The awesome power of ribosome profiling

RICHARD JACKSON and NANCY STANDART

Department of Biochemistry, University of Cambridge, Cambridge CB2 1QW, United Kingdom

Ribosome profiling has proved to be a fantastic innovation which has revealed so much about mRNA translation that had previously gone unrecognized. The RNA community owes Jonathan Weissman and Nick Ingolia a deep sense of gratitude for conceiving the approach in the first place, and for their detailed optimization of the methodology. The origins of the approach can be traced back to the discovery of polysomes over 50 years ago, and the finding that treatment with endoribonucleases converted them to monomeric ribosomes that were each bound to a short protected mRNA fragment. This was first exploited by Joan Steitz and Marilyn Kozak to identify and sequence the translation initiation sites of prokaryotic bacteriophage RNA and eukaryotic reovirus RNAs, respectively; and extended by Sandra Wolin and Peter Walter to reveal ribosome pausing at the initiation and termination codons of preprolactin mRNA, as well as at the point at which SRP interacts with the nascent protein. Because of the limitations of sequencing methods available in that era, these experiments necessarily used cell-free translation systems programmed preferably with a single mRNA species (or a maximum of 4 reovirus mRNAs).

It was the advent of modern deep sequencing technologies, and the associated methods of amplifying the short mRNA fragments to generate cDNA libraries, that opened up the possibility of using the approach to globally map the positions of ribosomes on mRNAs in intact cells. We consider one of the most important innovations of the method was to generate the protected mRNA fragments using *E. coli* RNase I, which has little (if any) sequence specificity, in contrast to the nucleases used previously (RNase A, RNase T1, and micrococcal nuclease). Such is the precision of RNase I cutting especially at the 5′-end, that the positions of ribosomes along the ORF can be determined with single nucleotide precision, and the footprints show a clear trinucleotide periodicity, which allows assignment of the translation reading frame and distinguishes footprints arising from translating ribosomes from RNA fragments that are protected for any other reason.

In most analyses, the length distribution of protected mRNA fragments shows a single symmetrical peak with a median of 28–29 nt (*S.cerevisiae*) or 30–31 nt (mammalian cells), which presumably reflects the larger size of mammalian 60S ribosomal subunit. However, protected fragments of ∼16 nt, ∼21 nt, and 40–65 nt were also obtained from *S. cerevisiae* by Nick Guydosh and Rachel Green. The 16 nt fragments originate from an 80S ribosome stalled at the physical 3′-end of a truncated (partially degraded) mRNA; and the 40–65 nt fragments were assigned to closely stacked 80S ribosomes that accumulate when the leading ribosome is stalled. The ∼21 nt fragments are more intriguing and appear to arise from an alternative (probably rotated) conformation of translating 80S ribosome: this form is stabilized by anisomycin, whereas cycloheximide (CHX) stabilizes the ∼30 nt species (Pat Brown group). It would be very interesting to compare the structures of these two forms by cryo-electron microscopy or X-ray crystallography.

In order to attach significance to the ribosome footprint signal strength on each mRNA species, we need to know the abundance of each mRNA, which is achieved by isolating total polyA+ RNA from the cell extracts, fragmenting it to a ∼30 nt length, and subjecting the fragments to deep sequencing. This allows the average ribosome density on each mRNA to be determined, but because it is only an average, it is unable to reveal situations in which the mRNA shows a bimodal distribution on polysome gradients, with some of it totally repressed and untranslated.

We list below what we consider to be the major unanticipated discoveries due to ribosome profiling, with particular emphasis on higher eukaryotes, except where otherwise stated.

1. Perhaps the most unexpected discovery has been the discovery, in mouse ES cells, of numerous translated short upstream ORFs (uORFs) with near-cognate initiation codons (CUG>GUG>UUG>ACG>others), which outnumber AUG-initiated uORFs by ∼4:1. There is absolutely no doubt that these uORFs are actually translated, but the cell harvesting procedure may have exaggerated the frequency of initiation at these near-cognate sites (see below). The precedent of yeast GCN4 mRNA has suggested that the role of such uORFs is to regulate translation of the

protein-coding ORF, rather than as a source of peptides with biological function, but it remains to be seen if this still holds for all these additional uORFs.

2. For many protein-coding ORFs, the annotated start codon was found to be not the only in-frame initiation site, and in some cases not even the main start site. Thus there are many more cases of mRNAs coding for N-terminally extended or truncated isoforms of the annotated protein than previously appreciated.

3. Particularly in the case of *Drosophila* embryos and S2 cells (but also *S. cerevisiae* and human fibroblasts), numerous cases of a low level of leaky termination have been found, resulting in protein isoforms with C-terminal extensions, which in some cases have amino acid sequences that are phylogenetically conserved.

4. The triplet periodicity of ribosome footprints, coupled with the use of harringtonine (see below), has enabled several cases of translation occurring in two relatively long overlapping reading frames to be identified. Moreover, Caia Duncan and Juan Mata found 19 examples of protein-coding mRNAs transcribed from both strands of the same genomic DNA segment in *S. pombe*.

5. Profiling has revealed that several capped and polyadenylated mRNAs in *S. cerevisiae*, *Drosophila* S2 cells, and vertebrates, which were previously classified as "(long) noncoding RNAs" (since they lack any ORF longer than ∼100 codons) are actually translated: they have short 5′-proximal ORFs which give footprints showing the characteristic triplet periodicity (except when translation occurs in two overlapping frames), and they are subject to nonsense mediated decay, which is dependent on translation. In some cases the peptide products have been detected either by mass spectrometry or epitope tagging, and some of these short ORFs are phylogenetically conserved, even between zebrafish embryos and human cells, implying that the encoded peptides are likely to have a biological function.

6. A highly ingenious method has been developed by the Weissman group for in vivo biotin-tagging the 60S subunit of yeast ribosomes that are in close proximity to either the ER or mitochondria, thereby allowing these ribosome sub-populations to be selected for profiling. For numerous secreted proteins, the profiling revealed the minimum length of nascent protein required for the ribosome to engage the Sec61/Sec63 or the Ssh1 translocons in the ER. The mitochondria studies revealed that most inner membrane proteins were invariably inserted cotranslationally, but cotranslational insertion of other proteins was greatly favored by CHX addition to allow more time for ribosomes to engage the mitochondria.

7. Treating mouse ES cells with harringtonine for various times enabled the rate of elongation to be determined as ∼330 codons per minute, with little variation between different mRNAs.

8. Examples of ribosome stalling or pausing, sometimes coupled with close ribosome stacking, were found at consec-utive proline codons in ES cells as well as in *S. cerevisiae*, and also at His codons in *S. cerevisiae* on starvation for histidine.

9. Nick Guydosh and Rachel Green found that ∼10% of *S. cerevisiae* mRNA species had several 3′-UTR ribosome footprints in all three reading frames, which did not show the usual triplet periodicity and were significantly increased in *dom34Δ* strains. They were thought to be due to a failure of the normal post-termination ribosome release (or disassembly) process catalyzed by ABCE1 (Rli1), resulting in 80S ribosomes that migrate along the 3′-UTR without translating it. In wild-type strains they would most probably be rescued (i.e., recycled) by the combined action of Dom34 and Hbs1 proteins.

One issue that has concerned us all along is the extent to which the cell harvesting and lysis procedure may distort the steady state ribosome distribution. Unless the material can be cooled to ∼0°C extremely rapidly, polysome run-off will occur, depleting ribosomes from all initiation sites, all short ORFs and the 5′-proximal part of long ORFs. To avoid this, many practitioners incubate the cells with CHX for up to 5 min, before cooling. This prevents run-off, but it will give footprints at initiation sites which are far greater than at steady state, because CHX doesn't inhibit scanning or initiation, and so an additional 80S initiation complex can be added to every mRNA which has a vacant initiation site at the time when the inhibitor starts to "bite." Although this problem has been recognized previously by some authors, no one has yet applied numbers to it. For highly efficient sites that capture virtually all scanning ribosomes, we reckon that 100% occupancy of vacant initiation sites will occur within <10 sec of the onset of inhibition by CHX. (This estimate is based on the fact that in reticulocyte cells, where ∼90% of translation is synthesis of the two globin chains [of average length 143 amino acids], and the average polysome size is 5 ribosomes, it takes ∼30 sec for a ribosome to synthesize a globin chain at 37°C, and so initiation events must be occurring on each globin mRNA at an average frequency of once every 6 sec.) It follows that if an initiation site is mutated to be so inefficient that it captures only 10% of scanning ribosomes, 100% occupancy of all vacant sites will still be achieved in <2 min after the onset of inhibition. So a 3–5 min incubation with CHX could result in footprints of equal magnitude at initiation sites that really differ in intrinsic efficiency by a factor of 10 or more! Added to this, there is a polarity effect in that an 80S ribosome stalled at a 5′-proximal site will prevent any subsequent scanning ribosome from reaching downstream sites, and this could result in the observed footprint signal at a uORF initiation site with mediocre context actually being greater than the signal at a highly efficient downstream start site.

Precisely the same problems arise if cells are incubated for 3–5 min with harringtonine, which inhibits neither initiation nor on-going elongation (or polysome run-off), but does block the transition from initiation to elongation, so that it

leaves clear footprints of 80S ribosomes stalled at every initiation site. This has proved to be an excellent way of identifying initiation sites, but, like CHX, it can completely obscure their relative utilization frequency under steady-state conditions.

Nevertheless, we appreciate the importance of keeping a proper sense of perspective, and realize that these reservations pale into relative insignificance compared with the tremendous achievements of ribosome profiling.