

best candidates for such antibodies or the subsequent routes to maturation.

Further technical developments of the method described by Mao *et al.*¹ should accommodate larger libraries than those used here while maintaining specific knowledge about each member. Declining costs of gene synthesis and advancing systems for miniaturized expression, purification and screening of multiple recombinant proteins in parallel should also help. Such advances would begin to bring the vast experience available in high-throughput screening of combinatorial libraries of small molecules to bear on the problem of identifying new therapeutic monoclonal antibodies.

COMPETING FINANCIAL INTERESTS

The author declares no competing financial interests.

1. Mao, H. *et al. Nat. Biotechnol.* **28**, 1195–1202 (2010).
2. Nelson, A.L., Dhimolea, E. & Reichert, J.M. *Nat. Rev. Drug Discov.* **9**, 767–774 (2010).
3. Hoogenboom, H.R. *Nat. Biotechnol.* **23**, 1105–1116 (2005).
4. Kindt, T.J., Osborne, B.A. & Goldsby, R.A. *Kuby Immunology*, edn. 6 (W.H. Freeman, 2006).
5. Suntharalingam, G. *et al. N. Engl. J. Med.* **355**, 1018–1028 (2006).
6. Weiner, L.M. & Carter, P. *Nat. Biotechnol.* **23**, 556–557 (2005).
7. Xu, J.L. & Davis, M.M. *Immunity* **13**, 37–45 (2000).
8. Weinstein, J.A., Jiang, N., White, R.A., Fisher, D.S. & Quake, S.R. *Science* **324**, 807–810 (2009).
9. Boyd, S.D. *et al. Sci. Transl. Med.* **1**, 12ra23 (2009).
10. Kwong, P.D. & Wilson, I.A. *Nat. Immunol.* **10**, 573–578 (2009).

regions (UTRs). The start and stop codons, which define the protein open reading frame, tend to occur in relatively unstructured regions, a feature that may facilitate access by the translation machinery. The relative lack of secondary structure in the 3' UTR may ensure that factors that regulate mRNA translation, stability and localization have access to their binding sites.

The finding that open reading frames are more structured than the flanking regions is consistent with a study from our laboratory³ showing that the structured regions of the HIV-1 genome occur preferentially downstream of sequences encoding individual protein domains, perhaps facilitating protein folding during translation. Also in accordance with earlier results⁴, Kertesz *et al.*² found that highly translated messages appear to have less defined structure near the translational start site than those with lower translation levels, suggesting that mRNA structure modulates ribosome activity.

Characterizing the extent of double- and single-stranded structure in the yeast transcriptome represents an important first step toward the goal of determining the complete three-dimensional structures of these RNAs. However, generating accurate nucleotide-resolution structures from RNase cleavage data is very challenging⁵. Moreover, *in vitro* folding, as used in PARS, is not likely to fully recapitulate folding in the cellular environment. The ability of any approach to guide RNA structure prediction is best assessed by comparing predicted models to RNA structures known to be in a functional conformation. Although very few yeast RNAs have been studied over significant lengths, the secondary structure of the 18S ribosomal RNA from the small (40S) ribosomal subunit is well characterized⁶. We found that the PARS-assisted, secondary-structure prediction for the 18S RNA² contains only half of the accepted base pairs for this RNA and thus does not resemble the biologically functional state in many regions. In addition, the PARS-assisted secondary structure for 9 of 14 tRNAs with read depths greater than 1 were predicted incorrectly. Thus, at this juncture, PARS results do not reproduce these physiologically relevant RNA structures and should be interpreted cautiously.

In contrast, previous, lower-throughput approaches have yielded very high-resolution structural information under defined, controlled conditions in which the RNAs retain most of their native structure. In the mid-1980s, Harry Noller's laboratory demonstrated that primer extension could identify RNA nucleotides modified by small, structure-selective, organic chemicals. Using these reagents,

Toward global RNA structure analysis

David M Mauger & Kevin M Weeks

Deep sequencing provides a first view of the RNA structures in a eukaryotic transcriptome.

The structures of RNA molecules, even in complex environments, have been interrogated for many years by studying how individual nucleotides react with enzymatic or chemical probes¹. However, these methods have largely been limited to studying single RNAs or small fragments of large RNAs. Writing in *Nature*, Kertesz *et al.*² have succeeded in analyzing much of the yeast transcriptome by melding a traditional biochemical method for probing RNA structure with the power of highly parallel DNA sequencing in an approach called parallel analysis of RNA structure (PARS). Their results provide new insight into the role of mRNA structure in gene expression, confirming, in particular, that RNA structure regulates protein synthesis, most probably by controlling the accessibility of mRNAs to the translational machinery.

Kertesz *et al.*² extracted total RNA from *Saccharomyces cerevisiae*, enriched it for mRNA, refolded the RNA, and treated it separately with S1 nuclease and RNase V1, nucleases specific for single- and double-stranded RNA, respectively (Fig. 1). The resulting

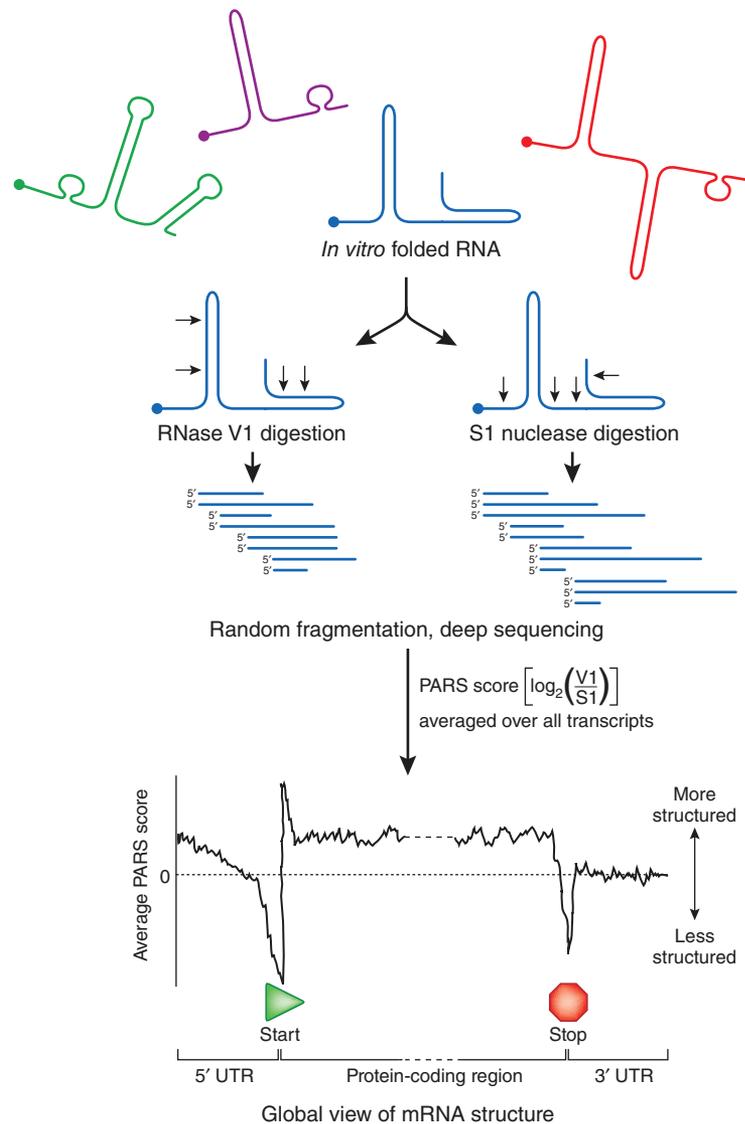
structure-selective cleavage fragments were converted to double-stranded DNAs and sequenced using the SOLiD system, generating tens of millions of reads. Sequences within less abundant transcripts and those that were more difficult to convert from RNA to double-stranded DNA were largely undetected. The authors then determined individual RNase cleavage sites from the sequencing reads and compared the digestion frequencies of the two RNases at each nucleotide. This comparison, termed a PARS score, provides a measure of the single-stranded or double-stranded character of each nucleotide in each RNA with sufficient read coverage. By characterizing thousands of yeast transcripts, the authors generated structural data for approximately half of the *S. cerevisiae* transcriptome.

These data are especially valuable when analyzed at a global level (Fig. 1). In aggregate, they lend strong support to the general hypothesis that the information encoded in large mRNAs is regulated and organized by RNA structure. In several cases, the authors were able to confirm and extend models for the roles of RNA structure previously developed through detailed analyses of individual RNAs. The data revealed that the prototypical yeast mRNA consists of regions with distinct characteristics: the open reading frame is more structured than either the 5' or 3' untranslated

David M. Mauger and Kevin M. Weeks are in the Department of Chemistry, University of North Carolina, Chapel Hill, North Carolina, USA.

e-mail: weeks@unc.edu

Figure 1 Overview of the PARS approach for measuring RNA structure on a transcriptome scale. RNAs are refolded *in vitro*, cleaved with structure-selective ribonucleases, fragmented and subjected to highly parallel sequencing. The PARS score gives a measure of double versus single strandedness. One use of the PARS scores is to average these values over many RNAs to yield a global view of RNA structure. PARS suggests that most *S. cerevisiae* transcripts have less secondary structure in untranslated regions (UTRs) than in coding regions, and that mRNAs with structured translation start sites are translated with lower efficiency than those with unstructured translation start sites.



Noller and colleagues examined the structure of the *Escherichia coli* 16S ribosomal RNA and detected fine-scale perturbations resulting from the binding of proteins and drugs^{7,8}. We developed a descendant of this technology, called selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE), which makes use of broadly reactive, yet structure-selective, small-molecule reagents. Using SHAPE, it is often possible to predict RNA structures with high accuracy⁹.

As next-generation technologies for probing RNA structure are further refined, it will be important to combine such innovations in high-throughput data acquisition with bioinformatic approaches for accurately interpreting complex data sets. Eventually, it may be possible to carry out global analyses of how RNA structures change in response to environmental stimuli and developmental cues. The ability to accurately determine the structures of and interactions between components of entire transcriptomes will contribute substantially to our understanding of the biological roles of RNA, including in RNA processing, translation and non-coding RNA function. This understanding should facilitate the use of molecular biology tools to manipulate the cellular state and develop RNA-directed therapeutics.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Zaug, A.J. & Cech, T.R. *RNA* **1**, 363–374 (1995).
2. Kertesz, M. *et al.* *Nature* **467**, 103–107 (2010).
3. Watts, J.M. *et al.* *Nature* **460**, 711–716 (2009).
4. Kozak, M. *Gene* **361**, 13–37 (2005).

5. Mathews, D.H., Sabina, J., Zuker, M. & Turner, D.H. *J. Mol. Biol.* **288**, 911–940 (1999).
6. Gutell, R.R. *Nucleic Acids Res.* **21**, 3051–3054 (1993).
7. Moazed, D., Stern, S. & Noller, H.F. *J. Mol. Biol.* **187**, 399–416 (1986).
8. Moazed, D. & Noller, H.F. *Nature* **327**, 389–394 (1987).
9. Deigan, K.E., Li, T.W., Mathews, D.H. & Weeks, K.M. *Proc. Natl. Acad. Sci. USA* **106**, 97–102 (2009).