

Programmed ribosomal frameshifting in the expression of the regulator of intestinal stem cell proliferation, adenomatous polyposis coli (APC)

Pavel V. Baranov,^{1,†,*} Norma M. Wills,^{2,†} Kathy A. Barriscale,^{1,3} Andrew E. Firth,⁴ Molly C. Jud,² Anthea Letsou,² Gerard Manning⁵ and John F. Atkins^{2,6,*}

¹Biochemistry Department; ⁶BioSciences Institute; University College Cork; Cork, Ireland; ²Department of Human Genetics; University of Utah; Salt Lake City, UT USA;

³Ph.D., Scholars Programme in Cancer Biology; University College Cork; Cork, Ireland; ⁴Department of Pathology; University of Cambridge; Cambridge, UK;

⁵Razavi Newman Center for Bioinformatics; Salk Institute for Biological Studies; La Jolla; CA USA

[†]These authors contributed equally to this work.

Key words: adenomatous polyposis coli, programmed ribosomal frameshifting, recoding, RNA pseudoknot, translational control, alternative decoding

Abbreviations: APC, adenomatus polypolis coli; PRF, programmed ribosomal frameshifting (or frameshift); ORF, open reading frame

A programmed ribosomal frameshift (PRF) in the decoding of APC (adenomatous polyposis coli) mRNA has been identified and characterized in Caenorhabditis worms, Drosophila and mosquitoes. The frameshift product lacks the C-terminal approximately one-third of the product of standard decoding and instead has a short sequence encoded by the -1 frame which is just 13 residues in *C. elegans*, but is 125 in *D. melanogaster*. The frameshift site is A AAA AAC in Caenorhabditids, fruit flies and the mosquitoes studied while a variant A AAA AAA is found in some other nematodes. The predicted secondary RNA structure of the downstream stimulators varies considerably in the species studied. In the twelve sequenced Drosophila genomes, it is a long stem with a four-way junction in its loop. In the five sequenced Caenorhabditis species, it is a short RNA pseudoknot with an additional stem in loop 1. The efficiency of frameshifting varies significantly, depending on the particular stimulator within the frameshift cassette, when tested with reporter constructs in rabbit reticulocyte lysates. Phylogenetic analysis of the distribution of APC programmed ribosomal frameshifting cassettes suggests it has an ancient origin and raises questions about the possibility of synthesis of alternative protein products during expression of APC in other organisms such as humans. The origin of APC as a PRF candidate emerged from a prior study of evolutionary signatures derived from comparative analysis of the 12 fly genomes. Three other proposed PRF candidates (Xbp1, CG32736, CG14047) with switches in conservation of reading frames are likely explained by mechanisms other than PRF.

Introduction

Diversion of a proportion of translating ribosomes to an alternative reading frame is utilized in the decoding of a substantial number of mobile genetic elements including viruses.^{1,2} This frameshifting is generally programmed in the sense that it is evolutionarily conserved and required for the synthesis of a particular functional protein or a variant protein with a distinct function. The efficiency of programmed ribosomal frameshifting (PRF) is often elevated above the background level by stimulatory elements, frequently called recoding signals, embedded in the mRNA. Many examples of programmed ribosomal frameshifting (PRF) are known in the decoding of viral genes and mobile chromosomal elements such

as in several families of bacterial insertion sequences,³ in the *S. cerevisiae* Ty elements,⁴ as well as in diverse higher eukaryotic retrotransposons.⁵ However, only few cases of PRF are known in the decoding of non-mobile cellular chromosomal genes.

The characteristics of frameshift sites and recoding signals are relevant to identifying new cases of PRF. Most -1 PRF sites are easier to identify than their +1 counterparts due to commonalities in the frameshift site that usually consists of a heptameric nucleotide sequence that allows dissociation of codon: anticodon pairing in both the ribosomal P- and A-sites and reassociation to mRNA in the -1 frame.⁶⁻⁹ However, some exceptions are known when the shift site involves just a tetrameric¹⁰ or a hexameric nucleotide sequence.¹¹

*Correspondence to: Pavel V. Baranov and John F. Atkins; Email: brave.oval.pan@gmail.com and atkins@genetics.utah.edu

Submitted: 01/16/11; Revised: 03/03/11; Accepted: 03/04/11

DOI: 10.4161/rna.8.4.15395

Such patterns are often accompanied by specific 3' stimulatory signals in the form of secondary RNA structures such as stem loops or pseudoknots, separated from the shift site by 5 to 9 bases.^{12,13} In some instances, however, a distal component also participates in conjunction with the 3' proximal sequence.¹⁴ A number of computational approaches for the identification of PRF have been developed.¹⁵ Those that predict PRF in homologs of known PRF genes achieve high sensitivity.^{16,17} However, elucidation of novel instances of PRF (where no PRF homologs are known) remains difficult. A direct search for sequence patterns known to be shift-prone may reveal numerous novel sites of ribosomal frameshifting, as has been shown in both prokaryotes¹⁸ and eukaryotes.¹⁹ While such instances of frameshifting may be biologically relevant (e.g., for fine-tuning mRNA turnover^{20,21}), we use the term 'programmed ribosomal frameshifting' to refer to cases where frameshifting has a more specific function (e.g., required for synthesis of a distinct functional product) and is subject to purifying selection as evidenced by phylogenetic conservation of the frameshift site. 'Unprogrammed', but nonetheless operational, sites of ribosomal frameshifting have been shown to have destabilizing effects on mRNAs.²¹ However it remains to be determined whether this is also the case where ribosomal frameshifting is 'programmed'. Such cases may indeed affect mRNA stability at some level or, alternatively, the corresponding mRNAs may have evolved specific stabilizing elements that negate the potential destabilizing effects of frameshifting.²²

Computational identification of PRF events is facilitated when both overlapping ORFs are long. In such cases, both ORFs can be easily identified, and the estimation of their coding potential is feasible. However, several PRF cases involve one very short ORF. Examples where the first ORF is short include the autoregulatory PRF in bacterial release factor 2,^{23,24} and antizyme²⁵ decoding. Short second ORFs are present in *E. coli dnaX*,^{26,27} and mammalian MA3,²⁸ frameshifting. The greatest phylogenetic range of these cases is seen for antizyme +1 frameshifting which acts as a sensor and effector for the control of intracellular polyamine levels from yeast to mammals.

Comparative sequence analysis is important for identifying novel occurrences of PRF as exemplified by recent work on alphaviruses and flaviviruses.^{29,30} Purifying selection acts on the evolution of most protein-coding genes and results in characteristic patterns of sequence change, where substitutions to synonymous codons dominate over changes to nonsynonymous codons. This can be used to identify the frame and extent of protein coding sequences.³¹ Using this principle, Lin and colleagues³² developed a computational approach based on two quantitative metrics: reading frame conservation and codon substitution frequencies, for identifying protein coding regions based on the analysis of multiple nucleotide alignments of closely related sequences. They applied this technique to re-evaluate *D. melanogaster* genome annotation using genomic alignments of 12 *Drosophila* species and identified four genes whose evolutionary signatures for protein coding regions switch from one translational frame into another. They proposed that these genes are candidates for PRF.³² They are Xbp1, CG32736, CG14047 and APC (adenomatous polyposis coli). We have computationally re-examined these

candidates to gain insight into what mechanisms may be responsible for the apparent transition of evolutionary signatures between different reading frames. This analysis has confirmed the prediction of APC as a gene with PRF, while transitions between frames in the other three cases are more likely due to other mechanisms. Further, we confirmed and experimentally characterized APC PRF on the frameshift cassettes from several species.

APC functions in the canonical Wnt/*Wingless* (*Wg*) signaling pathway in mammals and *Drosophila*³³ (see **Figure 1** for APC domain organization). APC homologs are found exclusively and universally in animals, along with other members of the Wnt pathway.³⁴ The Wnt pathway is implicated in many developmental processes, such as cell proliferation, polarity and migration.³⁵ Two paralogs of APC exist in mammals, APC and APC2, while two paralogs also exist in *Drosophila*, *dAPC* and *dAPC2*, although these are not direct orthologs of mammalian APC/APC2.^{36,37} Mutation of both the mammalian and *Drosophila* APC genes results in aberrant Wnt/*Wg* signaling and mutants of mammalian APC are an important direct cause of colon cancer.³¹ With respect to controlling cell growth and proliferation, however, a role for *dAPC* has recently been described in regulating the intestinal stem cell proliferation.³⁸ The *C. elegans* APC homolog *apr-1* seems to have a similar role as part of the Wnt pathway in regulating proliferation of the stem cell-like seam cells.^{39,40}

Results

Computational sequence analysis of APC PRF in insects. One of the four candidates identified by Lin et al.,²⁶ the shift in Xbp1, is due to a conserved nonstandard splicing event. In both yeast and human, ER stress activates the RNase domain of the IRE-1 protein causing the excision of a 23-base non-canonical intron in the Xbp1 transcript. This causes a shift in reading frame, leading to a longer protein encoding an active transcription factor. A pair of stem-loop structures is present at the splice sites in yeast and human and is conserved among insect orthologs (**Sup. Fig. 1**). *Drosophila* ESTs encode both spliced and unspliced forms, and splicing is induced by ER stress.⁴¹ Details of the analysis of Xbp1 and the following two candidates are in Supplemental material. The mRNA of one of them, CG32736, is probably bicistronic. Both ORFs are well conserved throughout coelomates, but are not well-annotated nor are they members of larger gene families. Although both ORFs are found in the same transcript in most coelomates (including invertebrate chordates and the lamprey, an early branching vertebrate), they are not always in the same reading frame. All other vertebrates have the two ORFs at different chromosomal locations. In contrast, the strong predicted frameshift in CG14047 is not easily explained and, though CG14047 is conserved throughout insects, its function is unknown. The nucleotide sequence of the shift region is inconsistent with known models of programmed ribosomal frameshifting. Moreover, the behavior cannot be explained by a single frameshifting event.

The shift in evolutionary signatures for the APC gene, the fourth candidate identified by Lin et al.,²⁶ is at position chr3R:24,661,410 (here and further genomic coordinates are given for 2006 BDGP R5 assembly). A nucleotide alignment

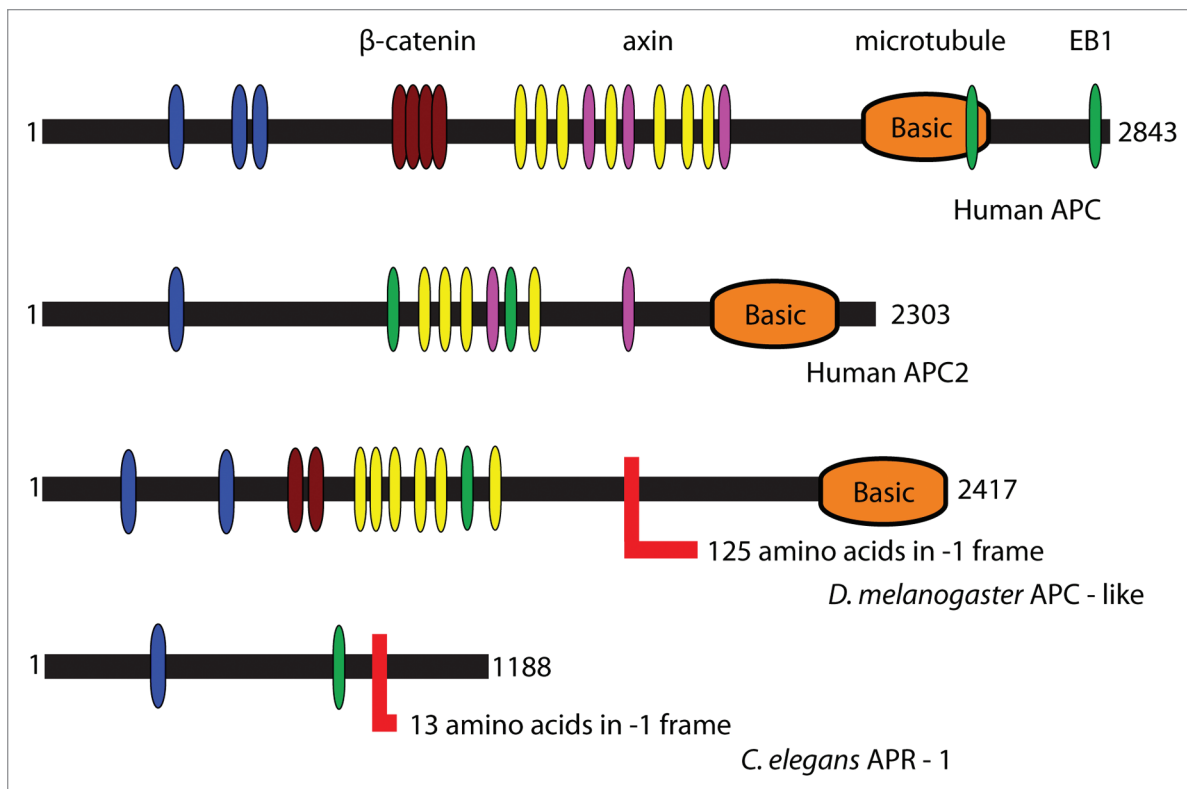


Figure 1. Domain organization of APC proteins. Highlighted are conserved PFAM domains identified by a PFAM search.⁷⁶ The color codes are as follows: blue-PF00514 (armadillo), brown-PF05972, yellow-PF05923 (cysteine-rich regions), purple-PF05924 (SAMP), orange-PF05956 (basic) and green-PF05937 (EB1 binding domain). Frameshift sites are shown with an L-like red shape and indicate the length of the C-terminal extension after PRF in each case.

of APC from the 12 sequenced *Drosophila* species genomes and that of the malaria mosquito *Anopheles gambiae*, is shown in **Figure 2**. The organization of the frameshift cassette is very similar to those that are used in many viral genomes to promote -1 frameshifting.¹² It consists of a tandem slippery site A_AA.A_AA.C (underscores separate codons in the initial frame and dots separate codons in the new -1 frame) that, following anticodon: codon dissociation, allows two tRNAs to re-pair to the mRNA in the -1 frame. The shift sequence is followed by predicted conserved mRNA secondary structures that may stimulate the frameshifting level. The tandem slippery site also has a conserved CC at its 5' end (in accordance with the hypothesis that the E-site is involved in modulating -1 frameshifting⁴²) and has a conserved G at its 3'-end (in accordance with an observation that the stacking potential of the third A-site codon base and its 3' adjacent base affects frameshifting efficiency⁴³). The predicted secondary structure is a four-way junction stem-loop that is conserved across all *Drosophila* species (**Fig. 3A** shows the predicted structure for *D. melanogaster*). Interestingly, mosquitoes have different predicted 3' stimulatory structures. In the malaria mosquito, *Anopheles gambiae*, there is a predicted RNA pseudoknot, while in the house mosquito, *Culex quinquefasciatus*, there is a strong RNA stem loop (**Fig. 3**). The Culicinae and Anophelinae clades are reported to have diverged about 150 Mya⁴⁴ and therefore a considerable sequence divergence is not surprising. The location of the shift in coding potential between frames detected by

Lin et al.³² is 143 nucleotides downstream from the predicted location of the actual frameshifting site described above—chr3R:24,661,696 genomic coordinate. The reasons for this are easy to understand. Since the likely frameshift cassette is highly conserved, there are few informative codons that can be used for the analysis of protein coding evolutionary signatures. So it is not surprising that Lin et al.³² predicted the shift location 3' of the bona fide frameshift site. Lin et al.³² reported a 3' switch of the evolutionary signature back to the zero frame, and this corresponds to the location of the stop codon in the -1 frame segment of the overlap (beyond which, there should be no purifying selection in the -1 frame).

The sequences of APC homologs from various species are highly divergent. This obstructs correct identification of the full APC CDS in some species. A further impediment to constructing correct APC gene models is its low expression level and consequent paucity of EST sequences. We failed to identify frameshifting in APC decoding in three hymenopteran genomes, those of the honey bee *Apis mellifera* and the ants *Harpegnathos saltator* and *Camponotus floridanus*. The genomic region between the two exons that share similarity with dAPC upstream and downstream of the PRF site lack the A AAA AAC motif. Either PRF is not used during expression of hymenopteran APC or the PRF cassettes diverge so much that we are not able to recognize it.

Identification of an APC frameshifting cassette in nematodes. A meaningful comparative sequence analysis similar to

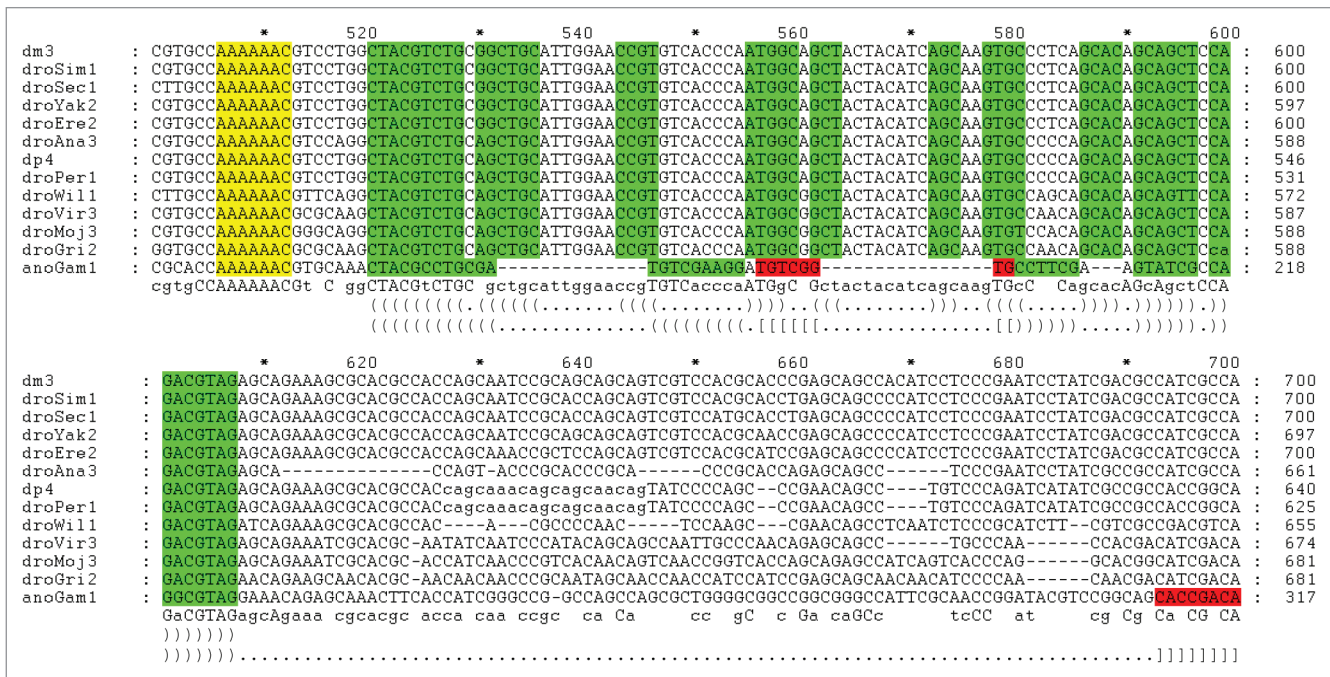


Figure 2. Multiple alignment of fly and malaria mosquito APC gene sequences corresponding to the region translated in the -1 frame. Heptameric frameshift sites are shown in yellow. Nucleotides involved in base-pairing interactions within the same double-stranded regions of the predicted secondary structure are differentially colored, e.g., the second stem of the predicted pseudoknot in the malaria mosquito (anoGam1) is shown in red. Secondary structures are also shown in bracket format below the alignment. The first bracket row corresponds to the four-way junction in fruit flies. The second bracket row corresponds to the malaria mosquito pseudoknot.

the one that revealed APC as a PRF candidate in *Drosophila* species³² is practical only for sequences that are relatively closely related. Evolutionary distances between APC sequences from nematodes of the Caenorhabditis genus allow such an analysis. To analyze evolutionary behavior of APC genes from Caenorhabditis their protein products were aligned, and then back-translated nucleotide alignments were subjected to analysis by the MLOGD program.⁴⁵ MLOGD uses nucleotide and amino acid substitution matrices to model sequence evolution and estimate the likelihood, in a sliding window, that an alignment is coding in each possible reading frame. This analysis did not reveal a significantly long region of the APC alignment in which substitutions would indicate evolutionary conservation of protein coding sequence in an alternative reading frame. However, a slight increase in the signal for the -1 frame was observed for a very short fragment of mRNA (Fig. 4). We further analyzed the alignment by examining conservation at synonymous sites using the method described in reference 30. Locally enhanced nucleotide conservation at synonymous positions is indicative of overlapping elements such as frameshift-stimulating motifs. This revealed a highly significant conservation peak that coincided with the short MLOGD coding signal for the -1 reading frame (Fig. 4). The sequence alignment around this peak shows a conserved motif of A AAA AAC, identical to the frameshift site of *Drosophila* APC (alignment, Fig. 5). Conservation of the sequence 3' from the shift site is apparently due to the existence of an RNA pseudoknot that likely plays a stimulatory role in the frameshifting. However,

the secondary structure involved is dramatically different from those used by fruit flies and mosquitoes (Fig. 3). Frameshifting at the A AAA AAC in the *C. elegans* APC gene would give a product with a C-terminal “tail” of 13 amino acids encoded by the -1 frame, whose last 7 amino acids and stop site are absolutely conserved within the 5 Caenorhabditids.

Among other nematode APC sequences, *Pristionchus pacificus* (whose last common ancestor with *C. elegans* lived 200–300 million years ago) appears to have an A AAA AAC shift site followed 6 nts 3' by a stem loop structure, while *Loa loa* (“eye worm”, the cause of a filariasis) and *Brugia malayi* (a cause of elephantiasis or lymphatic filariasis), likely have an A AAA AAA frameshift site (Fig. 5) but cassettes with the relevant sequences have not been tested.

Expression of APC frameshift cassettes in reticulocyte lysates. We tested -1 ribosomal frameshifting on the *Drosophila* APC sequence *in vitro*. The proposed frameshift heptanucleotide flanked by 15 nt 5' and 104 nt 3' (Fig. 6A) was cloned between the renilla and firefly luciferase genes in the pDluc vector⁴⁶ such that expression of the renilla-APC-firefly luciferase fusion requires -1 frameshifting. Expressing this construct in a rabbit reticulocyte lysate *in vitro* translation system resulted in two firefly luciferase-containing products (Fig. 6B, lane 2); one co-migrates with the in-frame control for -1 frameshifting at A AAA AAC (Fig. 6B, lane 1), while the other migrates more slowly. The primary frameshift at A AAA AAC occurs 1.7% of the time, and the secondary frameshift occurs at 0.3%. To delimit the origin of the second firefly luciferase-containing

product, stop codons were created in the -1 reading frame or in the zero reading frame, 5' or 3' of the shifty heptanucleotide, respectively. Occurrence of -1 frameshifting 5' of the new -1 stop would result in elimination of the secondary product. However, if it were to occur 3' of the shift site, then the introduced zero frame stop codon should preclude production of the secondary product. The results indicate that the secondary frameshift occurs 3' of the primary A AAA AAC site (Fig. 6B, lanes 3 and 4). [All subsequent constructs contain the zero frame stop codon to focus on the primary frameshift event; the nature of the secondary frameshift event is currently under investigation.]

Involvement of the predicted shift site in frameshifting was tested by changing the nucleotide sequence to disrupt tRNA:mRNA base pairing in the -1 frame (C AAG AAC) while maintaining the same amino acid sequence. As expected, this abolished the observed frameshift (Fig. 6B, lane 5). Several 3' deletions were designed to determine the importance of particular features of the proposed secondary structure (Fig. 6A). When the 3' sequence was limited to +83 nt, +62 nt, +44 nt, or +8 nt (Fig. 6B, lanes 6–9), frameshifting was decreased to less than 0.4% in all cases. These results are consistent with the importance of the 16 base-pair stem for frameshifting (it has one mismatch and its 3' end is 94 nts 3' of the shift site, Fig. 3). [A construct including ~800 additional nt of APC sequence 3' of the proposed stem-loop structure did not affect the level of frameshifting (data not shown) indicating the lack of long-range stimulatory signals.] Disruption of potential base pairing by mutations in the 5' or 3' part of the stem decreased frameshifting to 0.4% (Fig. 6B, lanes 10 and 11). However, the combination of the complementary 5' and 3' stem mutations, predicted to restore base-pairing (Fig. 6B, lane 12), restored frameshifting to 1.7%, the same level as the parent construct containing the zero frame stop codon 3' of the shift site.

The predicted frameshift signals in *A. gambiae* and *C. elegans* APCs (Fig. 3) were tested as described above for *D. melanogaster*. The sequences encompassing the proposed structures promoted 17% frameshifting in *Anopheles gambiae* APC and 11.5% frameshifting in *C. elegans* APC, both appreciably higher than their counterpart in *Drosophila melanogaster* (Fig. 6C). In the *C. elegans* sequence, the first stop codon in the -1 frame after the shift site is involved in base pairing in a stem of a proposed pseudoknot structure so that -1 frameshifting results in the production of a renilla luciferase-APC fusion (smaller product) while translation in the zero frame results in the renilla luciferase-APC-firefly luciferase fusion (larger product).

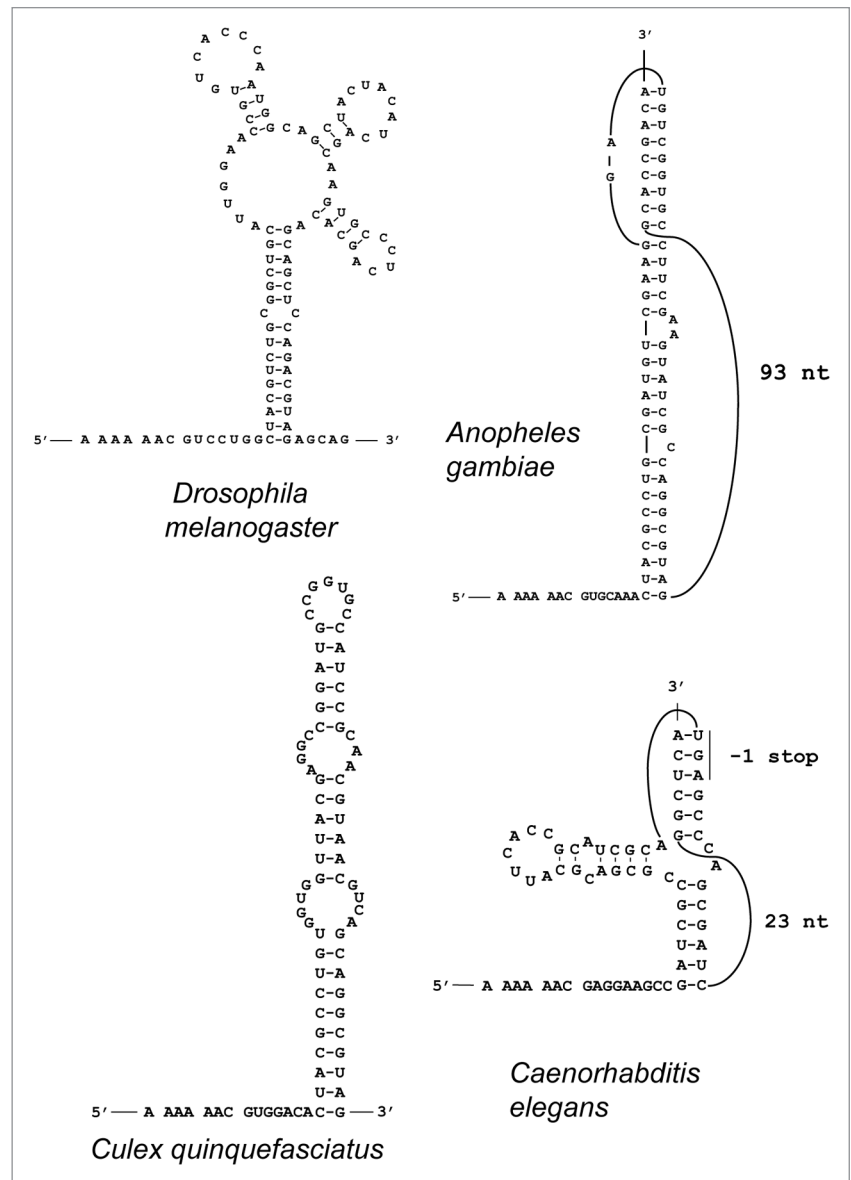


Figure 3. Predicted stimulatory PRF secondary RNA structures in flies and mosquitoes. The sequences shown include the A AAA AAC slippery pattern (with codons in the zero frame separated by spaces) at the 5' end and the entire mRNA regions forming the predicted stimulatory structures 3' from the shift site. Latin names of the organisms are indicated.

Discussion

Programmed ribosomal frameshifting is used in the decoding of the APC gene in *Caenorhabditis* nematodes, mosquitoes and one of the two paralogs in *Drosophila*. Phylogenetic analysis shows that the APC gene has undergone two independent duplications, in vertebrates and in insects (Fig. 7). The second *Drosophila* copy is genomically truncated at almost exactly the frameshift region and shows no sign of a frameshift signal. While we did not find the frameshift motif in some other insects, its conservation in nematodes suggests that it had an ancient origin. The poor sequence conservation downstream of the shift site makes it difficult to detect frameshift signals in other metazoan genomes,

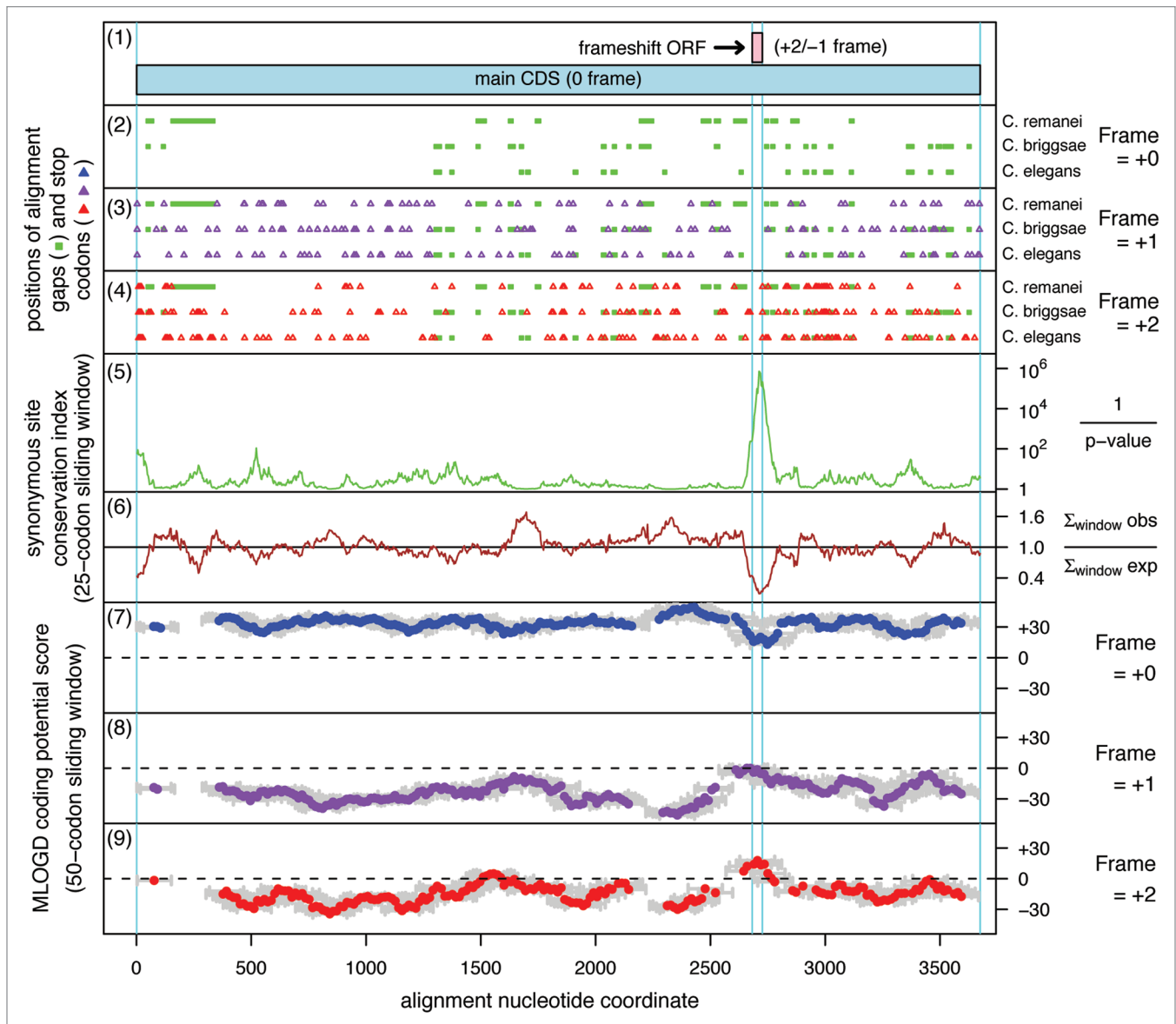


Figure 4. Coding potential statistics for *Caenorhabditis* APC. (1) Map of the APC coding sequence showing the predicted frameshift ORF. (2-4) The positions of stop codons in each of the three forward reading frames. The +0 frame corresponds to APC and is therefore devoid of stop codons. (5 and 6) Conservation at synonymous sites within APC (see ref. 30 for details). (5) depicts the probability that the degree of conservation within a given window could be obtained under a null model of neutral evolution at synonymous sites, while (6) depicts the absolute amount of conservation as represented by the ratio of the observed number of substitutions within a given window to the number expected under the null model. (7-9) MLOGD sliding window plots (see ref. 61 for details). The null model, in each window, is that the sequence is non-coding, while the alternative model is that the sequence is coding in the given reading frame. Positive scores favor the alternative model and, as expected, in the +0 frame (7) there is a strong coding signature throughout APC. In the +1 and +2 frames (8 and 9), scores are generally negative with some random scatter into low positive scores, except in the region occupied by the predicted frameshift ORF, where there is a strong positive coding signal in the +2/-1 reading frame. (Due to the limited sequence data available, a 50-codon sliding window was used to increase the signal-to-noise ratio throughout the alignment, notwithstanding that this is expected to dilute the signal from the overlapping ORF itself—since, at ~16 codons, the overlapping ORF is much shorter than the 50-codon window). Note that, regardless of the sign (either positive or negative), the magnitude of MLOGD scores tends to be lower within the overlap region (7-9) due to there being fewer substitutions with which to discriminate the null model from the alternative model in this region of above-average nucleotide conservation.

suggesting that directed cDNA cloning of APC homologs may be required to fully address the origin and extent of this mechanism. However, its existence in flies and nematodes implies that exploitation of this mechanism for the expression of APC protein variants is beneficial. Therefore, its existence in other animal lineages is not unexpected, but its location within APC and its sequence

organization could vary considerably, thus identification of PRF in distant sequences may not be an ordinary task.

The PRF product is shorter than the standard product and shares about two thirds of the same sequence at its N-terminal region (Fig. 1). The proximal gene segment encoding the shared N-terminal part is significantly more conserved than the distal

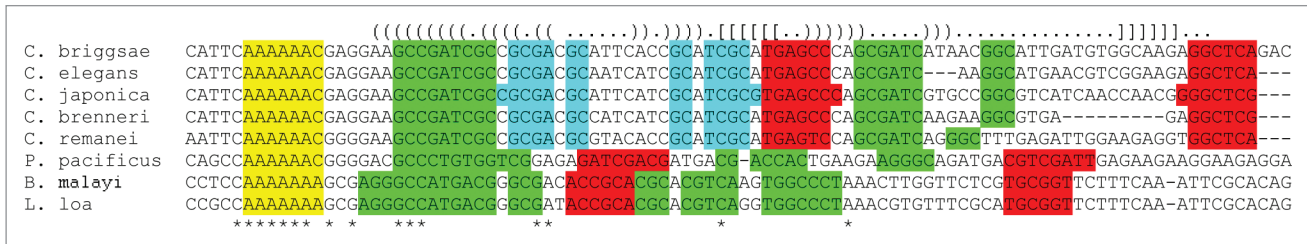


Figure 5. Multiple alignment of genomic sequences in the region of the PRF cassette from seven nematodes. Heptameric frameshift sites are shown in yellow. Nucleotides involved in predicted base-pairing interactions within the same double stranded regions of the secondary structure are differentially colored (e.g., the first stem of the pseudoknot is green, the second is red and the stem in the first loop of the pseudoknot is in blue). The secondary structure of the RNA pseudoknot in the *C. elegans* sequence is also shown in bracket format above the alignment and its diagram is shown in Figure 2.

part of the gene. While APC is a multifunctional protein, it is the central region of the protein that is most directly implicated in Wingless signaling. Interestingly, there is a strong tendency for removal of the carboxyl terminus of the protein in mutations that lead to colorectal cancer in humans, mostly (60%) by point mutations leading to frameshift or nonsense mutations. Based on these findings, it is thought that a truncated form of APC, which lacks axin/conductin interaction sites, leads to the constitutive Wg/Wnt signaling and tumorigenesis.⁴⁷ This suggests that the frameshifted products may have distinct functions to the full length protein.

Multiple functions have been attributed to APC and its interacting partners. While several domains are required for its best characterized function as a protein that binds to β -catenin and thus downregulates Wnt/Wg signaling, APC also interacts with several proteins with established roles in cytoskeleton dynamics, in particular those associated with either cell migration (ASEF, APC-stimulated guanine nucleotide exchange factor; IQGAPI, IQ domain-containing GTP-ase activating protein; KAP3, kinesin super-family associated protein 3; mDIA, diaphanous) or polarized cell division (EB1, end binding 1; hDLG, human Disc Large; Scribble) (reviewed in ref. 48). The C-terminal part of mammalian APC has two domains involved in microtubule binding: a large region containing positively charged amino acids (basic domain),⁴⁹ and in APC1, at least, an EB1 binding domain.⁵⁰ Though not sufficient on its own, the core sequence of an EB1 binding site is SxIP.⁵¹ Curiously, the *Drosophila* APC and *C. elegans* APC studied here have just one SxIP sequence, and it is encoded 5' of the frameshift site while human APC-2 has two SxIP sequences both of which are in the middle of the protein. The possible significance of these, including juxtaposition to the frameshift site, is unknown. It is notable that interactions specific for polarized cell division are confined to the conserved region that is lost in the ribosomal frameshift derived dAPC products. Considering these data, along with the knowledge that a truncated form of APC is functional in tumorigenesis, we suspect that ribosomal frameshifting limits the way in which APC functions *in vivo*. Both fly and worm systems will provide good genetic models in which to test the physiologically relevant role(s) of frameshifted APC. This being said, our initial attempts to generate isoform-specific peptide antibodies to delimit the expression and roles of frameshifted and unframeshifted APC isoforms in *Drosophila* were not successful (data not shown).

The recent demonstration that *Drosophila* APC plays an important role in maintenance of *Drosophila* intestinal stem cells is particularly intriguing with respect to our understanding of the role of human APC as a tumor suppressor.³⁸ Best known for its role in colorectal cancer, several *in vivo* studies in mouse have shown that properly regulated Wnt/Wg signaling is essential for normal proliferation of the precursor cells of the intestinal epithelium.⁵²⁻⁵⁴ Thus, the functions of fly and human APC in regulating cell growth and proliferation are likely conserved. Consistent with this idea is the demonstration that *Drosophila* APC can interact with the Scribbles and Dlg proteins (which result in disc overgrowth phenotypes when mutated), even though *Drosophila* APC has not itself been shown in genetic studies to function as a tumor suppressor. Notably, the Scribbles and Dlg interaction sites are not present in the frameshifted dAPC isoform.⁴⁸

Since the focus of selection in the region immediately 3' of what is now known to be the frameshifting site in *Drosophila* is the -1 rather than the zero frame, presumably the segment of the shorter protein encoded by the -1 frame (the C-terminal segment) is likely functional also. However, the corresponding region in *C. elegans* APC is very short, and we cannot exclude the possibility that selection there has only been operating to synthesize a truncated version of the product of standard decoding—one that lacks functions associated with the full-length product.

Whether decoding of either human APC gene yields a frameshift-derived product is unknown, but alternative forms of APC are known to be synthesized. Three alternative transcripts have been identified. (Internal IRES-mediated initiation has also been reported in ref. 55 and ref. 56.) One of the alternative transcripts (Refseq: NM_001127511) produces short protein isoform A (Refseq: NP_001120983), and two others (NM_000038 and NM_001127510) produce long isoform B (NP_001120983). Isoform A and B differ by a single exon encoding 27 amino acid residues in the N-terminal part of APC.⁵⁷

Interestingly, while the frameshift site is A AAA AAC in the *C. elegans*, *D. melanogaster* and mosquito cassettes tested here, the stimulatory secondary RNA structures are highly variable (see Fig. 3). The predicted structures in worms and the malaria mosquito (*A. gambiae*) are H-type pseudoknots and there is no significant sequence similarity between the worm and mosquito versions of the pseudoknot. Moreover, different worms appear to have strikingly different pseudoknots (see Fig. 5). In flies and

efficiency was tested in a heterologous mammalian in vitro system, these efficiencies may not correctly reflect the natural level of the frameshifting. Despite the low efficiency of 1.7% found with the *D. melanogaster* cassette, the strong evolutionary conservation suggests the frameshift product plays an important role. Presumably either only a very small amount is required or the endogenous levels in the cells where it is normally expressed is higher, conceivably due to a trans-acting factor. Also, it is important to keep in mind that the efficiency of frameshifting could be affected by expression levels⁵⁹ and translation initiation rates⁶⁰ of the corresponding mRNA, so that the frameshifting efficiencies determined by the use of artificial genetic constructs may not be truly representative of endogenous frameshifting levels.

The organization of ORFs involved in the production of the alternative PRF APC protein product is unusual and suggests that there might be similar instances of PRF in other chromosomal genes that cannot be readily identified by computational methods. The main reason why PRF was identified in fruit flies was the existence of strong evolutionary selection acting on the protein sequence encoded by the alternative frame while the selection on the standard frame was more relaxed. This created the sequence region with the major protein-coding evolutionary signature corresponding to the alternative frame, as detected by Kellis and colleagues³² and MLOGD.⁶¹ Interestingly the selection is much stronger in the -1 frame even though the product of the frameshifting is minor in comparison with the product of standard translation. Although in worms the evolutionary signatures corresponding to the alternative frame are also detectable, the additional peptide generated by the sequence downstream of the frameshift site is much shorter than its fly counterpart. In principle, the situation with APC may well be a fortunate exception. In general, when the functional importance of a frameshift product depends simply on its truncated nature rather than the actual peptide encoded by the new frame, or where selection acting on the protein sequence encoded in the zero frame is similar or stronger than in the alternative frame, frameshifting would not be easily detectable by methods based on identifying protein coding evolutionary signatures. This applies in decoding the *E. coli dnaX* gene, where the 50% of ribosomes that shift frame at the programmed frameshift site only translate one codon in the new frame before terminating.²⁶ In fact, we were able to identify PRF in *Caenorhabditis* nematodes foremost because we expected it to be there and specifically searched for it using different methods. Most likely the frame switch between evolutionary signatures would not be detectable in a genome-wide analysis similar to the one performed by Lin et al.³² for fruit flies. Thus, on one hand, identification of PRF in APC illustrates the applicability of comparative sequence analysis for the identification of novel instances of programmed ribosomal frameshifting. On the other hand, it suggests that many instances of PRF would not be detectable by such methods. To accurately investigate the occurrence and utilization of ribosomal frameshifting on a whole genome scale, it is necessary to develop high-throughput techniques for the identification of non-standard decoding. One possibility for the development of such methods is utilization of the ribosomal profiling technique that allows measurement of the density and location of ribosomes on mRNA with sub-codon precision.⁶²

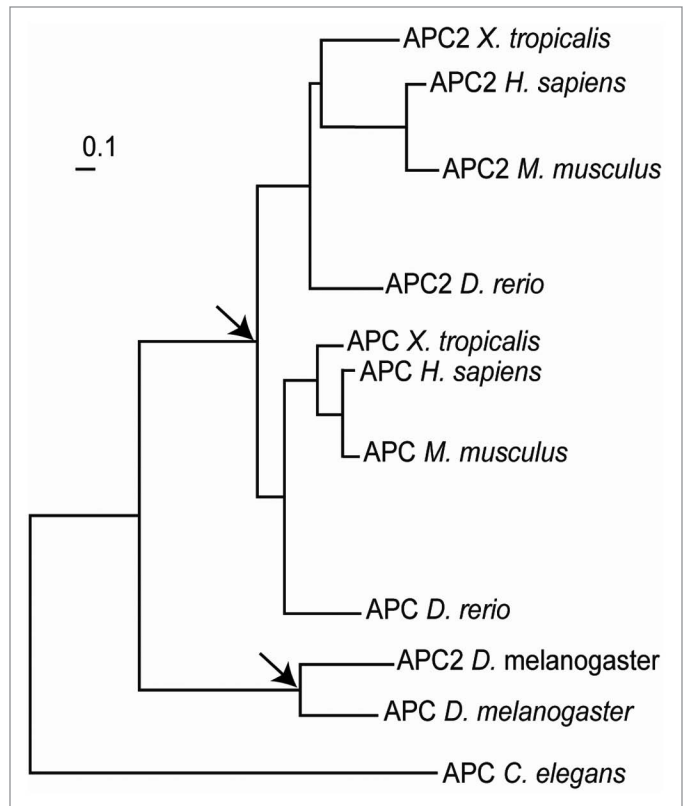


Figure 7. Phylogenetic tree of APC genes extracted from TreeFam.⁷⁰ The nodes corresponding to the events of independent APC gene duplications are indicated with arrows.

Materials and Methods

Software, computational methods and databases. The coordinates for predicted switches of reading frames in the *D. melanogaster* genome were obtained from the online supplemental materials for Lin et al.³² at www.broadinstitute.org/~mlin/fly_genes/dmel_frameshifts.dat. Conversions between coordinates from different assemblies were made with the FlyBase LiftOver tool.⁶³ APC homologs were identified in the NCBI NR and dbEST databases by Blast, and were repredicted using GeneWise⁶⁴ and genomic sequences from the UCSC genome browser,⁶⁵ using a variety of homology templates. MultiZ alignments⁶⁶ of genomic sequences used in this study were obtained from the UCSC conservation tracks for 12 Flies, Mosquito, Honeybee and Beetle MultiZ Alignments.⁶⁵ To generate multiple alignments of CDSs for the analysis by MLOGD⁶¹ and synonymous site conservation,³⁰ nucleotide sequences were translated into protein sequences and aligned using MUSCLE⁶⁷ and then the resulting protein alignments were backtranslated to nucleotide alignments using t-coffee.⁶⁸ CDS prediction for worm APC genes was made with the help of the GeneWise program.⁶⁴ Visualization of sequences for the manual analysis was done with the assistance of Gendoc v2.6.002 (www.nrbcs.org/gfx/genedoc/),⁶⁹ pDraw32 v1.1.107 (www.acaclone.com) and SequenceCruncher v. 1.09 (gesteland.genetics.utah.edu/bmoore/Downloads/SeqCrn.html). Phylogenetic trees of

APC genes were obtained from TreeFam⁷⁰ and visualized using TreeView.⁷¹ Initial secondary RNA structure predictions were carried out with MFOLD⁷² and SFOLD,⁷³ and then manually reevaluated to examine the potential existence of RNA pseudoknots and for verification via compensatory substitutions supporting the structure. Visualization of secondary RNA structures was performed with PseudoViewer 3.⁷⁴

Analysis of frameshifting efficiency in WT and mutant PRF cassettes. *Construction of plasmids.* The sequences encompassing the predicted frameshift site and the 3' stimulatory structure (frameshift cassette) for *Drosophila melanogaster*, *Anopheles gambiae* and *C. elegans* APCs were generated using overlapping synthetic oligonucleotides and cloned into the XhoI and BglII sites in vector pDluc,⁴⁶ a derivative of the p2luc vector.⁷⁵ The firefly luciferase gene is in the -1 reading frame relative to the upstream renilla luciferase gene such that -1 frameshifting within the frameshift cassette results in a renilla luciferase-APC-firefly luciferase fusion product. A *Drosophila melanogaster* APC insert including ~800 nt 3' of the predicted stem-loop structure was synthesized by GenScript and cloned into the HindIII and BglII sites of pDluc (due to an internal XhoI site in the APC sequence). All plasmids were verified by DNA sequencing.

Frameshift assays. Plasmid DNAs (0.2 µg) were used as templates in 10 µl reactions of the rabbit reticulocyte lysate TNT[®] T7 Quick Coupled Transcription/Translation system (Promega).³⁵ S-methionine (Perkin Elmer) was included in the reactions and protein products were separated by SDS-PAGE. Dried gels were

analyzed using a Typhoon PhosphorImager (GE Healthcare) and the amount of radioactivity in each product was determined using the ImageQuant 5.2 program (Molecular Dynamics). After normalization for the number of methionine residues in termination and frameshift products (9 and 22, respectively, for *Drosophila* APC), the frameshift efficiencies were calculated [frameshift/(frameshift + termination)]. Note that in the *C. elegans* APC sequence, the -1 termination codon is located within the predicted stimulatory structure, so the smaller product represents the frameshift product.

Acknowledgements

The authors would like to thank (1) Dr. Avril Coghlan for useful suggestions related to identification of APC sequences in nematodes and for her help with TreeFam and (2) Dr. Erik Jorgensen for drawing our attention to the location (with respect to the frameshift site) of the SxIP motif in *C. elegans* and its potential implication in binding EB1.

Financial Support

This work was supported by Science Foundation Ireland (grant number 06/IN.1/B81 to P.V.B., 08/IN.1/B1889 to J.F.A.), the Wellcome Trust (grant number 088789 to A.E.F.), and NIH (grant number HG004164) to G.M.

Note

Supplemental materials can be found at: www.landesbioscience.com/journals/rnabiology/article/15395

References

- Bekaert M, Firth AE, Zhang Y, Gladyshev VN, Atkins JF, Baranov PV. Recode-2: new design, new search tools, and many more genes. *Nucleic Acids Res* 2010; 38:D69-74.
- Atkins JF, Gesteland, RF, eds. *Recoding: Expansion of Decoding Rules Enriches Gene Expression*. New York and Heidelberg: Springer, 2010.
- Baranov PV, Fayet O, Hendrix RW, Atkins JF. Recoding in bacteriophages and bacterial IS elements. *Trends Genet* 2006; 22:174-81.
- Stahl G, Ben Salem S, Li Z, McCarty G, Raman A, Shah M, et al. Programmed +1 translational frameshifting in the yeast *Saccharomyces cerevisiae* results from disruption of translational error correction. *Cold Spring Harbor Symp Quant Biol* 2001; 66:249-58.
- Gao X, Havecker ER, Baranov PV, Atkins JF, Voytas DF. Translational recoding signals between gag and pol in diverse LTR retrotransposons. *RNA* 2003; 9:1422-30.
- Jacks T, Power MD, Masiarz FR, Luciw PA, Barr PJ, Varmus HE. Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. *Nature* 1988; 331:280-3.
- Weiss RB, Dunn DM, Shuh M, Atkins JF, Gesteland RF. *E. coli* ribosomes re-phase on retroviral frameshift signals at rates ranging from 2 to 50 percent. *New Biologist* 1989; 1:159-69.
- Dinman JD, Icho T, Wickner RB. A -1 ribosomal frameshift in a double-stranded RNA virus of yeast forms a gag-pol fusion protein. *Proc Natl Acad Sci USA* 1991; 88:174-8.
- Brierley I, Jenner AJ, Inglis SC. Mutational analysis of the "slippery-sequence" component of a coronavirus ribosomal frameshifting signal. *J Mol Biol* 1992; 227:463-79.
- Sekine Y, Eisaki N, Ohtsubo E. Translational control in production of transposase and in transposition of insertion sequence IS3. *J Mol Biol* 1994; 235:1406-20.
- Licznar P, Mejlhede N, Prere ME, Wills N, Gesteland RF, Atkins JF, et al. Programmed translational -1 frameshifting on hexanucleotide motifs and the wobble properties of tRNAs. *EMBO J* 2003; 22:4770-8.
- Brierley I, Pennell S. Structure and function of the stimulatory RNAs involved in programmed eukaryotic-1 ribosomal frameshifting. *Cold Spring Harb Symp Quant Biol* 2001; 66:233-48.
- Giedroc DP, Cornish PV. Frameshifting RNA pseudoknots: structure and mechanism. *Virus Res* 2009; 139:193-208.
- Barry JK, Miller WA. A -1 ribosomal frameshift element that requires base pairing across four kilobases suggests a mechanism of regulating ribosome and replicase traffic on a viral RNA. *Proc Natl Acad Sci USA* 2002; 99:11133-8.
- Firth AE, Bekaert M, Baranov PV. Computational resources for studying recoding. In: Atkins JF, Gesteland, RF, eds. *Recoding: Expansion of Decoding Rules Enriches Gene Expression*. New York and Heidelberg: Springer, 2010:281-300.
- Bekaert M, Atkins JF, Baranov PV. ARFA: a program for annotating bacterial release factor genes, including prediction of programmed ribosomal frameshifting. *Bioinformatics* 2006; 22:2463-5.
- Bekaert M, Ivanov IP, Atkins JF, Baranov PV. Ornithine decarboxylase antizyme finder (OAF): fast and reliable detection of antizymes with frameshifts in mRNAs. *BMC Bioinformatics* 2008; 9:178.
- Gurvich OL, Baranov PV, Zhou J, Hammer AW, Gesteland RF, Atkins JF. Sequences that direct significant levels of frameshifting are frequent in coding regions of *Escherichia coli*. *EMBO J* 2003; 22:5941-50.
- Jacobs JL, Belew AT, Rakauskaitė R, Dinman JD. Identification of functional, endogenous programmed -1 ribosomal frameshift signals in the genome of *Saccharomyces cerevisiae*. *Nucleic Acids Res* 2007; 35:165-74.
- Plant EP, Wang P, Jacobs JL, Dinman JD. A programmed -1 ribosomal frameshift signal can function as a cis-acting mRNA destabilizing element. *Nucleic Acids Res* 2004; 32:784-90.
- Belew AT, Advani VM, Dinman JD. Endogenous ribosomal frameshift signals operate as mRNA destabilizing elements through at least two molecular pathways in yeast. *Nucleic Acids Res* 2011; 39:2799-808.
- Weil JE, Hadjithomas M, Beeman KL. Structural characterization of the Rous sarcoma virus RNA stability element. *J Virol* 2009; 83:2119-29.
- Craigie WJ, Caskey CT. Expression of peptide chain release factor 2 requires high-efficiency frameshift. *Nature* 1986; 322:273-5.
- Baranov PV, Gesteland RF, Atkins JF. Release factor 2 frameshifting sites in different bacteria. *EMBO Rep* 2002; 3:373-7.
- Ivanov IP, Matsufuji S. Autoregulatory frameshifting in antizyme gene expression governs polyamine levels from yeast to mammals. In: Atkins JF, Gesteland, RF, eds. *Recoding: Expansion of Decoding Rules Enriches Gene Expression*. New York and Heidelberg: Springer, 2010:281-300.
- Tsuchihashi Z, Brown PO. Sequence requirements for efficient translational frameshifting in the *Escherichia coli* dnaX gene and the role of an unstable interaction between tRNA(Lys) and an AAG lysine codon. *Genes Dev* 1992; 6:511-9.
- Larsen B, Wills NM, Gesteland RF, Atkins JF. tRNA-mRNA base pairing stimulates a programmed -1 ribosomal frameshift. *J Bacteriol* 1994; 176:6842-51.

28. Wills NM, Moore B, Hammer A, Gesteland RF, Atkins JF. A functional -1 ribosomal frameshift signal in the human paraneoplastic Ma3 gene. *J Biol Chem* 2006; 281:7082-8.
29. Firth AE, Chung BY, Fleeton MN, Atkins JF. Discovery of frameshifting in Alphavirus 6K resolves a 20-year enigma. *Virology* 2008; 5:108.
30. Firth AE, Atkins JF. A conserved predicted pseudoknot in the NS2A-encoding sequence of West Nile and Japanese encephalitis flaviviruses suggests NS1' may derive from ribosomal frameshifting. *Virology* 2009; 6:14.
31. Nekrutenko A, Makova KD, Li WH. The K(A)/K(S) ratio test for assessing the protein-coding potential of genomic regions: an empirical and simulation study. *Genome Res* 2002; 12:198-202.
32. Lin MF, Carlson JW, Crosby MA, Matthews BB, Yu C, Park S, et al. Revisiting the protein-coding gene catalog of *Drosophila melanogaster* using 12 fly genomes. *Genome Res* 2007; 17:1823-36.
33. Logan CY, Nusse R. The Wnt signaling pathway in development and disease. *Annu Rev Cell Dev Biol* 2004; 20:781-810.
34. Srivastava M, Simakov O, Chapman J, Fahey B, Gauthier ME, Mitros T, et al. The Amphimedon queenslandica genome and the evolution of animal complexity. *Nature* 2010; 466:720-6.
35. Cadigan KM, Nusse R. Wnt signaling: a common theme in animal development. *Genes Dev* 1997; 11:3286-305.
36. van Es JH, Kirkpatrick C, van de Wetering M, Molenaar M, Miles A, Kuipers J, et al. Identification of APC2, a homologue of the adenomatous polyposis coli tumour suppressor. *Curr Biol* 1999; 9:105-8.
37. Aoki K, Taketo MM. Adenomatous polyposis coli (APC): a multi-functional tumor suppressor gene. *J Cell Sci* 2007; 120:3327-35.
38. Lee WC, Beebe K, Sudmeier L, Michelli CA. Adenomatous polyposis coli regulates *Drosophila* intestinal stem cell proliferation. *Development* 2009; 136:2255-64.
39. Huang X, Tian E, Xu Y, Zhang H. The *C. elegans* engrailed homolog *ceh-16* regulates the self-renewal expansion division of stem cell-like seam cells. *Dev Biol* 2009; 333:337-47.
40. Ren H, Zhang H. Wnt signaling controls temporal identities of seam cells in *Caenorhabditis elegans*. *Dev Biol* 2010; 345:144-55.
41. Ryoo HD, Domingos PM, Kang MJ, Steller H. Unfolded protein response in a *Drosophila* model for retinal degeneration. *EMBO J* 2007; 26:242-52.
42. Leger M, Dulude D, Steinberg SV, Brakier-Gingras L. The three transfer RNAs occupying the A, P and E sites on the ribosome are involved in viral programmed -1 ribosomal frameshift. *Nucleic Acids Res* 2007; 35:5581-92.
43. Bertrand C, Prere MF, Gesteland RF, Atkins JF, Fayet O. Influence of the stacking potential of the base 3' of tandem shift codons on -1 ribosomal frameshifting used for gene expression. *RNA* 2002; 8:16-28.
44. Calvo E, Pham VM, Marinotti O, Andersen JF, Ribeiro JM. The salivary gland transcriptome of the neotropical malaria vector *Anopheles darlingi* reveals accelerated evolution of genes relevant to hematophagy. *BMC Genomics* 2009; 10:57.
45. Firth AE, Brown CM. Detecting overlapping coding sequences with pairwise alignments. *Bioinformatics* 2005; 21:282-92.
46. Fixsen SM, Howard MT. Processive selenocysteine incorporation during synthesis of eukaryotic selenoproteins. *J Mol Biol* 2010; 399:385-96.
47. Fearnhead NS, Britton MP, Bodmer WF. The ABC of APC. *Hum Mol Genet* 2001; 10:721-33.
48. Schneikert J, Behrens J. The canonical Wnt signalling pathway and its APC partner in colon cancer development. *Gut* 2007; 56:417-25.
49. Deka J, Kuhlmann J, Muller O. A domain within the tumor suppressor protein APC shows very similar biochemical properties as the microtubule-associated protein tau. *Eur J Biochem* 1998; 253:591-7.
50. Dikovskaya D, Li Z, Newton IP, Davidson I, Hutchins JR, Kalab P, et al. Microtubule assembly by the Apc protein is regulated by importin-beta-RanGTP. *J Cell Sci* 2010; 123:736-46.
51. Honnappa S, Gouveia SM, Weisbrich A, Damberger FF, Bhavesh NS, Jawhari H, et al. An EB1-binding motif acts as a microtubule tip localization signal. *Cell* 2009; 138:366-76.
52. Oshima M, Oshima H, Kitagawa K, Kobayashi M, Itakura C, Taketo M. Loss of Apc heterozygosity and abnormal tissue building in nascent intestinal polyps in mice carrying a truncated Apc gene. *Proc Natl Acad Sci USA* 1995; 92:4482-6.
53. Shibata H, Toyama K, Shioya H, Ito M, Hirota M, Hasegawa S, et al. Rapid colorectal adenoma formation initiated by conditional targeting of the Apc gene. *Science* 1997; 278:120-3.
54. Sansom OJ, Reed KR, Hayes AJ, Ireland H, Brinkmann H, Newton IP, et al. Loss of Apc in vivo immediately perturbs Wnt signaling, differentiation, and migration. *Genes Dev* 2004; 18:1385-90.
55. Heppner Goss K, Trzepacz C, Tuohy TM, Groden J. Attenuated APC alleles produce functional protein from internal translation initiation. *Proc Natl Acad Sci USA* 2002; 99:8161-6.
56. Kozak M. New ways of initiating translation in eukaryotes? *Mol Cell Biol* 2001; 21:1899-907.
57. Carson DJ, Santoro IM, Groden J. Isoforms of the APC tumor suppressor and their ability to inhibit cell growth and tumorigenicity. *Oncogene* 2004; 23:7144-8.
58. Baranov PV, Henderson CM, Anderson CB, Gesteland RF, Atkins JF, Howard MT. Programmed ribosomal frameshifting in decoding the SARS-CoV genome. *Virology* 2005; 332:498-510.
59. Gurvich OL, Baranov PV, Gesteland RF, Atkins JF. Expression levels influence ribosomal frameshifting at the tandem rare arginine codons AGG-AGG and AGA-AGA in *Escherichia coli*. *J Bacteriol* 2005; 187:4023-32.
60. Gendron K, Charbonneau J, Dulude D, Heveker N, Ferbeyre G, Brakier-Gingras L. The presence of the TAR RNA structure alters the programmed -1 ribosomal frameshift efficiency of the human immunodeficiency virus type 1 (HIV-1) by modifying the rate of translation initiation. *Nucleic Acids Res* 2008; 36:30-40.
61. Firth AE, Brown CM. Detecting overlapping coding sequences in virus genomes. *BMC Bioinformatics* 2006; 7:75.
62. Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* 2009; 324:218-23.
63. Drysdale R. FlyBase : a database for the *Drosophila* research community. *Methods in Molecular Biology* (Clifton, NJ) 2008; 420:45-59.
64. Birney E, Clamp M, Durbin R. GeneWise and Genomewise. *Genome Res* 2004; 14:988-95.
65. Rhead B, Karolchik D, Kuhn RM, Hinrichs AS, Zweig AS, Fujita PA, et al. The UCSC Genome Browser database: update 2010. *Nucleic Acids Res* 2010; 38:D613-9.
66. Blanchette M, Kent WJ, Riemer C, Elnitski L, Smit AF, Roskin KM, et al. Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res* 2004; 14:708-15.
67. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004; 32:1792-7.
68. Notredame C. Computing multiple sequence/structure alignments with the T-coffee package. *Curr Protoc Bioinformatics* 2010; Chapter 3:Unit 3 8 1-25.
69. Nicholas KB, Nicholas HB, Deerfield DW, 2nd. GeneDoc: a tool for editing and annotating multiple sequence alignments. *EMBNEWNEWS* 1997; 4:14.
70. Ruan J, Li H, Chen Z, Coghlan A, Coin LJ, Guo Y, et al. TreeFam: 2008 Update. *Nucleic Acids Res* 2008; 36:D735-40.
71. Page RD. Visualizing phylogenetic trees using TreeView. *Curr Protoc Bioinformatics* 2002; Chapter 6:Unit 6 2.
72. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 2003; 31:3406-15.
73. Ding Y, Chan CY, Lawrence CE. Sfold web server for statistical folding and rational design of nucleic acids. *Nucleic Acids Res* 2004; 32:W135-41.
74. Byun Y, Han K. PseudoViewer: web application and web service for visualizing RNA pseudoknots and secondary structures. *Nucleic Acids Res* 2006; 34:W416-22.
75. Grentzmann G, Ingram JA, Kelly PJ, Gesteland RF, Atkins JF. A dual-luciferase reporter system for studying recoding signals. *RNA* 1998; 4:479-86.
76. Finn RD, Mistry J, Tate J, Coghill P, Heger A, Pollington JE, et al. The Pfam protein families database. *Nucleic Acids Res* 2010; 38:D211-22.