorders, autoimmune disorders, hypoxia, metabolic diseases and viral infections, thus understanding how the "readers" "writers" and "erasers" inherent to RNA modification control such disease processes is critical. The chapter of Yu et al. [8] presents several sensitive and efficient methods and protocols for detecting and quantifying two of the most abundant modifications, namely post-transcriptional 2′-O-methylation and pseudouridylation. Finally, technologies to label RNA *in vivo* by fluorescently labeling proteins that bind and enable its visualization at sites of transcription of a gene in living cells are becoming increasingly popular and have significantly advanced our understanding of transcription dynamics. Methods described herein by Palangat et al. [9] allow visualization of nascent RNA synthesis from a single genomic locus using a single-gene two-color transcription assay system.

The role of computation in RNA structure prediction and analysis should not be underestimated. Given the plethora of issues that need to be modeled and the cost of experiments, the development of computational techniques can and has served the vital interests of the RNA structure community, ultimately forming a synergy between

computational and experimental techniques. Several chapters in this edition deal with this subject. The article of Lorenz et al. [10] reviews the development of algorithms for the prediction or RNA secondary structure. It takes the reader from the development of the first dynamic programming algorithms for minimal free energy structure prediction through the concept of determining ensembles of structures to enable the calculation of base pair probabilities and finally through the use of sequence alignment methods and the use of external probes in conjunction with prediction algorithms to help to constrain the predictions to improve accuracy. An ultimate goal in RNA structure prediction is determining the 3D structure or structures given an RNA sequence and/or a secondary structure. In order to accomplish this task, some approaches make use of known structural motifs that are derived from various experimental techniques (e.g. NMR, X-ray crystallography). Databases of these motifs have been constructed and serve as input into the structure prediction algorithms. Perlea et al. [11] describe efforts to computationally extract, compile and classify a database of three-dimensional RNA motifs currently focused on recurrent hairpin and internal loop structures. This database can be used for RNA nanotechnology, RNA structure prediction from sequence, and enhancing our understanding of RNA structure in general. One can utilize databases of motifs to help predict the three-dimensional structure of RNAs from sequence data. A program utilizing this idea is described by Biesiada et al. [12], who have assembled a database of 3D RNA motif fragments and match these with a secondary structure that is input by the user. These motifs are then merged to form potential 3D models of the given secondary structure. Distance constraints can also be utilized to further refine the predictions. RNA nanostructures can be created utilizing this methodology applying single sequence structure generation.

The design of artificial RNA constructs can serve as platforms to deliver drugs to combat various diseases including cancer. A computational procedure described by Parlea et al. [13] utilizes software that was specifically developed for generating RNA-based nanoparticles that can self-assemble from multiple sequences. The methodology works in reverse of conventional RNA folding programs by first designing a 3D RNA architecture, which in turn is used to generate the set of sequences needed for this architecture to self-assemble. To build the architectures, RNA motifs are used like building blocks and are connected. This methodology was used to combinatorially search for and generate a catalog of ring-like RNA nanostructures that can be used for further nano development. As a proof of principle the experimental assembly of a predicted five-stranded RNA nanoring is presented. Understanding the characteristics of RNA base, sugar and backbone interactions in three-dimensional RNA modeling and folding is crucial to ultimately predicting how RNAs fold and how their dynamics affect function. However, the ability to simulate folding and RNA interactions can be quite computationally time consuming. In the article of Dawson et al. [14], both coarse grained and all atom physical representations of RNAs are discussed. In this context, representations and the determination of force fields become important for accurate models. Coarse-grained representations can accelerate computations, but special techniques are required for the development of force fields. Rather than using pure physical characteristics, the use of databases of structures to derive the statistics of the interactions, a knowledge-based approach is also presented. The issues associated with these different techniques are presented as well as some needs for the future.

A final section of this edition addresses the notion that experimental methodologies used to determine RNA structure and function might be leveraged for therapeutic purposes. One such approach involves the use of short hairpin RNA (shRNA) which, In view of its ability to provide specific, long-lasting, gene silencing has shown considerable promise. However, off-target effects have hindered shRNA usage in many applications, and numerous failures are triggered by pitfalls in shRNA design that are often associated with impoverished biogenesis. Accordingly, the article of Bofill-De Ros et al. [15] discusses principles of shRNA design (pre-miRNA-like, pri-miRNA-like and Ago-shRNA) with emphasis on their structure. Aptamers are highly-structured, single-stranded nucleic acid ligands whose binding properties are comparable to those of antibody/antigen interactions. Their utility as therapeutics has primarily involved the tendency of aptamers to inhibit their target upon binding. In addition, they can be used as targeting agents to deliver therapeutic cargos to cells. Similar to antibodies, aptamers can be designed with high affinity and specificity for their target. However, their relatively small size generally supports better tissue penetration and superior target-to-noise ratios compared to their protein counterparts. In addition, chemically-optimized aptamers are non-immunogenic. A series of articles describes methods for rapid development of aptamers against recombinant, purified protein targets [16], a computational modeling methodology that, when coupled to a functional assay, can help determine key sequence and structural motifs of an RNA aptamer [17] and an aptamer fluorescence binding and internalization (AFBI) assay that facilitates their rapid and high throughput testing on cells [18]. Finally, recognizing that microarray-based ligand screening provides a rapid, straightforward and high-throughput platform complementary to conventional phenotypic high throughput screening, Abulwerdi et al. [19] describe how this screening platform can be adapted to identify small molecule, peptoid and peptide-based RNA-binding ligands.

References

[1]. Liu Yu, Yu Ping, Dyba Marzena, Sousa Rui, Stagno Jason, Wang Yun-Xing, Applications of PLOR in labeling large RNAs at specific sites, Methods 103 (2016) 4–10. [PubMed: 27033177]

[2]. Longhini Andrew P., Regan M LeBlanc, T. Kwaku Dayie, Chemo-enzymatic Labeling for Rapid Assignment of RNA Molecules, Methods 103 (2016) 11–17. [PubMed: 27090003]

[3]. Bhandari Yuba, Bhandari Yuba R., Jiang Wei, Stahlberg Eric A., Stagno Jason R., Wang Yun-Xing, Modeling RNA topological structures using small angle X-ray scattering, Methods 103 (2016) 18–24. [PubMed: 27090001]

[4]. Schön Peter, Imaging and force probing RNA by atomic force microscopy, Methods 103 (2016) 25–33. [PubMed: 27222101]

[5]. Watters Kyle E., Yu Angela M., Strobel Eric J., Settle Alex H., Lucks Julius, Characterizing RNA structures in vitro and in vivo with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq), Methods 103 (2016) 34–48. [PubMed: 27064082]

[6]. Hulscher Ryan, Bohon Jen, Rappé Mollie C., Gupta Sayan, D'Mello Rhijuta, Sullivan Michael, Ralston Corie Y., Chance Mark R., Woodson Sarah A., Probing the structure of ribosome assembly intermediates in vivo using DMS and hydroxyl radical footprinting, Methods 103 (2016) 49–56. [PubMed: 27016143]

[7]. Stephenson James D., Kenyon Julia C., Symmons Martyn F., Lever Andrew, Characterizing 3D RNA structure by single molecule FRET, Methods 103 (2016) 57–67. [PubMed: 26853327]

[8]. Yu Yi-Tao, Huang Chao, Karijolich John, Detection and quantification of RNA 2′-O methylation and pseudouridylation, Methods 103 (2016) 68–76. [PubMed: 26853326]

[9]. Palangat Murali, Larson Daniel, Single-gene dual-color reporter cell line to analyze rna synthesis in vivo, Methods 103 (2016) 77–85. [PubMed: 27068658]

[10]. Lorenz Ronny, Wolfinger Michael T., Tanzer Andrea, Hofacker Ivo L., Predicting RNA secondary structures from sequence and probing data, Methods 103 (2016) 86–98. [PubMed: 27064083]

[11]. Parlea Lorena G., Sweeney Blake S., Maryam Hosseini Asanjan Craig L. Zirbel, Leontis Neocles B., The RNA 3D motif atlas: computational methods for extraction, organization and evaluation of RNA motifs, Methods 103 (2016) 99–119. [PubMed: 27125735]

[12]. Biesiada Marcin, Pachulska-Wieczorek Katarzyna, Adamiak Ryszard W., Purzycka Katarzyna, RNAComposer and RNA 3D structure prediction for nanotechnology, Methods 103 (2016) 120–127. [PubMed: 27016145]

[13]. Parlea Lorena, Bindewald Eckart, Sharan Rishabh, Bartlett Nathan, Moriarty Daniel, Oliver Jerome, Afonin Kirill, Shapiro Bruce A., Ring Catalog: a resource for designing Self-assembling RNA Nanostructures, Methods 103 (2016) 128–137. [PubMed: 27090005]

[14]. Dawson Wayne K., Maciejczyk Maciej, Jankowska Elzbieta J., Bujnicki Janusz M., Coarse-grained modeling of RNA 3D structure, Methods 103 (2016) 138–156. [PubMed: 27125734]

[15]. Xavier Bofill-De Ros Shuo Gu, Guidelines for the optimal design of miRNA-based shRNAs, Methods 103 (2016) 157–166. [PubMed: 27083402]

[16]. Urak Kevin T., Shore Sabrina, Rockey William M., Chen Shi-Jie, McCaffrey Anton P., Giangrande Paloma, In vitro RNA SELEX for the generation of chemically-optimized therapeutic RNA drugs, Methods 103 (2016) 167–174. [PubMed: 26972786]

[17]. Xiaojun Xu, Dickey David D., Chen Shi-Jie, Giangrande Paloma, Structural computational modeling of RNA aptamers, Methods 103 (2016) 175–179. [PubMed: 26972787]

[18]. Thiel William, Giangrande Paloma H., AFBI assay – aptamer fluorescence binding and internalizationassay for cultured adherent cells, Methods 103 (2016) 180–187. [PubMed: 26972784]

[19]. Abulwerdi Fardokht A., Schneekloth Jay, Microarray-based technologies for the discovery of selective. RNA binding molecules, Methods 103 (2016) 188–195. [PubMed: 27109057]