RECODE: a database of frameshifting, bypassing and
codon redefinition utilized for gene expression

Pavel V. Baranov, Olga L. Gurvich, Olivier Fayet1, Marie Françoise Prère1, W. Allen Miller2, 
Raymond F. Gesteland, John F. Atkins* and Michael C. Giddings

Department of Human Genetics, University of Utah, 15N 2030E Room 7410, Salt Lake City, UT 84112-5330, USA, 
1Microbiologie et Génétique Moléculaire, CNRS, 118 route de Narbonne, 31062, Toulouse Cedex, France and 
2Plant Pathology Department and L.H. Baker Center for Bioinformatics and Biological Statistics, Iowa State 
University, Ames 50011-1020, USA

Received August 31, 2000; Revised and Accepted November 1, 2000

ABSTRACT

The RECODE database is a compilation of ‘programmed’
translational recoding events taken from the scientific 
literature and personal communications. The database 
deals with programmed ribosomal frameshifting, 
codon redefinition and translational bypass occurring 
in a variety of organisms. The entries for each event 
include the sequences of the corresponding genes, 
their encoded proteins for both the normal and 
alternate decoding, the types of the recoding events 
involved, trans-factors and cis-elements that 
influence recoding. The database is freely available 
at http://recode.genetics.utah.edu/.

INTRODUCTION

Recoding is the reprogramming of mRNA translation by localized 
alterations in the standard translational rules. Recoding is 
utilized in the expression of a minority of genes in probably all 
organisms, though the extent of occurrence within organisms is 
unknown. In known cases the product of recoding is functionally 
important, but there may exist cases where the importance of 
recoding does not lie in the encoded protein. Three classes of 
recoding are known.

(i) Frameshifting at a particular site can yield two protein 
products from one coding sequence or one protein product 
from two overlapping open reading frames (ORFs). In some cases 
a set ratio of two products is important and in other cases 
frameshifting has a regulatory purpose. In the latter the level of 
frameshifting is influenced by the concentration of proteins or 
other factors present during translation. The known cases of 
frameshifting where the product is utilized involve shifts of 
one base, either +1 or −1 (but shifts of two bases have been 
demonstrated in artificial systems and changes of reading 
frame may also occur with bypassing).

(ii) Bypassing (hopping) occurs when a block of nucleotides 
within a coding sequence is not translated. Translation is 
temporarily suspended, ribosomes traverse the coding gap and 
protein synthesis resumes to yield a single protein. This allows 
the coupling of two ORFs separated on an mRNA by a coding 
gap, and is frame-independent.

(iii) Codon redefinition involves site-specific alteration of 
codon meaning. All the included cases involve redefinition of 
a stop codon to specify an amino acid, often glutamine, trypto-
phan or selenocysteine. (The altered meaning of certain codons 
when they function as an initiation codon is not included in this 
compendium.)

An example of each type is illustrated in Figure 1 and 
and all string-based fields accept the wildcards '*' (match 0 or 
more characters) and '?' (match any single character). Entries 
are also possible. Search fields left blank are taken as wildcards,
and all string-based fields accept the wildcards '*' (match 0 or 
more characters) and '?' (match any single character). Entries 
have the following fields.

(i) Gene, common gene name.

*To whom correspondence should be addressed. Tel: +1 801 585 3434; Fax: +1 801 585 3910; Email: john.atkins@genetics.utah.edu
Table 1. An overview of gene types that utilize recoding for their expression

<table>
<thead>
<tr>
<th>Genes/Proteins</th>
<th>Occurrence</th>
<th>RECODING type</th>
<th>Stimulators/Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>oaz</td>
<td><em>S. pombe</em> to vertebrates</td>
<td>+1 FS</td>
<td>YCC UGA or UUU UGA</td>
</tr>
<tr>
<td>oaz1/2</td>
<td>Mammals</td>
<td>+1 FS</td>
<td>UCC UGA</td>
</tr>
<tr>
<td>oaz3</td>
<td>Mammals</td>
<td>+1 FS</td>
<td>UCC UGA</td>
</tr>
<tr>
<td>p45</td>
<td><em>Euplotes</em>, a Ciliate</td>
<td>+1 FS</td>
<td>AAA UAA</td>
</tr>
<tr>
<td>est3</td>
<td><em>S. cerevisiae</em></td>
<td>+1 FS</td>
<td>CUU AGU</td>
</tr>
<tr>
<td>Retrotransposons Ty1, Ty2, Ty4</td>
<td>+1 FS</td>
<td>GCG AGU U</td>
<td>Hungry codon, short 3' sequence; “Once-only” pairing/occlusion</td>
</tr>
<tr>
<td>Retrotransposon Ty3</td>
<td>+1 FS</td>
<td>UGA</td>
<td>Very early in gene; Shifted ribosomes terminate; Non-functional product</td>
</tr>
<tr>
<td><em>pyfB</em></td>
<td>Most bacteria</td>
<td>+1 FS</td>
<td>CUU UGA C</td>
</tr>
<tr>
<td>dnaX</td>
<td><em>E. coli</em> &amp; relatives</td>
<td>+1 FS</td>
<td>AAAA AAG</td>
</tr>
<tr>
<td>subunits γ &amp; α</td>
<td><em>(T. thermophila)</em></td>
<td>+1 FS</td>
<td>(TTTTTTTTTTTTT)</td>
</tr>
<tr>
<td><em>edd</em></td>
<td><em>B. subtilis</em></td>
<td>+1 FS</td>
<td>AGA AGG</td>
</tr>
<tr>
<td>arg1</td>
<td><em>E. coli</em></td>
<td>+1 FS</td>
<td>UUC</td>
</tr>
<tr>
<td>Insertion Sequences (IS1, IS3 family, Mobile elements)</td>
<td>Various</td>
<td>(Clostridia IS120)</td>
<td>(tr. sl.)</td>
</tr>
<tr>
<td>kel</td>
<td><em>D. melanogaster</em></td>
<td>RT</td>
<td>UGA</td>
</tr>
<tr>
<td>oaf</td>
<td>Out at first</td>
<td>RT</td>
<td>UGA</td>
</tr>
<tr>
<td>hde</td>
<td>Headcase protein</td>
<td>UAA</td>
<td>Only <em>B. firmus</em> has a premature stop codon</td>
</tr>
<tr>
<td>ropA</td>
<td>DNA topoisomerase 1</td>
<td>RT</td>
<td>UGA</td>
</tr>
<tr>
<td>Adhesion factors</td>
<td>Enterotoxicogenic <em>E. coli</em></td>
<td>RT</td>
<td>UAG</td>
</tr>
<tr>
<td>Genes encoding selenocysteine containing proteins</td>
<td><em>Bactria</em></td>
<td>Se I</td>
<td>UGA</td>
</tr>
<tr>
<td></td>
<td>Archaean</td>
<td>UGA</td>
<td>5' or 3' UTR SECIS, special EF &amp; tRNA</td>
</tr>
<tr>
<td></td>
<td>Worms to mamm.</td>
<td>UGA</td>
<td>3' UTR SECIS &amp; its binding factor SPB2, special elongation factor eEFsec &amp; tRNA</td>
</tr>
<tr>
<td>gag-pol or gag-pro-pol</td>
<td><em>HIV, MMTV</em></td>
<td>-1 FS</td>
<td>XXY YYY</td>
</tr>
<tr>
<td>genes of retroviruses other than Spumaretroviruses</td>
<td><em>MuLV</em></td>
<td>RT</td>
<td>UAG</td>
</tr>
<tr>
<td>pol</td>
<td>RNA replicase</td>
<td>-1 FS</td>
<td>GGGU UUU</td>
</tr>
<tr>
<td>pol</td>
<td>DNA replicase</td>
<td>+1 FS</td>
<td>UUUA AAC</td>
</tr>
<tr>
<td>pol</td>
<td>RNA replicase</td>
<td>-1 FS</td>
<td>UUUA AAC</td>
</tr>
<tr>
<td>pol</td>
<td>RNA replicase</td>
<td>RT</td>
<td>UAG</td>
</tr>
<tr>
<td>gene 60</td>
<td>Topoisomerase subunit</td>
<td>RT</td>
<td>GGA,(47),GGA</td>
</tr>
<tr>
<td>Coat protein replicase</td>
<td>RNA phage QB</td>
<td>RT</td>
<td>UGA</td>
</tr>
<tr>
<td>Coat lysis hybrid</td>
<td>RNA phage MS2</td>
<td>+1 FS</td>
<td>UGA</td>
</tr>
<tr>
<td>Capsid-RNA replicase</td>
<td>Sindbis</td>
<td>RT</td>
<td>UGA</td>
</tr>
<tr>
<td>Genes <em>g-t</em> tail assembly protein</td>
<td><em>Lamphoid phages</em></td>
<td>-1 FS</td>
<td>GGA AAG</td>
</tr>
<tr>
<td>Genes 10</td>
<td>Major coat protein</td>
<td>Phase T7</td>
<td>-1 FS</td>
</tr>
</tbody>
</table>

References are in reviews (1–3) and in the database.
FS, frameshifting; RT, readthrough (codon redefinition); Se I, selenocysteine insertion (a special case of redefinition); Byp, translational bypassing (the matched takeoff and landing site codons, GGA, in T4 gene 60 bypassing are separated by 47 nt, indicated in parenthesis); PK, RNA pseudoknot; EF, elongation factor. SD 3bp5 indicates that the distance between Shine–Dalgarno sequence, with which translating ribosomes interact to influence recoding, is 3 base pairs 5' of the frameshift site. Codons in the initial frame are separated by spaces and the codons at which the new frame is set are underlined. Two cases of transcription slippage (tr. sl.) are also included as they yield the same end result as recoding. In these cases the DNA sequence on which the RNA polymerase slips is given.
(ii) Type, type of recoding event (see Introduction).
(iii) Organism name, official name of the organism and other names used. Other names are represented according to information provided by the NCBI Taxonomy Database (4).
(iv) Cis-elements, characterized elements of primary, secondary or tertiary mRNA structure that are known to influence recoding.
(v) Trans-elements, cellular factors that influence recoding, e.g. proteins, small ligands etc.
(vi) Function of recoding, recoding can play various roles in cells. It is utilized for the regulation of gene expression level such as in the synthesis of bacterial release factors 2 (RF2) (5) or in eukaryotic antizymes (OAZ) (6). Recoding is also utilized for production of the proper ratio between two proteins; an example is the expression of Escherichia coli dnaX where two subunits of DNA polymerase III (γ and τ) are synthesized in a 1:1 ratio (7).
(vii) Product/function, information about a gene product(s) and its/their function(s).
(viii) Translation without recoding, sequence of the protein or polypeptide synthesized by standard translation.
(ix) Translation with recoding, sequence of the protein whose synthesis results from recoding.
(x) References, primary research papers that describe particular recoding events are cited with the corresponding hyperlink to their MEDLINE abstract.
(xi) Comments/notes, any important information about the entry that is not described in others fields.
(xii) Evidence, provides information on whether the (presumed) event has been demonstrated experimentally.

Figure 1. Three examples of recoding events. (A) Antizyme frameshifting. The +1 shift at the last codon (UCC) before the termination codon of ORF1 of human antizyme 1 is stimulated by polyamines and by a 5′ mRNA element and a 3′ pseudoknot. (B) Gene 60 bypassing. Fifty nucleotides between codons 47 and 48 of phage T4 gene 60 coding sequence are bypassed by half the ribosomes in response to matched takeoff and landing site codons, a stop codon directly after the take-off site in a stem–loop structure and a nascent peptide signal that acts within the ribosome. (C) Procaryotic selenocysteine insertion—redefinition. UGA codons in procaryotes that specify selenocysteine are directly followed by a stem structure whose apical loop is bound by a selenocysteine tRNA specific elongation factor SELB, resulting in a tethered aminoacylated tRNA poised for the oncoming ribosome. These figures are adapted from Atkins et al. (8).
Gene/mRNA, nucleotide sequence of the mRNA.

Because many mRNAs are very large, only the gene sequence of interest is given. For each gene the translation initiation codon as well as the mRNA elements thought to be important for recoding are marked. The following designations are used in the database: bold, any regions of mRNA which are important for recoding; underlined, recoding site (frameshift site for frameshift events, take-off and landing-site codons for translational bypass, redefined codon for codon redefinition); blue, start codon; red, stop codons important for recoding; brown, Shine–Dalgarno sequence; green and violet, double-stranded RNAs that play a role in frameshifting, violet is for stem 2 of pseudoknots and kissing loops; sequences in italics, those whose importance for recoding is at the level of their peptide product. The logo of the database (Fig. 2) can be used as a key for these designations.

AVAILABILITY AND SUBMISSIONS

RecodeWeb is freely available at http://recode.genetics.utah.edu/.

The authors welcome submission of new examples or additional information about already entered recoding events. Submission should be via electronic form at http://recode.genetics.utah.edu/submission/.

ACKNOWLEDGEMENTS

We are grateful to Drs Ivaylo Ivanov and Stefan Aigner for providing data before publication, and the following sources of support for this work: National Cancer Center fellowship to P.V.B.; Thomas Dee fellowship to O.L.G.; Centre National de la Recherche Scientifique, the Université Paul Sabatier and a grant from the Programme de Recherche Fondamentale en Microbiologie et maladies infectieuses et parasitaires to O.F.; grant (DEFG03-99ER62732) from Department of Energy to R.F.G; NIH grant GM 48152 to J.F.A. and NHGRI Genome Scholar award (K22 HG000401) to M.C.G.

REFERENCES