Nucleic Acids Research, 2003, Vol. 31, No. 1 87–89
DOI: 10.1093/nar/gkg024

RECODE 2003

Pavel V. Baranov, Olga L. Gurvich, Andrew W. Hammer, Raymond F. Gesteland and John F. Atkins*

Department of Human Genetics, University of Utah, 15N 2030E Room 7410, Salt Lake City, UT 84112-5330, USA

Received September 12, 2002; Accepted September 20, 2002

ABSTRACT

The RECODE database is a compilation of translational recoding events (programmed ribosomal frameshifting, codon redefinition and translational bypass). The database provides information about the genes utilizing these events for their expression, recoding sites, stimulatory sequences and other relevant information. The Database is freely available at http://recode.genetics.utah.edu/.

The RECODE database was created in the year 2000 (1). It provides information about genes utilizing recoding and *cis*-signals or *trans*-factors involved. Recoding is non-standard translation of mRNA utilized for gene expression. Normally recoding is programmed in mRNA by special *cis*-signals [for the most recent detailed reviews on recoding see (2–5)]. The database currently deals with the three main types of recoding:

1. Frameshifting. Ribosome shifts frame at a particular mRNA site to yield a protein encoded by two overlapping open reading frames.
2. Programmed bypassing (hopping). During bypassing translation is suspended at a particular codon and is resumed at a non-overlapping downstream codon.
3. Codon redefinition. Localized alterations of codon meaning. Currently known cases involve redefinition of stop codons to selenocysteine (selenocysteine incorporation) or to a standard amino acid (readthrough). Specialized components of translational machinery are required for selenocysteine incorporation.

Several types of recoding discovered recently are not covered by RECODE database, but are being under consideration for future updates:

Trans-translation. tmRNA targeting of several proteins is programmed by specific signals in their mRNA. Although some cases are very well documented (6–8), the biological implication is understood for only a few of them (9). For more information on tmRNA see (10,11).

'Intra-ribosomal' cleavage of nascent peptide. Only one case has been reported currently (12).

Pyrrolysine incorporation (13–14). It is not known yet, whether this is a case of recoding or codon reassignment (15).

The database is regularly updated. Currently the database constitutes a wealth of information on the 363 genes known to utilize recoding for their expression from 422 related publications. The data are organized and stored in a similar manner to those in the initial RECODE database (1). The main change concerns the description of stimulatory *cis*-elements.

In the existing version of the database stimulatory signals are classified as follows:

1. UGA, UAA or UAG stop codons. Stop codons are assigned as stimulatory signals if their identity has an important effect on the efficiency of recoding. However, they are not considered as stimulatory signals in those cases where they form the recoding site. For instance, with codon redefinition those 'stop codons' that encode an amino acid (standard one or selenocysteine) are not considered to be stimulatory signals. However, if stop codons partially overlap a frameshifting site they are considered as stimulatory signals. For example, the release factor 2 (RF2) frameshifting cassette contains the sequence CUU UGA. Leu-tRNA slips +1 from CUU to UUU (i.e. the frameshift site is CUU U). The first nucleotide of the UGA stop codon is a part of the frameshifting site, however, it serves to stimulate frameshifting by being an inefficient termination signal and it plays an important role in regulation of RF2 biosynthesis. Thus, it is annotated as a stimulatory signal. Stop codons that play a role in recoding are shown in red within the annotated sequences.
2. 3′ RNA pseudoknot (kissing loops). These signals are known to stimulate a variety of different types of recoding (frameshifting in the plus and minus directions, codon redefinition). An RNA pseudoknot is a stem-loop RNA structure where nucleotides in the loop participate in complementary interactions with a region of RNA downstream of this stem-loop. The distance between the signal and frameshifting sequence is highly important. See Atkins *et al*. (16) for a detailed review on this type of stimulator. Nucleotides that form double helices of stem-loop structures are shown in green within the annotated sequences while nucleotides that participate in complementary interactions between loops are shown in violet.
3. 5′ Shine-Dalgarno sequence. An internal Shine-Dalgarno sequence that may not be used for the initiation of translation stimulates recoding through interactions with

*To whom correspondence should be addressed. Tel: +1 801 585 3434; Fax: +1 801 585 3910; Email: atkins@howard.genetics.utah.edu

the anti Shine-Dalgarno in the RNA of small ribosomal subunits of translating ribosomes (i.e. mRNA:rRNA interactions). This signal is only operative in bacteria since eukaryotic ribosomes lack an anti Shine-Dalgarno. It stimulates frameshifting in either direction, plus or minus. The distance between Shine-Dalgarno and frameshifting sequence is important, short (3–4 nts) stimulates frameshifting in the plus direction and long (8–14 nts) stimulates frameshifting in the minus direction (16). Shine-Dalgarno sequences used for stimulation of frameshifting are shown in brown within the annotated sequences.

4. 3′ stem loop structure. It probably acts in a similar way to RNA pseudoknots or kissing loops. However, it is important for recoding only at the level of a simple hairpin structure. Nucleotides that form the stem of a hairpin are shown in green.
5. 3′ nascent peptide. The sequence upstream of the recoding site is important at the level of encoded peptide, for example synonymous mutations do not have a significant effect on the efficiency of recoding. The sequences encoding stimulatory nascent peptides are in italics within the annotated sequences.
6. SECIS element. Specific RNA structure that reprograms UGA stop codons to selenocysteine codons. SECIS elements are indicated within annotated sequences as other RNA structures, e.g. stem loops.
7. 5′ other. An RNA element located upstream of recoding site whose identity is important for the efficiency of frameshifting, but the nature of its effect is unknown. An example is the 5′ stimulatory signal in antizyme mRNA (17).
8. 3′ nucleotide flanking stop codons. The identity of the nucleotide immediately downstream of the stop codon is very important for the efficiency of this stop codon as termination signal.
9. 3′ quadruplet flanking a frameshifting site. In some cases, +1 frameshifting is stimulated by the sequence downstream of a shifty P-site codon. The stimulatary quadruplet consists of two overlapping codons. The codon in the zero frame is decoded by sparse tRNA, while the codon in the new frame is decoded by abundant tRNA. This combination shifts the equilibrium between the P-site tRNA occupancy of the original P-site codon and the codon in a new frame in favor of the latter. See (18) for the detailed description of this mechanism.
10. 3′ linear sequence is a downstream RNA sequence that is not involved in the formation of intra-RNA secondary structures.
11. 3′ distant region is a sequence important for the efficiency of recoding located over 1000 nts downstream of the recoding site (19–20).
12. 3′ repeat. It is a downstream sequence important for recoding that contains repetitive elements (21).

There are an increasing number of publications on biologically functional alterations in standard rules of genetic decoding. Due to their high divergence the current version of the database cannot cover all these cases. The lack of consistency in the description of recoding events in the literature makes highly problematic their steady description in the RECODE database. We plan to modify the database so that it will be able to suit rapid progress in the field of translational recoding. We would highly appreciate a feedback on this issue from all scientists involved in study of intriguing phenomena (recoding) in order to make the database more cohesive and unified.

ACKNOWLEDGEMENTS

We are grateful to Drs W. Allen Miller, Olivier Fayet and Michael Giddings for their significant contributions in the development of the initial version of the RECODE database. This work has been supported by NIH grants R0-1GM48152 to J.F.A. and 5R0-1GM061200-03 to R.F.G.

REFERENCES

1. 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.
2. Gesteland,R.F. and Atkins,J.F. (1996) Recoding: dynamic reprogramming of translation. *Annu. Rev. Biochem.*, **65**, 741–768.
3. Brierley,I. and Pennell,S. (2001) Structure and function of the stimulatory RNAs involved in programmed eukaryotic −1 ribosomal frameshifting. *Cold Spring Harb. Symp. Quant. Biol.*, **66**, 233–248.
4. Baranov,P.V., Gesteland,R.F. and Atkins,J.F. (2002) Recoding: translational bifurcations in gene expression. *Gene*, **286**, 187–201.
5. Stahl,G., McCarty,G.P. and Farabaugh,P.J. (2002) Ribosome structure: revisiting the connection between translational accuracy and unconventional decoding. *Trends Biochem. Sci.*, **27**, 178–183.
6. Collier,J., Binet,E. and Bouloc,P. (2002) Competition between SsrA tagging and translational termination at weak stop codons in *Escherichia coli*. *Mol. Microbiol.*, **45**, 745–754.
7. Hayes,C.S., Bose,B. and Sauer,R.T. (2002) Stop codons preceded by rare arginine codons are efficient determinants of SsrA tagging in *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **99**, 3440–3445.
8. Ueda,K., Yamamoto,Y., Ogawa,K., Abo,T., Inokuchi,H. and Aiba,H. (2002) Bacterial SsrA system plays a role in coping with unwanted translational readthrough caused by suppressor tRNAs. *Genes Cells*, **7**, 509–519.
9. Abo,T., Inada,T., Ogawa,K. and Aiba,H. (2000) SsrA-mediated tagging and proteolysis of LacI and its role in the regulation of *lac* operon. *EMBO J.*, **19**, 3762–3769.
10. Knudsen,B., Wower,J., Zwieb,C. and Gorodkin,J. (2001) tmRDB (tmRNA database). *Nucleic Acids Res.*, **29**, 171–172.
11. Williams,K. (2002) The tmRNA Website: invasion by an intron. *Nucleic Acids Res.*, **30**, 179–182.
12. Donnelly,M.L., Luke,G., Mehrotra,A., Li,X., Hughes,L.E., Gani,D. and Ryan,M.D. (2001) Analysis of the aphthovirus 2A/2B polyprotein 'cleavage' mechanism indicates not a proteolytic reaction, but a novel translational effect: a putative ribosomal 'skip'. *J. Gen. Virol.*, **82**, 1013–1025.
13. Srinivasan,G., James,C.M. and Krzycki,J.A. (2002) Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. *Science*, **296**, 1459–1462.
14. Hao,B., Gong,W., Ferguson,T.K., James,C.M., Krzycki,J.A. and Chan,M.K. (2002) A new UAG-encoded residue in the structure of a methanogen methyltransferase. *Science*, **296**, 1462–1466.
15. Atkins,J.F. and Gesteland,R. (2002) Biochemistry. The 22nd amino acid. *Science*, **296**, 1409–1410.
16. 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., Wills,N.M., Zhou,J. and Gesteland,R.F. (2001) Overriding standard decoding: Implications of recoding for ribosome function and enrichment of gene expression. *Cold Spring Harb. Symp. Quant. Biol.*, **66**, 217–232.
17. Ivanov,I.P., Gesteland,R.F. and Atkins,J.F. (2000) Antizyme expression: a subversion of triplet decoding, which is remarkably conserved by

evolution, is a sensor for an autoregulatory circuit. *Nucleic Acids Res.*, **28**, 3185–3196.
18. Baranov,P.V., Gesteland,R.F. and Atkins,J.F. (2002) P-site tRNA as a key player in ribosomal frameshifting. *J. Mol. Biol.*, submitted.
19. Paul,C.P., Barry,J.K., Dinesh-Kumar,S.P., Brault,V. and Miller,W.A. (2001) A sequence required for −1 ribosomal frameshifting located four kilobases downstream of the frameshift site. *J. Mol. Biol.*, **310**, 987–999.
20. 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.
21. Brown,C.M., Dinesh-Kumar,S.P. and Miller,W.A. (1996) Local and distant sequences are required for efficient readthrough of the barley yellow dwarf virus PAV coat protein gene stop codon. *J. Virol.*, **70**, 5884–5892.