Published online 25 September 2009 Nucleic Acids Research, 2010, Vol. 38, Database issue D69–D74
doi:10.1093/nar/gkp788

Recode-2: new design, new search tools, and many more genes

Michaël Bekaert[1], Andrew E. Firth[2], Yan Zhang[3], Vadim N. Gladyshev[3], John F. Atkins[2,4] and Pavel V. Baranov[5,*]

[1]School of Biology and Environmental Science, University College Dublin, [2]BioSciences Institute, University College Cork, Ireland, [3]Division of Genetics, Department of Medicine, Brigham & Women's Hospital and Harvard Medical School, Boston, MA 02115, USA, [4]Human Genetics Department, University of Utah, UT 84112 USA and [5]Biochemistry Department, University College Cork, Ireland

Received August 15, 2009; Accepted September 4, 2009

ABSTRACT

'Recoding' is a term used to describe non-standard read-out of the genetic code, and encompasses such phenomena as programmed ribosomal frameshifting, stop codon readthrough, selenocysteine insertion and translational bypassing. Although only a small proportion of genes utilize recoding in protein synthesis, accurate annotation of 'recoded' genes lags far behind annotation of 'standard' genes. In order to address this issue, provide a service to researchers in the field, and offer training data for developers of gene-annotation software, we have gathered together known cases of recoding within the Recode database. Recode-2 is an improved and updated version of the database. It provides access to detailed information on genes known to utilize translational recoding and allows complex search queries, browsing of recoding data and enhanced visualization of annotated sequence elements. At present, the Recode-2 database stores information on approximately 1500 genes that are known to utilize recoding in their expression—a factor of approximately three increase over the previous version of the database. Recode-2 is available at http://recode.ucc.ie

INTRODUCTION

The term 'translational recoding' describes the utilization of non-standard decoding during protein synthesis and encompasses such processes as ribosomal frameshifting, codon redefinition, translational bypassing and StopGo (1–7). What is often considered as a decoding error—e.g. a frameshifting error or mistranslation of a particular codon—may occasionally benefit the organism by increasing its fitness and survival. In such instances the propensity for the decoding 'error' may be selected for during evolution, leading to the formation of a particular sequence context that elevates the frequency of the 'error'. To discriminate such cases of programmed decoding 'misbehaviour' from promiscuous translational errors or translational noise, the term recoding is used. The position within an mRNA where a recoding event takes place is termed the 'recoding site'. Sequence elements responsible for increasing the efficiency of recoding events are termed 'recoding stimulatory signals', and a minimal sequence fragment that allows recoding to take place at the natural efficiency (i.e. relative to the level of standard decoding at the recoding site) is termed a 'recoding cassette'.

Recoding can benefit gene expression in a number of ways. It can regulate gene expression by being part of a sensor for particular cellular conditions. Prominent examples include ribosomal frameshifting in bacterial release factor 2 (RF2) and eukaryotic antizyme mRNAs. In both instances, ribosomal frameshifting is required for the production of the corresponding active full-length protein products. In the RF2 mRNA, the efficiency of frameshifting is negatively regulated by the cellular concentration of its product, RF2, providing an autoregulatory circuit for its biosynthesis (8–10). In the antizyme mRNA, the efficiency of frameshifting is modulated by cellular levels of polyamines, whose concentration in turn is controlled by antizyme (11,12). Thus, this mechanism ensures the maintenance of antizyme production at the levels required to support physiologically appropriate concentrations of polyamines. Recoding can also be used for the diversification of protein products encoded by a single gene. An illustrative example is in bacterial *dnaX* mRNA, where frameshifting allows synthesis of two different protein subunits—sharing the same N-terminal part—from a single open reading frame (ORF) in its mRNA (13–15). A presumed constant ratio of frameshifting in *dnaX* ensures a fixed stoichiometric balance between these two subunits (16). This balance,

*To whom correspondence should be addressed. Tel: +353 (0) 21 4904212; Fax: +353 (0) 21 4904259; Email: p.baranov@ucc.ie

then, is independent of the absolute levels of *dnaX* transcription and translational initiation on its mRNA. Similarly, in many viruses recoding is responsible for setting a ratio between protein products (such as those encoded by *gag–pro–pol* genes in retroviruses) produced from a single mRNA (17). Recoding also provides RNA viruses with a mechanism for the translation of downstream ORFs on polycistronic RNAs [other mechanisms include leaky scanning, shunting, reinitiation, IRESs and the production of subgenomic RNAs (18)] and may also be involved in global regulation mechanisms, such as mediating the switch between translation and replication on the same genomic RNA (19). Finally, recoding provides a way for the incorporation of non-standard amino acids—e.g. amino acids that share their codons with termination signals (the most prominent example of which is selenocysteine, encoded by UGA) (20–22). For further information on the diverse variety of recoding functions, see recent reviews (1,3,7,23,24).

Recoding cassettes may be composed of a variety of diverse sequence elements. For example, primary nucleotide sequences may promote re-arrangements of tRNA molecules relative to their codons in mRNA inside the ribosome or affect recognition of tRNAs or release factors in the ribosomal A-site. On the other hand, many recoding signals act in the form of RNA secondary structures, such as simple stem-loops, or more complex pseudoknots, kissing stem-loops and other structures that involve interactions between considerably distant RNA regions (19,25–28). *Trans*-acting RNA signals affecting ribosomal decoding through complementary interactions with ribosomal RNA (29–32), or through the nascent peptide acting within the ribosome exit tunnel (6,33,34), are also known. Some recoding events—such as selenocysteine insertion—require the presence of additional specialized machinery such as selenocysteine tRNAs, selenocysteine-specific translation factors and several other components of the selenocysteine biosynthesis and insertion pathway (20,35–37). Recent reviews on stimulatory signals involved in the modulation of recoding events and molecular mechanisms of recoding provide further details (7,25,27,38,39).

Despite considerable progress in the development of computational tools for the prediction of protein coding genes in sequenced genomes, the identification and annotation of recoded genes lags far behind. The hurdle lies not so much in the fact that recoded genes do not obey standard rules of genetic readout but, rather, in the considerable diversity of recoded genes and sequence elements responsible for recoding. Even among evolutionarily related genes, all utilizing recoding, the diversity of recoding signals can be considerable. An extreme example is when orthologous genes utilize recoding at different stages of gene expression to achieve the same goal. An example is in *dnaX*, where ribosomal frameshifting is employed by *enterobacteria*, but transcriptional slippage is used in *Thermus thermophilus* (40). A similar situation occurs in bacterial insertion sequence (IS) elements, where a certain group of IS elements utilizes transcriptional slippage to produce ORFA–ORFB fusions, while many other IS elements utilize ribosomal frameshifting for the same purpose (41). The diversity of recoding functions, combined with the wide spectrum of unrelated sequence elements involved in recoding, makes the design of a uniform model of recoding intractable. Nonetheless, in recent years, we have witnessed the development of specialized models and computational tools for the identification of particular subsets of recoding cassettes, or tools that are specific to recoding events in particular groups of homologous genes (42–45).

These developments, at least partially, were facilitated by the availability of a compiled dataset of known recoded genes collected together in the Recode database (http://recode.genetics.utah.edu), which was initially launched 9 years ago (46,47). To facilitate further development of computational tools for the prediction of recoded genes in the ever faster growing body of sequence data, as well as to provide bench researchers with up-to-date information on recoding, an efficient means of Recode database population and annotation are now required. In this article, we describe the incarnation of the database, Recode-2. The major advances of Recode-2 (hosted in a new location http://recode.ucc.ie) over previous versions include a new web design allowing enhanced visualization of stimulatory signals, a uniform RecodeML format for the annotation of recoded genes, and a significantly larger number of entries—including many recently identified cases—that altogether have more than doubled the size of the database since its last published update.

DATABASE ORGANIZATION AND USAGE

The data are stored in a local PostgreSQL database that is queried by PHP scripts embedded in the web interface. The schema of the PostgreSQL database is shown in Figure 1. The database stores information on individual genes that utilize recoding, the mechanisms and stimulatory signals involved, and references to the original literature sources that describe the recoding events. In order to facilitate the uniform annotation of recoding events, we have designed an XML-based format for the annotation of recoded genes, RecodeML. The document type definition for RecodeML is available at the Recode-2 web site at http://recode.ucc.ie/dtd The extensibility of the RecodeML format will allow incorporation of new annotation, if required, for newly discovered types of recoding, and the associated features, as they are being discovered. The database handles batch importation of properly designed RecodeML entries into the PostgreSQL database, thus facilitating rapid population of the database with new data.

The data in the database may be explored in two ways. They may be browsed by one of the three categories: kingdom (archaea, bacteria, eukaryotes and viruses), organism and type of recoding. The data may also be searched directly by key words that can be inserted into the search field. Searches that use regular expressions are allowed. The output of a database search is a list of Recode-2 entries in a short format that includes organism name, kingdom, genus, type of recoding event, status of

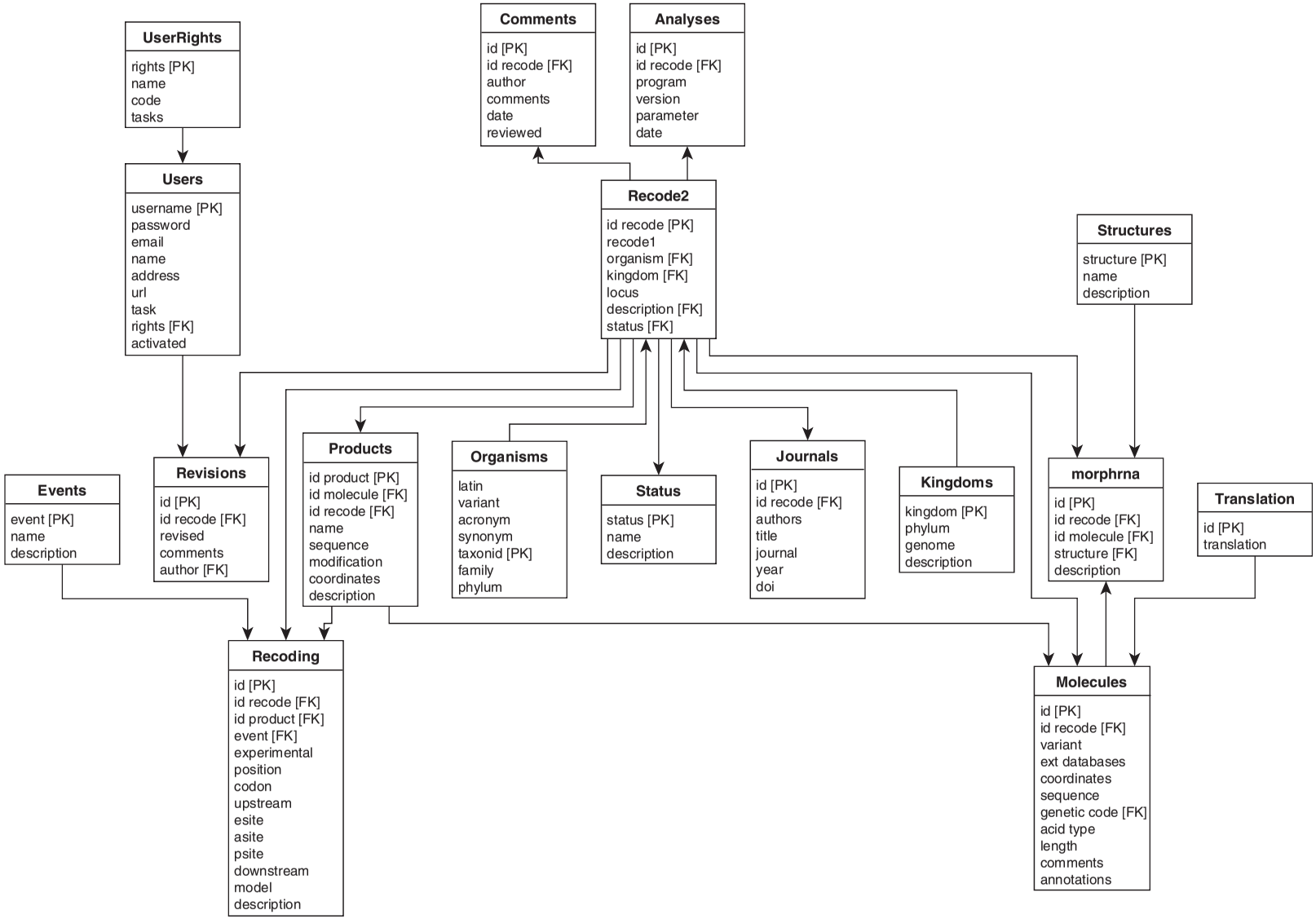

Figure 1. Entity-Relationship diagram for the Recode-2 database. The database schematic shows the relationships between the database tables. Each Recode-2 entry is composed of one entry in the Recode-2 table and a variable number of entries in related tables. PK (Primary Key) indicates the row selected as the unique identifier for a table. FK (Foreign Key) indicates rows of a table whose values match those in the primary key of a related table. (This ensures the constancy of the database.) The arrows show the connection relationships between the various tables (PK to FK).

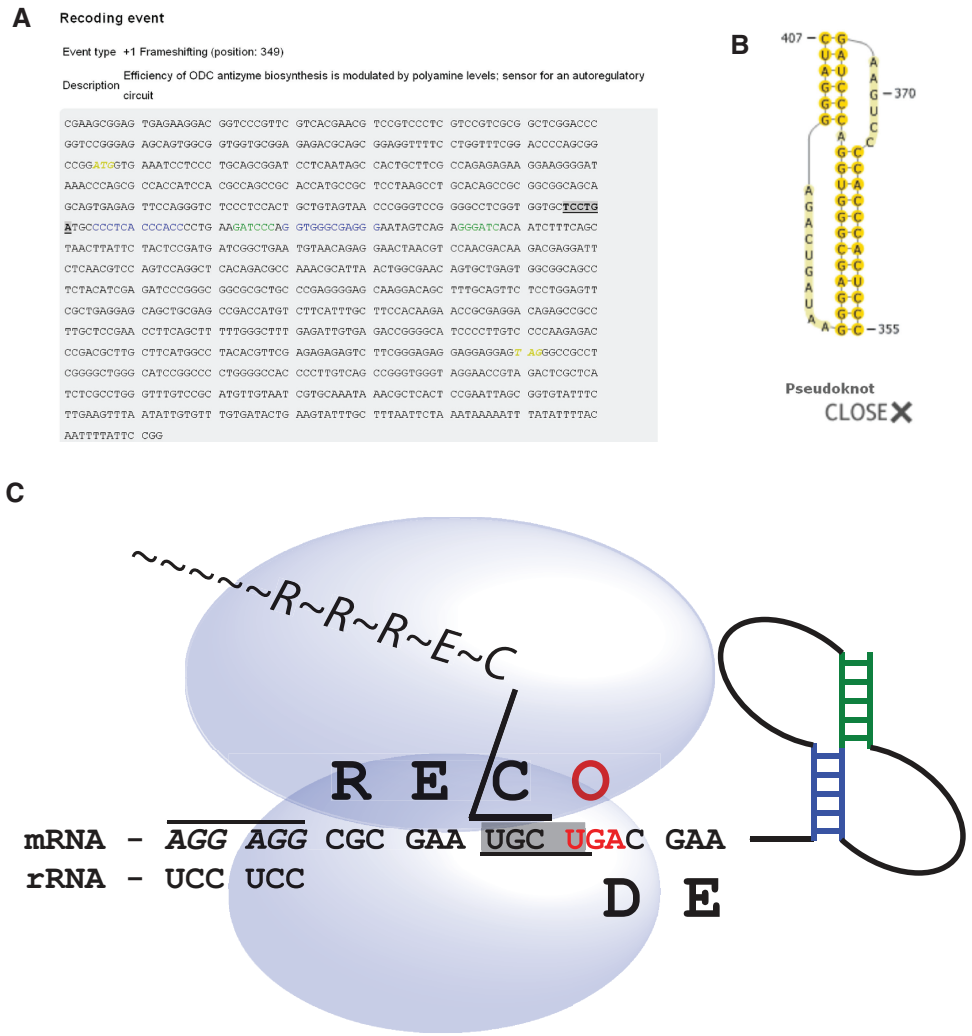

Figure 2. Screenshots from the Recode-2 database for the entry describing +1 frameshifting in human antizyme 1. (**A**) Annotated sequence of the human antizyme 1 gene. Colouring and formatting of the text highlights functional elements according to a scheme described on the Help page of the Recode-2 database and within the Recode-1 logo. (**B**) Structure of the stimulatory RNA pseudoknot. (**C**) The Recode1 logo.

the entry in the database and a link to the full database entry. The full description includes the following additional information: (i) the common name of the gene and the validation status of the recoding event; (ii) the organism description, giving the organism name and a link to the NCBI Taxonomy Browser (48); (iii) the sequence description, giving the Genbank (49) accession numbers for matching sequences (with links to Genbank) and links to detailed annotations of the sequences and to diagrams of RNA secondary structures involved in stimulation of the recoding event; (iv) information on the protein sequence generated as a result of recoding; (v) comments on the function of the recoding event and any additional notes and (vi) references to relevant literature (with links to corresponding abstracts in PubMed). The detailed sequence annotation appears in the form of text decorations that are described in the Help page of the database and are also illustrated within the Recode-1 logo itself (which can be used for rapid intuitive decoding of the text decorations and their associations with the mechanistic ways by which different sequences affect ribosome functions). To generate RNA secondary structure diagrams, PseudoViewer3 (50) is used, since it can handle complex pseudoknot RNA structures correctly. Figure 2 shows an example of sequence annotation for the human oaz1 gene, alongside a diagram of a stimulatory RNA secondary structure, and the Recode-1 logo.

NOVEL DATA

Unlike Recode-1, where all data on recoding events were introduced manually, Recode-2 also utilizes automated identification of recoding events by the recently developed computer programs ARFA (43) and OAF (44), that are able to identify and annotate +1 frameshifting events in mRNAs of bacterial RF2s and eukaryotic antizyme (OAZs), respectively. However, a significant source of recoding events remains to be serendipitous discoveries by experimental studies that sometimes are complemented

by more systematic studies of large groups of similar genes (51,52). Therefore, a large proportion of new data are still populated manually or semi-manually. To ease manual population of recoding events, a special form has been designed that is available in the database upon user registration. User registration needs to be approved by one of the database contributors. The novel data in the database include 249 RF2 mRNAs identified by ARFA, 152 events identified by OAF, 200 new selenoprotein genes (53–56) and ~200 new viral annotations (57) including the newly discovered frameshift cassettes in potyviruses (58), alphaviruses (59) and the Japanese encephalitis group of flaviviruses (60).

FUTURE DEVELOPMENT

The database will expand in accordance with the growth of available sequence information that will be scanned by one of the existing programs for recode annotation. We also plan to continue developing tools for the automatic identification of recoding events from nucleotide sequences. As the field grows and the number of recoded genes progressively increases, it becomes harder to extract data from the relevant literature and a number of novel recoded genes may escape the database. Therefore, we encourage users and researchers in the field to submit their data directly to the Recode-2 database. We are also willing to provide help with the analysis of potential new recoding events.

ACKNOWLEDGEMENTS

We would like to express our appreciation to the colleagues who have contributed data for the previous versions of the database.

FUNDING

Science Foundation Ireland (SFI) grants (to P.V.B. and J.F.A.); National Institutes of Health grants (to J.F.A. and V.N.G.). Funding for open access charge: Science Foundation Ireland.

Conflict of interest statement. None declared.

REFERENCES

1. Namy,O., Rousset,J.P., Napthine,S. and Brierley,I. (2004) Reprogrammed genetic decoding in cellular gene expression. *Mol. Cell*, **13**, 157–168.
2. Farabaugh,P.J. (1996) Programmed translational frameshifting. *Annu. Rev. Genet.*, **30**, 507–528.
3. Baranov,P.V., Gesteland,R.F. and Atkins,J.F. (2002) Recoding: translational bifurcations in gene expression. *Gene*, **286**, 187–201.
4. Dinman,J.D. (2006) Programmed ribosomal frameshifting goes beyond viruses: organisms from all three kingdoms use frameshifting to regulate gene expression, perhaps signaling a paradigm shift. *Microbe*, **1**, 521–527.
5. Atkins,J.F., Wills,N.M., Loughran,G., Wu,C.Y., Parsawar,K., Ryan,M.D., Wang,C.H. and Nelson,C.C. (2007) A case for "StopGo": reprogramming translation to augment codon meaning of GGN by promoting unconventional termination (Stop) after addition of glycine and then allowing continued translation (Go). *RNA*, **13**, 803–810.
6. Herr,A.J., Atkins,J.F. and Gesteland,R.F. (2000) Coupling of open reading frames by translational bypassing. *Annu. Rev. Biochem.*, **69**, 343–372.
7. Atkins,J.F. and Gesteland,R.F. (2010) *Recoding: expansion of decoding rules enriches gene expression*. Springer, New York.
8. Craigen,W.J. and Caskey,C.T. (1986) Expression of peptide chain release factor 2 requires high-efficiency frameshift. *Nature*, **322**, 273–275.
9. Craigen,W.J. and Caskey,C.T. (1987) The function, structure and regulation of E. coli peptide chain release factors. *Biochimie*, **69**, 1031–1041.
10. Baranov,P.V., Gesteland,R.F. and Atkins,J.F. (2002) Release factor 2 frameshifting sites in different bacteria. *EMBO Rep.*, **3**, 373–377.
11. Ivanov,I.P. and Atkins,J.F. (2007) Ribosomal frameshifting in decoding antizyme mRNAs from yeast and protists to humans: close to 300 cases reveal remarkable diversity despite underlying conservation. *Nucleic Acids Res.*, **35**, 1842–1858.
12. Matsufuji,S., Matsufuji,T., Miyazaki,Y., Murakami,Y., Atkins,J.F., Gesteland,R.F. and Hayashi,S. (1995) Autoregulatory frameshifting in decoding mammalian ornithine decarboxylase antizyme. *Cell*, **80**, 51–60.
13. Flower,A.M. and McHenry,C.S. (1990) The gamma subunit of DNA polymerase III holoenzyme of *Escherichia coli* is produced by ribosomal frameshifting. *Proc. Natl Acad. Sci. USA*, **87**, 3713–3717.
14. Tsuchihashi,Z. and Kornberg,A. (1990) Translational frameshifting generates the gamma subunit of DNA polymerase III holoenzyme. *Proc. Natl Acad. Sci. USA*, **87**, 2516–2520.
15. Blinkowa,A.L. and Walker,J.R. (1990) Programmed ribosomal frameshifting generates the *Escherichia coli* DNA polymerase III gamma subunit from within the tau subunit reading frame. *Nucleic Acids Res.*, **18**, 1725–1729.
16. Larsen,B., Gesteland,R.F. and Atkins,J.F. (1997) Structural probing and mutagenic analysis of the stem-loop required for *Escherichia coli* dnaX ribosomal frameshifting: programmed efficiency of 50%. *J. Mol. Biol.*, **271**, 47–60.
17. Brierley,I. and Dos Ramos,F.J. (2006) Programmed ribosomal frameshifting in HIV-1 and the SARS-CoV. *Virus Res.*, **119**, 29–42.
18. Thiebeauld,O., Pooggin,M.M. and Ryabova,L.A. (2007) Alternative translation strategies in plant viruses. *Plant Viruses*, **1**, 1–20.
19. Miller,W.A. and White,K.A. (2006) Long-distance RNA–RNA interactions in plant virus gene expression and replication. *Annu. Rev. Phytopathol.*, **44**, 447–467.
20. Squires,J.E. and Berry,M.J. (2008) Eukaryotic selenoprotein synthesis: mechanistic insight incorporating new factors and new functions for old factors. *IUBMB Life*, **60**, 232–235.
21. Allmang,C. and Krol,A. (2006) Selenoprotein synthesis: UGA does not end the story. *Biochimie*, **88**, 1561–1571.
22. Hatfield,D.L., Berry,M.J. and Gladyshev,V.N. (2006) *Selenium: Its Molecular Biology and Role in Human Health*. Springer, New York.
23. Baranov,P.V., Fayet,O., Hendrix,R.W. and Atkins,J.F. (2006) Recoding in bacteriophages and bacterial IS elements. *Trends Genet.*, **22**, 174–181.
24. Plant,E.P. and Dinman,J.D. (2008) The role of programmed-1 ribosomal frameshifting in coronavirus propagation. *Front Biosci.*, **13**, 4873–4881.
25. Giedroc,D.P. and Cornish,P.V. (2009) Frameshifting RNA pseudoknots: structure and mechanism. *Virus Res.*, **139**, 193–208.
26. Giedroc,D.P., Theimer,C.A. and Nixon,P.L. (2000) Structure, stability and function of RNA pseudoknots involved in stimulating ribosomal frameshifting. *J. Mol. Biol.*, **298**, 167–185.
27. Brierley,I., Gilbert,R.J. and Pennell,S. (2008) RNA pseudoknots and the regulation of protein synthesis. *Biochem. Soc. Trans.*, **36**, 684–689.
28. Barry,J.K. and Miller,W.A. (2002) 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*, **99**, 11133–11138.
29. Weiss,R.B., Dunn,D.M., Atkins,J.F. and Gesteland,R.F. (1987) Slippery runs, shifty stops, backward steps, and forward hops: -2,

-1, +1, +2, +5, and +6 ribosomal frameshifting. *Cold Spring Harb. Symp. Quant. Biol.*, **52**, 687–693.
30. Larsen,B., Peden,J., Matsufuji,S., Matsufuji,T., Brady,K., Maldonado,R., Wills,N.M., Fayet,O., Atkins,J.F. and Gesteland,R.F. (1995) Upstream stimulators for recoding. *Biochem. Cell Biol.*, **73**, 1123–1129.
31. Atkins,J.F., Baranov,P.V., Fayet,O., Herr,A.J., Howard,M.T., Ivanov,I.P., Matsufuji,S., Miller,W.A., Moore,B., Prere,M.F. *et al.* (2001) Overriding standard decoding: implications of recoding for ribosome function and enrichment of gene expression. *Cold Spring Harb. Symp. Quant. Biol.*, **66**, 217–232.
32. Curran,J.F. and Yarus,M. (1988) Use of tRNA suppressors to probe regulation of Escherichia coli release factor 2. *J. Mol. Biol.*, **203**, 75–83.
33. Wills,N.M., O'Connor,M., Nelson,C.C., Rettberg,C.C., Huang,W.M., Gesteland,R.F. and Atkins,J.F. (2008) Translational bypassing without peptidyl-tRNA anticodon scanning of coding gap mRNA. *EMBO J.*, **27**, 2533–2544.
34. Weiss,R.B., Huang,W.M. and Dunn,D.M. (1990) A nascent peptide is required for ribosomal bypass of the coding gap in bacteriophage T4 gene 60. *Cell*, **62**, 117–126.
35. Lescure,A., Fagegaltier,D., Carbon,P. and Krol,A. (2002) Protein factors mediating selenoprotein synthesis. *Curr. Protein Pept. Sci.*, **3**, 143–151.
36. Walczak,R., Hubert,N., Carbon,P. and Krol,A. (1997) Solution structure of SECIS, the mRNA element required for eukaryotic selenocysteine insertion–interaction studies with the SECIS-binding protein SBP. *Biomed. Environ. Sci.*, **10**, 177–181.
37. Commans,S. and Bock,A. (1999) Selenocysteine inserting tRNAs: an overview. *FEMS Microbiol. Rev.*, **23**, 335–351.
38. Baranov,P.V., Gesteland,R.F. and Atkins,J.F. (2004) P-site tRNA is a crucial initiator of ribosomal frameshifting. *RNA*, **10**, 221–230.
39. Liao,P.Y., Gupta,P., Petrov,A.N., Dinman,J.D. and Lee,K.H. (2008) A new kinetic model reveals the synergistic effect of E-, P- and A-sites on +1 ribosomal frameshifting. *Nucleic Acids Res.*, **36**, 2619–2629.
40. Larsen,B., Wills,N.M., Nelson,C., Atkins,J.F. and Gesteland,R.F. (2000) Nonlinearity in genetic decoding: homologous DNA replicase genes use alternatives of transcriptional slippage or translational frameshifting. *Proc. Natl Acad. Sci. USA*, **97**, 1683–1688.
41. Baranov,P.V., Hammer,A.W., Zhou,J., Gesteland,R.F. and Atkins,J.F. (2005) Transcriptional slippage in bacteria: distribution in sequenced genomes and utilization in IS element gene expression. *Genome Biol.*, **6**, R25.
42. Theis,C., Reeder,J. and Giegerich,R. (2008) KnotInFrame: prediction of -1 ribosomal frameshift events. *Nucleic Acids Res.*, **36**, 6013–6020.
43. Bekaert,M., Atkins,J.F. and Baranov,P.V. (2006) ARFA: a program for annotating bacterial release factor genes, including prediction of programmed ribosomal frameshifting. *Bioinformatics*, **22**, 2463–2465.
44. Bekaert,M., Ivanov,I.P., Atkins,J.F. and Baranov,P.V. (2008) Ornithine decarboxylase antizyme finder (OAF): fast and reliable detection of antizymes with frameshifts in mRNAs. *BMC Bioinformatics*, **9**, 178.
45. Moon,S., Byun,Y., Kim,H.J., Jeong,S. and Han,K. (2004) Predicting genes expressed via -1 and +1 frameshifts. *Nucleic Acids Res.*, **32**, 4884–4892.
46. Baranov,P.V., Gurvich,O.L., Fayet,O., Prere,M.F., Miller,W.A., Gesteland,R.F., Atkins,J.F. and Giddings,M.C. (2001) RECODE: a database of frameshifting, bypassing and codon redefinition utilized for gene expression. *Nucleic Acids Res.*, **29**, 264–267.
47. Baranov,P.V., Gurvich,O.L., Hammer,A.W., Gesteland,R.F. and Atkins,J.F. (2003) Recode 2003. *Nucleic Acids Res.*, **31**, 87–89.
48. Sayers,E.W., Barrett,T., Benson,D.A., Bryant,S.H., Canese,K., Chetvernin,V., Church,D.M., DiCuccio,M., Edgar,R., Federhen,S. *et al.* (2009) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **37**, D5–15.
49. Benson,D.A., Karsch-Mizrachi,I., Lipman,D.J., Ostell,J. and Sayers,E.W. (2009) GenBank. *Nucleic Acids Res.*, **37**, D26–D31.
50. Byun,Y. and Han,K. (2009) PseudoViewer3: generating planar drawings of large-scale RNA structures with pseudoknots. *Bioinformatics*, **25**, 1435–1437.
51. Gurvich,O.L., Baranov,P.V., Zhou,J., Hammer,A.W., Gesteland,R.F. and Atkins,J.F. (2003) Sequences that direct significant levels of frameshifting are frequent in coding regions of *Escherichia coli*. *EMBO J.*, **22**, 5941–5950.
52. Xu,J., Hendrix,R.W. and Duda,R.L. (2004) Conserved translational frameshift in dsDNA bacteriophage tail assembly genes. *Mol. Cell*, **16**, 11–21.
53. Zhang,Y. and Gladyshev,V.N. (2009) Comparative genomics of trace elements: emerging dynamic view of trace element utilization and function. *Chem. Rev.*, in press.
54. Zhang,Y., Romero,H., Salinas,G. and Gladyshev,V.N. (2006) Dynamic evolution of selenocysteine utilization in bacteria: a balance between selenoprotein loss and evolution of selenocysteine from redox active cysteine residues. *Genome Biol.*, **7**, R94.
55. Zhang,Y. and Gladyshev,V.N. (2008) Trends in selenium utilization in marine microbial world revealed through the analysis of the global ocean sampling (GOS) project. *PLoS Genet.*, **4**, e1000095.
56. Kim,H.Y., Zhang,Y., Lee,B.C., Kim,J.R. and Gladyshev,V.N. (2009) The selenoproteome of Clostridium sp. OhILAs: characterization of anaerobic bacterial selenoprotein methionine sulfoxide reductase A. *Proteins*, **74**, 1008–1017.
57. Bekaert,M. and Rousset,J.P. (2005) An extended signal involved in eukaryotic -1 frameshifting operates through modification of the E site tRNA. *Mol. Cell*, **17**, 61–68.
58. Chung,B.Y., Miller,W.A., Atkins,J.F. and Firth,A.E. (2008) An overlapping essential gene in the Potyviridae. *Proc. Natl Acad. Sci. USA*, **105**, 5897–5902.
59. Firth,A.E., Chung,B.Y., Fleeton,M.N. and Atkins,J.F. (2008) Discovery of frameshifting in Alphavirus 6K resolves a 20-year enigma. *Virol. J.*, **5**, 108.
60. Firth,A.E. and Atkins,J.F. (2009) A conserved predicted pseudoknot in the NS2A-encoding sequence of West Nile and Japanese encephalitis flaviviruses suggests NS1' may derive from ribosomal frameshifting. *Virol. J.*, **6**, 14.