

Programmed ribosomal frameshifting: Much ado about knotting!

S. L. Alam, J. F. Atkins, and R. F. Gesteland

Department of Human Genetics, University of Utah, 15N 2030E, Salt Lake City, UT 84112-5330

Rules of the genetic code can be temporarily suspended to make room for a variety of specific tricks that direct synthesis of additional proteins from mRNAs that would not be predicted by their sequences. This “recoding” includes specification of the 21st encoded amino acid, selenocysteine (1, 2). In another form, recoding itself provides a sensing mechanism to regulate expression (3). In other cases mRNA can direct synthesis of a set ratio of two proteins that share some amino acid sequences. Kim, Su, Maas, O’Neill, and Rich, in this issue of PNAS (4), describe experiments that show the mRNA structural features important for one such case of ratio setting.

We know little about the global contribution that recoding makes to the complexity of proteomes. It could be that many mRNAs encode, in addition to their standard products, minor products that so far would have escaped detection. These products are likely to have distinct functions and contribute to the biological complexity of the proteome. Clearly the complexity of the human proteome is far beyond the more than 100,000 human genes. The total mRNA population might be 250,000 from alternate splicing, editing, and use of alternate promoters. How much this complexity will be expanded by protein modification, protein splicing, and recoding is far from clear but a total of 500,000 might not be surprising. Certainly the contribution from recoding is completely unknown; however, specific examples being studied give some guide to what is to come.

Recoding by redefinition of codons and translational bypassing are being studied in detail but the majority of known examples are programmed shifts in reading frame (3, 5). Depending on the shift site ribosomes can be instructed to slip into +1 or the –1 frame. Examples are found in all well-studied organisms from viruses to bacteria to humans, but in the case of human gene expression only one family of genes, antizyme, is known to use frameshifting and this is +1 (6, 7). Many more cases of –1 frameshifting are known partly because of their use by members of

large families of viruses and bacterial insertion sequences. Programmed –1 frameshifting occurs at “slippery” shift sites. At these sites, initial codon-anticodon pairing in the zero frame is followed by dissociation and re-pairing of the tRNA anticodon to mRNA at an overlapping codon. The great majority of cases involve realignment of the codon anticodon pairing of not just one tRNA but of two adjacent tRNAs in the ribosome (peptidyl tRNA and aminoacyl tRNA). Even though the pairing rules for re-pairing are more relaxed than those for initial pairing, tandem slippage gives signature shift sites of the general form X XXY YYZ where X, Y, and Z can even be the same nucleotides. Tandem slippage was first recognized for the programmed frameshifting that occurs close to the 3’ end of the *gag* gene of many retroviruses (8). This frameshifting is required for synthesis of the GagPol fusion polyprotein as there is no independent ribosome entry to the *pol* gene. [In retroviruses that have a separate protease gene between *gag* and *pol* two frameshift events are required to synthesize the GagProPol fusion polyprotein (9, 10).]

The potential for codon:anticodon alignment is not the only important feature for programmed –1 frameshifting. Frameshifting is stimulated by discrete mRNA structures that in nearly all cases are formed from contiguous sequence starting 5–9 nt 3’ of the shift site. Although in HIV (11) and *Escherichia coli dnaX* (12, 13) this stimulating structure is a simple stem loop, and in IS911 it is a three-way junction (14), in most cases it is a pseudoknot as first recognized in studies on viruses (15–17).

There are various types of pseudoknots, but the classical H-(hairpin) pseudoknot (18) requires pairing of the loop of a hairpin loop with downstream sequences, resulting in two stems that are connected by two loops. RNA pseudoknots are widely occurring structural motifs that are involved in various RNA functions (19). Pseudoknots were first recognized experimentally from studying the folding of the 3’ end of the turnip yellow mosaic virus

RNA (for review see ref. 20). They have since been found in many classes of RNA including: ribosomal RNA (21), catalytic and self-splicing RNA (22, 23), tmRNA (24–26), internal ribosome entry sites of some picornaviruses (27), and translational repression sites on mRNAs for some ribosomal proteins (28) and for certain T-even bacteriophage proteins (29, 30). However, the pseudoknots involved in recoding are unique in that, as they play their role as a structure, they are immediately unfolded and their now linear sequence serves as a template for decoding.

Pseudoknots cause ribosomal pausing, raising the possibility that this delay at the crucial site allows more time for anticodon:mRNA realignment to occur. This hypothesis has been tested in two cases with the conclusion that pausing may be necessary for –1 frameshifting but it is not sufficient (31, 32)—something else must be going on. Much remains to be learned, including when the pseudoknot is melted out and with what the pseudoknot might interact (see ref. 33 for discussion).

Kim *et al.* (4) describe experiments that functionally test the unique structural features of a frameshift stimulating pseudoknot from the plant virus beet western yellow virus (BWYV), a luteovirus. Replication of BYMV genomic RNA requires two virus-encoded functions, P1 and P2 (for review see ref. 34). The polymerase, P2, is expressed only as an extension of P1 when a few percent of the ribosomes, three-quarters of the way through the coding sequence for P1, shift reading frame into the overlapping ORF for P2 (16, 35, 36). This means that the P1 protein and the P1-P2 frameshifted protein share their first 461 aa. Then, each uses a different frame to decode the next 146 aa, and P1-P2 then continues for another 429 unique aa. The frameshift event is a –1 tandem shift at a G GGG AAC site (the same shift site used by simian retrovirus 1; ref. 37), stimulated by a pseudoknot 6 nt downstream (Fig. 1).

Rich’s group (38) recently described a crystal structure of the BWYV pseudoknot.

See companion article on page 14234.

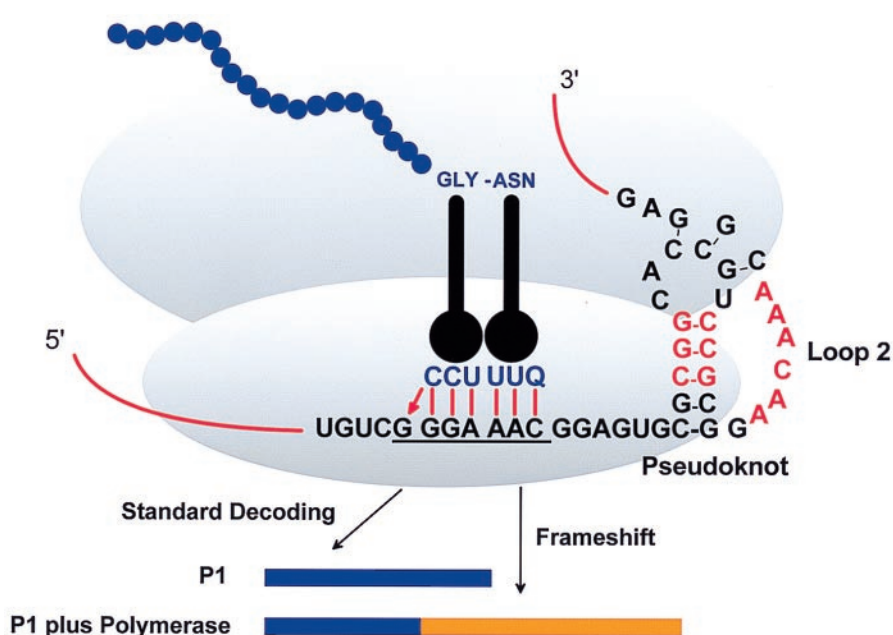


Fig. 1. Model for BWYV – 1 frameshift. Bases in red are conserved in all known luteoviruses.

The structure shows that the two stems are not coaxially stacked because of a 48° rotation between the two pairs of bases at the junction and because of unusual conformations of the junctional U and A bases. The result is an overall bend angle between the stems of 25°. There is a quadruple base interaction between a nucleotide in loop 1 and stem regions. And perhaps most significantly loop 2 forms a triple helix through interactions with the minor groove of stem 1. The experiments reported by Kim *et al.* (4) test the functional importance of key nucleotides in this structure (38) for stimulation of frameshifting.

The mutational results support the complex structure—residues predicted to have key interactions are crucial for activity and those that are not interacting are insensitive to change. The unusual structural features of the interactions between loop 2 and stem 1 are functionally important as are the specific interactions around the stem junctions. A few bases are unexpectedly sensitive to mutation and the suggestion is that these may make key contacts with ribosomes or perhaps some

other factor important to the process. Interestingly, some mutational changes increase the relatively low efficiency of this shift site to a higher level, implying that selection has led to a system tuned to the most useful efficiency.

From these results, and the other well-studied examples of pseudoknots that promote frameshifting, it is clear that more than one structure can suffice (39–41) but not all pseudoknots work (42).

The differences between active pseudoknots seem particularly apparent in the importance of loop 2 sequences. In BWYV, loop 2 plays a crucial structural and functional role. With infectious bronchitis virus (17) and simian retrovirus-1 (37, 43) pseudoknots, loop 2 sequences have great latitude as if their role is merely to connect stems. However, identities of some bases are crucial in loop 2 of the pseudoknot in murine leukemia virus that stimulates, not frameshifting, but in-frame read-through of a stop codon. (The mouse mammary tumor virus pseudoknot cannot substitute for its murine leukemia virus counterpart in stimulating read-through; ref. 44.) Presumably, in the cases where

the identities of loop 2 bases are unimportant, the sequences are not forming part of an essential structure, nor are they forming key interactions with ribosomes or other possible factors.

Both the structure and the function studies with BWYV point to the crucial role of the stem-junction region. This is also the case with the mouse mammary tumor virus pseudoknot, but here the junction is very different. The NMR-based structure shows an angle of 60° between the stems. This is caused by a “wedge” base on one side at the junction (39, 45, 46), and functional tests imply that this base, and hence the bend, is important (47).

However, it would be comforting and perhaps even instructive to have both x-ray and NMR structures for the same pseudoknot.

The spacer region between the hepta-nucleotide shift site and the pseudoknot is also important for function. Changing the spacer length generally decreases frameshifting. Presumably the spacer length alters the position of the paused ribosome and influences the probability of shifting. Sequence identity of the spacer has not been sufficiently tested for frameshifting cases. It is known that the identity of some nucleotides is essential for pseudoknot-dependent read-through in murine leukemia virus (44, 48)—these bases could be interacting either with the pseudoknot to form a more complicated structure or perhaps with translational components. The spacer sequences in BWYV should be tested both for functional importance and for possible influences on the pseudoknot structure. At the same time it would be worth testing to make sure that the stop codon inserted for convenience immediately after the shift site does not influence the results (see ref. 49).

The crystal structure of the BWYV combined with the new mutational results clearly shows the complexity and sophistication of RNA structures that in this case form transiently to promote alteration of decoding. This is now a firm footing from which to determine just how this structural information of pseudoknots is used to influence the mechanism of frameshifting. It is equally a challenge to discover the exploitation of these mechanisms for expression of the mammalian proteome.

- Baron, C. & Böck, A. (1995) in *tRNA, Structure, Biosynthesis, and Function*, eds. Söll, D. & Raj-Bhandary, U. L. (Am. Soc. Microbiol., Washington DC), pp. 529–544.
- Low, S. C. & Berry, M. J. (1996) *Trends Biochem. Sci.* **21**, 203–208.
- Gesteland, R. F. & Atkins, J. F. (1996) *Annu. Rev. Biochem.* **65**, 741–768.
- Kim, Y.-G., Su, L., Maas, S., O’Neill, A. & Rich, A. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 14234–14239.
- Atkins, J. F., Böck, A., Matsufuji, S. & Gesteland, R. F. (1999) in *The RNA World*, eds. Gesteland,

- R. F., Cech, T. R. & Atkins, J. F. (Cold Spring Harbor Lab. Press, Plainview, NY), 2nd Ed., pp. 637–673.
- Matsufuji, S., Matsufuji, T., Miyazaki, Y., Murakami, Y., Atkins, J. F., Gesteland, R. F. & Hayashi, S. (1995) *Cell* **80**, 51–60.
- Ivanov, I. P., Gesteland, R. F. & Atkins, J. F. (1998) *Genomics* **52**, 119–129.
- Jacks, T., Madhani, H. D., Masiarz, R. F. & Varmus, H. E. (1988) *Cell* **55**, 447–458.
- Moore, R., Dixon, M., Smith, R., Peters, G. & Dickson, C. (1987) *J. Virol.* **61**, 480–490.
- Jacks, T., Townsley, K., Varmus, H. E. & Majors,

- J. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 4298–4302.
- Parkin, N. T., Chamorro, M. & Varmus, H. E. (1992) *J. Virol.* **66**, 5147–5151.
- Tsuchihashi, Z. (1991) *Nucleic Acids Res.* **19**, 2457–2462.
- Larsen, B., Gesteland, R. F. & Atkins, J. F. (1997) *J. Mol. Biol.* **271**, 47–60.
- Rettberg, C. C., Prère, M. F., Gesteland, R. F., Atkins, J. F. & Fayet, O. (1999) *J. Mol. Biol.* **286**, 1365–1378.
- Brierley, I., Digard, P. & Inglis, S. C. (1989) *Cell* **57**, 537–547.

16. ten Dam, E. B., Pleij, C. W. A. & Bosch, L. (1990) *Virus Genes* **4**, 121–136.
17. Brierley, I., Rolley, N. J., Jenner, A. J. & Inglis, S. C. (1991) *J. Mol. Biol.* **220**, 889–902.
18. Gulyaev, A., van Batenburg, F. H. D. & Pleij, C. W. A. (1999) *RNA* **5**, 609–617.
19. Deiman, B. A. L. M. & Pleij, C. W. A. (1997) *Semin. Virol.* **8**, 166–175.
20. Pleij, C. W. A., Rietveld, K. & Bosch, L. (1985) *Nucleic Acids Res.* **13**, 1717–1731.
21. Moazed, D. & Noller, H. F. (1987) *Nature (London)* **327**, 389–394.
22. Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Kundrot, C. E., Cech, T. R. & Doudna, J. A. (1996) *Science* **273**, 1678–1685.
23. Ferré-D'Amaré, A. R., Zhou, K. H. & Doudna, J. A. (1998) *Nature (London)* **395**, 567–574.
24. Williams, K. P. & Bartel, D. P. (1996) *RNA* **2**, 1306–1310.
25. Felden, B., Himeno, H., Muto, A., McCutcheon, J. P., Atkins, J. F. & Gesteland, R. F. (1997) *RNA* **3**, 89–103.
26. Nameki, N., Chattopadhyay, P., Himeno, H., Muto, A. & Kawai, G. (1999) *Nucleic Acids Res.* **27**, 3667–3675.
27. Rijnbrand, R., van der Straaten, T., van Rijn, P. A., Spaan, W. J. & Bredenbeek, P. J. (1997) *J. Virol.* **71**, 451–457.
28. Tang, C. K. & Draper, D. E. (1989) *Cell* **57**, 531–536.
29. Shamoo, Y., Tam, A., Konigsberg, W. H. & Williams, K. R. (1993) *J. Mol. Biol.* **232**, 89–104.
30. Du, Z. & Hoffman, D. W. (1997) *Nucleic Acids Res.* **25**, 1130–1135.
31. Tu, C., Tzeng, T.-H. & Bruenn, J. A. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 8636–8640.
32. Somogyi, P., Jenner, A. J., Brierley, I. & Inglis, S. C. (1993) *Mol. Cell. Biol.* **13**, 6931–6940.
33. Atkins, J. F. & Gesteland, R. F. (1999) *Nat. Struct. Biol.* **6**, 206–207.
34. Miller, W. A., Dinesh-Kumar, S. P. & Paul, C. P. (1995) *Crit. Rev. Plant Sci.* **14**, 179–211.
35. Veidt, I., Lot, H., Leiser, M., Scheidecker, D., Guilley, H., Richards, K. & Jonard, G. (1988) *Nucleic Acids Res.* **16**, 9917–9932.
36. Garcia, A., van Duin, J. & Pleij, C. W. A. (1993) *Nucleic Acids Res.* **21**, 401–406.
37. ten Dam, E. B., Verlaan, P. W. G. & Pleij, C. W. A. (1996) *RNA* **1**, 146–154.
38. Su, L., Chen, L., Egli, M., Berger, J. M. & Rich, A. (1999) *Nat. Struct. Biol.* **6**, 285–292.
39. Gonzalez, R. L. & Tinoco, I. (1999) *J. Mol. Biol.* **289**, 1267–1282.
40. Naphthine, S., Liphardt, J., Bloys, A., Routledge, S. & Brierley, I. (1999) *J. Mol. Biol.* **288**, 305–320.
41. Liphardt, J., Naphthine, S., Kontos, H. & Brierley, I. (1999) *J. Mol. Biol.* **288**, 321–335.
42. Kang, H., Hines, J. V. & Tinoco, I. (1996) *J. Mol. Biol.* **259**, 135–147.
43. Du, Z., Holland, J. A., Hansen, M. R., Giedroc, D. P. & Hoffman, D. W. (1997) *J. Mol. Biol.* **270**, 464–470.
44. Wills, N. M., Gesteland, R. F. & Atkins, J. F. (1994) *EMBO J.* **13**, 4137–4144.
45. Shen, L. & Tinoco, I. (1995) *J. Mol. Biol.* **247**, 963–978.
46. Kang, H. & Tinoco, I. (1997) *Nucleic Acids Res.* **25**, 1943–1949.
47. Chen, X., Kang, H., Shen, L. X., Chamorro, M., Varmus, H. E. & Tinoco, I. (1996) *J. Mol. Biol.* **260**, 479–483.
48. Alam, S. L., Wills, N. M., Ingram, J. A., Atkins, J. F. & Gesteland, R. F. (1999) *J. Mol. Biol.* **288**, 837–852.
49. Horsfield, J. A., Wilson, D. N., Mannering, S. A., Adamski, F. M. & Tate, P. T. (1995) *Nucleic Acids Res.* **23**, 1487–1494.