

The underworld of RNA

**The RNA World (3rd edition)**

**Edited by Raymond F Gesteland,
Thomas R Cech & John F Atkins**

Cold Spring Harbor Laboratory Press, 2006
768 pp. hardcover, \$135
ISBN 0879697393

Reviewed by John S Mattick

This is a big book, full of information. Although it is a compendium of chapters by different authors, it manages a surprisingly smooth flow and consistency of style.

The book is adorned with a foreword by Francis Crick and a prologue by James Watson. These now-historical reflections remind us of the simple initial questions about the transmission of genetic information, the role of RNA and the nature of the 'genetic code'—the answers to which we now take for granted, perhaps prematurely. Embedded in these early perspectives is the view that RNA is primarily the middleman between DNA and protein, encapsulated in Watson's recollection of having jotted down the line 'DNA>RNA>protein' in late 1952. This became the so-called 'central dogma', with its implicit and still widely held assumption that most genes encode proteins.

The book begins with the hypothesis that RNA, because it has both informational and catalytic capacity, was the ancestor of both DNA and proteins, albeit with sobering reminders in subsequent chapters that the evolution of these capacities in RNA itself, and the devolution of many of its presumed primordial functions (such as replication and metabolism) to proteins, was not at all trivial and is in itself problematic. The book then proceeds to discuss the modern manifestations of functional RNA that provide evidence of its capability. These include 'riboswitches' (RNA structures that bind ligands to effect allosteric changes to control gene expression) and the versatility of RNA 'aptamers' that can be evolved *in vitro*. It also includes self-cleaving ribozymes and group I and II introns, the latter being the precursors of the pre-mRNA introns that occupy over one-third of our genome. As the book shows, the reverse transcription of RNA also underpins the retroposition of short interspersed nuclear elements (SINEs), long interspersed nuclear elements (LINEs) and retroviral-like elements that constitute most of the rest of our genome.

The book also discusses ribonucleoprotein complexes, particularly the ribosome and the spliceosome, the latter of which evolved in eukaryotes ostensibly to rid pre-mRNA transcripts of troublesome introns. RNA catalysis lies at the heart of both of these machines. However, there is no chapter on introns themselves, despite the fact that they comprise over 95% of pre-mRNA transcripts in humans, nor any discussion of the functions, if any, of the excised intronic RNAs, especially in view of the fact that many

miRNAs and snoRNAs are encoded within them.

In addition, there is neither a discussion of the thousands of non-protein coding RNAs (ncRNAs) that have been recently identified in cDNA and genome tiling array transcriptomic studies nor any treatment of the now-significant subset that has been functionally studied in various contexts. Admittedly this field has exploded in the past year or two, and may not have had time to be considered for this edition. The section on long ncRNAs is limited to regulation of sex chromosome dosage and imprinting, important and intriguing topics in their own right and likely indicative of a broader phenomenon of interallelic communication.

Also disappointingly, the book does not have a section devoted to the expanding field of RNA editing, except for the baroque case of uridine insertion and deletion in trypanosomes. It does cover known RNAs, including 'small nucleolar RNAs' (snoRNAs, which are involved in two forms of editing) and the related telomerase RNA, 'small nuclear' largely spliceosomal RNAs (snRNAs) and microRNAs (miRNAs). The latter are clearly central to normal and abnormal development (as in cancer) and already comprise a significant and rapidly growing fraction of known human 'genes'.

With the exception of miRNAs, which were first discovered by sensitive genetic screens, the majority of the well-known ncRNAs are relatively abundant. By contrast, because of their small quantities and complexity, most miRNAs have (since) been identified bioinformatically and then validated *post hoc*, which raises the specter that there may well be many more miRNAs, as well as sno-like RNAs and other classes of yet-to-be-discovered regulatory RNAs derived from noncoding transcripts, which span at least 70% of the mammalian genome.

These vast tracts of transcribed but untranslated sequences have been dismissed as evolutionary relics and transcriptional noise. This seems to be because the alternative explanation (that these RNAs may themselves be functional) challenges the protein-centric orthodoxy of molecular biology, despite the fact that there are more conserved noncoding than coding sequences in the mammalian genome, and that chromatin modification (epigenetic memory) and alternative splicing, the hallmarks of differentiation and development in complex organisms, have no satisfactory explanation in terms of protein-based sequence recognition and control. Interestingly, although it is accepted that DNA encodes information digitally, it is assumed that its outputs are mainly analog (that is, proteins). Little consideration has been given to the possibility that some and perhaps most of the information transmitted from the genome (particularly in higher organisms) may be itself digital—RNA communication and control signals such as miRNAs—although the potential of RNA as a sequence-specific regulatory molecule was canvassed by Jacob and Monod in 1961.

The current edition concludes with chapters on predicting and modeling RNA structure, as well as methods for imaging and tracking RNA dynamics, a field that will burgeon, along with RNA informatics, as it becomes obvious that we have hardly scratched the surface of the modern RNA world.

This is no coffee table book, but rather a good reference for anyone who wishes to begin to understand the diverse structures, functions and roles of RNA in cell biology. Although perhaps already out of date in some important and potentially revolutionary respects, it is excellent in most others. Unfortunately, its penetration may be limited by its unavailability in electronic format.

John Mattick is at the Institute for Molecular Bioscience, University of Queensland, St. Lucia, Queensland 4072, Australia.
e-mail: j.mattick@imb.uq.edu.au