Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

The *where*, *what* and *how* of ribosomal frameshifting in retroviral protein synthesis

Dolph Hatfield and Stephen Oroszlan

The *gag* and *pol* genes of most retroviruses occur in different reading frames and their translation as a single polypeptide is carried out by ribosomal frameshifting in the –1 direction. The alignment of the different reading frames occurs by overlapping reading in response to at least two signals within the RNA: one is a heptanucleotide stretch at the frameshift site and the other is a stem–loop structure which occurs just downstream of the first signal.

ALTHOUGH, IN THEORY, an RNA sequence can be decoded in any of three reading frames, each of which would produce a completely different protein, most mRNAs can only be translated in one reading frame. This is set during the initiation phase of protein synthesis, when the ribosomal complex recognizes and assembles at the first codon (AUG) of the encoded protein. Translation then procedes until a stop codon is reached. Some mRNAs, however, encode two or more genes in different and overlapping reading frames and thus alternative strategies are required for expression of these additional genes (for a review, see Ref. 1). One such mechanism is known as ribosomal frameshifting in which the reading frame of a mRNA is changed at a specific site (or sites) during translation, resulting in the expression of a single protein from two (or multiple) overlapping genes. In bacteria and yeast, ribosomal frameshifting may operate by a variety of mechanisms such that the reading frame is changed in the 5' or the 3' direction (see Refs 2–5 and references therein). In yeast, the virus-like particle, L-A[5], and the retrovirus-like transposon, Ty[6], both contain two large overlapping reading frames. Interestingly, the different reading frames are aligned in L-A by a frameshift of one nucleotide in the 5' (or –1) direction, while in Ty they are aligned by a frameshift of one nucleotide in the 3' (or +1) direction. Reviews on these systems have been published recently[2,5]. In higher eukaryotes, however, only a shift by one nucleotide in the –1 direction has been described. Most of our understanding of frameshifting in higher eukaryotes has been obtained through studies with vertebrate viruses, primarily retroviruses, which will form the focus of this review.

Our understanding of frameshifting has increased substantially in the last five years since it was first demonstrated that Rous sarcoma virus (RSV) utilizes ribosomal frameshifting to align two large, briefly overlapping reading frames[7]. Previously, a popular concept was that alignment of the different reading frames occurred by a small splice in the mRNA, but this possibility was ruled out when the alignment was shown to occur by ribosomal frameshifting[7–9]. In fact, we now know *where* the frameshift occurs. We also know a great deal about *what* information is present in RNA for aligning the different reading frames. In addition, we know *how* the different reading frames are aligned and we have some insight into *how* the frameshift is accomplished; i.e. *how* the fidelity of translation is maintained during the repositioning of the ribosome on the mRNA.

D. Hatfield is at the Laboratory of Experimental Carcinogenesis, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA. **S. Oroszlan** is at the Laboratory of Molecular Virology and Carcinogenesis, BRI, Basic Research Program, National Cancer Institute-Frederick Cancer Research Facility, Frederick, MD 21701, USA.

Arrangement and translation of the retroviral *gag*, *pro* and *pol* genes

Overlapping genes occur in the *gag-pro-pol* translational unit of the genome-sized mRNA of many retroviruses. These genes are all expressed from the same AUG initiation codon. The *gag* gene encodes the structural proteins of the virus and the *pol* gene encodes the replication enzymes, reverse transcriptase and integrase. The replication enzymes are usually synthesized at a 5–10% level of the Gag proteins. The *pro* gene, which encodes the viral protease, can be expressed in the same reading frame as *gag* and *pol* or in a separate reading frame overlapping *gag* and/or *pol*. For example, in mammalian type C retroviruses, the *pro* and *pol* genes are expressed through suppression of the termination codon at the end of *gag* by a glutamine tRNA[10]; and thus, the same reading frame is maintained in translation of *gag* and *pol*. This means of expressing the Gag–Pol fusion protein is designated as in-frame readthrough (Fig. 1). In other retroviruses, the genes downstream of *gag* occur in different reading frames. Expression of the Gag-Pro-Pol fusion protein in some of these retroviruses requires a single frameshift event in the –1 direction to align the different reading frames, whereas expression in others requires two such events, one between *gag* and *pro* and the other between *pro* and *pol* (Fig. 1).

The frameshift site

The frameshift occurs within the region known as the overlapping reading frame. The boundaries of this region (also designated as the overlap or frameshift window) are established, for example, in retroviruses requiring a single frameshift event, by the termination codon at the end of *gag* (the 0 reading frame) and the first upstream termination codon in the –1 frame (the *pol* reading frame). The size of the overlap may vary from a few nucleotides (e.g. 13 in the *pro–pol* frameshift window of the mouse mammary tumor virus [MMTV])[12,13] to more than a hundred nucleotides (e.g. 178 in the *pro–pol* frameshift window of human T-cell lymphotrophic virus-1 [HTLV-1])[14]. In each overlap sequenced to date one of three common sequences occurs, A AAC, U UUA or U UUU (where asparagine [AAC], leucine [UUA] or phenylalanine [UUU] are read in the 0 frame), whose role is to signal the

 0376–5067/90/$02.00

frameshift event[15]. Mutagenesis studies within and around the frameshift region show that the signal actually consists of a heptanucleotide sequence[15,16]. Examples of frameshift signals are shown in Table I.

The heptanucleotide frameshift signal may occur anywhere within the overlap from the extreme 3' end (e.g. just before the *gag* termination codon in RSV)[15] to near the 5' end (e.g. just inside the 5' boundary of the window in human immunodeficiency virus-1 [HIV-1] which is 234 nucleotides upstream of the *gag* termination codon)[17]. The site of the frameshift is the 3' base at the end of the heptanucleotide signal (designated with an arrow in Fig. 2). This was demonstrated by sequencing the transframe protein or peptide, synthesized either *in vivo*[8] or *in vitro* (Refs 15 and 18, and S. H. Nam, T. D. Copeland, M. Hatanaka and S. Oroszlan, unpublished), through the frameshift site (see Fig. 2). The data show, for example, that leucine and isoleucine are generated from the UU*AUA* sequence in RSV (where UU*A* is the 3' terminal codon of the frameshift signal)[15]; and isoleucine is read in the –1 frame by *AUA* and leucine in the 0 frame by either UU (if the frames are aligned by two out of three base reading) or UU*A* (if the frames are aligned by overlapping reading). Similarly, asparagine and proline are generated from the AA*CCA* sequence in the *pro–pol* HTLV-1 overlapping reading frame (where AA*C* is the 3' terminal codon of the frameshift signal); and proline is read in the –1 frame by *CCA* and asparagine in the 0 frame by either AA or AA*C*. Thus, these studies define the frameshift site, but do not demonstrate whether the different reading frames are aligned as a result of two out of three base reading (doublet decoding) or overlapping reading as originally proposed by Hizi *et al.*[8]. We will return to this question in a later section.

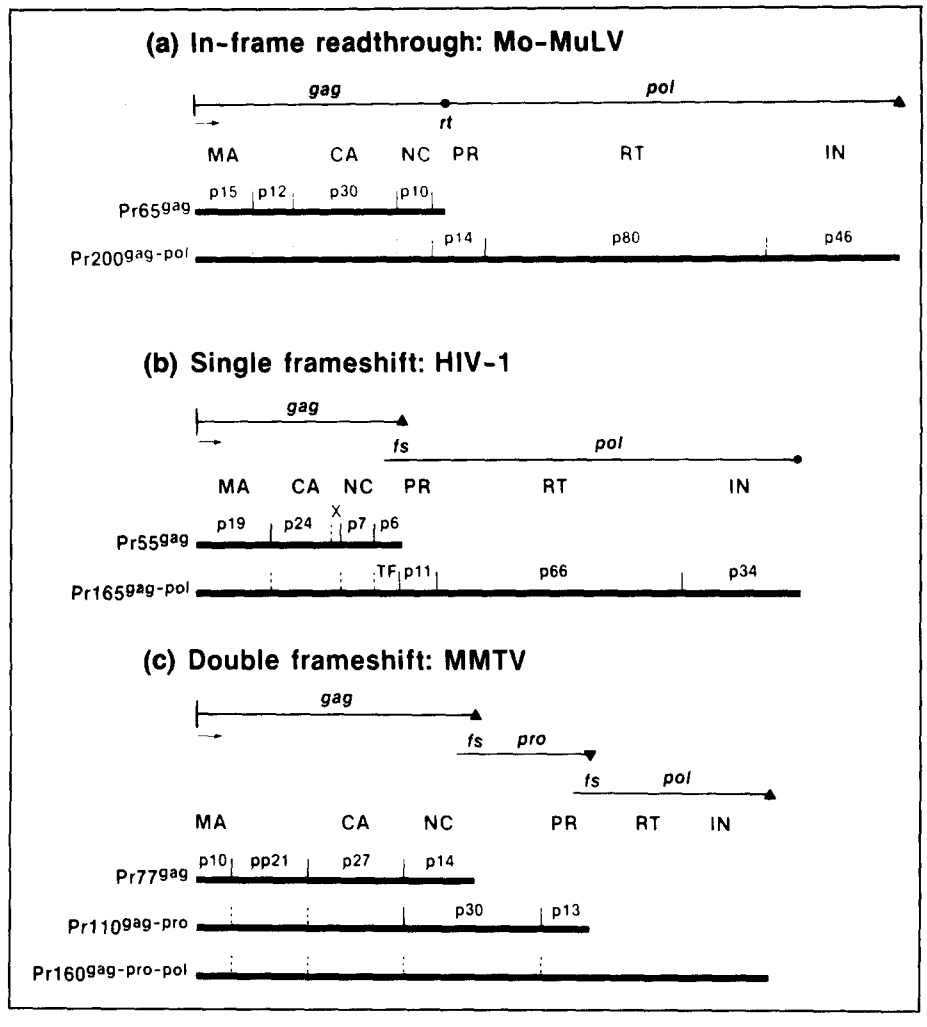

Figure 1

Mechanisms of expression of retroviral *pol* and *pro* genes from a single *gag-pro-pol* translational unit. The three known mechanisms are: (a) in-frame readthrough, (b) single frameshift and (c) double frameshift as illustrated in the figure for Moloney (Mo)-MuLV, HIV-1 and MMTV, respectively. The *gag*, *pro* and *pol* open reading frames are shown together with symbols for termination codons: ●, UAG, ▲, UAA and ▼, UGA. The horizontal bars in bold represent the primary translational products (Pr) which are processed into smaller functional units as indicated. The protein nomenclature used is that of Leis *et al.*[11]: Abbreviations are: rt, readthrough; fs, frameshift; MA, matrix; CA, capsid; NC, nucleocapsid; PR, protease; RT, reverse transcriptase; and IN, integrase. Location of transframe protein, TF, is shown within $Pr165^{gag-pol}$. The TF identified in MMTV at the *gag–pro* junction is p30. The numbers indicate the approximate molecular weight of the proteins. Arrows indicate the site of translation initiation.

Information present in RNA for frameshifting

There are at least two types of information present in RNA which signal the frameshift event. One type is encoded in the heptanucleotide signal discussed above. As noted, mutagenesis studies show that only the seven nucleotide stretch is required at the frameshift site[15,16] for the shift to occur. This conclusion is supported by other studies in which efficient frameshifting is maintained when the codon immediately upstream (in the –1 frame) and that immediately downstream (in the 0 frame) of the heptanucleotide frameshift signal in coronavirus infectious bronchitis virus (IBV) are replaced with nonsense codons (i.e. the frameshift window consists of only a seven nucleotide stretch; S. Inglis, pers. commun.).

The second type of information which has an important role in frameshifting is RNA secondary structure. Stem–loop structures occur just downstream of the *gag–pol* and of the *gag–pro* and *pro–pol* retroviral frameshift sites[7,13,15,18,19] as well as just downstream of the frameshift site in IBV[20,21]. In RSV, disruption of base pairings within the stem by generating specific stem-destabilizing mutations results in a decrease in frameshifting, while restoring these base pairings by generating specific stem-restabilizing mutations rescues frameshifting[15]. Deletion of certain bases further downstream of the stem–loop structure in IBV also inhibits frameshifting suggesting that many of these downstream bases interact with the stem–loop resulting in a tertiary structure known as a pseudoknot[21]. The occurrence and role of pseudoknots in retroviruses and in

Table I. Examples of frameshift signals

Frameshift signals	Viruses[a]	Location
A AAA AAC	BLV, HTLV-1 and -2, MMTV, STLV-1	*gag–pro*
U UUA AAC	BLV, HTLV-1 and -2, STLV-1, IBV	*pro–pol*
G GGA AAC	MPMV, SRV-1 and -2	*gag–pro*
A AAU UUU	MPMV, SRV-1 and -2	*pro–pol*
A AAU UUA	RSV	*gag–pol*
U UUU UUA	HIV-1 and -2, SIV	*gag–pol*
G GAU UUA	MMTV	*pro–pol*

[a] Abbreviations: BLV, bovine leukemia virus; HTLV, human T-cell lymphotrophic virus; MMTV, mouse mammary tumour virus; STLV-1, simian T-cell leukemia virus; IBV, infectious bronchitis virus; MPMV, Mason-Pfizer monkey virus; SRV-1 and -2, simian acquired immunodeficiency virus syndrome (SAIDS, designated SRV); RSV, Rous sarcoma virus; HIV, human immunodeficiency virus; and SIV, simian immunodeficiency virus. References to viral sequences are given in Jacks *et al.*[15] and Hatfield *et al.*[28]

other RNA structures have been reviewed recently[22,23]. Examples of stem–loop structures are shown in Fig. 3.

Location of the stem–loop structure relative to the frameshift site is also important since altering the distance between these two elements by as few as three nucleotides in either direction inhibits frameshifting[21]. Furthermore, the stem–loop structures are thermodynamically highly stable and their involvement in frameshifting is statistically relevant relative to other such configurations which may occur within several hundred nucleotides upstream or downstream of the frameshift site[24]. It seems, therefore, that the ribosomal frameshift event requires a carefully positioned downstream stem–loop structure for efficient frameshifting which may function to pause translation long enough for the shift to occur.

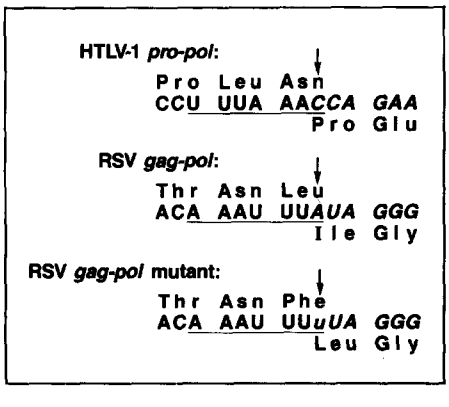

Figure 2

Sequence of the transframe protein at the frameshift site. Underlined letters show the frameshift signals; normal letters, the 0 reading frame; letters in italics, the -1 reading frame; an arrow indicates the frameshift site and *u* signifies the site of the mutation.

Alignment of the different reading frames

Mutations in the 3' base at the end of the heptanucleotide signal involving UUU or UUA codons do not inhibit frameshifting[15,16]. Since the sequence of the transframe peptide shows that the shift has to occur at the codon of which this is the third base, this suggests that the alignment must occur by doublet decoding. However, as sound as this argument may seem, the sequence of a transframe peptide generated from a frameshift signal containing a mutation in the 3' base demonstrates that the alignment occurs by overlapping reading. A single base change at the 3' end of the frameshift signal results in two new amino acids in the transframe protein. That is, alteration of the RSV UU*AUA* sequence, which codes for leucine (UU*A*) in the 0 frame and isoleucine (*AUA*) in the -1 frame, to UU*UUA* (where *U* represents the altered base which occurs at the 3' end of the frameshift signal) results in the occurrence of phenylalanine (UU*U*) and leucine (*U*UA) in the transframe protein[15] (see Fig. 2). This observation demonstrates that the base at the 3' end of the frameshift signal must be read twice; once in the 0 frame and once in the -1 frame[8]. Thus, alignment of the different reading frames occurs by overlapping reading of the 3' nucleotide at the end of the frameshift signal (Fig. 2).

The simultaneous-slippage model of frameshifting

The frameshift event involves translocation of the aminoacyl-tRNA and the peptidyl-tRNA (which are present in the ribosomal A- and P-sites, respectively) by one nucleotide in the 5' direction. We do not know precisely the means by which the frameshift is carried out, but the simultaneous-slippage model proposed by Jacks *et al.*[15] provides the best explanation of the frameshift event at present. In this model, the peptidyl-tRNA located in the P-site and the aminoacyl-tRNA located in the A-site are proposed to slip simultaneously by one base in the 5' direction resulting in both tRNAs misreading the 3' base or reading two out of three bases in the corresponding mRNA. The slippage prepares the ribosome to read the –1 frame. In the next step, normal transfer of the growing polypeptide to the aminoacyl-tRNA and its translocation to the P-site would bring the first codon in the –1 frame to the A-site. The –1 reading is then consummated with normal decoding of the A-site and transfer of the nascent peptide to the incoming aminoacyl-tRNA.

If the slippage model is correct, we would expect that the codon : anticodon interactions of tRNAs in the ribosomal A- and P-sites would not be altered significantly by the shift from the 0 frame (the frame in which the tRNAs were decoded) to the –1 reading frame (the frame in which the tRNAs now have a new set of codons). Otherwise, the ribosome:codon:anticodon complex may be destabilized and fall apart. Within the heptanucleotide signal, the base in the 3' position of the upstream codon in the 0 frame (i.e. U, A or G) is identical to the bases in the first two positions of the downstream codon (i.e. UU, AA or GG, respectively; see columns A–C, Fig. 4). Thus, the shift to the new reading frame maintains similar codon:anticodon interactions provided the isoacceptors in the A- and P-sites misread the base in the 3' position of the codon or read only two out of three bases. It should be noted that two out of three base reading cannot be distinguished from misreading the 3' base of the codon in many cases. There is a direct relationship between promotion of frameshifting and misreading or reading two out of three bases, since the same tRNAs are involved in both processes (i.e. tRNAs that promote frameshifting must then, after the frameshift event, misread the 3' base of the corresponding new set of codons). Interestingly, tRNAs which lack a modified base in their anticodon loop are known to promote frameshifting[25] and, for that matter, misreading[26,27]. A recent analysis of the tRNA in HIV infected cells showed that most of the Phe-tRNA lacked the highly modified

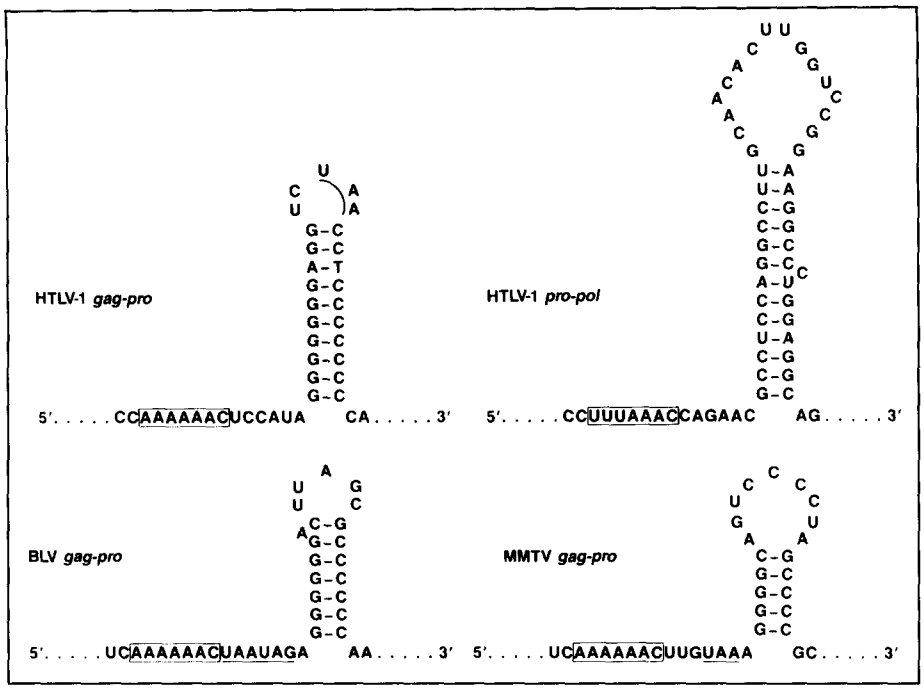

Figure 3

The predicted RNA stem–loop structures which occur just downstream of the ribosomal frameshift sites in HTLV-1[14] and of the *gag–pro* ribosomal frameshift sites in BLV[19] and MMTV[12]. The frameshift sites are boxed and the termination codons are underlined.

Wye base in its anticodon loop, while in bovine leukemia virus (BLV) and HTLV-1 infected cells most of the Asn-tRNA lacked the highly modified Q base in its anticodon loop[28]. This study showed a correlation between the occurrence of hypomodified tRNAs in retroviral infected cells and their utilization in translating codons within the respective frameshift signals. The lack of a hypermodified base in the anticodon loop of tRNA would create more space in and around the frameshift site and greater flexibility of movement of the tRNA anticodon would be expected in the absence of a highly modified base. Thus, decoding of a hypomodified tRNA at the frameshift site may be a requirement for promoting frameshifting[28].

Within the frameshift signals analysed to date, one exception to the occurrence of the same base in the 3' position and first two 5' positions of the upstream codons is the presence of an Asp codon (GAU) in the *pro–pol* frameshift signal of MMTV. A shift to the –1 reading frame would result in Asp-tRNA decoding GGA (a glycine codon) (column D, Fig. 4). Thus it would seem that, following the shift to the –1 reading frame, mismatching between the middle and third positions of the codon, and the first and second positions of the corresponding Asp-tRNA anticodon are permitted provided the simultaneous-slippage model is correct.

Conclusions

Our knowledge of the frameshift event in vertebrate viruses has increased substantially since it was first demonstrated by Jacks and Varmus[7] to be utilized in RSV as a means of aligning different reading frames. We now know that frameshifting in the –1 direction occurs by overlapping reading at the 3' terminal base within the frameshift signal. It is also apparent that a stem–loop structure, which occurs immediately downstream of the frameshift signal and which may also exist as a pseudoknot, is required for efficient frameshifting.

There are several areas in retroviral frameshifting of which further study is required. We need to know more about the status of the tRNAs involved in this process. We also need to know more about how the frameshift occurs and the reason it occurs at only a moderate to low level. We need to know more about the role of the ribosome in this process and if a specific protein may be required to bring about frameshifting. Additionally, it is important to know if ribosomal frameshifting in the +1 direction also occurs in higher eukaryotes. A central question to resolve, however, is whether frameshifting in the –1 direction is a requirement of the host cell. If frameshifting is not required by the host, then the frameshift event should provide a target for inhibiting expression of viruses utilizing this regulatory mechanism of gene expression in translation.

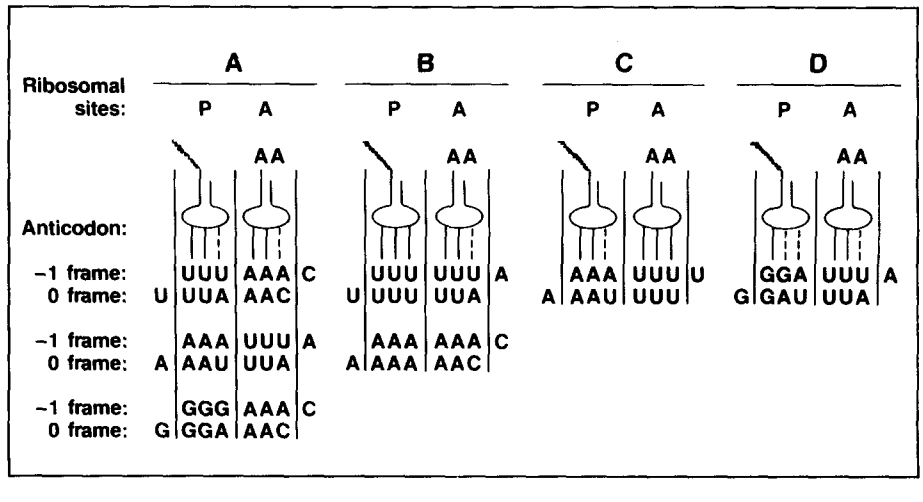

Figure 4

Bases read at the ribosomal A- and P-sites within the frameshift signal after the shift of the reading frame. The sequences shown represent a summary of the ribosomal frameshift signals determined in vertebrate viruses (see Fig. 1 and Refs 15 and 28). They are arranged in four classes (columns A–D) depending on the codon:anticodon interaction after the frameshift event as follows: Shift from the 0 to the –1 frame results in misreading, or reading two out of three bases (see text) in both the ribosomal A- and P-sites (A), just the A-site (B), or just the P-site (C); in column D, a shift to the –1 frame results in reading only one base in the P-site and two bases in the A-site by the standard Watson–Crick base pairings. Squiggly lines show the nascent polypeptides attached to tRNAs in the P-sites; AA, the amino acid attached to tRNA in the A-site, and the dashed line, mismatching in codon:anticodon interactions between standard Watson–Crick base pairs.

Acknowledgements

We thank Drs Reed Wickner, Samuel Wilson and Pat Becerra for helpful suggestions and critical evaluation of the manuscript and Dr Stephen Inglis for permission to quote his unpublished data on IBV. This work was sponsored in part by NCI, DHHS under contract number NO1-CO-74101 with BRI.

References

1 Cattaneo, R. (1989) *Trends Biochem. Sci.* 14, 165–167
2 Craigen, W. J. and Caskey, C. T. (1987) *Cell* 50, 1–2
3 Weiss, R. B., Dunn, J. F., Atkins, J. F. and Gesteland, R. F. (1987) *Cold Spring Harbor Symp. Quant. Biol.* 52, 687–693
4 Hughes, D., Thompson, S., O'Connor, M., Tuohy, T., Nichols, B. and Atkins, J. (1989) *J. Bacteriol.* 171, 1028–1036
5 Wickner, R. B. (1989) *FASEB J.* 3, 2257–2265
6 Mellor, J., Fulton, A. M., Dobson, M. J., Wilson, W., Kingsman, S. M. and Kingsman, A. J. (1985) *Nature* 313, 243–246
7 Jacks, T. and Varmus, H. E. (1985) *Science* 230, 1237–1242
8 Hizi, A., Henderson, L. E., Copeland, T. D., Sowder, R. C., Hixson, C. V. and Oroszlan, S. (1987) *Proc. Natl Acad. Sci. USA* 84, 7041–7045
9 Nam, S. H., Kidokoro, M., Shida, H. and Hatanaka, M. (1988) *J. Virol.* 62, 3718–3728
10 Yoshinaka, Y., Katoh, I., Copeland, T. D. and Oroszlan, S. (1985) *Proc. Natl Acad. Sci. USA* 82, 1618–1622
11 Leis, J., Baltimore, D., Bishop, J. M., Coffin, J., Fleissner, E., Goff, S. P., Oroszlan, S., Robinson, H., Skalka, A. M., Temin, H. M. and Vogt, V. (1988) *J. Virol.* 62, 1808–1809
12 Moore, R., Dixon, M., Smith, R., Peters, G. and Dickson, C. (1987) *J. Virol.* 61, 480–490
13 Jacks, T., Townsley, K., Varmus, H. E. and Majors, J. (1987) *Proc. Natl Acad. Sci. USA* 84, 4298–4302
14 Inoue, J-I., Watanabe, T., Sato, M., Oda, A., Toyoshima, K., Yoshida, M. and Seiki, M. (1986) *Virology* 150, 187–195
15 Jacks, T., Madhani, H. D., Masiarz, F. R. and Varmus, H. E. (1988) *Cell* 55, 447–458
16 Wilson, W., Braddock, M., Adams, S. E., Rathjen, P. D., Kingsman, S. M. and Kingsman, A. J. (1988) *Cell* 55, 1159–1169
17 Ratner, L., Haseltine, W., Patarca, R., Livak, K. J., Starcich, B., Josephs, S. F., Doran, E. R., Rafalski, J. A., Whitehorn, E. A., Baumeister, K., Ivanoff, L., Petteway, S. R., Pearson, M. L., Lautenberger, J. A., Papas, T. S., Ghrayeb, J., Chang, N. T., Gallo, R. C. and Wong-Staal, F. (1985) *Nature* 313, 277–284
18 Jacks, T., Power, M. D., Masiarz, F. R., Luciw, P. A., Barr, P. J. and Varmus, H. E. (1988) *Nature* 331, 280–283
19 Rice, N. R., Stephens, R. M., Burny, A. and Gilden, R. V. (1985) *Virology* 142, 357–377
20 Brierley, I., Boursnell, M. E. G., Binns, M. M., Bilimoria, B., Blok, V. C., Brown, T. D. K. and Inglis, S. G. (1987) *EMBO J.* 6, 3779–3785
21 Brierley, I., Digard, P. and Inglis, S. C. (1989) *Cell* 57, 537–547
22 Schimmel, P. (1989) *Cell* 58, 9–12
23 Pleij, C. W. A. (1990) *Trends Biochem. Sci.* 15, 143–147
24 Le, S-Y., Chen, J-H. and Maizel, J. V. (1989) *Nucleic Acids Res.* 17, 6143–6152
25 Björk, G. R., Wilkström, P. M. and Byström, A. S. (1989) *Science* 244, 986–989
26 Björk, G. R., Erickson, J. U., Gustafsson, C. E., Hagervall, T. G., Jönsson, Y. H. and Wilkström, P. M. (1987) *Annu. Rev. Biochem.* 56, 263–287
27 Wilson, R. K. and Roe, B. A. (1989) *Proc. Natl Acad. Sci. USA* 86, 409–413
28 Hatfield, D., Feng, Y-X., Lee, B. J., Rein, A., Levin, J. G. and Oroszlan, S. (1989) *Virology* 173, 736–742