

## Circular code motifs near the ribosome decoding center



Karim El Soufi, Christian J. Michel\*

Theoretical Bioinformatics, ICube, University of Strasbourg, CNRS, 300 Boulevard Sébastien Brant, 67400 Illkirch, France

### ARTICLE INFO

#### Article history:

Received 12 June 2015

Accepted 13 July 2015

Available online 14 September 2015

#### Keywords:

Circular code motif

Ribosomal RNA

Transfer RNA

Translation code

### ABSTRACT

A maximal  $C^3$  self-complementary trinucleotide circular code  $X$  is identified in genes of bacteria, eukaryotes, plasmids and viruses (Michel, 2015; Arquès and Michel, 1996). A translation (framing) code based on the circular code was proposed in Michel (2012) with the identification of several  $X$  circular code motifs ( $X$  motifs shortly) in both ribosomal RNAs (rRNAs) and their decoding center, and transfer RNAs (tRNAs). We extended these results in two ways. First, three universal  $X$  motifs were determined in the ribosome decoding center: the  $X$  motif  $m_{AA}$  containing the conserved nucleotides A1492 and A1493, the  $X$  motif  $m_C$  containing the conserved nucleotide G530 and the  $X$  motif  $m$  with unknown biological function (El Soufi and Michel, 2014). Secondly, statistical analysis of  $X$  motifs of greatest lengths performed on different and large tRNA populations according to taxonomy, tRNA length and tRNA score showed that these  $X$  motifs have occurrence probabilities in the 5' and/or 3' regions of 16 isoaccepting tRNAs of prokaryotes and eukaryotes greater than the random case (Michel, 2013). We continue here the previous works with the identification of  $X$  motifs in rRNAs of prokaryotes and eukaryotes near the ribosome decoding center. Seven  $X$  motifs  $PrRNAXm$  conserved in 16S rRNAs of prokaryotes  $P$  and four  $X$  motifs  $ErRNAXm$  conserved in 18S rRNAs of eukaryotes  $E$  are identified near the ribosome decoding center. Furthermore, four very large  $X$  motifs of length greater than or equal to 20 nucleotides, 14 large  $X$  motifs of length between 16 and 19 nucleotides and several  $X$  motifs of length greater or equal to 9 nucleotides are found in tRNAs of prokaryotes. Some properties of these  $X$  motifs in tRNAs are described. These new results strengthen the concept of a translation code based on the circular code (Michel, 2012).

© 2015 Elsevier Ltd. All rights reserved.

### 1. Introduction

The ribosome is a complex ribonucleoprotein particle responsible for the synthesis of the cell protein by translating messenger RNA (mRNA). Ribosomes are composed of two subunits, a large subunit and a small subunit. Each subunit is formed by ribosomal RNAs (rRNAs) and proteins. A ribosome contains three transfer RNA (tRNA) binding sites: A-site (aminoacyl), P-site (peptidyl) and E-site (exit). During translation, the aminoacyl tRNA binds to the A-site where the decoding center containing the universally conserved dinucleotide AA (A1492 and A1493) distinguishes cognate from non-cognate tRNAs by anticodon-codon interactions (Wilson, 2014). The transfer of the amino acid from the P-site to the A-site results in the peptide-bond forming between the carboxyl group at the P-site and the newly arrived amino acid at the A-site. As the ribosome progresses by three nucleotides, the peptidyl tRNA moves from the A-site to the P-site. Finally the unloaded tRNA moves from the P-site to the E-site.

In 1996, a statistical analysis of occurrence frequencies of the 64 trinucleotides {AAA, . . . , TTT} in the three frames 0, 1 and 2 of genes of both prokaryotes and eukaryotes showed that the trinucleotides are not uniformly distributed in these three frames (Arquès and Michel, 1996). By convention here, the frame 0 is the reading frame in a gene and the frames 1 and 2 are the reading frame 0 shifted by one and two nucleotides in the 5'–3' direction, respectively. By excluding the four periodic permuted trinucleotides {AAA, CCC, GGG, TTT} and by assigning each trinucleotide to a preferential frame (frame of its highest occurrence frequency), three subsets  $X = X_0, X_1$  and  $X_2$  of 20 trinucleotides are found in the frames 0, 1 and 2, respectively, simultaneously of two large gene populations (protein coding regions): eukaryotes (26,757 sequences, 11,397,678 trinucleotides) and prokaryotes (13,686 sequences, 4709,758 trinucleotides) (Arquès and Michel, 1996). This set  $X$  contains the 20 following trinucleotides:

$$X = \{AAC, AAT, ACC, ATC, ATT, CAG, CTC, CTG, GAA, GAC, GAG, GAT, GCC, GGC, GGT, GTA, GTC, GTT, TAC, TTC\}. \quad (1)$$

The two sets  $X_1$  and  $X_2$ , of 20 trinucleotides each, in the shifted frames 1 and 2 of genes can be deduced from  $X$  by the circular permutation map (see below). These three trinucleotide sets

\* Corresponding author.

E-mail addresses: [kelsoufi@unistra.fr](mailto:kelsoufi@unistra.fr) (K. El Soufi), [c.michel@unistra.fr](mailto:c.michel@unistra.fr) (C.J. Michel).

present several strong mathematical properties, particularly the fact that  $X$  is a maximal  $C^3$  self-complementary trinucleotide circular code (Arquès and Michel, 1996). A trinucleotide circular code has the fundamental property to always retrieve the reading frame in any position of any sequence generated with the circular code. In particular, initiation and stop trinucleotides as well as any frame signals are not necessary to define the reading frame. Indeed, a window of 12 nucleotide length positioned anywhere in a sequence generated with the circular code  $X$  always retrieves the reading frame (Tables 2 and 3 in Michel, 2012).

A translation (framing) code based on the circular code was proposed in Michel (2012) with the identification of  $X$  circular code motifs ( $X$  motifs shortly) in the bacterial ribosomal RNA of *Thermus thermophilus*, in particular in the ribosome decoding center which recognizes the codon-anticodon helix in A-tRNA, and its tRNA of phenylalanine. A 3D visualization of  $X$  motifs in the ribosome shows several spatial configurations involving  $X$  motifs of mRNA, tRNA and rRNA. These results were extended in two ways. Firstly, three universal  $X$  motifs were identified in the ribosome decoding center of all the studied rRNAs from different kingdoms: bacteria *Escherichia coli* and *T. thermophilus*, archaea *Pyrococcus furiosus*, nuclear eukaryotes *Saccharomyces cerevisiae*, *Triticum aestivum* and *Homo sapiens*, and chloroplast *Spinacia oleracea*. These three  $X$  motifs are  $m_{AA}$  containing the conserved nucleotides A1492 and A1493,  $m_C$  containing the conserved nucleotide G530 and  $m$  whose biological function is unknown (El Soufi and Michel, 2014). Secondly, a statistical analysis of  $X$  motifs of greatest lengths performed on different and large tRNA populations according to the taxonomy, tRNA length and tRNA score showed that these  $X$  motifs have occurrence probabilities in the 5' and/or 3' regions of 16 isoaccepting tRNAs of prokaryotes and eukaryotes greater than the random case (Michel, 2013). By developing a search algorithm of  $X$  motifs in a DNA global multiple sequence alignment, we extend here the previous works to the identification of  $X$  motifs in rRNAs of prokaryotes and eukaryotes near the ribosome decoding center. Furthermore, in contrast to the statistical analysis of the distribution of  $X$  motifs of greatest lengths (Michel, 2013), a detailed analysis of  $X$  motifs is performed in the 20 isoaccepting tRNAs of bacteria. Several properties of  $X$  motifs are described according to (i) their type; (ii) their length: very large with a length greater than 20 nucleotides, large with a length between 16 and 19 nucleotides, otherwise with a length between 9 and 15 nucleotides ( $X$  motifs of lengths equal to 9 nucleotides retrieve the reading frame with a probability of 99.9%, Table 3 and Fig. 4 in Michel, 2012); (iii) their position in the tRNAs: 5' regions, anticodon regions, 3' regions; and (iv) their relations:  $X$  motifs shifted in frame by +1 or +2 nucleotides from other  $X$  motifs or the anticodons of tRNAs.

## 2. Method

### 2.1. Recall

The definitions of code, trinucleotide code, trinucleotide circular code, self-complementary trinucleotide circular code,  $C^3$  trinucleotide circular code and  $C^3$  self-complementary trinucleotide circular code related to the  $X$  motifs, i.e. motifs from the circular code  $X$  (Eq. (1)), are given in Michel (2012, 2013) and El Soufi and Michel (2014).

The trinucleotide set  $X$  (Eq. (1)) coding the reading frames in eukaryotic and prokaryotic genes is a maximal (20 trinucleotides)  $C^3$  self-complementary trinucleotide circular code with a window length equal to 12 nucleotides for retrieving the reading frame. The fundamental property of a circular code is the ability to retrieve the reading (original or constructed) frame of any sequence generated with this circular code. A circular code is a set of words over an

alphabet such that any sequence written on a circle (the next letter after the last letter of the sequence being the first letter) has a unique decomposition (factorization) into words of the circular code (Fig. 1 in Michel, 2012; for a graphical representation of the circular code definition and Fig. 2 in Michel, 2012; for an example). The reading frame in a sequence (a gene) is retrieved after the reading of a certain number of letters (nucleotides), called the window of the circular code. The length of this window for retrieving the reading frame is the letter length of the longest ambiguous words which can be read in at least two frames, plus one letter (Fig. 3 in Michel, 2012; for an example). For the circular code  $X$ , this window needs a length of 12 nucleotides as the two longest ambiguous words GGTAATTACCA and GGTAATTACCT of  $X$  have 11 nucleotides (Tables 2 and 3 in Michel, 2012).

In this paper, we study  $X$  circular code motifs ( $X$  motifs shortly) near the ribosome decoding center. It is important to remind the reader that there are two concepts: (i) the circular code  $X$ , which is a set of 20 trinucleotides (Eq. (1)); and (ii)  $X$  motifs, which are motifs (words) obtained with the circular code  $X$ . We give here a few examples of  $X$  motifs: AAC,AAT (a concatenation of the 1st and 2nd trinucleotides of  $X$ , the commas showing the adopted decomposition), TTC,TAC,AAC (a concatenation of the 20th, 19th and 1st trinucleotides of  $X$ ), AG,AAC,AAT (a concatenation of the suffix AG of the 6th or 11th trinucleotides of  $X$ , and the 1st and 2nd trinucleotides of  $X$ ), AG,AAC,AAT,AC (a concatenation of the suffix of the 6th or 11th trinucleotides of  $X$ , the 1st and 2nd trinucleotides of  $X$ , and the prefix AC of the 3rd trinucleotide of  $X$ ), etc. The motifs, for example, AAC,AAT,AAG, CA,AAC,AAT and AAC,AAT,AG are not  $X$  motifs.

### 2.2. Search algorithm of $X$ motifs in a DNA global multiple sequence alignment

We present here a search algorithm of  $X$  motifs of lengths greater than a given number of nucleotides in a DNA global multiple sequence alignment (global MSA with the program ClustalW2). It will identify common  $X$  motifs in multiple aligned RNA sequences. The algorithm is presented with DNA sequences, i.e. on the 4-letter alphabet  $A_4 = \{A, C, G, T\}$ , its extension on RNA sequences, i.e. on the 4-letter alphabet  $\{A, C, G, U\}$ , being trivial.

Let a trinucleotide  $t$  of the circular code  $X$  defined in Eq. (1) be the three letters  $t = l_1 l_2 l_3 \in A_4^3 = \{AAA, \dots, TTT\}$ . Let  $\text{Pref}_{\text{let}}(X)$  be the set containing the letters  $l_1 \in A_4$  of  $X$  and  $\text{Pref}_{\text{dilet}}(X)$  be the set containing the dileters  $l_1 l_2 \in A_4^2 = \{AA, \dots, TT\}$  of  $X$ . Then, by inspection of  $X$ , we have:

$$\text{Pref}_{\text{let}}(X) = \{A, C, G, T\} = A_4, \quad (2)$$

$$\text{Pref}_{\text{dilet}}(X) = \{AA, AC, AT, CA, CT, GA, GC, GG, GT, TA, TT\}. \quad (3)$$

Remark 1.  $\text{Card}(\text{Pref}_{\text{let}}(X)) = 4$  and  $\text{Card}(\text{Pref}_{\text{dilet}}(X)) = 11$  (among 16 dinucleotides).

The algorithm uses the following classical notions of language theory. Let  $x$  be a word (sequence) on  $A_4$  of length  $|x|$ .  $x[i]$  denotes the letter at index  $i$  of  $x$  and  $x[i..j]$  denotes the factor of  $x$  defined by  $x[i]x[i+1]..x[j]$  of length  $j - i + 1$ .

The function  $X\text{motif}$  searches an  $X$  motif at a given position  $\text{startX}$  (input parameter in integer) in a DNA sequence  $\text{seq}$  of length  $|\text{seq}|$  on  $A_4$  or  $A_4 \cup \{-\}$  (an aligned sequence with gaps) and returns its end position  $\text{endX}$  (output parameter in integer).

The function  $\text{Search\_Xmotif\_seq}$  searches all the  $X$  motifs in a DNA sequence  $\text{seq}$  (input parameter in string) of length  $|\text{seq}|$  on  $A_4$  or  $A_4 \cup \{-\}$  which are greater or equal to a minimum number of nucleotides in the  $X$  motif, named  $\text{lgMinX}$  (input parameter in integer), and returns a list  $\text{listXMotif}$  (output parameter) of  $X$

---

```

Xmotif[startX]
1. endX = startX
2. iX = 1 // index on X
3. testX = true
4. while testX
5. {
6.   if {seq[endX]  $\cap$  A4}  $\neq$  {} then
7.   {
8.     if iX = 1[3] then // Case 1 modulo 3: Pref1let
9.     {
10.      if endX  $\leq$  |seq| and {seq[endX]  $\cap$  Pref1let}  $\neq$  {} then
11.      {
12.       iX++
13.       endX++
14.      }
15.     }
16.     else testX = false
17.   }
18.   if iX = 2[3] then // Case 2 modulo 3: Prefdilet
19.   {
20.     if endX  $\leq$  |seq| and {seq[endX-1..endX]  $\cap$  Prefdilet}  $\neq$  {} then
21.     {
22.      iX++
23.      endX++
24.     }
25.     else testX = false
26.   }
27.   if iX = 0[3] then // Case 0 modulo 3: X
28.   {
29.     if endX  $\leq$  |seq| and {seq[endX-2..endX]  $\cap$  X}  $\neq$  {} then
30.     {
31.      iX++
32.      endX++
33.     }
34.     else testX = false
35.   }
36. }
37. }
38. return endX--

```

---

motifs of lengths greater or equal to lgMinX, otherwise an empty list. Each X motif is also a list containing the series of nucleotides of the X motif, its length and its start and end positions in the DNA sequence seq.

Let multipleSeqAlign (input parameter) be a list of cardinal card [multipleSeqAlign] aligned sequences (string) of length |seqAlign|. Let seqCommon be the common sequence (string) of length |seqAlign| containing the nucleotide which is identical (or almost

---

#### **Search\_Xmotif\_seq[seq,lgMinX]**

```

1. listXMotif  $\leftarrow$  {}
2. for start  $\leftarrow$  1 to |seq| step +1 do
3. {
4.   end  $\leftarrow$  Xmotif[start] // start: start position of Xmotif in seq
5.   lg  $\leftarrow$  end-start+1
6.   if lg  $\geq$  lgMinX then listXMotif  $\leftarrow$  {seq[start..end],lg,start,end}
7. }
8. return listXMotif

```

---

identical) at the same position in all sequences of MSA. The Algorithm\_Search\_Xmotif\_seqCommon constructs the common sequence seqCommon containing the universally conserved nucleotides in MSA otherwise a character different from A<sub>4</sub> (here the wild character “\*”) and then applies the function

---

```

Algorithm_Search_Xmotif_seqCommon[multipleSeqAlign,lgMinX]
1. for i ← 1 to |seqAlign| step +1 do
2. {
3.   seqCommon[i] ← "*"
4.   letter ← {}
5.   for j ← 1 to card[multipleSeqAlign] step +1 do
6.     letter ← letter U multipleSeqAlign[i,j]
7.   if card[letter] = 1 then seqCommon[i] ← multipleSeqAlign[i,j]
8. }
9. return Search_Xmotif_seq[seqCommon,lgMinX]

```

---

Search\_Xmotif\_seq to the constructed seqCommon for identifying the common X motifs in MSA.

An X motif returned by the Algorithm\_Search\_Xmotif\_seqCommon beginning at position *b* and ending at position *e* is characterized by either the integer interval [*b..e*] or the triplet (*b, e, l*) where the nucleotide length  $l = e - b + 1$ .

### 2.3. Coverage of X motifs in a DNA region

Let a set of *m* X motifs occurring in the position set  $P = \{[b_1..e_1], \dots, [b_m..e_m]\}$  in a DNA region  $R = [a..b]$ , e.g. the 5' or 3' region in a tRNA (extension on {A, C, G, U}). Note that the intersection of intervals [*b<sub>i</sub>..e<sub>i</sub>*] are not necessarily empty. Then, Interval(*P*) = {[min<sub>1</sub>..max<sub>1</sub>], ..., [min<sub>*n*</sub>..max<sub>*n*</sub>]} is the union of the ranges *b*<sub>1</sub> to *e*<sub>1</sub>, ..., *b*<sub>*m*</sub> to *e*<sub>*m*</sub>. Note that the intersection of intervals [min<sub>*i*</sub>..max<sub>*i*</sub>] are always empty. Therefore, Coverage(*X,R*) giving the probability of sites of the DNA region *R* occupied by *X* motifs is simply defined by:

$$\text{Coverage}(X, R) = \frac{1}{b-a+1} \sum_{i=1}^n (\max_i - \min_i + 1). \quad (4)$$

### 2.4. Crystallographic data

The Algorithm\_Search\_Xmotif\_seqCommon is applied to global multiple sequence alignment of rRNAs of different organisms whose crystallographic structures are known and available in the Protein Data Bank (PDB, [www.rcsb.org/pdb/home/home.do](http://www.rcsb.org/pdb/home/home.do), February 2015). The selected PDB entries have necessarily a bacterial 16S rRNA or a eukaryotic 18S rRNA, and if possible in addition an mRNA and/or tRNA to visualize the spatial interaction of their X motifs. PDB entries containing synthetic chains were excluded. The studied PDB crystallographic structures are for bacteria: *E. coli* (Brilot et al., 2013) and *T. thermophilus* (Jenner et al., 2010); for archaea: *P. furiosus* (Armache et al., 2013); for (nuclear) eukaryotes: *S. cerevisiae* (Armache et al., 2010a), *T. aestivum* (Armache et al., 2010a,b; Gogala et al., 2014) and *H. sapiens* (Anger et al., 2013). Table 1 summarizes the main features of the crystallographic structures studied.

### 2.5. Scripts in Jmol language

Jmol is an open-source Java viewer for chemical structures in 3D (<http://www.jmol.org/>). It allows the reading of a variety of file

formats and high-performance 3D rendering with no hardware requirements. Several scripts were written in Jmol for each PDB entry to visualize the X motifs in the messenger, transfer and ribosomal RNAs. They are not detailed here.

## 3. Results

### 3.1. Identification of conserved X circular code motifs in rRNAs near the ribosome decoding center

#### 3.1.1. Identification of seven conserved X circular code motifs PrRNAX<sub>*m*</sub> in 16S rRNAs of prokaryotes P near the ribosome decoding center

By using the Algorithm\_Search\_Xmotif\_seqCommon (Section 2.2) with the global multiple sequence alignment ClustalW2, seven X circular code motifs PrRNAX<sub>*m*</sub> are identified which are conserved in the prokaryotic 16S rRNAs of bacteria *E. coli* (3J5T) and *T. thermophilus* (3I8G), and archaea *P. furiosus* (3J20) (Table 2):

- (i) PrRNAX<sub>1</sub>(*E. coli*, 537, 545, 9) = PrRNAX<sub>1</sub>(*T. thermophilus*, 517, 525, 9) = G,GAG,GGY,GC of nine nucleotides starts with the nucleotide G suffix of CAG, CTG, GAG ∈ X, has two trinucleotides GAG, GGY ∈ X where Y=T in *E. coli* and Y=C in *T. thermophilus*, and ends with the dinucleotide GC prefix of GCC ∈ X. The large X motif PrRNAX<sub>1</sub>(*P. furiosus*, 480, 497, 18) = GC,GGT,AAT,ACC,GGC,GGC,C of 18 nucleotides starts with the dinucleotide GC suffix of GGC ∈ X, has five trinucleotides GGT, AAT, ACC, GGC, GGC ∈ X and ends with the nucleotide C prefix of CAG, CTC, CTG ∈ X. PrRNAX<sub>1</sub> of *E. coli* and *T. thermophilus* are partial suffixes of PrRNAX<sub>1</sub> of *P. furiosus*.
- (ii) PrRNAX<sub>2</sub>(*E. coli*, 688, 697, 10), PrRNAX<sub>2</sub>(*T. thermophilus*, 668, 677, 10) and PrRNAX<sub>2</sub>(*P. furiosus*, 643, 654, 12) have the common X motif GGT,GAA,AT of eight nucleotides with GGT, GAA ∈ X. PrRNAX<sub>2</sub> of *E. coli* and *T. thermophilus* start with the dinucleotide GC suffix of GGC ∈ X and end with the dinucleotide AT prefix of ATC, ATT ∈ X. PrRNAX<sub>2</sub> of *P. furiosus* starts with the nucleotide G suffix of CAG, CTG, GAG ∈ X and ends with the dinucleotide CT prefix of CTC, CTG ∈ X.
- (iii) PrRNAX<sub>3</sub>(*E. coli*, 714, 736, 23) and PrRNAX<sub>3</sub>(*T. thermophilus*, 694, 713, 20) have the large common X motif G,AAY,R<sub>1</sub>CC, GR<sub>2</sub>T,GGC,GAA,GGC of 19 nucleotides starting with the nucleotide G suffix of CAG, CTG, GAG ∈ X followed by six trinucleotides AAY, R<sub>1</sub>CC, GR<sub>2</sub>T, GGC, GAA, GGC ∈ X where Y=T, R<sub>1</sub>=A and R<sub>2</sub>=G in *E. coli* while Y=C, R<sub>1</sub>=G and R<sub>2</sub>=A in *T. thermophilus*. PrRNAX<sub>3</sub> of *E. coli* ends with the nucleotide C prefix of CAG, CTC, CTG ∈ X whereas PrRNAX<sub>3</sub> of *T. thermophilus* ends with the nucleotide A prefix of

AAC, AAT, ACC, ATC, ATT  $\in X$ . The  $X$  motif  $PrRNAXm_3(P. furiosus, 676, 690, 15) = GT, GGC, GAA, GGC, GCC, C$  of 15 nucleotides is a conserved suffix of  $PrRNAXm_3$  of *E. coli* (14 identical letters among 15) starting with the dinucleotide GT suffix of GGT in  $PrRNAXm_3$  of *E. coli*.

- (iv) The large  $X$  motif  $PrRNAXm_4(E. coli, 789, 807, 19) = TA, GAT, ACC, CTG, GTA, GTC, CA$  of 19 nucleotides starts with the dinucleotide TA suffix of  $GTA \in X$ , has five trinucleotides GAT, ACC, CTG, GTA, GTC  $\in X$  and ends with the dinucleotide CA prefix of  $CAG \in X$ .  $PrRNAXm_4(T. thermophilus, 769, 777, 9) = PrRNAXm_4(P. furiosus, 743, 751, 9) = TA, GAT, ACC, C$  of nine nucleotides is a prefix of  $PrRNAXm_4$  of *E. coli* ending with the nucleotide C prefix of CTG in  $PrRNAXm_4$  of *E. coli*.
- (v)  $PrRNAXm_5(E. coli, 1186, 1197, 12)$ ,  $PrRNAXm_5(T. thermophilus, 1164, 1174, 11)$  and  $PrRNAXm_5(P. furiosus, 1146, 1156, 11)$  have the common  $X$  motif G, GRY<sub>1</sub>, GAC, GKY<sub>2</sub>, W of 11 nucleotides starting with the nucleotide G suffix of CAG, CTG, GAG  $\in X$  followed by three trinucleotides GRY<sub>1</sub>, GAC, GKY<sub>2</sub>  $\in X$  where  $R=A, Y_1=T, K=T, Y_2=C$  and  $W=A$  in *E. coli*,  $R=A, Y_1=C, K=T, Y_2=C$  and  $W=T$  in *T. thermophilus* while  $R=G, Y_1=C, K=G, Y_2=T$  and  $W=A$  in *P. furiosus*.  $PrRNAXm_5$  of *E. coli* ends with the dinucleotide AA prefix of AAC, AAT  $\in X$ ,  $PrRNAXm_5$  of *T. thermophilus* ends with the nucleotide T prefix of TAC, TTC  $\in X$  and  $PrRNAXm_5$  of *P. furiosus* ends with the nucleotide A prefix of AAC, AAT, ACC, ATC, ATT  $\in X$ .
- (vi) The large common  $X$  motif  $PrRNAXm_6(E. coli, 1211, 1228, 18) = PrRNAXm_6(T. thermophilus, 1189, 1206, 18) = T, TAC, GRC, CWG, GGC, KAC, AC$  of 18 nucleotides starts with the nucleotide T suffix of AAT, ATT, GAT, GGT, GTT  $\in X$ , has five trinucleotides TAC, GRC, CWG, GGC, KAC  $\in X$  where  $R=A, W=A$  and  $K=T$  in *E. coli* while  $R=G, W=T$  and  $K=G$  in *T. thermophilus*, and ends with the dinucleotide AC prefix of ACC  $\in X$ .  $PrRNAXm_6(P. furiosus, 1180, 1188, 9) = G, GGC, TAC, AC$  of nine nucleotides is a suffix of  $PrRNAXm_6$  of *E. coli* starting with the nucleotide G suffix of CAG in  $PrRNAXm_6$  of *E. coli*.
- (vii) The common  $X$  motif  $PrRNAXm_7(E. coli, 1368, 1382, 15) = PrRNAXm_7(T. thermophilus, 1347, 1361, 15) = PrRNAXm_7(P. furiosus, 1328, 1342, 15) = RC, GGY, GAA, TAC, GTY, C$  of 15 nucleotides start with the dinucleotide AC ( $R=A$ ) suffix of AAC, GAC, TAC  $\in X$  in *E. coli* and with the dinucleotide GC ( $R=G$ ) suffix of GGC  $\in X$  in *T. thermophilus* and *P. furiosus*, has four trinucleotides GGY, GAA, TAC, GTY  $\in X$  where  $Y=T$  in *E. coli* and *T. thermophilus* while  $Y=C$  in *P. furiosus*, and ends with the nucleotide C prefix of CAG, CTC, CTG  $\in X$ .

These seven prokaryotic rRNA  $X$  motifs have different primary structures. Thus, the classical bioinformatics methods, such as sequence alignment or phylogenetic inference, are not able to find these motifs which are only revealed by the circular code theory.

Figs. 1–3 show that the prokaryotic rRNA  $X$  motifs  $PrRNAXm_1$  in maroon,  $PrRNAXm_2$  in pink,  $PrRNAXm_3$  in gold,  $PrRNAXm_4$  in orange,  $PrRNAXm_5$  in navy,  $PrRNAXm_6$  in purple and  $PrRNAXm_7$  in red of *E. coli*, *T. thermophilus* and *P. furiosus* are near the ribosome decoding

center (5' regions of tRNAs in dark blue, 3' regions of tRNAs in clearer blue and anticodons of tRNAs in black).

### 3.1.2. Identification of four conserved $X$ circular code motifs $ErRNAXm$ in 18S rRNAs of (nuclear) eukaryotes E near the ribosome decoding center

By using the Algorithm\_Search\_Xmotif\_seqCommon (Section 2.2) with the global multiple sequence alignment ClustalW2, four  $X$  circular code motifs  $ErRNAXm$  are identified which are conserved in the eukaryotic 18S rRNAs of *S. cerevisiae* (3IZE), *T. aestivum* (3J5Z) and *H. sapiens* (3J3D) (Table 3):

- (i)  $ErRNAXm_1(S. cerevisiae, 900, 911, 12) = ErRNAXm_1(T. aestivum, 905, 916, 12) = ErRNAXm_1(H. sapiens, 957, 968, 12) = A, GGT, GAA, ATT, CT$  of 12 nucleotides starts with the nucleotide A suffix of GAA, GTA  $\in X$ , has three trinucleotides GGT, GAA, ATT  $\in X$  and ends with the dinucleotide CT prefix of CTC, CTG  $\in X$ .
- (ii) The large common  $X$  motif  $ErRNAXm_2(S. cerevisiae, 987, 1004, 18) = ErRNAXm_2(T. aestivum, 992, 1009, 18) = ErRNAXm_2(H. sapiens, 1044, 1061, 18) = G, NTC, GAA, GAY, GAT, CAG, AT$  of 18 nucleotides starts with the nucleotide G suffix of CAG, CTG, GAG  $\in X$ , has five trinucleotides NTC, GAA, GAY, GAT, CAG  $\in X$  where  $N=A$  and  $Y=T$  in *S. cerevisiae*,  $N=C$  and  $Y=C$  in *T. aestivum*, while  $N=T$  and  $Y=C$  in *H. sapiens*, and ends with the dinucleotide AT prefix of ATC, ATT  $\in X$ .
- (iii)  $ErRNAXm_3(S. cerevisiae, 1189, 1197, 9) = ErRNAXm_3(T. aestivum, 1193, 1201, 9) = ErRNAXm_3(H. sapiens, 1246, 1254, 9) = A, CTC, AAC, AC$  of nine nucleotides starts with the nucleotide A suffix of GAA, GTA  $\in X$ , has two trinucleotides CTC, AAC  $\in X$  and ends with the dinucleotide AC prefix of ACC  $\in X$ .
- (iv) The large  $X$  motif  $ErRNAXm_4(S. cerevisiae, 1564, 1582, 19) = TC, TTC, AAC, GAG, GAA, TTC, CT$  of 19 nucleotides starts with the dinucleotide TC suffix of ATC, CTC, GTC, TTC  $\in X$ , has five trinucleotides TTC, AAC, GAG, GAA, TTC  $\in X$  and ends with the dinucleotide CT prefix of CTC, CTG  $\in X$ .  $ErRNAXm_4(T. aestivum, 1575, 1596, 12) = TC, AAC, GAG, GAA, T$  is a factor of  $ErRNAXm_4$  of *S. cerevisiae* starting with the dinucleotide TC suffix of the 1st TTC  $\in X$  in  $ErRNAXm_4$  of *S. cerevisiae* and ending with the nucleotide T prefix of the 2nd TTC  $\in X$  in  $ErRNAXm_4$  of *S. cerevisiae*. The  $X$  motif  $ErRNAXm_4(H. sapiens, 1631, 1645, 15) = TG, AAC, GAG, GAA, TTC, C$  is an almost exact suffix of  $ErRNAXm_4$  of *S. cerevisiae*.

As with the prokaryotic rRNA  $X$  motifs  $PrRNAXm$ , the four eukaryotic rRNA  $X$  motifs  $ErRNAXm$  have completely different primary structures and thus, they can only be revealed by the circular code theory.

Figs. 4–6 show that the (nuclear) eukaryotic rRNA  $X$  motifs  $ErRNAXm_1$  in purple,  $ErRNAXm_2$  in pink,  $ErRNAXm_3$  in red and  $ErRNAXm_4$  in orange of *S. cerevisiae*, *T. aestivum* and *H. sapiens* are near the ribosome decoding center (5' regions of tRNAs in dark blue, 3' regions of tRNAs in clearer blue and anticodons of tRNAs in black), except  $ErRNAXm_1$  in *S. cerevisiae* and *T. aestivum*.

**Table 1**

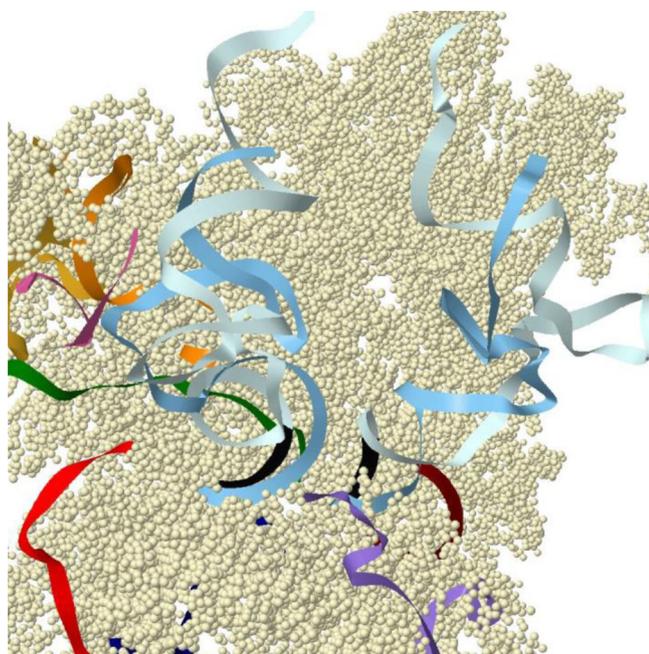
$X$  circular code motifs studied in six crystallographic structures of the Protein Data Bank PDB. The main features of the studied crystallographic structures are given: PDB identification (ID), kingdom, organism, type (16S for prokaryotes, 18S for eukaryotes) and base length (b) of rRNA, mRNA (Yes for available, No for unavailable), type of tRNA for the A, P and E sites (No for unavailable).

PDB ID	Kingdom	Organism	rRNA	mRNA	A-tRNA	P-tRNA	E-tRNA
3J5T	Bacteria	<i>Escherichia coli</i>	16S (1542 b)	Yes	Phe	Phe	No
3I8G	Bacteria	<i>Thermus thermophilus</i>	16S (1516 b)	Yes	Phe	Phe	Phe
3J20	Archaea	<i>Pyrococcus furiosus</i>	16S (1495 b)	No	No	Phe	Phe
3IZE	Eukaryote, nuclear	<i>Saccharomyces cerevisiae</i>	18S (1800 b)	Yes	No	Asp	No
3J5Z	Eukaryote, nuclear	<i>Triticum aestivum</i>	18S (1810 b)	Yes	No	Asp	No
3J3D	Eukaryote, nuclear	<i>Homo Sapiens</i>	18S (1869 b)	No	No	No	Met

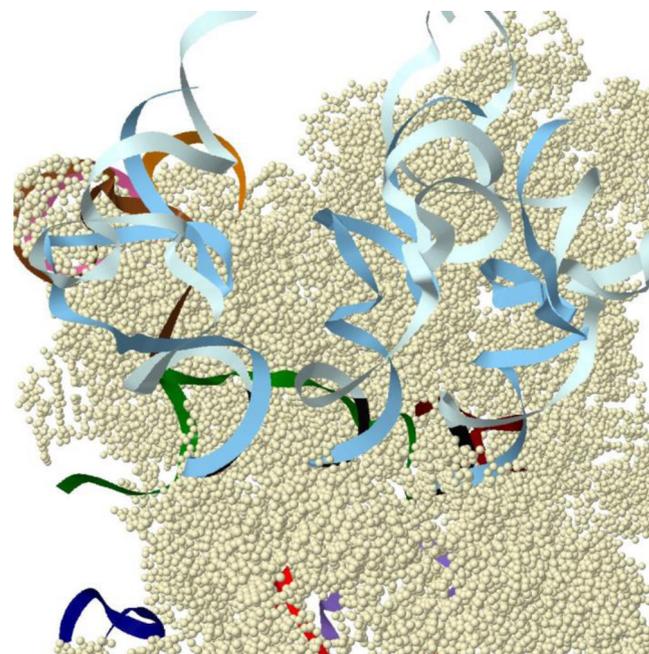
**Table 2**

Identification of seven X circular code motifs PrRNAXm in 16S rRNAs of prokaryotes P (bacteria *E. coli* and *T. thermophilus*, archaea *P. furiosus*) near the ribosome decoding center.

Symbol	X circular code motifs	Prokaryotic rRNA (PDB ID)	Start	End	Length
PrRNAXm <sub>1</sub>	G, GAG, GGT, GC	<i>E. coli</i> (3J5T)	537	545	9
	G, GAG, GGC, GC	<i>T. thermophilus</i> (3I8G)	517	525	9
	GC, GGT, AAT, ACC, GGC, GGC, C * **	<i>P. furiosus</i> (3J20)	480	497	18
PrRNAXm <sub>2</sub>	GC, GGT, GAA, AT	<i>E. coli</i> (3J5T)	688	697	10
	GC, GGT, GAA, AT	<i>T. thermophilus</i> (3I8G)	668	677	10
	G, GGT, GAA, ATC, CT *** ** *	<i>P. furiosus</i> (3J20)	643	654	12
PrRNAXm <sub>3</sub>	G, AAT, ACC, GGT, GGC, GAA, GGC, GGC, C	<i>E. coli</i> (3J5T)	714	736	23
	G, AAC, GCC, GAT, GGC, GAA, GGC, A	<i>T. thermophilus</i> (3I8G)	694	713	20
	GT, GGC, GAA, GGC, GCC, C * *** ** *	<i>P. furiosus</i> (3J20)	676	690	15
PrRNAXm <sub>4</sub>	TA, GAT, ACC, CTG, GTA, GTC, CA	<i>E. coli</i> (3J5T)	789	807	19
	TA, GAT, ACC, C	<i>T. thermophilus</i> (3I8G)	769	777	9
	TA, GAT, ACC, C ** *** ** *	<i>P. furiosus</i> (3J20)	743	751	9
PrRNAXm <sub>5</sub>	G, GAT, GAC, GTC, AA	<i>E. coli</i> (3J5T)	1186	1197	12
	G, GAC, GAC, GTC, T	<i>T. thermophilus</i> (3I8G)	1164	1174	11
	G, GGC, GAC, GGT, A * * *** *	<i>P. furiosus</i> (3J20)	1146	1156	11
PrRNAXm <sub>6</sub>	T, TAC, GAC, CAG, GGC, TAC, AC	<i>E. coli</i> (3J5T)	1211	1228	18
	T, TAC, GGC, CTG, GGC, GAC, AC	<i>T. thermophilus</i> (3I8G)	1189	1206	18
	G, GGC, TAC, AC * *** ** **	<i>P. furiosus</i> (3J20)	1180	1188	9
PrRNAXm <sub>7</sub>	AC, GGT, GAA, TAC, GTT, C	<i>E. coli</i> (3J5T)	1368	1382	15
	GC, GGT, GAA, TAC, GTT, C	<i>T. thermophilus</i> (3I8G)	1347	1361	15
	GC, GGC, GAA, TAC, GTC, C * ** *** ** *	<i>P. furiosus</i> (3J20)	1328	1342	15



**Fig. 1.** X circular code motifs near the bacterial ribosome decoding center of *Escherichia coli* (crystallographic structure PDB 3J5T): the mRNA (green), the rRNA X motifs PrRNAXm<sub>1</sub>(537, 545, 9) (maroon), PrRNAXm<sub>2</sub>(688, 697, 10) (pink), PrRNAXm<sub>3</sub>(714, 736, 23) (gold), PrRNAXm<sub>4</sub>(789, 807, 19) (orange), PrRNAXm<sub>5</sub>(1186, 1197, 12) (navy), PrRNAXm<sub>6</sub>(1211, 1228, 18) (purple), PrRNAXm<sub>7</sub>(1368, 1382, 15) (red), and the tRNA (5' region in dark blue, 3' region in clearer blue and anticodon in black). The remaining rRNA (lemonchiffon) is outside the neighborhood of these X motifs. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



**Fig. 2.** X circular code motifs near the bacterial ribosome decoding center of *Thermus thermophilus* (crystallographic structure PDB 3I8G): the mRNA (green), the rRNA X motifs PrRNAXm<sub>1</sub>(517, 525, 9) (maroon), PrRNAXm<sub>2</sub>(668, 677, 10) (pink), PrRNAXm<sub>3</sub>(694, 713, 20) (gold), PrRNAXm<sub>4</sub>(769, 777, 9) (orange), PrRNAXm<sub>5</sub>(1164, 1174, 11) (navy), PrRNAXm<sub>6</sub>(1189, 1206, 18) (purple), PrRNAXm<sub>7</sub>(1347, 1361, 15) (red), and the tRNA (5' region in dark blue, 3' region in clearer blue and anticodon in black). The remaining rRNA (lemonchiffon) is outside the neighborhood of these X motifs. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

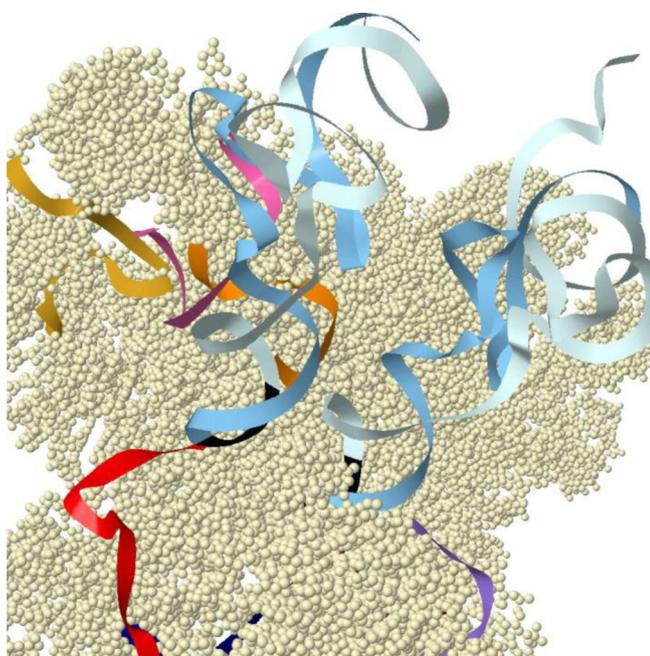
**Table 3**  
Identification of four X circular code motifs *ErRNAXm* in 18S rRNAs of (nuclear) eukaryotes *E* (*S. cerevisiae*, *T. aestivum*, *H. sapiens*) near the ribosome decoding center.

Symbol	X circular code motifs	Eukaryotic rRNA (PDB ID)	Start	End	Length
<i>ErRNAXm<sub>1</sub></i>	A, GGT, GAA, ATT, CT	<i>S. cerevisiae</i> (3IZE)	900	911	12
	A, GGT, GAA, ATT, CT	<i>T. aestivum</i> (3J5Z)	905	916	12
	A, GGT, GAA, ATT, CT * *** **	<i>H. sapiens</i> (3J3D)	957	968	12
<i>ErRNAXm<sub>2</sub></i>	G, ATC, GAA, GAT, GAT, CAG, AT	<i>S. cerevisiae</i> (3IZE)	987	1004	18
	G, CTC, GAA, GAC, GAT, CAG, AT	<i>T. aestivum</i> (3J5Z)	992	1009	18
	G, TTC, GAA, GAC, GAT, CAG, AT * ** *** ** *** **	<i>H. sapiens</i> (3J3D)	1044	1061	18
<i>ErRNAXm<sub>3</sub></i>	A, CTC, AAC, AC	<i>S. cerevisiae</i> (3IZE)	1189	1197	9
	A, CTC, AAC, AC	<i>T. aestivum</i> (3J5Z)	1193	1201	9
	A, CTC, AAC, AC * *** **	<i>H. sapiens</i> (3J3D)	1246	1254	9
<i>ErRNAXm<sub>4</sub></i>	TC, TTC, AAC, GAG, GAA, TTC, CT	<i>S. cerevisiae</i> (3IZE)	1564	1582	19
	TC, AAC, GAG, GAA, T	<i>T. aestivum</i> (3J5Z)	1575	1596	12
	TG, AAC, GAG, GAA, TTC, C * *** ** ** *	<i>H. sapiens</i> (3J3D)	1631	1645	15

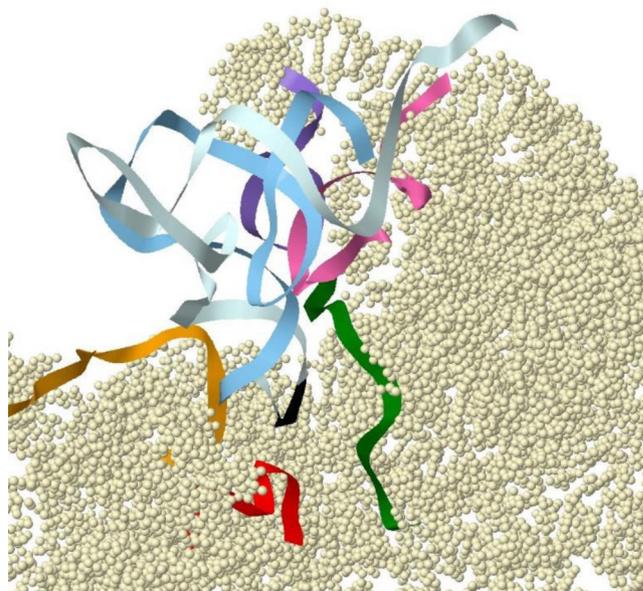
### 3.2. Identification of X circular code motifs in tRNAs of prokaryotes

We give the main features of X motifs for each isoaccepting tRNA of prokaryotes (bacteria *E. coli* and *T. thermophilus*, archaea *P. furiosus*), more details on the X motifs are found in Tables 4a–u. We use the classical genetic alphabet convention to be able to engulf X motifs that are fairly similar within a unified pattern. Let  $R = \{A, G\}$ ,  $Y = \{C, T\}$ ,  $S = \{C, G\}$ ,  $W = \{A, T\}$ ,  $K = \{G, T\}$ ,  $M = \{A, C\}$  and  $N = \{A, C, G, T\}$ . Furthermore, in term of X motif length, we are distinguishing three classes of X motifs: very large X motifs greater or equal to 20 nucleotides (remember that the average lengths of prokaryotic tRNAs range typically from 71 to 91 nucleotides for Cys and Ser,

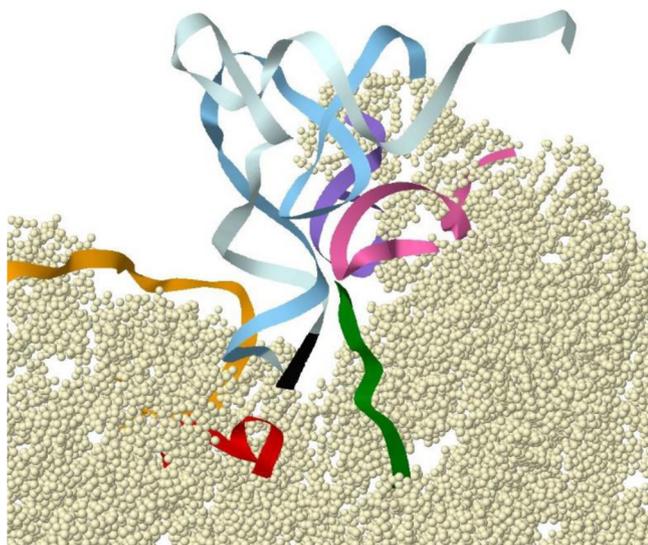
respectively, see Section 2.4.1 and Fig. 2 in Michel, 2013), large X motifs between 16 and 19 nucleotides and X motifs between 9 and 15 nucleotides. X motifs of lengths equal to 9 nucleotides already retrieve the reading frame with a probability of 99.9% and X motifs of lengths greater or equal to 12 nucleotides always retrieve, by definition, the reading frame, i.e. with a probability of 100% (Table 3 and Fig. 4 in Michel, 2012). Moreover, the underline in an X motif signifies that the underlined nucleotides are in common with one or more other motifs. Finally, the X motifs are studied according to three regions of tRNAs: X motifs between the 5' ends of tRNAs and the anticodon (called here 5' regions), X motifs having at least one nucleotide in the anticodon (anticodon regions) and X motifs between the anticodon and the 3' ends of tRNAs (3' regions). The



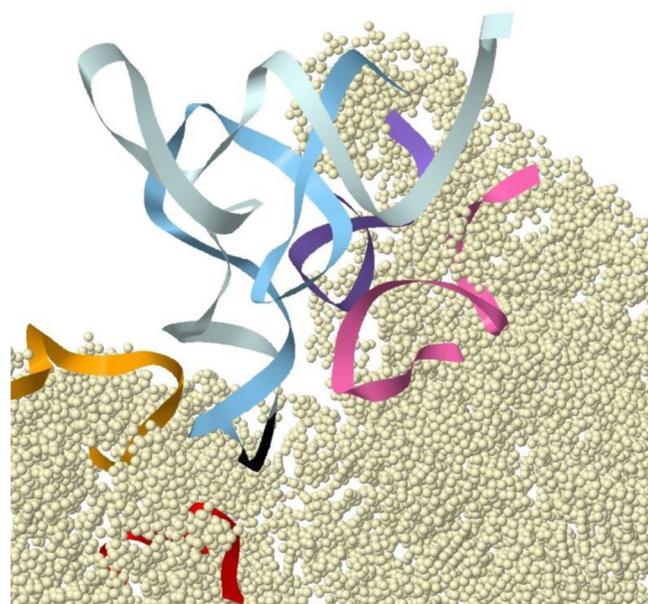
**Fig. 3.** X circular code motifs near the archaeal ribosome decoding center of *Pyrococcus furiosus* (crystallographic structure PDB 3J20): the rRNA X motifs *PrRNAXm<sub>1</sub>*(480, 497, 18) (maroon), *PrRNAXm<sub>2</sub>*(643, 654, 12) (pink), *PrRNAXm<sub>3</sub>*(676, 690, 15) (gold), *PrRNAXm<sub>4</sub>*(743, 751, 9) (orange), *PrRNAXm<sub>5</sub>*(1146, 1156, 11) (navy), *PrRNAXm<sub>6</sub>*(1180, 1188, 9) (purple), *PrRNAXm<sub>7</sub>*(1328, 1342, 15) (red), and the tRNA (5' region in dark blue, 3' region in clearer blue and anticodon in black). The remaining rRNA (lemonchiffon) is outside the neighborhood of these X motifs and the mRNA is missing (Table 1). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



**Fig. 4.** X circular code motifs near the (nuclear) eukaryotic ribosome decoding center of *Saccharomyces cerevisiae* (crystallographic structure PDB 3IZE): the mRNA (green), the rRNA X motifs *ErRNAXm<sub>1</sub>*(900, 911, 12) (purple), *ErRNAXm<sub>2</sub>*(987, 1004, 18) (pink), *ErRNAXm<sub>3</sub>*(1189, 1197, 9) (red), *ErRNAXm<sub>4</sub>*(1564, 1582, 19) (orange), and the tRNA (5' region in dark blue, 3' region in clearer blue and anticodon in black). The remaining rRNA (lemonchiffon) is outside the neighborhood of these X motifs. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



**Fig. 5.** X circular code motifs near the (nuclear) eukaryotic ribosome decoding center of *Triticum aestivum* (crystallographic structure PDB 3J5Z): the mRNA (green), the rRNA X motifs *ErRNAXm*<sub>1</sub>(905, 916, 12) (purple), *ErRNAXm*<sub>2</sub>(992, 1009, 18) (pink), *ErRNAXm*<sub>3</sub>(1193, 1201, 9) (red), *ErRNAXm*<sub>4</sub>(1575, 1596, 12) (orange), and the tRNA (5' region in dark blue, 3' region in clearer blue and anticodon in black). The remaining rRNA (lemonchiffon) is outside the neighborhood of these X motifs. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



**Fig. 6.** X circular code motifs near the (nuclear) eukaryotic ribosome decoding center of *Homo Sapiens* (crystallographic structure PDB 3J3D): the rRNA X motifs *ErRNAXm*<sub>1</sub>(957, 968, 12) (purple), *ErRNAXm*<sub>2</sub>(1044, 1061, 18) (pink), *ErRNAXm*<sub>3</sub>(1246, 1254, 9) (red), *ErRNAXm*<sub>4</sub>(1631, 1645, 15) (orange), and the tRNA (5' region in dark blue, 3' region in clearer blue and anticodon in black). The remaining rRNA (lemonchiffon) is outside the neighborhood of these X motifs and the mRNA is missing (Table 1). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

results below will identify X motifs, a few of them being very large, and different relations, in particular a shifting by 0, +1 or +2 mod 3 nucleotides with other X motifs or the anticodon.

### 3.2.1. X circular code motifs in Ala-tRNAs (Table 4a)

- (i) 5' regions of Ala-tRNAs: The X motif T,CAG,CTG,GG and the class of X motifs GC,CTG,GWA,K are shifted in frame (modulo 3 according to their suffix-prefix). The class of X motifs GC,GCC,GCC,YT occurs before (5') the anticodons GGC and TGC.
- (ii) Anticodon regions of Ala-tRNAs: The X motif GC,GCC,GCC,CTC, GC is in a different frame than the anticodon CGC. The very large X motif *Ala-tRNAXm*<sub>1</sub> GC,CTC,AAT,GGC,ATT,GAG,GAG, GTC,A of 24 nucleotides in *T. thermophilus* is in frame with the anticodon GGC. The X motif GC,CTG,AAT,C which is prefix of *Ala-tRNAXm*<sub>1</sub> is in frame with the anticodon CGC.
- (iii) 3' regions of Ala-tRNAs: The class of X motifs *Ala-tRNAXm*<sub>2</sub> K, SAG,GAG,GTC,W is suffix of *Ala-tRNAXm*<sub>1</sub>. The class of X motifs R,GAG,GYC,R is suffix of *Ala-tRNAXm*<sub>2</sub>. Two X motifs are also observed: A,GGT,CAG,GG and CC,CTC,GGC,T.

### 3.2.2. X circular code motifs in Arg-tRNAs (Table 4b)

- (i) 5' regions of Arg-tRNAs: The X motif CC,GGT,GGC,CT is found. The large X motif *Arg-tRNAXm*<sub>1</sub> CC,GTA,GTT,CAG,CTG,GAT,A of 18 nucleotides is identified in *E. coli*. The class of X motifs S, GTA,GYY,TA is prefix of *Arg-tRNAXm*<sub>1</sub>, and the class of X motifs T,CAG,CWG,GAT,A and *Arg-tRNAXm*<sub>2</sub> RS,CWG,GAT,R are suffix of *Arg-tRNAXm*<sub>1</sub>. The X motif AT,AAC,GAG,C and *Arg-tRNAXm*<sub>2</sub> are shifted in frame. The X motif TA,GAG,TAC,T is observed. The class of X motifs G,GGC,GYC,GGC,CT and the X motif GC,AAC, GAC,CT occur before the anticodons TCG and TCT.
- (ii) Anticodon regions of Arg-tRNAs: The large X motif *Arg-tRNAXm*<sub>3</sub> A,GAG,AAC,GCC,GCC,CTC,C of 17 nucleotides in *P. furiosus* is in a different frame than the anticodon CCG. The X motif GC,GTC,GGC,CTC,C which is suffix of *Arg-tRNAXm*<sub>3</sub> is in a different frame than the anticodon CCG. The large X motif *Arg-tRNAXm*<sub>4</sub> AG,GGC,GGC,GGC,CTC,CT of 16 nucleotides in *P.*

*furiosus* is in a different frame than the anticodon CCT. The X motif G,GGC,TTC,CT which is suffix of *Arg-tRNAXm*<sub>4</sub> is in a different frame than the anticodon CCT. The X motif *Arg-tRNAXm*<sub>5</sub> A,GCC,GCC,TCT,TA is in a different frame than the anticodon TCT. The class of X motifs G,RCC,TCT,TA which is suffix of *Arg-tRNAXm*<sub>5</sub> is in a different frame than the anticodon TCT. The X motif G,GCC,TTC,GG is in a different frame than the anticodon TCG. The large X motif *Arg-tRNAXm*<sub>6</sub> A,GTA,CTC,GGC,TAC,GAA,C of 17 nucleotides in *E. coli* is in a different frame than the anticodon ACG. The X motif T,GAC, TAC,GG is in a different frame than the anticodon ACG. The class of X motifs G,GAY,CWS,SAG,GTC,S is in frame with the anticodons ACG and GCG.

- (iii) 3' regions of Arg-tRNAs: The X motifs A,GCC,GAA,GGT,CAG,A and the class of X motifs A,WTC,CTG,CAG,GG are observed.

### 3.2.3. X circular code motifs in Asn-tRNAs (Table 4c)

- (i) 5' regions of Asn-tRNAs: The X motif *Asn-tRNAXm*<sub>1</sub> GCC,GCC, GTA,GC is observed. The X motif *Asn-tRNAXm*<sub>2</sub> T,GTA,GTT,CAG, T and *Asn-tRNAXm*<sub>1</sub> are shifted in frame. The large X motif *Asn-tRNAXm*<sub>3</sub> T,CAG,CAG,GTA,GAG,CAG,C of 17 nucleotides in *T. thermophilus* and *Asn-tRNAXm*<sub>2</sub> are shifted in frame. The X motif AG,AAC,GGC,GG is shifted by +2 nucleotides from *Asn-tRNAXm*<sub>3</sub> (underlined nucleotides).
- (ii) Anticodon regions of Asn-tRNAs: The X motif *Asn-tRNAXm*<sub>4</sub> TA, ACC,GGT,A is in a different frame than the anticodon GTT.
- (iii) 3' regions of Asn-tRNAs: The X motif CC,GGC,GGT,C and *Asn-tRNAXm*<sub>4</sub> are shifted in frame. The X motif G,GGC,GGC,GGC, GCC is observed.

### 3.2.4. X circular code motifs in Asp-tRNAs (Table 4d)

- (i) 5' regions of Asp-tRNAs: The X motif *Asp-tRNAXm*<sub>1</sub> G,GGT,GGT, GTA,GCC,C is observed. The large X motif *Asp-tRNAXm*<sub>2</sub> GT,GGT,

**Table 4a**Identification of X circular code motifs *Ala-tRNAXm* in tRNAs of alanine (Ala) in prokaryotes (bacteria *E. coli* and *T. thermophilus*, archaea *P. furiosus*).

AC Ala	AC pos	5' region	AC	3' region	Organism ID	Start	End	Size
TGC	34	T, CAG, CTG, GG			<i>E. coli C08004469</i>	12	20	9
TGC	34	T, CAG, CTG, GG			<i>E. coli C08004522</i>	12	20	9
GGC	34	T, CAG, CTG, GG			<i>E. coli C08004535</i>	12	20	9
TGC	36	GC, CTG, GTA, T			<i>P. furiosus At1825</i>	15	23	9
GGC	36	GC, CTG, GTA, T			<i>P. furiosus At1834</i>	15	23	9
CGC	35	GC, CTG, GAA, GAG, C			<i>P. furiosus At1833</i>	15	26	12
TGC	36	GC, GCC, GCC, CT			<i>P. furiosus At1825</i>	26	35	10
GGC	36	GC, GCC, GCC, TT			<i>P. furiosus At1834</i>	26	35	10
CGC	35	GC, GCC, GCC, CT	<b>C, GC</b>		<i>P. furiosus At1833</i>	25	37	13
GGC	34	GC, CTC, AAT,	<b>GGC,</b>	ATT, GAG, GAG, GTC, A	<i>T. thermophilus C025943</i>	26	49	24
CGC	34	GC, CTG, AAT,	<b>C</b>		<i>T. thermophilus C025964</i>	26	34	9
TGC	34			G, CAG, GAG, GTC, T	<i>E. coli C08004469</i>	39	49	11
TGC	34			G, CAG, GAG, GTC, T	<i>E. coli C08004522</i>	39	49	11
CGC	34			T, CAG, GAG, GTC, A	<i>T. thermophilus C025964</i>	39	49	11
GGC	34			AA, GAG, GTC, A	<i>E. coli C08004535</i>	41	49	9
CGC	35			G, GAG, GCC, GC	<i>P. furiosus At1833</i>	43	51	9
GGC	34			A, GGT, CAG, GG	<i>T. thermophilus C025943</i>	44	52	9
CGC	34			A, GGT, CAG, GG	<i>T. thermophilus C025964</i>	44	52	9
GGC	34			CC, CTC, GGC, T	<i>T. thermophilus C025943</i>	62	70	9
CGC	34			CC, CTC, GGC, T	<i>T. thermophilus C025964</i>	62	70	9

**Table 4b**Identification of X circular code motifs *Arg-tRNAXm* in tRNAs of arginine (Arg) in prokaryotes (bacteria *E. coli* and *T. thermophilus*, archaea *P. furiosus*).

AC Arg	AC pos	5' region	AC	3' region	Organism ID	Start	End	Size
TGC	36	G, GCC, GGT, GGC, CT			<i>P. furiosus At1842</i>	2	13	12
GGC	36	CC, GGT, GGC, CT			<i>P. furiosus At1850</i>	4	13	10
ACG	35	CC, GTA, GTT, CAG, CTG, GAT, A			<i>E. coli C08004532</i>	5	22	18
TCT	35	CC, GTA, GCC, TA			<i>P. furiosus At1826</i>	5	14	10
CCT	33	G, GTA, GCC, TA			<i>P. furiosus At1852</i>	6	14	9
CCG	36	G, GTA, GTT, TA			<i>P. furiosus At1832</i>	6	14	9
CCG	35	T, CAG, CTG, GAT, A			<i>E. coli C08004500</i>	12	22	11
ACG	35	T, CAG, CTG, GAT, A			<i>E. coli C08004529</i>	12	22	11
ACG	35	T, CAG, CTG, GAT, A			<i>T. thermophilus C025939</i>	12	22	11
CCG	35	T, CAG, CTG, GAT, A			<i>T. thermophilus C025957</i>	12	22	11
CCT	35	T, CAG, CAG, GAT, A			<i>T. thermophilus C025920</i>	12	22	11
TCT	35	AG, CAG, GAT, A			<i>P. furiosus At1826</i>	14	22	9
CCG	36	GC, CAG, GAG, A			<i>P. furiosus At1832</i>	15	23	9
CCT	33	GC, CAG, GAT, A			<i>P. furiosus At1852</i>	15	23	9
TCG	36	GC, CTG, GAT, GG			<i>P. furiosus At1842</i>	15	24	10
GGC	36	GC, CTG, GAT, A			<i>P. furiosus At1850</i>	15	23	9
CCT	31	AT, AAC, GAG, C			<i>E. coli C08004487</i>	21	29	9
ACG	35	TA, GAG, TAC, T			<i>E. coli C08004529</i>	21	29	9
ACG	35	TA, GAG, TAC, T			<i>E. coli C08004532</i>	21	29	9
TCT	35	AG, GGC, GCC, GGC, CT			<i>P. furiosus At1826</i>	22	34	13
TCG	36	G, GGC, GTC, GGC, CT			<i>P. furiosus At1842</i>	24	35	12
TCT	35	GC, AAC, GAC, CT			<i>E. coli C08004473</i>	25	34	10
CCG	36	A, GAG, AAC, GCC, GCC, CT	<b>C, C</b>		<i>P. furiosus At1832</i>	21	37	17
CCG	35	GC, GTC, GGC, CT	<b>C, C</b>		<i>T. thermophilus C025957</i>	25	36	12
CCT	36	AG, GGC, GGC, GGC, CT	<b>C, CT</b>		<i>P. furiosus At1852</i>	23	38	16
CCT	35	G, GGC, TT	<b>C, CT</b>		<i>T. thermophilus C025920</i>	29	37	9
TCT	34	A, GCC, GCC, T	<b>TC, T</b>	A	<i>T. thermophilus C025953</i>	26	37	12
TCT	35	G, GCC, T	<b>TC, T</b>	A	<i>P. furiosus At1826</i>	30	38	9
TCT	35	G, ACC, T	<b>TC, T</b>	A	<i>E. coli C08004473</i>	30	38	9
TCG	36	G, GCC, T	<b>TC, G</b>	A	<i>P. furiosus At1842</i>	31	39	9
ACG	35	A, GTA, CTC, GGC, T	<b>AC, G</b>	AA, C	<i>E. coli C08004529</i>	24	40	17
ACG	35	A, GTA, CTC, GGC, T	<b>AC, G</b>	AA, C	<i>E. coli C08004532</i>	24	40	17
ACG	35	T, GAC, T	<b>AC, G</b>	G	<i>T. thermophilus C025939</i>	30	38	9
GGC	36		<b>G,</b>	GAC, CTC, GAG, GTC, C	<i>P. furiosus At1850</i>	38	51	14
ACG	35		<b>G,</b>	GAT, CAG, CAG, GTC, GG	<i>T. thermophilus C025939</i>	37	51	15
CCT	31			TG, CAG, GTT, C	<i>E. coli C08004487</i>	47	55	9
CCG	35			A, GCC, GAA, GGT, CAG, A	<i>T. thermophilus C025957</i>	39	52	14
TCG	36			A, GCC, GAA, GGT, C	<i>P. furiosus At1842</i>	40	50	11
CCT	31			A, TTC, CTG, CAG, GG	<i>E. coli C08004487</i>	57	68	12
TCT	35			A, ATC, CTG, CAG, GGC, GC	<i>E. coli C08004473</i>	59	73	15

**Table 4c**Identification of X circular code motifs *Asn-tRNAXm* in tRNAs of asparagine (Asn) in prokaryotes (bacteria *E. coli* and *T. thermophilus*, archaea *P. furiosus*).

AC Asn	AC pos	5' region	AC	3' region	Organism ID	Start	End	Size
GTT	34	GCC, GCC, GTA, GC			<i>P. furiosus At1856</i>	1	11	11
GTT	34	T, GTA, GTT, CAG, T			<i>E. coli C08004483</i>	6	16	11
GTT	34	T, GTA, GTT, CAG, T			<i>E. coli C08004537</i>	6	16	11
GTT	34	T, CAG, CAG, GTA, GAG, CAG, C			<i>T. thermophilus C025962</i>	12	28	17
GTT	34	AG, AAC, GGC, GG			<i>E. coli C08004483</i>	21	30	10
GTT	34	AG, AAC, GGC, GG			<i>E. coli C08004537</i>	21	30	10
GTT	34		<b>T</b>	A, ACC, GGT, A	<i>T. thermophilus C025962</i>	36	44	9
GTT	34			CC, GGC, GGT, C	<i>P. furiosus At1856</i>	40	48	9
GTT	34			G, GGC, GGC, GGC, GCC	<i>P. furiosus At1856</i>	63	75	13

**Table 4d**Identification of X circular code motifs *Asp-tRNA<sup>Asp</sup>* in tRNAs of aspartic acid (Asp) in prokaryotes (bacteria *E. coli* and *T. thermophilus*, archaea *P. furiosus*).

AC Asp	AC pos	5' region	AC	3' region	Organism	Start	End	Size
GTC	37	G, GGT, GGT, GTA, GCC, C			<i>P. furiosus At1869</i>	5	18	14
GTC	35	GT, GGT, GTA, GTT, GGT, TA			<i>T. thermophilus C025932</i>	7	22	16
GTC	35	G, GTA, GTT, CAG, T			<i>E. coli C08004470</i>	6	16	11
GTC	35	A, GTC, GGT, TA			<i>E. coli C08004470</i>	14	22	9
GTC	35	TG, GTT, AAC, AC			<i>T. thermophilus C025932</i>	17	26	10
GTC	35	TA, GAA, TAC, CTG, C			<i>E. coli C08004470</i>	21	32	12
GTC	35	A, GAA, TAC, CT			<i>E. coli C08004470</i>	22	30	9
GTC	35			G, GAG, ATC, GC	<i>T. thermophilus C025932</i>	43	51	9

**Table 4e**Identification of X circular code motifs *Cys-tRNA<sup>Cys</sup>* in tRNAs of cysteine (Cys) in prokaryotes (bacteria *E. coli* and *T. thermophilus*, archaea *P. furiosus*).

AC Cys	AC pos	5' region	AC	3' region	Organism	Start	End	Size
GCA	33	GGC, GCC, GTA, GCC, AA			<i>T. thermophilus C025925</i>	1	14	14
GCA	33	GC, GTT, AAC, AA			<i>E. coli C08004540</i>	4	13	10
GCA	33	TA, GAG, GCC, A			<i>P. furiosus At1859</i>	13	21	9
GCA	33	A, GGC, CAG, GC			<i>P. furiosus At1859</i>	16	24	9
GCA	33	A, GGT, CT	<b>G, CA</b>		<i>T. thermophilus C025925</i>	27	35	9
GCA	33	A, CT	<b>G, CA</b>	G, ATC, C	<i>P. furiosus At1859</i>	30	40	11
GCA	33		<b>A</b>	A, AAC, CTC, CA	<i>T. thermophilus C025925</i>	35	44	10
GCA	33			A, TTC, GCC, GGT, T	<i>T. thermophilus C025925</i>	44	54	11
GCA	33			G, GCC, GGC, GCC, T	<i>T. thermophilus C025925</i>	62	72	11

**Table 4f**Identification of X circular code motifs *Gln-tRNA<sup>Gln</sup>* in tRNAs of glutamine (Gln) in prokaryotes (bacteria *E. coli* and *T. thermophilus*, archaea *P. furiosus*).

AC Gln	AC pos	5' region	AC	3' region	Organism ID	Start	End	Size
TTG	33	G, GGT, ATC, GCC, AA			<i>E. coli C08004549</i>	3	14	12
CTG	33	G, GGT, ATC, GCC, AA			<i>E. coli C08004552</i>	3	14	12
CTG	33	G, GGT, GTC, GTC, TA			<i>T. thermophilus C025940</i>	3	14	12
TTG	33	G, GGC, GTC, GTC, TA			<i>T. thermophilus C025948</i>	3	14	12
TTG	35	GT, GGT, GTA, GC			<i>P. furiosus At1827</i>	7	16	10
CTG	35	GT, GGT, GTA, GC			<i>P. furiosus At1849</i>	7	16	10
TTG	33	AA, GGC, ACC, GGT, TT			<i>E. coli C08004549</i>	20	32	13
CTG	33	AA, GGC, ACC, GG			<i>E. coli C08004552</i>	20	29	10
CTG	33	G, ATT,	<b>CTG,</b>	ATT, C	<i>E. coli C08004552</i>	29	39	11
TTG	33		<b>TT, G</b>	AT, ACC, GGC, ATT, C	<i>E. coli C08004549</i>	33	47	15
CTG	33			CC, GGC, ATT, C	<i>E. coli C08004552</i>	39	47	9
TTG	33			CC, GCC, GGT, GGT, T	<i>T. thermophilus C025948</i>	39	53	15
CTG	33			CC, GCC, GGT, C	<i>T. thermophilus C025940</i>	39	47	9
TTG	33			CC, CTG, GTT, C	<i>E. coli C08004549</i>	47	55	9
CTG	33			CC, GAG, GTT, C	<i>E. coli C08004552</i>	47	55	9
TTG	33			A, ATC, CAG, GTA, C	<i>E. coli C08004549</i>	57	67	11
CTG	33			A, ATC, CTC, GTA, C	<i>E. coli C08004552</i>	57	67	11

**Table 4g**Identification of X circular code motifs *Glu-tRNA<sup>Glu</sup>* in tRNAs of glutamic acid (Glu) in prokaryotes (bacteria *E. coli* and *T. thermophilus*, archaea *P. furiosus*).

AC Glu	AC pos	5' region	AC	3' region	Organism ID	Start	End	Size
CTC	35	CC, GGT, GGT, GTA, GCC, C			<i>P. furiosus At1830</i>	4	18	15
TTC	37	CC, GGT, GGT, GTA, GCC, C			<i>P. furiosus At1847</i>	4	18	15
TTC	35	CC, TTC, GTC, TA			<i>E. coli C08004497</i>	5	14	10
TTC	35	CC, TTC, GTC, TA			<i>E. coli C08004534</i>	5	14	10
TTC	34	CC, ATC, GAC, TA			<i>T. thermophilus C025941</i>	5	14	10
CTC	32	CC, ATC, GTC, TA			<i>T. thermophilus C025923</i>	5	14	10
CTC	33	CC, ATC, GTC, TA			<i>T. thermophilus C025942</i>	5	14	10
TTC	35	TA, GAG, GCC, CAG, GAC, ACC, GCC, CT			<i>E. coli C08004497</i>	13	34	22
TTC	35	TA, GAG, GCC, CAG, GAC, ACC, GCC, CT			<i>E. coli C08004534</i>	13	34	22
CTC	33	TA, GAG, GCC, TA			<i>T. thermophilus C025942</i>	13	22	10
TTC	34	AG, GTC, ACC, GGC, CT			<i>T. thermophilus C025941</i>	21	33	13
TTC	34			AA, GCC, GGC, GGC, GG	<i>T. thermophilus C025941</i>	37	49	13
TTC	35			AC, GGC, GGT, AAC, A	<i>E. coli C08004497</i>	38	49	12
TTC	35			AC, GGC, GGT, AAC, A	<i>E. coli C08004534</i>	38	49	12
CTC	32			AG, GCC, GAA, AC	<i>T. thermophilus C025923</i>	38	47	10
CTC	33			AG, GCC, GAG, AC	<i>T. thermophilus C025942</i>	39	48	10

**Table 4h**Identification of X circular code motifs Gly-tRNA<sup>Xm</sup> in tRNAs of glycine (Gly) in prokaryotes (bacteria *E. coli* and *T. thermophilus*, archaea *P. furiosus*).

AC Gly	AC pos	5' region	AC	3' region	Organism ID	Start	End	Size
CCC	37	GC, GGT, GGT, A			<i>P. furiosus</i> At1837	1	9	9
TCC	37	GC, GGT, GGT, A			<i>P. furiosus</i> At1844	1	9	9
GCC	37	GC, GGT, GGT, A			<i>P. furiosus</i> At1851	1	9	9
TCC	34	G, GGC, ATC, GTA, TA			<i>E. coli</i> C08004509	3	14	12
CCC	33	G, GGC, GTA, GTT, CA			<i>E. coli</i> C08004527	3	14	12
CCC	37	TG, GTA, GTC, TA			<i>P. furiosus</i> At1837	5	14	10
TCC	37	TG, GTA, GTC, TA			<i>P. furiosus</i> At1844	5	14	10
GCC	37	TG, GTA, GTC, TA			<i>P. furiosus</i> At1851	5	14	10
CCC	33	AG, TTC, AAT, GGT, A			<i>E. coli</i> C08004527	9	20	12
TCC	34	AT, AAT, GGC, TA			<i>E. coli</i> C08004509	12	21	10
CCC	37	GC, CTG, GTC, TA			<i>P. furiosus</i> At1837	15	24	10
TCC	37	GC, CTG, GTC, TA			<i>P. furiosus</i> At1844	15	24	10
GCC	37	GC, CTG, GTC, TA			<i>P. furiosus</i> At1851	15	24	10
CCC	33	TG, GTA, GAA, C			<i>E. coli</i> C08004527	16	24	9
GCC	34	TG, GTA, GAG, CA			<i>E. coli</i> C08004512	17	26	10
GCC	34	TG, GTA, GAG, CA			<i>E. coli</i> C08004539	17	26	10
CCC	34	TG, GTA, GAG, CA			<i>T. thermophilus</i> C025960	17	26	10
TCC	35	TG, GTA, GAG, CA			<i>T. thermophilus</i> C025952	18	27	10
GCC	34	G, GTA, GAG, CA			<i>T. thermophilus</i> C025918	18	26	9
CCC	33	AG, AAC, GAG, A			<i>E. coli</i> C08004527	20	28	9
TCC	34	AT, TAC, CTC, A			<i>E. coli</i> C08004509	21	29	9
TCC	37	AG, GAC, GCC, GGC, CT			<i>P. furiosus</i> At1844	24	36	13
TCC	35	GC, ATC, GGC, CT			<i>T. thermophilus</i> C025952	25	34	10
CCC	37	AG, GAC, GCC, GGC, CT	C, C		<i>P. furiosus</i> At1837	24	38	15
CCC	34	GC, ATC, GGC, TT	C, C		<i>T. thermophilus</i> C025960	24	35	12
GCC	37	AG, GAC, GCC, ACC, CT	G, C		<i>P. furiosus</i> At1851	24	38	15
TCC	34	A, GCC, T	TC, C	A	<i>E. coli</i> C08004509	29	37	9
TCC	35	G, GCC, T	TC, C	A	<i>T. thermophilus</i> C025952	30	38	9
CCC	34	AA, GCC, GAG, GGT, C			<i>T. thermophilus</i> C025960	37	48	12
TCC	35	AA, GCC, GAG, GGT, C			<i>T. thermophilus</i> C025952	38	49	12
CCC	37	AA, GCC, GGC, GAC, C			<i>P. furiosus</i> At1837	40	51	12
TCC	37	A, GCC, GGC, GAC, C			<i>P. furiosus</i> At1844	41	51	11
TCC	34	T, GAT, GAT, GC			<i>E. coli</i> C08004509	41	49	9
CCC	33	TA, TAC, GAG, GGT, T			<i>E. coli</i> C08004527	42	53	12
GCC	37	TG, GAG, ACC, C			<i>P. furiosus</i> At1851	44	52	9
GCC	34	GC, GAG, TTC, GAG, T			<i>E. coli</i> C08004512	49	60	12
GCC	34	GC, GAG, TTC, GAG, T			<i>E. coli</i> C08004539	49	60	12
GCC	34	GT, CTC, GTT, T			<i>E. coli</i> C08004512	59	67	9
GCC	34	GT, CTC, GTT, T			<i>E. coli</i> C08004539	59	67	9
CCC	33	CC, TTC, GCC, C			<i>E. coli</i> C08004527	60	68	9
CCC	37	G, GCC, ACC, GC			<i>P. furiosus</i> At1837	66	74	9
TCC	37	G, GCC, ACC, GC			<i>P. furiosus</i> At1844	66	74	9
GCC	37	G, GCC, ACC, GC			<i>P. furiosus</i> At1851	66	74	9

**Table 4i**Identification of X circular code motifs His-tRNA<sup>Xm</sup> in tRNAs of histidine (His) in prokaryotes (bacteria *E. coli* and *T. thermophilus*, archaea *P. furiosus*).

AC His	AC pos	5' region	AC	3' region	Organism ID	Start	End	Size
GTG	36	G, GGT, GGT, GTA, GCC, T			<i>P. furiosus</i> At1843	5	18	14
GTG	35	T, CAG, CTG, GTT, A			<i>T. thermophilus</i> C025927	12	22	11
GTG	36	GC, CTG, GTT, A			<i>P. furiosus</i> At1843	15	23	9
GTG	34	TG, GTA, GAG, C			<i>E. coli</i> C08004501	17	25	9
GTG	34	A, GCC, CTG, GAT, T			<i>E. coli</i> C08004501	23	33	11
GTG	34		A, TTC, CAG, TT		<i>E. coli</i> C08004501	37	45	9
GTG	34		A, GTT, GTC, GT		<i>E. coli</i> C08004501	42	50	9
GTG	36		CC, CTG, GCC, C		<i>P. furiosus</i> At1843	43	51	9

**Table 4j**Identification of X circular code motifs Ile-tRNA<sup>Xm</sup> in tRNAs of isoleucine (Ile) in prokaryotes (bacteria *E. coli* and *T. thermophilus*, archaea *P. furiosus*).

AC Ile	AC pos	5' region	AC	3' region	Organism ID	Start	End	Size
GAT	35	G, GGC, GAT, TA			<i>T. thermophilus</i> C025963	1	9	9
GAT	35	T, CAG, CTG, GTT, A			<i>T. thermophilus</i> C025963	12	22	11
GAT	36	GC, CTG, GTC, A			<i>P. furiosus</i> At1831	15	23	9
GAT	35	A, GGT, GGT, TA			<i>E. coli</i> C08004468	14	22	9
GAT	35	A, GGT, GGT, TA			<i>E. coli</i> C08004521	14	22	9
GAT	35		AG, GGT, GAG, GTC, GGT, GGT, T		<i>E. coli</i> C08004468	39	56	18
GAT	35		AG, GGT, GAG, GTC, GGT, GGT, T		<i>E. coli</i> C08004521	39	56	18
GAT	35		GT, GAG, GTC, GGT, GGT, T		<i>T. thermophilus</i> C025963	42	56	15
GAT	36		G, TTC, GAA, GCC, C		<i>P. furiosus</i> At1831	55	65	11
GAT	35		CC, ACC, ATC, GCC, CA		<i>T. thermophilus</i> C025963	62	74	13
GAT	35		T, CAG, GCC, TAC		<i>E. coli</i> C08004468	66	75	10

**Table 4k**Identification of X circular code motifs *Leu-tRNA<sub>Xm</sub>* in tRNAs of leucine (Leu) in prokaryotes (bacteria *E. coli* and *T. thermophilus*, archaea *P. furiosus*).

AC Leu	AC pos	5' region	AC	3' region	Organism ID	Start	End	Size
CAG	35	GC, GAA, GGT, GGC, GG			<i>E. coli</i> C08004502	1	13	13
CAG	35	GC, GAA, GGT, GGC, GG			<i>E. coli</i> C08004516	1	13	13
CAG	35	GC, GAA, GGT, GGC, GG			<i>E. coli</i> C08004517	1	13	13
TAG	35	GC, GAG, GAT, GGC, GG			<i>T. thermophilus</i> C025928	1	13	13
TAA	34	G, GGT, GGC, GG			<i>T. thermophilus</i> C025929	5	13	9
CAG	35	G, GGT, GGC, GG			<i>T. thermophilus</i> C025947	5	13	9
CAA	35	G, GGT, GGC, GG			<i>T. thermophilus</i> C025954	5	13	9
GAG	35	A, GGT, GGT, GG			<i>E. coli</i> C08004524	5	13	9
GAG	35	G, GGT, GGT, GG			<i>T. thermophilus</i> C025921	5	13	9
TAA	35	G, GAT, GGT, GG			<i>E. coli</i> C08004541	5	13	9
CAG	37	G, GTT, GCC, GAG, C			<i>P. furiosus</i> At1867	6	16	11
TAG	37	G, GTT, GCC, GAG, C			<i>P. furiosus</i> At1838	6	16	11
CAA	37	G, GTT, GCC, GAG, C			<i>P. furiosus</i> At1848	6	16	11
GAG	37	G, GTT, GCC, GAG, C			<i>P. furiosus</i> At1853	6	16	11
CAA	35	GT, GGC, GAA, ATC, GGT, A			<i>E. coli</i> C08004515	7	21	15
TAG	35	GT, GGC, GAA, ATT, GGT, A			<i>E. coli</i> C08004548	7	21	15
GAG	35	TG, GAA, CTG, GTA, GAC, AC			<i>T. thermophilus</i> C025921	11	26	16
CAG	37	GC, CTG, GTC, AA			<i>P. furiosus</i> At1867	15	24	10
TAG	37	GC, CTG, GTC, AA			<i>P. furiosus</i> At1838	15	24	10
CAA	37	GC, CTG, GTC, AA			<i>P. furiosus</i> At1848	15	24	10
GAG	37	GC, CTG, GTC, AA			<i>P. furiosus</i> At1853	15	24	10
TAA	37	GC, CTG, GCC, AA			<i>P. furiosus</i> At1862	15	24	10
CAG	35	TG, GTA, GAC, GC			<i>E. coli</i> C08004502	17	26	10
CAG	35	TG, GTA, GAC, GC			<i>E. coli</i> C08004516	17	26	10
CAG	35	TG, GTA, GAC, GC			<i>E. coli</i> C08004517	17	26	10
GAG	35	TG, GTA, GAC, AC			<i>E. coli</i> C08004524	17	26	10
TAG	35	TG, GTA, GAC, GC			<i>E. coli</i> C08004548	17	26	10
CAG	35	TG, GTA, GAC, GC			<i>T. thermophilus</i> C025947	17	26	10
TAA	34	G, GTA, GAC, GC			<i>T. thermophilus</i> C025929	17	25	9
CAA	35	G, GTA, GAC, GC			<i>E. coli</i> C08004515	18	26	9
TAA	35	G, GTA, GAC, AC			<i>E. coli</i> C08004541	18	26	9
TAG	35	G, GTA, GAC, GC			<i>T. thermophilus</i> C025928	18	26	9
CAA	35	G, GTA, GAC, GC			<i>T. thermophilus</i> C025954	18	26	9
GAG	35	AC, GCC, ATC, TT			<i>T. thermophilus</i> C025921	25	34	10
TAG	35	AC, CAG, ATT, <b>TA</b>			<i>E. coli</i> C08004548	27	36	10
CAG	35	TG, ATT, <b>CAG</b> , GGT, CA			<i>T. thermophilus</i> C025947	30	42	13
CAG	37	G, ATT, <b>CAG</b> , GGT, C			<i>P. furiosus</i> At1867	33	43	11
GAG	37	G, ATT, <b>GAG</b> , GGT, C			<i>P. furiosus</i> At1853	33	43	11
TAG	35	<b>AG</b> , GTT, CTG, GC			<i>E. coli</i> C08004548	36	45	10
CAA	35	<b>AA</b> , AAT, CTG, CTG, T			<i>T. thermophilus</i> C025954	36	47	12
CAA	35	<b>AA</b> , ATC, AAC, C			<i>E. coli</i> C08004515	37	45	9
CAA	35	A, ACC, GTA, GAA, AT			<i>E. coli</i> C08004515	42	53	12
TAA	35	CC, CTC, GGC, GTT, C			<i>E. coli</i> C08004541	41	52	12
TAG	35	T, GGC, GCC, GC			<i>E. coli</i> C08004548	42	50	9
GAG	35	G, GGT, GGT, GCC, C			<i>T. thermophilus</i> C025921	39	49	11
TAA	37	CC, GGT, GCC, GTA, GG			<i>P. furiosus</i> At1862	43	55	13
CAA	35	GT, GCC, GGT, T			<i>E. coli</i> C08004515	56	64	9
TAG	35	GC, GAG, TTC, AA			<i>E. coli</i> C08004548	58	67	10
TAG	35	GT, CTC, GCC, T			<i>E. coli</i> C08004548	68	76	9
CAA	35	G, GCC, TTC, GGC, ACC			<i>E. coli</i> C08004515	72	84	13
TAG	35	G, CTC, CTC, GC			<i>T. thermophilus</i> C025928	73	81	9

**Table 4l**Identification of X circular code motifs *Lys-tRNA<sub>Xm</sub>* in tRNAs of lysine (Lys) in prokaryotes (bacteria *E. coli* and *T. thermophilus*, archaea *P. furiosus*).

AC Lys	AC pos	5' region	AC	3' region	Organism ID	Start	End	Size
CTT	34	AG, CTC, AAC, T			<i>T. thermophilus</i> C025946	9	17	9
TTT	33	T, CAG, CTG, GC			<i>T. thermophilus</i> C025922	12	20	9
TTT	35	GC, CTG, GTT, A			<i>P. furiosus</i> At1828	15	23	9
CTT	36	GC, CTG, GTT, A			<i>P. furiosus</i> At1858	15	23	9
CTT	34	AA, CTG, GTA, GAG, CA			<i>T. thermophilus</i> C025946	14	26	13
TTT	33	TG, GTA, GAG, CAG, TT			<i>E. coli</i> C08004474	17	29	13
TTT	33	A, GTT, GAC, <b>TT</b>			<i>E. coli</i> C08004474	26	34	9
TTT	33	A, ACC, GAC, <b>TT</b>			<i>T. thermophilus</i> C025922	26	34	9
CTT	34	<b>T</b> A, ATC, GGT, GG			<i>T. thermophilus</i> C025946	36	45	10
TTT	33	TA, ATC, GGT, A			<i>T. thermophilus</i> C025922	36	44	9
TTT	33	TA, ATC, AAT, T			<i>E. coli</i> C08004474	36	44	9
CTT	34	TG, GGT, TAC, A			<i>T. thermophilus</i> C025946	43	51	9
TTT	35	CC, GGT, GGT, C			<i>P. furiosus</i> At1828	42	50	9
CTT	34	TA, CAG, GTT, C			<i>T. thermophilus</i> C025946	48	56	9
CTT	36	G, CAG, GTC, GG			<i>P. furiosus</i> At1858	44	52	9
TTT	33	A, ATC, CTG, CA			<i>E. coli</i> C08004474	58	66	9
TTT	33	A, ATC, CTG, CA			<i>T. thermophilus</i> C025922	58	66	9

GTA, GTT, GGT, TA of 16 nucleotides in *T. thermophilus* and *Asp-tRNA<sub>Xm1</sub>* are shifted in frame. The X motif *Asp-tRNA<sub>Xm3</sub>* G, GTA, GTT, CAG, T is suffix of *Asp-tRNA<sub>Xm2</sub>*. The X motif *Asp-tRNA<sub>Xm4</sub>* A, GTC, GGT, TA is shifted by +1 nucleotide from *Asp-tRNA<sub>Xm3</sub>*. The X motif TG, GTT, AAC, AC is shifted by +2 nucleotides from *Asp-tRNA<sub>Xm4</sub>*. The X motif TA, GAA, TAC, CTG, C is observed.

(ii) 3' regions of *Asp-tRNA<sub>Xm</sub>*: The X motif G, GAG, ATC, GC is observed.

### 3.2.5. X circular code motifs in *Cys-tRNA<sub>Xm</sub>* (Table 4e)

(i) 5' regions of *Cys-tRNA<sub>Xm</sub>*: The X motif *Cys-tRNA<sub>Xm1</sub>* GGC, GCC, GTA, GCC, AA is observed. The X motif GC, GTT, AAC, AA is suffix of *Cys-tRNA<sub>Xm1</sub>*. The X motif A, GGC, CAG, GC is shifted by +1 nucleotide from TA, GAG, GCC, A.

(ii) Anticodon regions of *Cys-tRNA<sub>Xm</sub>*: The X motif *Cys-tRNA<sub>Xm2</sub>* A, GGT, CTG, CA is in a different frame than the anticodon GCA. The X motif *Cys-tRNA<sub>Xm3</sub>* A, CTG, CAG, ATC, C and *Cys-tRNA<sub>Xm2</sub>* are shifted in frame, thus *Cys-tRNA<sub>Xm3</sub>* is in a different frame

**Table 4m**Identification of X circular code motifs *Met-tRNAXm* in tRNAs of methionine (Met) in prokaryotes (bacteria *E. coli* and *T. thermophilus*, archaea *P. furiosus*).

AC Met	AC pos	5' region	AC	3' region	Organism ID	Start	End	Size
CAT	35	GGC, TAC, GTA, GC			<i>E. coli</i> C08004547	1	11	11
CAT	35	GGC, GGC, GTA, GC			<i>T. thermophilus</i> C025934	1	11	11
CAT	35	TG, GAG, CAG, C			<i>E. coli</i> C08004492	8	16	9
CAT	35	TG, GAG, CAG, C			<i>E. coli</i> C08004525	8	16	9
CAT	35	TG, GAG, CAG, C			<i>T. thermophilus</i> C025937	8	16	9
CAT	34	AG, CTC, AAC, GGT, CAG, A			<i>T. thermophilus</i> C025955	9	23	15
CAT	35	A, GGT, GGT, CAG, A			<i>T. thermophilus</i> C025934	14	24	11
CAT	35	A, GTT, GGT, TA			<i>E. coli</i> C08004547	14	22	9
CAT	36	GC, CTG, GTC, AA			<i>P. furiosus</i> At1857	15	24	10
CAT	36	GC, CTG, GTC, A			<i>P. furiosus</i> At1855	15	23	9
CAT	35	GC, CTG, GTA, GC			<i>E. coli</i> C08004492	15	24	10
CAT	35	GC, CTG, GTA, GC			<i>E. coli</i> C08004525	15	24	10
CAT	35	GC, CTG, GTA, GC			<i>T. thermophilus</i> C025937	15	24	10
CAT	34	TA, GAG, CAG, GC			<i>E. coli</i> C08004495	20	29	10
CAT	34	TA, GAG, CAG, GC			<i>E. coli</i> C08004533	20	29	10
CAT	35	AG, CTC, GTC, GG			<i>E. coli</i> C08004492	22	31	10
CAT	35	AG, CTC, GTC, GG			<i>E. coli</i> C08004525	22	31	10
CAT	35	AG, CTC, GTC, GG			<i>T. thermophilus</i> C025937	22	31	10
CAT	35		<b>AT</b>	AAT, GAT, GG	<i>E. coli</i> C08004547	36	45	10
CAT	34		<b>T</b>	A, ACC, GGT, A	<i>T. thermophilus</i> C025955	36	44	9
CAT	35			CC, GAA, GGT, C	<i>E. coli</i> C08004492	41	49	9
CAT	35			CC, GAA, GGT, C	<i>T. thermophilus</i> C025937	41	49	9
CAT	35			CC, GAA, GAT, C	<i>E. coli</i> C08004525	41	49	9
CAT	35			GT, GGT, GTC, GT	<i>T. thermophilus</i> C025934	42	51	10
CAT	35			AG, GTC, GTC, GGT, T	<i>E. coli</i> C08004492	45	56	12
CAT	35			AG, ATC, GTC, GGT, T	<i>E. coli</i> C08004525	45	56	12
CAT	34			TG, CAG, GTT, C	<i>T. thermophilus</i> C025955	48	56	9
CAT	36			CC, GAG, GTT, CA	<i>P. furiosus</i> At1846	50	59	10
CAT	34			G, CTG, GTT, CA	<i>E. coli</i> C08004495	49	57	9
CAT	34			G, CTG, GTT, CA	<i>E. coli</i> C08004533	49	57	9
CAT	36			G, TTC, GAA, GCC, C	<i>P. furiosus</i> At1857	55	65	11
CAT	34			AA, GTC, CAG, CAG, GG	<i>E. coli</i> C08004495	57	69	13
CAT	34			AA, GTC, CAG, CA	<i>E. coli</i> C08004533	57	66	10
CAT	36			AA, ATC, CTC, GGC, C	<i>P. furiosus</i> At1846	59	70	12
CAT	34			A, ATC, CTC, CA	<i>T. thermophilus</i> C025955	58	66	9
CAT	35			CC, GTC, GTA, GCC	<i>E. coli</i> C08004547	63	73	11
CAT	35			CC, ACC, GCC, GCC, ACC	<i>T. thermophilus</i> C025934	63	76	14

**Table 4n**Identification of X circular code motifs *Phe-tRNAXm* in tRNAs of phenylalanine (Phe) in prokaryotes (bacteria *E. coli* and *T. thermophilus*, archaea *P. furiosus*).

AC Phe	AC pos	5' region	AC	3' region	Organism ID	Start	End	Size
GAA	34	GCC, GAG, GTA, GC			<i>T. thermophilus</i> C025933	1	11	11
GAA	34	TG, GTA, GAG, CA			<i>T. thermophilus</i> C025933	17	26	10
GAA	34	G, GTA, GAG, CAG, GG			<i>E. coli</i> C08004519	18	29	12
GAA	34		G, ATT,	<b>GAA</b> , AAT, C	<i>E. coli</i> C08004519	30	40	11
GAA	35			G, GGT, GTC, GG	<i>P. furiosus</i> At1845	43	51	9
GAA	34			GT, GTC, GGC, GGT, T	<i>T. thermophilus</i> C025933	44	55	12
GAA	34			CC, CTC, GGC, ACC	<i>T. thermophilus</i> C025933	65	75	11

**Table 4o**Identification of X circular code motifs *Pro-tRNAXm* in tRNAs of proline (Pro) in prokaryotes (bacteria *E. coli* and *T. thermophilus*, archaea *P. furiosus*).

AC Pro	AC pos	5' region	AC	3' region	Organism ID	Start	End	Size
TGG	35	C, GGC, GAG, TA			<i>E. coli</i> C08004503	1	9	9
TGG	35	G, GGC, GTA, GC			<i>T. thermophilus</i> C025951	3	11	9
CGG	37	G, GCC, GTA, GG			<i>P. furiosus</i> At1870	3	11	9
TGG	37	G, GCC, GTA, GG			<i>P. furiosus</i> At1864	3	11	9
TGG	35	G, CAG, GCC, GGT, A			<i>T. thermophilus</i> C025951	12	22	11
GGG	35	GC, CTG, GTA, GC			<i>E. coli</i> C08004486	15	24	10
CGG	35	GC, CTG, GTA, GC			<i>E. coli</i> C08004520	15	24	10
GGG	35	GC, CTG, GTA, GC			<i>T. thermophilus</i> C025919	15	24	10
CGG	37	CC, ATC, CTG, C			<i>P. furiosus</i> At1870	22	30	9
TGG	37	CC, ATC, CTG, C			<i>P. furiosus</i> At1864	22	30	9
GGG	35	GC, ACC, GTC, AT			<i>E. coli</i> C08004486	25	34	10
CGG	35	AC, TTC, GTT,	<b>C</b>		<i>E. coli</i> C08004520	27	35	9
CGG	35	AC, CTC, GTT,	<b>C</b>		<i>T. thermophilus</i> C025931	27	35	9
TGG	35	AA, CTG, GTT,	<b>T</b>		<i>E. coli</i> C08004503	27	35	9
GGG	36	CC, GGC, CT	<b>G, GG</b>		<i>P. furiosus</i> At1866	29	38	10
CGG	37	G, GGC, TT	<b>C, GG</b>		<i>P. furiosus</i> At1870	31	39	9
GGG	35		<b>G, G</b>	GT, GTC, GG	<i>E. coli</i> C08004486	36	44	9
GGG	36		<b>G, G</b>	GC, GCC, GG	<i>P. furiosus</i> At1866	37	45	9
CGG	35		<b>G,</b>	GAC, GAA, GG	<i>E. coli</i> C08004520	37	45	9
CGG	35		<b>G,</b>	GAC, GAG, GG	<i>T. thermophilus</i> C025931	37	45	9
TGG	35		<b>G,</b>	GAG, CAG, GG	<i>T. thermophilus</i> C025951	37	45	9
GGG	35			G, GTC, GTC, GGT, T	<i>T. thermophilus</i> C025919	46	56	11
GGG	35			G, GAG, GTT, CA	<i>E. coli</i> C08004486	50	58	9
CGG	35			G, CTG, GTT, CA	<i>T. thermophilus</i> C025931	50	58	9
GGG	35			AA, ATC, CTC, T	<i>E. coli</i> C08004486	58	66	9
CGG	35			AA, ATC, CAG, T	<i>T. thermophilus</i> C025931	58	66	9
CGG	35			A, ATC, CTC, TA	<i>E. coli</i> C08004520	59	67	9
GGG	35			CC, GGC, CTC, C	<i>T. thermophilus</i> C025919	62	70	9
GGG	36			CC, GGC, GGC, C	<i>P. furiosus</i> At1866	63	71	9
CGG	37			CC, GGC, GGC, C	<i>P. furiosus</i> At1870	64	72	9
TGG	37			CC, GGC, GGC, C	<i>P. furiosus</i> At1864	64	72	9
TGG	35			T, CTC, GCC, GAC	<i>E. coli</i> C08004503	66	75	10

**Table 4p**Identification of X circular code motifs *Sec-tRNA<sup>Sec</sup>* in tRNAs of selenocysteine (Sec) in bacteria of *E. coli*.

AC SeC	AC pos	5' region	AC	3' region	Organism ID	Start	End	Size
TCA	35	AG, ATC, GTC, GTC, T			<i>E. coli C08004496</i>	4	15	12
TCA	35	CC, GGT, GAG, GC			<i>E. coli C08004496</i>	16	25	10
TCA	35	G, CTG, GAC, T	<b>TC, A</b>	A	<i>E. coli C08004496</i>	27	38	12
TCA	35		<b>A</b>	A, ATC, CAG, TT	<i>E. coli C08004496</i>	37	46	10
TCA	35			TG, ATC, TTC, C	<i>E. coli C08004496</i>	83	91	9

**Table 4q**Identification of X circular code motifs *Ser-tRNA<sup>Ser</sup>* in tRNAs of serine (Ser) in prokaryotes (bacteria *E. coli* and *T. thermophilus*, archaea *P. furiosus*).

AC Ser	AC pos	5' region	AC	3' region	Organism ID	Start	End	Size
CGA	35	A, GAT, GCC, GG			<i>E. coli C08004538</i>	5	13	9
CGA	35	G, GGT, GCC, GG			<i>T. thermophilus C025935</i>	5	13	9
CGA	35	G, GTA, GCC, TA			<i>P. furiosus At1829</i>	6	14	9
GGA	35	G, GTA, GCC, TA			<i>P. furiosus At1839</i>	6	14	9
GCT	35	G, GTA, GCC, TA			<i>P. furiosus At1860</i>	6	14	9
GCT	35	TG, GCC, GAG, A			<i>E. coli C08004528</i>	8	16	9
TGA	35	TG, GCC, GAG, C			<i>E. coli C08004545</i>	8	16	9
TGA	35	TG, GCC, GAG, C			<i>T. thermophilus C025944</i>	8	16	9
GCT	35	TG, GCC, GAG, T			<i>T. thermophilus C025945</i>	8	16	9
TGA	35	GC, CTG, GTA, GG			<i>P. furiosus At1868</i>	15	24	10
GCT	35	GC, CTG, GTA, GG			<i>P. furiosus At1860</i>	15	24	10
CGA	35	G, CTG, AAC, GG			<i>E. coli C08004538</i>	18	26	9
TGA	35	G, GTT, GAA, GGC, GGC, GGT, CT			<i>T. thermophilus C025944</i>	17	34	18
TGA	35	G, GTT, GAA, GGC, ACC, GGT, CT			<i>E. coli C08004545</i>	17	34	18
CGA	35	G, GAA, GGC, GC			<i>P. furiosus At1829</i>	20	28	9
GGA	35	G, GAA, GGC, GC			<i>P. furiosus At1839</i>	20	28	9
GCT	35	T, GAA, GGC, GC			<i>E. coli C08004528</i>	20	28	9
GCT	35	TG, GTC, GAA, GGC, GGC, ACC, CT	<b>G, CT</b>		<i>T. thermophilus C025945</i>	16	37	22
GCT	35	AG, GGC, GCC, GGC, CT	<b>G, CT</b>		<i>P. furiosus At1860</i>	22	37	16
CGA	35	G, ACC, GGT, CT	<b>C, GA</b>	A, AAC, C	<i>E. coli C08004538</i>	26	42	17
CGA	35	G, GCC, GGT, CT	<b>C, GA</b>	A, AAC, C	<i>T. thermophilus C025935</i>	26	42	17
CGA	35	A, CT	<b>C, GA</b>	G, ATC, C	<i>P. furiosus At1829</i>	32	42	11
GGA	35	G, GGC, CT	<b>G, GA</b>	G, A	<i>P. furiosus At1839</i>	29	39	11
GGA	35	GC, CT	<b>G, GA</b>	A, A	<i>E. coli C08004544</i>	31	39	9
GGA	35	A, CT	<b>G, GA</b>	A, ATC, GT	<i>T. thermophilus C025950</i>	32	43	12
TGA	35	T	<b>T, GA</b>	A, AAC, C	<i>E. coli C08004545</i>	34	42	9
TGA	35	T	<b>T, GA</b>	A, AAC, C	<i>T. thermophilus C025944</i>	34	42	9
CGA	35	AA, ACC, GGT, A			<i>T. thermophilus C025935</i>	38	46	9
GCT	35	AA, GCC, GGT, GG			<i>P. furiosus At1860</i>	38	47	10
GCT	35	AA, GGT, GTT, GC			<i>T. thermophilus C025945</i>	38	47	10
TGA	35	AA, ACC, GGC, GAC, C			<i>E. coli C08004545</i>	38	49	12
GGA	35	GT, GTA, TAC, GGC, AAC, GTA, T			<i>E. coli C08004544</i>	42	59	18
CGA	35	TG, GGC, GTT, C			<i>P. furiosus At1829</i>	45	53	9
GGA	35	TG, GGC, GTT, C			<i>P. furiosus At1839</i>	45	53	9
TGA	35	TG, GGC, GTT, T			<i>P. furiosus At1868</i>	45	53	9
GGA	35	G, TTC, GCC, CA			<i>P. furiosus At1839</i>	50	58	9
TGA	35	GC, GAA, GCC, CA			<i>T. thermophilus C025944</i>	52	61	10
CGA	35	AA, CTC, TAC, C			<i>E. coli C08004538</i>	54	62	9
GCT	35	AA, ACC, GGT, GCC, GC			<i>T. thermophilus C025945</i>	55	67	13
GGA	35	TA, AAC, CTC, C			<i>T. thermophilus C025950</i>	55	63	9
CGA	35	G, GGC, CTC, GC			<i>T. thermophilus C025935</i>	58	66	9
TGA	35	A, GAG, TTC, GAA, T			<i>E. coli C08004545</i>	62	72	11
TGA	35	CC, CTC, ACC, CTC, C			<i>T. thermophilus C025944</i>	76	87	12
GCT	35	CC, GCC, CTC, T			<i>T. thermophilus C025945</i>	79	87	9
GGA	35	CC, GCC, CTC, T			<i>T. thermophilus C025950</i>	80	88	9

than the anticodon GCA. The X motifs AA, AAC, CTC, CA and *Cys-tRNA<sup>Ser</sup>*<sub>3</sub> are shifted in frame, thus they are in a different frame than the anticodon GCA.

- (iii) 3' regions of *Cys-tRNA<sup>Ser</sup>*: Two X motifs A, TTC, GCC, GGT, T and G, GCC, GGC, GCC, T are observed.

### 3.2.6. X circular code motifs in *Gln-tRNA<sup>Ser</sup>* (Table 4f)

- (i) 5' regions of *Gln-tRNA<sup>Ser</sup>*: The class of X motifs *Gln-tRNA<sup>Ser</sup>*<sub>1</sub> G, GGY, RTC, GYC, WA is identified. The X motif GT, GGT, GTA, GC is shifted by +1 nucleotide from the class of X motifs G, GGY, GTC, GTC, IA belonging to *Gln-tRNA<sup>Ser</sup>*<sub>1</sub>. The X motif AA, GGC, ACC, GGT, TT occurs before the anticodon.
- (ii) Anticodon regions of *Gln-tRNA<sup>Ser</sup>*: The X motif G, ATT, CTG, ATT, C is in frame with the anticodon CTG. The X motif *Gln-tRNA<sup>Ser</sup>*<sub>2</sub> TT, GAT, ACC, GGC, ATT, C is in a different frame than the anticodon TTG.

- (iii) 3' regions of *Gln-tRNA<sup>Ser</sup>*: The X motif CC, GCC, GGT, GGT, GT, T and *Gln-tRNA<sup>Ser</sup>*<sub>2</sub> are shifted in frame. Two classes of X motifs CC, SWG, GTT, C and A, ATC, CWS, GTA, C are observed.

### 3.2.7. X circular code motifs in *Glu-tRNA<sup>Ser</sup>* (Table 4g)

- (i) 5' regions of *Glu-tRNA<sup>Ser</sup>*: The X motif *Glu-tRNA<sup>Ser</sup>*<sub>1</sub> CC, GGT, GGT, GTA, GCC, C is observed. The class of X motifs CC, WTC, GWC, TA is shifted by +2 nucleotides from *Glu-tRNA<sup>Ser</sup>*<sub>1</sub>. A very large X motif *Glu-tRNA<sup>Ser</sup>*<sub>2</sub> TA, GAG, GCC, CAG, GAC, ACC, GCC, CT of 22 nucleotide in *E. coli* and *Glu-tRNA<sup>Ser</sup>*<sub>1</sub> are shifted in frame. The X motif TA, GAG, GCC, TA is prefix of *Glu-tRNA<sup>Ser</sup>*<sub>2</sub> and the X motif AG, GTC, ACC, GGC, CT is suffix of *Glu-tRNA<sup>Ser</sup>*<sub>2</sub>.
- (ii) 3' regions of *Glu-tRNA<sup>Ser</sup>*: The X motifs AA, GCC, GGC, GGC, GG and *Glu-tRNA<sup>Ser</sup>*<sub>3</sub> AC, GGC, GGT, AAC, A occur after (3') the anticodon TTC. The class of X motifs AG, GCC, GAR, AC is shifted by +2 nucleotides from *Glu-tRNA<sup>Ser</sup>*<sub>3</sub>.

**Table 4r**Identification of X circular code motifs *Thr-tRNAXm* in tRNAs of threonine (Thr) in prokaryotes (bacteria *E. coli* and *T. thermophilus*, archaea *P. furiosus*).

AC Thr	AC pos	5' region	AC	3' region	Organism ID	Start	End	Size	
GGT	35	CC, GGT, GGC, T			<i>P. furiosus</i> At1863	4	12	9	
CGT	33	GCC, GGT, GTA, GC			<i>T. thermophilus</i> C025936	1	11	11	
TGT	35	G, GTA, GCC, TA			<i>P. furiosus</i> At1865	6	14	9	
TGT	34	AG, CTC, AAC, C			<i>T. thermophilus</i> C025958	9	17	9	
GGT	34	T, CAG, CAG, GTA, GAG, CA			<i>T. thermophilus</i> C025961	12	26	15	
GGT	35	GC, CTG, GTA, GAG, C			<i>P. furiosus</i> At1863	15	26	12	
CGT	34	TG, GTA, GAG, CAG, C			<i>E. coli</i> C08004472	17	28	12	
GGT	34	TG, GTA, GAG, C			<i>E. coli</i> C08004510	17	25	9	
GGT	34	TG, GTA, GAG, C			<i>E. coli</i> C08004523	17	25	9	
TGT	34	AG, GTA, GAG, CA			<i>E. coli</i> C08004507	17	26	10	
CGT	33	G, GTA, GAG, CA			<i>T. thermophilus</i> C025936	17	25	9	
CGT	33		GC, CT	<b>C, GT</b> A, A	<i>T. thermophilus</i> C025936	29	37	9	
CGT	34		A, TT	<b>C, GT</b> A, AT	<i>E. coli</i> C08004472	31	39	9	
CGT	35		A, CT	<b>C, GT</b> A, ATC, C	<i>P. furiosus</i> At1861	32	42	11	
TGT	34		T	<b>T, GT</b> A, ATC, A	<i>E. coli</i> C08004507	33	41	9	
TGT	35		T	<b>T, GT</b> A, ATC, C	<i>P. furiosus</i> At1865	34	42	9	
TGT	34		T	<b>T, GT</b> A, ATC, GG	<i>T. thermophilus</i> C025958	33	42	10	
TGT	34		<b>GT</b>	AAT, CAG, TA	<i>E. coli</i> C08004507	35	44	10	
GGT	34			AG, GGT, GAG, GTC, GCC, GGT, T	<i>T. thermophilus</i> C025961	38	55	18	
GGT	34			AG, GGT, GAG, GTC, GGC, A	<i>E. coli</i> C08004510	38	52	15	
GGT	34			AG, GGT, GAG, GTC, C	<i>E. coli</i> C08004523	38	49	12	
CGT	34			GC, GAA, GGT, C	<i>E. coli</i> C08004472	40	48	9	
TGT	34			A, GTA, GGT, CA	<i>E. coli</i> C08004507	41	49	9	
CGT	35			CC, CAG, GTC, C	<i>P. furiosus</i> At1861	42	50	9	
TGT	35			CC, CAG, GTC, GC	<i>P. furiosus</i> At1865	42	51	10	
TGT	34			AG, GTC, ACC, A	<i>E. coli</i> C08004507	44	52	9	
CGT	34			AG, GTC, GTA, GGT, T	<i>E. coli</i> C08004472	44	55	12	
CGT	33			AG, GCC, GCC, GGT, T	<i>T. thermophilus</i> C025936	43	54	12	
TGT	34			AC, CAG, TTC, GAT, T	<i>E. coli</i> C08004507	49	60	12	
GGT	34			CC, CAG, TTC, GAC, T	<i>E. coli</i> C08004523	49	60	12	
GGT	34			G, CAG, TTC, GAA, T	<i>E. coli</i> C08004510	50	60	11	
TGT	34				CC, CTG, GGT, GGC, T	<i>T. thermophilus</i> C025958	62	73	12
GGT	34				T, CTG, GGT, ATC, A	<i>E. coli</i> C08004523	60	70	11
CGT	33				G, GCC, ACC, GGC, T	<i>T. thermophilus</i> C025936	62	72	11
CGT	34				T, ATT, ATC, GGC, ACC	<i>E. coli</i> C08004472	63	75	13
TGT	34				G, GTA, GTC, GGC, ACC	<i>E. coli</i> C08004507	63	75	13

**Table 4s**Identification of X circular code motifs *Trp-tRNAXm* in tRNAs of tryptophan (Trp) in prokaryotes (bacteria *E. coli* and *T. thermophilus*, archaea *P. furiosus*).

AC Trp	AC pos	5' region	AC	3' region	Organism ID	Start	End	Size
CCA	34	G, GGC, GTA, GTT, CA			<i>E. coli</i> C08004499	3	14	12
CCA	37	GT, GGT, GTA, GCC, T			<i>P. furiosus</i> At1854	7	18	12
CCA	34	AG, TTC, AAT, T			<i>E. coli</i> C08004499	9	17	9
CCA	34	AG, CTC, AAC, T			<i>T. thermophilus</i> C025938	9	17	9
CCA	37	GC, CTG, GTC, CA			<i>P. furiosus</i> At1854	15	24	10
CCA	34	TG, GTA, GAG, CA			<i>E. coli</i> C08004499	17	26	10
CCA	37	CC, ATC, ATC, GC			<i>P. furiosus</i> At1854	22	31	10
CCA	34	GC, ACC, GGT, CT	<b>C, CA</b>		<i>E. coli</i> C08004499	24	36	13
CCA	34	GC, ACC, GGT, CT	<b>C, CA</b>		<i>T. thermophilus</i> C025938	24	36	13
CCA	37	G, CT	<b>C, CA</b>	G, ACC, C	<i>P. furiosus</i> At1854	34	44	11
CCA	34			G, GGT, GTT, GG	<i>E. coli</i> C08004499	42	50	9
CCA	34			TG, GAG, GTT, C	<i>T. thermophilus</i> C025938	48	56	9
CCA	34			G, GAG, TTC, GAG, T	<i>E. coli</i> C08004499	50	60	11

**Table 4t**Identification of X circular code motifs *Tyr-tRNAXm* in tRNAs of tyrosine (Tyr) in prokaryotes (bacteria *E. coli* and *T. thermophilus*, archaea *P. furiosus*).

AC Tyr	AC pos	5' region	AC	3' region	Organism ID	Start	End	Size
GTA	36	G, GTA, GCC, TA			<i>P. furiosus</i> At1836	6	14	9
GTA	36	GC, CTG, GTA, GT			<i>P. furiosus</i> At1836	15	24	10
GTA	36	GT, GGC, GGC, GG			<i>P. furiosus</i> At1836	23	32	10
GTA	35	G, GAG, CAG, AC			<i>E. coli</i> C08004508	25	33	9
GTA	35	G, GAG, CAG, AC			<i>E. coli</i> C08004542	25	33	9
GTA	35	G, GAC, GGT, CT	<b>G, TA</b>		<i>T. thermophilus</i> C025959	26	37	12
GTA	35	T, <b>GTA</b> ,	AAT, CTG, C		<i>E. coli</i> C08004508	34	44	11
GTA	35	T, <b>GTA</b> ,	AAT, CTG, C		<i>E. coli</i> C08004542	34	44	11
GTA	35		AA, ACC, GTT, GGC, GTA, T		<i>T. thermophilus</i> C025959	38	52	15
GTA	35		T, GCC, GTC, ATC, GAC, TTC, GAA, GGT, T		<i>E. coli</i> C08004542	42	64	23
GTA	35		T, GCC, GTC, AC		<i>E. coli</i> C08004508	42	50	9
GTA	35		A, GAC, TTC, GAA, GGT, T		<i>E. coli</i> C08004508	51	64	14
GTA	35		AT, GCC, TTC, GC		<i>T. thermophilus</i> C025959	51	60	10

**Table 4u**Identification of X circular code motifs Val-tRNA<sub>Xm</sub> in tRNAs of valine (Val) in prokaryotes (bacteria *E. coli* and *T. thermophilus*, archaea *P. furiosus*).

AC Val	AC pos	5' region	AC	3' region	Organism ID	Start	End	Size
TAC	34	G, GGT, GAT, TA			<i>E. coli</i> C08004475	1	9	9
TAC	35	G, GGC, GGC, TA			<i>T. thermophilus</i> C025924	1	9	9
CAC	33	G, GGC, GGC, TA			<i>T. thermophilus</i> C025949	1	9	9
TAC	34	T, CAG, CTG, GG			<i>E. coli</i> C08004475	12	20	9
TAC	35	T, CAG, CTG, GCC, A			<i>T. thermophilus</i> C025924	12	22	11
TAC	35	AG, CTG, GTT, AT			<i>P. furiosus</i> At1841	14	23	10
GAC	35	A, GTT, GGT, TA			<i>E. coli</i> C08004481	14	22	9
GAC	35	A, GTT, GGT, TA			<i>E. coli</i> C08004482	14	22	9
GAC	35	A, GGT, GGC, TA			<i>T. thermophilus</i> C025926	14	22	9
GAC	36	A, CTG, GTT, AT			<i>P. furiosus</i> At1835	16	24	9
CAC	36	A, CTG, GTT, AT			<i>P. furiosus</i> At1840	16	24	9
TAC	35	AT, GAC, GCC, GCC, CT			<i>P. furiosus</i> At1841	22	34	13
GAC	35	GC, ACC, ACC, TT			<i>E. coli</i> C08004481	25	34	10
GAC	35	GC, ACC, ACC, TT			<i>E. coli</i> C08004482	25	34	10
CAC	36	AT, GAC, GCC, GCC, CT	<b>C, AC</b>		<i>P. furiosus</i> At1840	23	38	16
GAC	36	AT, GAC, GCC, ACC, CT	<b>G, AC</b>		<i>P. furiosus</i> At1835	23	38	16
TAC	35	A, CTC, GCC, T	<b>T</b>		<i>T. thermophilus</i> C025924	27	35	9
TAC	35	T, <b>TAC</b>	GAG, GC		<i>P. furiosus</i> At1841	34	42	9
CAC	33			A, GAG, GTC, GTA, GGT, T	<i>T. thermophilus</i> C025949	41	54	14
GAC	36			TG, GAG, GTC, C	<i>P. furiosus</i> At1835	43	51	9
TAC	35			A, GAG, GTT, CA	<i>T. thermophilus</i> C025924	50	58	9
TAC	34			G, GTC, GGC, GGT, T	<i>E. coli</i> C08004475	45	55	11
GAC	35			G, GTC, GTT, GGT, T	<i>E. coli</i> C08004481	46	56	11
GAC	35			G, GTC, GGT, GGT, T	<i>E. coli</i> C08004482	46	56	11
GAC	35			G, GTC, GGT, GGT, T	<i>T. thermophilus</i> C025926	46	56	11
TAC	35			AA, GTC, CTC, T	<i>T. thermophilus</i> C025924	58	66	9
TAC	34			CC, GTC, ATC, ACC, CA	<i>E. coli</i> C08004475	61	73	13
CAC	33			CC, TAC, GCC, GCC, CA	<i>T. thermophilus</i> C025949	60	72	13
TAC	35			T, GCC, GCC, CA	<i>T. thermophilus</i> C025924	66	74	9

### 3.2.8. X circular code motifs in Gly-tRNAs (Table 4h)

- (i) 5' regions of Gly-tRNAs: Several classes of X motifs are identified: GC,GGT,GGT,A, *Gly-tRNA*<sub>Xm1</sub> G,GGC,RTM,GTW,YA, TG,GTA,GTC,TA suffix of *Gly-tRNA*<sub>Xm1</sub>, WY,AAT,GGY,W, GC,CTG, GTC,TA, G,GTA,GAR,C and AK,WAC,SWS,A. The two X motifs *Gly-tRNA*<sub>Xm2</sub> AG,GAC,GCC,GGC,CT and *Gly-tRNA*<sub>Xm3</sub> GC,ATC, GGC,CT occur before the anticodon TCC.
- (ii) Anticodon regions of Gly-tRNAs: The two X motifs AG,GAC, GCC,GGC,CTC,C whose prefix is *Gly-tRNA*<sub>Xm2</sub> and GC,ATC,GGC, TTC,C whose prefix is *Gly-tRNA*<sub>Xm3</sub> are in a different frame than the anticodon CCC. The X motif AG,GAC,GCC,ACC,CTG,C is in a different frame than the anticodon GCC. The class of X motifs R,GCC,TTC,CA is in a different frame than the anticodon TCC.
- (iii) 3' regions of Gly-tRNAs: The class of X motifs AA,GCC,GRS,GRY, C occurs after the anticodons CCC and TCC. Several X motifs are also found: TA,TAC,GAG,GGT,T, TG,GAG,ACC,C, GC,GAG,TTC, GAG,T, GT,CTC,GTT,T, CC,TTC,GCC,C and G,GCC,ACC,GC where no obvious relation could have been identified between them so far.

### 3.2.9. X circular code motifs in His-tRNAs (Table 4i)

- (i) 5' regions of His-tRNAs: Several classes of X motifs are identified: G,GGT,GGT,GTA,GCC,T, *His-tRNA*<sub>Xm1</sub> RS,CTG,GTT,A, TG,GTA,GAG,C shifted in frame with *His-tRNA*<sub>Xm1</sub> and A,GCC, CTG,GAT,T.
- (ii) 3' regions of His-tRNAs: The X motif *His-tRNA*<sub>Xm2</sub> A,GTT,GTC, GT is shifted by +1 nucleotide from A,TTC,CAG,TT. The X motif CC,CTG,GCC,C is shifted in frame with *His-tRNA*<sub>Xm2</sub>.

### 3.2.10. X circular code motifs in Ile-tRNAs (Table 4j)

- (i) 5' regions of Ile-tRNAs: Several X motifs are identified: G,GGC, GAT,TA and *Ile-tRNA*<sub>Xm1</sub> RS,CTG,GTY,A. The X motif A,GGT,GGT, TA is shifted by +1 nucleotide from *Ile-tRNA*<sub>Xm1</sub>.

- (ii) 3' regions of Ile-tRNAs: A large X motif *Ile-tRNA*<sub>Xm2</sub> AG,GGT, GAG,GTC,GGT,GGT,T of 18 nucleotides is identified in *E. coli*. A suffix of *Ile-tRNA*<sub>Xm2</sub> of 15 nucleotides is found in *T. thermophilus*. The X motif G,TTC,GAA,GCC,C is shifted by +1 nucleotide from *Ile-tRNA*<sub>Xm2</sub>. The X motifs are observed: CC, AAC,ATC,GCC,CA and T,CAG,GCC,TAC.

### 3.2.11. X circular code motifs in Leu-tRNAs (Table 4k)

- (i) 5' regions of Leu-tRNAs: Several classes of X motifs are shifted in frame in series: GC,GAR,GRT,GGC,GG, R,GRT,GGY,GG, G,GTT, GCC,GAG,C, GT,GGC,GAA,ATY,GGT,A, a large X motif *Leu-tRNA*<sub>Xm1</sub> TG,GAA,CTG,GTA,GAC,AC of 16 nucleotides in *T. thermophilus*, GC,CTG,GYC,AA, G,GTA,GAC,RC and AC,GCC,ATC, TT which occurs before the anticodon GAG.
- (ii) Anticodon regions of Leu-tRNAs: The X motif *Leu-tRNA*<sub>Xm2</sub> AC, CAG,ATT,TA is in frame with the anticodon TAG. The class of X motifs G,ATT,SAG,GGT,C in frame with *Leu-tRNA*<sub>Xm2</sub> except with its suffix TA of the anticodon TAG, is in frame with both the anticodons CAG and GAG. The X motifs AG,GTT,CTG,GC, AA, AAT,CTG,CTG,T and *Leu-tRNA*<sub>Xm3</sub> AA,ATC,AAC,C are in frame with the anticodons TAG, CAA and CAA, respectively.
- (iii) 3' regions of Leu-tRNAs: The X motif *Leu-tRNA*<sub>Xm4</sub> A,ACC,GTA, GAA,AT is shifted by +2 nucleotides from *Leu-tRNA*<sub>Xm3</sub>. Several classes of X motifs are shifted in frame: *Leu-tRNA*<sub>Xm4</sub>, CC,CTC,GGC,GTT,C, T,GGC,GCC,GC and SY,GGT,GCC,S. The X motif GC,GAG,TTC,AA is shifted by +1 nucleotide from GT,GCC, GGT,T. The X motifs GT,CTC,GCC,T and G,GCC,TTC,GGC,ACC are shifted in frame.

### 3.2.12. X circular code motifs in Lys-tRNAs (Table 4l)

- (i) 5' regions of Lys-tRNAs: The X motif *His-tRNA*<sub>Xm1</sub> T,CAG,CTG, GC is shifted by +1 nucleotide from AG,CTC,AAC,T. The X motifs *His-tRNA*<sub>Xm1</sub>, GC,CTG,GTT,A, AA,CTG,GTA,GAG,CA and *His-tRNA*<sub>Xm2</sub> TG,GTA,GAG,CAG,TT are shifted in frame.
- (ii) Anticodon regions of Lys-tRNAs: The class of X motifs A,RYY, GAC,TT which is in frame with the anticodon TTT, is shifted by

+1 nucleotide from *His-tRNA<sub>Xm2</sub>*. The X motif *His-tRNA<sub>Xm3</sub>* TA, ATC,GGT,GG in *T. thermophilus* C025946 is in frame with the anticodon CTT.

- (iii) 3' regions of Lys-tRNAs: Interestingly, the X motifs TA,ATC,GGT,A in *T. thermophilus* C025922 which is identical to *His-tRNA<sub>Xm3</sub>* (except its last letter) and TA,ATC,AAT,T occur after the anticodon TTT and are not involved in the anticodon of Lys-tRNAs. Several classes of X motifs are observed: TG,GGT,TAC,A,CC,GGT,GGT,C, R,CAG,GTYS and A,ATC,CTG,CA.

### 3.2.13. X circular code motifs in Met-tRNAs (Table 4m)

- (i) 5' regions of Met-tRNAs: Several classes of X motifs are shifted in frame: GGC,KRC,GTA,GC, TG,GAG,CAG,C and *Met-tRNA<sub>Xm1</sub>* AG,CTC,AAC,GGT,CAG,A. The X motif *Met-tRNA<sub>Xm2</sub>* A,GGT,GGT,CAG,A is suffix of *Met-tRNA<sub>Xm1</sub>*. The X motif *Met-tRNA<sub>Xm3</sub>* A, GTT,GGT,TA is prefix of *Met-tRNA<sub>Xm2</sub>*. The class of X motifs *Met-tRNA<sub>Xm4</sub>* GC,CTG,GTM,R is shifted by +2 nucleotides from *Met-tRNA<sub>Xm3</sub>* A,GTI,GGT,TA. The class of X motifs *Met-tRNA<sub>Xm4</sub>*, TA,GAG,CAG,GC and AG,CTC,GTC,GG are shifted in frame.
- (ii) Anticodon regions of Met-tRNAs: The X motif AT,AAT,GAT,GC is in frame with the anticodon CAT. The X motif TA,ACC,GGT,A is in a different frame than the anticodon CAT.
- (iii) 3' regions of Met-tRNAs: Several classes of X motifs are shifted in series: CC,GGA,GRT,C, GT,GGT,GTC,GT, *Met-tRNA<sub>Xm5</sub>* AG,RTC,GTC,GGT,I, *Met-tRNA<sub>Xm6</sub>* S,SWG,GTI,C shifted by +2 nucleotides from *Met-tRNA<sub>Xm5</sub>*, G,TTC,GAA,GCC,C shifted by +1 nucleotide from *Met-tRNA<sub>Xm6</sub>*, AA,GTC,CAG,CAG,GG, A,ATC,CTS,S, CC,GTC,GTA,GCC and CC,ACC,GCC,GCC,ACC.

### 3.2.14. X circular code motifs in Phe-tRNAs (Table 4n)

- (i) 5' regions of Phe-tRNAs: Two X motifs GCC,GAG,GTA,GC and G, GTA,GAG,CA are observed.
- (ii) Anticodon regions of Phe-tRNAs: The X motif G,ATT,GAA,AAT,C is in frame with the anticodon GAA.
- (iii) 3' regions of Phe-tRNAs: The X motifs G,GGT,GTC,GG and GT, GTC,GGC,GGT,T are shifted in frame. The X motif CC,CTC,GGC, ACC is observed.

### 3.2.15. X circular code motifs in Pro-tRNAs (Table 4o)

- (i) 5' regions of Pro-tRNAs: Several classes of X motifs are shifted in series: S,GSC,GWR,K, *Pro-tRNA<sub>Xm1</sub>* G,CAG,GCC,GGT,A, GC,CTG,GTA,GC shifted by +1 nucleotide from *tRNA<sub>Xm1</sub>*, CC,ATC,CTG,C and GC,ACC,GTC,AT.
- (ii) Anticodon regions of Pro-tRNAs: The class of X motifs *Pro-tRNA<sub>Xm2</sub>* AM,YTS,GTT,Y is in frame with the anticodons CGG and TGG. The class of X motifs S,GGC,YTR,GG is in a different frame than the anticodons CGG and GGG. The class of X motifs G,GGY,GYC,GG is in a different frame than the anticodon GGG. The class of X motifs *Pro-tRNA<sub>Xm3</sub>* G,GAS,SAR,GG is in frame with the anticodons CGG and TGG. The two classes of X motifs *Pro-tRNA<sub>Xm2</sub>* and *Pro-tRNA<sub>Xm3</sub>* may derive from an ancestral class of X motifs constructed by the concatenation of *Pro-tRNA<sub>Xm2</sub>* and *Pro-tRNA<sub>Xm3</sub>* AM,YTS,GTT,Y-A-G,GAS,SAR,GG of 19 nucleotides. Indeed, CAG belongs to X (Eq. (1)). Then, the nucleotide A in the middle site of CAG has mutated to G for building the anticodon CGG.
- (iii) 3' regions of Pro-tRNAs: The class of X motifs G,SWG,GTI,CA is shifted by +2 nucleotides from the X motif G,GTC,GTG,GGT,I. Several classes of X motifs are observed: A,ATC,CWS,T, CC,GGC,SKC,C and T,CTC,GCC,GAC.

**Table 5**

Coverage (Eq. (4)) of X circular code motifs in the 5' and 3' regions of tRNAs of prokaryotes (bacteria *E. coli* and *T. thermophilus*, archaea *P. furiosus*). The lengths and the anticodon positions are average values in each isoaccepting tRNA. The coverage of X motifs in the 5' and 3' regions of tRNAs is given in rounded percentage.

tRNA	Length	Anticodon position	Coverage of X motifs	
			5' regions (%)	3' regions (%)
Ala	76	34	67	63
Arg	77	35	97	88
Asn	76	34	91	63
Asp	78	36	80	23
Cys	74	33	94	77
Gln	75	34	85	77
Glu	77	34	91	32
Gly	76	35	100	95
His	77	35	85	35
Ile	77	35	62	93
Leu	87	36	100	94
Lys	78	34	76	69
Met	77	35	91	98
Phe	76	34	85	70
Pro	78	36	100	93
SeC	95	35	88	31
Ser	89	35	88	92
Thr	77	34	100	95
Trp	77	35	94	57
Tyr	83	35	85	59
Val	78	35	94	90
Mean	79	35	88	71
Min	74	33	62	23
Max	95	36	100	98

### 3.2.16. X circular code motifs in SeC-tRNAs (Table 4p)

- (i) 5' regions of SeC-tRNAs: Two X motifs AG,ATC,GTC,GTC,T and CC,GGT,GAG,GC are observed.
- (ii) Anticodon regions of SeC-tRNAs: The X motif *SeC-tRNA<sub>Xm</sub>* G,CTG,GAC,TTC,AA is in a different frame than the anticodon TCA. The X motif AA,ATC,CAG,TT shifted by +1 nucleotide from *SeC-tRNA<sub>Xm</sub>* is also in a different frame than the anticodon TCA.
- (iii) 3' regions of SeC-tRNAs: One X motif TG,ATC,TTC,C is observed.

### 3.2.17. X circular code motifs in Ser-tRNAs (Table 4q)

- (i) 5' regions of Ser-tRNAs: Several classes of X motifs are shifted: R,GNW,GCC,KR, *Ser-tRNA<sub>Xm1</sub>* TG,GCC,GAG,C, GC,CTG,GTA,GG shifted by +1 nucleotide from *Ser-tRNA<sub>Xm1</sub>*, G,CTG,AAC,GG, the class of large X motifs *Ser-tRNA<sub>Xm2</sub>* G,GTT,GAA,GGC,RSC,GGT,CT of 18 nucleotides in *T. thermophilus* and *E. coli* and K,GAA,GGC,GC factor of *Ser-tRNA<sub>Xm2</sub>*.
- (ii) Anticodon regions of Ser-tRNAs: The very large X motif *Ser-tRNA<sub>Xm3</sub>* TG,GTC,GAA,GGC,GGC,ACC,CTG,CT of 22 nucleotides in *T. thermophilus* is in a different frame than the anticodon GCT. The large X motif *Ser-tRNA<sub>Xm4</sub>* AG,GGC,GCC,GGC,CTG,CT of 16 nucleotides in *P. furiosus* which is suffix of *Ser-tRNA<sub>Xm3</sub>*, is thus in a different frame than the anticodon GCT. The class of large X motifs *Ser-tRNA<sub>Xm5</sub>* G,RCC,GGT,CTC,GAA,AAC,C of 17 nucleotides in *E. coli* and *T. thermophilus* is in a different frame than the anticodon CGA. The X motif A,CTC,GAG,ATC,C which is suffix of *Ser-tRNA<sub>Xm5</sub>*, is thus in a different frame than the anticodon CGA. The class of X motifs GC,CTG,GAR,A and A,CTG,GAA,ATC,GT which are shifted in frame, are in a different frame than the anticodon GGA. The X motif *Ser-tRNA<sub>Xm6</sub>* TT,GAA,AAC,C is in a different frame than the anticodon TGA.
- (iii) 3' regions of Ser-tRNAs: The class of X motifs AA,RSY,GKY,R is shifted by +2 nucleotides from *Ser-tRNA<sub>Xm6</sub>* TT,GAA,AAC,C. A large X motif *Ser-tRNA<sub>Xm7</sub>* GT,GTA,TAC,GGC,AAC,GTA,T of 18

nucleotides is identified in *E. coli*. The X motif G,TTC,GCC,CA is shifted by +2 nucleotides from the class of X motifs TG,GGC,GTT,Y. Several classes of X motifs are observed: GC,GAA,GCC,CA, AA,MYC,KRY,S, R,RRS,YTC,S and YC,RCC,CTC,Y.

### 3.2.18. X circular code motifs in Thr-tRNAs (Table 4r)

- (i) 5' regions of Thr-tRNAs: Several classes of X motifs are shifted: CC,GGT,GKM,K, *Thr-tRNA*<sub>m1</sub> G,GTA,GCC,TA, AG,CTC,AAC,C shifted by +2 nucleotides from *Thr-tRNA*<sub>m1</sub>, T,CAG,CAG, GTA,GAG,CA and its suffix G,GTA,GAG,C.
- (ii) Anticodon regions of Thr-tRNAs: The class of X motifs *Thr-tRNA*<sub>m2</sub> M,YTC,GTA,A is in a different frame than the anticodon CGT. The X motif *Thr-tRNA*<sub>m3</sub> TT,GTA,ATC, shifted in frame with *Thr-tRNA*<sub>m2</sub>, is in a different frame than the anticodon TGT. Interestingly, the X motif GT,AAI,CAG,TA shifted by +1 nucleotide from *Thr-tRNA*<sub>m3</sub> TT,GTA,ATC, is, in contrast, in frame with the anticodon TGT.
- (iii) 3' regions of Thr-tRNAs: The large X motif *Thr-tRNA*<sub>m4</sub> AG,GGT,GAG,GTC,GCC,GGT,T of 18 nucleotides is identified in *T. thermophilus*. The class of X motifs AG,GGT,GAG,GTC,S is prefix of *Thr-tRNA*<sub>m4</sub>. Several classes of X motifs are observed: M, GWA,GGT,C, CC,CAG,GTC,S, AG,GYC,RYM,R, S,CAG,TTC,GAN,T, Y,CTG,GGT,RKC,W and K,RYN,RYC,GGC,W.

### 3.2.19. X circular code motifs in Trp-tRNAs (Table 4s)

- (i) 5' regions of Trp-tRNAs: The X motif GI,GGI,GTA,GCC,T is shifted by +1 nucleotide from G,GGC,GTA,GTT,CA. Several classes of X motifs are identified: AG,YTC,AAI,T, GC,CTG,GTC, CA, TG,GTA,GAG,CA and CC,ATC,ATC,GC.
- (ii) Anticodon regions of Trp-tRNAs: The X motif *Trp-tRNA*<sub>m</sub> GC, ACC,GGT,CTC,CA is in a different frame than the anticodon CCA. The X motif G,CTC,CAG,ACC,C shifted in frame with *Trp-tRNA*<sub>m</sub>, is also in a different frame than the anticodon CCA.
- (iii) 3' regions of Trp-tRNAs: The class of X motifs TG,GAG,KTY,S is shifted by +2 nucleotides from the X motif G,GGT,GTT,GG.

### 3.2.20. X circular code motifs in Tyr-tRNAs (Table 4t)

- (i) 5' regions of Tyr-tRNAs: The X motif G,GTA,GCC,TA is observed. The X motif GI,GGC,GGC,GG is shifted by +1 nucleotide from the X motif CC,CTG,GTA,GT. The X motif G,GAG,CAG,AC occurs before the anticodon GTA.
- (ii) Anticodon regions of Tyr-tRNAs: The X motif *Tyr-tRNA*<sub>m1</sub> G, GAC,GGT,CTG,TA is in a different frame than the anticodon GTA. Interestingly, the X motif T,GTA,AAI,CTG,C shifted by +1 nucleotide from *Tyr-tRNA*<sub>m1</sub> is, in contrast, in frame with the anticodon GTA.
- (iii) 3' regions of Tyr-tRNAs: The X motif AA,ACC,GTT,GGC,GTA,T is shifted by +2 nucleotides from the X motif T,GTA,AAI,CTG,C. A very large X motif *Tyr-tRNA*<sub>m2</sub> T,GCC,GTC,ATC,GAC,TTC,GAA,GGT,T of 23 nucleotides is identified in *E. coli*. The X motif T, GCC,GTC,AC is prefix of *Tyr-tRNA*<sub>m2</sub>, the X motif A,GAC,TTC, GAA,GGT,T is suffix of *Tyr-tRNA*<sub>m2</sub> and the X motif AT,GCC, TTC,GC is factor of *Tyr-tRNA*<sub>m2</sub>.

### 3.2.21. X circular code motifs in Val-tRNAs (Table 4u)

- (i) 5' regions of Val-tRNAs: Several classes of X motifs are shifted: G,GGY,GRY,TA, T,CAG,CTG,GS, *Val-tRNA*<sub>m1</sub> AG,CTG,GTT,AT, A, GTT,GGT,TA shifted by +1 nucleotide from *Val-tRNA*<sub>m1</sub>, A,GGT,GGC,TA, *Val-tRNA*<sub>m2</sub> A,CTG,GTT,AI, AI,GAC,GCC,GCC,CT

shifted by +1 nucleotide from *Val-tRNA*<sub>m2</sub> and GC,ACC,ACC, TT which occurs before the anticodon GAC.

- (ii) Anticodon regions of Val-tRNAs: The class of large X motifs *Val-tRNA*<sub>m3</sub> AT,GAC,GCC,RCC,CTS,AC of 16 nucleotides identified in *P. furiosus* is in a different frame than the anticodons CAC and GAC. The X motif *Val-tRNA*<sub>m4</sub> A,CTC,GCC,TT is in a different frame than the anticodon TAC. Interestingly, the X motif T,TAC,GAG,GC shifted by +2 nucleotides from *Val-tRNA*<sub>m4</sub> is, in contrast, in frame with the anticodon TAC.
- (iii) 3' regions of Val-tRNAs: Several classes of X motifs are observed: *Val-tRNA*<sub>m5</sub> A,GAG,GTC,GTA,GGT,T, R,GAG,GTI,C prefix of *Val-tRNA*<sub>m5</sub>, G,GTC,GKY,GGT,T suffix of *Val-tRNA*<sub>m5</sub>, AA,GTC,CTC,T, CC,KWC,RYC,RCC,CA and T,GCC,GCC,CA.

### 3.2.22. Coverage of X motifs in the tRNAs (Table 5)

Table 5 shows that the coverage (Eq. (4)) of X motifs is greater in the 5' regions of tRNAs in *E. coli*, *T. thermophilus* and *P. furiosus* (mean equal to 88%, min equal to 62%, max equal to 100%) compared to their 3' regions (mean equal to 71%, min equal to 23%, max equal to 98%). The coverage of X motifs is maximal (100%) in the 5' regions of tRNAs of Gly, Leu, Pro and Thr and minimal (around 30%) in the 3' regions of tRNAs of Asp, Glu, His and SeC.

## 4. Conclusion

Seven X circular code motifs *PrRNAXm* which are conserved in 16S rRNAs of bacteria *E. coli* and *T. thermophilus*, and archaea *P. furiosus* are identified near the ribosome decoding center (Figs. 1–3), in particular the large X motif *PrRNAXm1* GC,GGT,AAI,ACC,GGC,GGC,C of 18 nucleotides in *P. furiosus*, the large common X motif *PrRNAXm3* G,AAI,R1CC,GR2T,GGC,GAA,GGC of 19 nucleotides in *E. coli* and *T. thermophilus*, the large X motif *PrRNAXm4* TA,GAT,ACC,CTG,GTA,GTC,CA of 19 nucleotides in *E. coli*, the large common X motif *PrRNAXm6* T,TAC,GRC,CWG,GGC,KAC,AC of 18 nucleotides in *E. coli* and *T. thermophilus* (details in Section 3.1.1). Four X circular code motifs *ErRNAXm* which are conserved in 18S rRNAs of *S. cerevisiae*, *T. aestivum* and *H. sapiens* are found near the ribosome decoding center (Figs. 4–6), in particular the large common X motif *ErRNAXm2* G,NTC,GAA,GAY,GAT,CAG,AT of 18 nucleotides in *S. cerevisiae*, *T. aestivum* and *H. sapiens* and the large X motif *ErRNAXm4* TC,TTC,AAI,GAG,GAA,TTC,CT of 19 nucleotides in *S. cerevisiae* (details in Section 3.1.2).

The identification of X circular code motifs in tRNAs of *E. coli*, *T. thermophilus* and *P. furiosus* shows several new features for the structure of tRNAs. The high coverage of X motifs in the 5' and 3' regions of these tRNAs (88% and 71%, respectively; with the exception of the 3' regions of tRNAs of Asp, Glu, His and SeC; Table 5) means that tRNAs may be constructed by a concatenation of X motifs. This hypothesis is strengthened by the fact that four very large X motifs of length greater than or equal to 20 nucleotides are found in tRNAs having in average 79 nucleotides (Table 5): *Ala-tRNA*<sub>m1</sub> GC,CTC,AAI,GGC,ATT,GAG,GAG,GTC,A of 24 nucleotides in *T. thermophilus*, *Glu-tRNA*<sub>m2</sub> TA,GAG,GCC,CAG,GAC,ACC,GCC,CT of 22 nucleotides in *E. coli*, *Ser-tRNA*<sub>m3</sub> TG,GTC,GAA,GGC,GGC,ACC,CTG,CT of 22 nucleotides in *T. thermophilus* and *Tyr-tRNA*<sub>m2</sub> T, GCC,GTC,ATC,GAC,TTC,GAA,GGT,T of 23 nucleotides in *E. coli*, and 14 large X motifs of lengths 16–19 nucleotides: *Arg-tRNA*<sub>m1</sub>, *Arg-tRNA*<sub>m3</sub>, *Arg-tRNA*<sub>m4</sub>, *Arg-tRNA*<sub>m6</sub>, *Asn-tRNA*<sub>m3</sub>, *Asp-tRNA*<sub>m2</sub>, *Ile-tRNA*<sub>m2</sub>, *Leu-tRNA*<sub>m1</sub>, *Ser-tRNA*<sub>m2</sub>, *Ser-tRNA*<sub>m4</sub>, *Ser-tRNA*<sub>m5</sub>, *Ser-tRNA*<sub>m7</sub>, *Thr-tRNA*<sub>m4</sub> and *Val-tRNA*<sub>m3</sub>. Remember that X motifs of lengths equal to 9 nucleotides retrieve the reading frame with a probability of 99.9% and X motifs of lengths greater than or equal to 12 nucleotides retrieve, by definition, the reading frame with a probability of 100% (Table 3 and Fig. 4 in

Michel, 2012). We also note that the coverage and the length of the X motifs could be greater if substitutions in X motifs were considered.

We mention some of the properties of X motifs found in tRNAs (Section 3.2):

- (i) An X motif can occur at the same position in the same isoaccepting tRNA of different species, e.g. the X motif CAG, GAG, GTC is at position 40 in Ala-tRNA of *E. coli* and *T. thermophilus* (Table 4a), the X motif T, CAG, CTG, GAT, A is at position 12 in Arg-tRNA of *E. coli* and *T. thermophilus* (Table 4b), etc.;
- (ii) an X motif can occur at the same position in different isoaccepting tRNAs, e.g. the X motif T, CAG, CTG, G is at position 12 in Ala-tRNA of *E. coli*, Arg-tRNA of *E. coli* and *T. thermophilus*, His-tRNA of *T. thermophilus*, Ile-tRNA of *T. thermophilus* and Lys-tRNA of *T. thermophilus*, Val-tRNA of *E. coli* and *T. thermophilus* (Table 4a, b, i, j, l, u), the X motif GTA, GTT, CAG is at the same position 7 in Arg-tRNA, Asn-tRNA, Asp-tRNA, Gly-tRNA, Trp-tRNA of *E. coli* (Table 4b–d, h, s), etc.;
- (iii) an X motif can be shifted by 0, +1 or +2 mod 3 nucleotides from another X motif in the same species or in different species;
- (iv) an X motif can be in the same frame as the anticodon, e.g. the very large X motif *Ala-tRNAX<sub>m1</sub>* GC, CTC, AAT, GGC, ATT, GAG, GAG, GTC, A of 24 nucleotides in Ala-tRNA of *T. thermophilus* is in frame with the anticodon GGC (Table 4a), the X motif G, ATT, CTG, ATT, C in Gln-tRNA of *E. coli* is in frame with the anticodon CTG (Table 4f), etc.;
- (v) an X motif can be in a different frame than the anticodon with a shift of one nucleotide, e.g. the X motif *Cys-tRNAX<sub>m3</sub>* A, CTG, CAG, ATC, C in Cys-tRNA of *P. furiosus* is shifted by +1 nucleotide relative to the anticodon GCA (Table 4e), etc.;
- (vi) an X motif can be in a different frame than the anticodon with a shift of two nucleotides, e.g. the large X motif *Arg-tRNAX<sub>m6</sub>* A, GTA, CTC, GGC, TAC, GAA, C of 17 nucleotides in Arg-tRNA of *E. coli* is shifted by +2 (−1) nucleotides relative to the anticodon ACG (Table 4b), etc.

The X circular code motifs containing the universally conserved nucleotides G530, A1492 and A1493 in rRNAs of the ribosome decoding center (Michel, 2012; El Soufi and Michel, 2014) and the X circular code motifs described here in rRNAs near the ribosome decoding center and in tRNAs allow pairing with the X circular code

motifs of mRNAs. Indeed, the X motifs have the circular code property for retrieving the reading frame, the  $C^3$  property for retrieving the two shifted frames and the complementary property for pairing. The results obtained here strengthen the concept of translation code based on the circular code proposed in Michel (2012).

## Acknowledgment

The first author thanks the Lebanese Association for Scientific Research (LASer) for funding.

## References

- Anger, A.M., Armache, J.P., Berninghausen, O., Habeck, M., Subklewe, M., Wilson, D.N., Beckmann, R., 2013. Structures of the human and Drosophila 80S ribosome. *Nature* 497, 80–85.
- Armache, J.P., Anger, A.M., Marquez, V., Franckenberg, S., Frohlich, T., Villa, E., Wilson, D.N., 2013. Promiscuous behaviour of archaeal ribosomal proteins: implications for eukaryotic ribosome evolution. *Nucleic Acids Res.* 41, 1284–1293.
- Armache, J.P., Jarasch, A., Anger, A.M., Villa, E., Becker, T., Bhushan, S., Beckmann, R., 2010a. Cryo-EM structure and rRNA model of a translating eukaryotic 80S ribosome at 5.5-Å resolution. *Proc. Natl. Acad. Sci. U. S. A.* 107, 19748–19753.
- Armache, J.P., Jarasch, A., Anger, A.M., Villa, E., Becker, T., Bhushan, S., Beckmann, R., 2010b. Localization of eukaryote-specific ribosomal proteins in a 5.5-Å cryo-EM map of the 80S eukaryotic ribosome. *Proc. Natl. Acad. Sci. U. S. A.* 107, 19754–19759.
- Arquès, D.G., Michel, C.J., 1996. A complementary circular code in the protein coding genes. *J. Theor. Biol.* 182, 45–58.
- Brilot, A.F., Korostelev, A.A., Ermolenko, D.N., Grigorieff, N., 2013. Structure of the ribosome with elongation factor G trapped in the pretranslocation state. *Proc. Natl. Acad. Sci. U. S. A.* 110, 20994–20999.
- Gogala, M., Becker, T., Beatrix, B., Armache, J.P., Barrio-Garcia, C., Berninghausen, O., Beckmann, R., 2014. Structures of the Sec61 complex engaged in nascent peptide translocation or membrane insertion. *Nature* 506, 107–110.
- El Soufi, K., Michel, C.J., 2014. Circular code motifs in the ribosome decoding center. *Comput. Biol. Chem.* 52, 9–17.
- Jenner, L.B., Demeshkina, N., Yusupova, G., Yusupov, M., 2010. Structural aspects of messenger RNA reading frame maintenance by the ribosome. *Nat. Struct. Mol. Biol.* 17, 555–560.
- Michel, C.J., 2012. Circular code motifs in transfer and 16S ribosomal RNAs: a possible translation code in genes. *Comput. Biol. Chem.* 37, 24–37.
- Michel, C.J., 2013. Circular code motifs in transfer RNAs. *Comput. Biol. Chem.* 45, 17–29.
- Michel, C.J., 2015. The maximal  $C^3$  self-complementary trinucleotide circular code X in genes of bacteria eukaryotes, plasmids and viruses. *J. Theor. Biol.* 380, 156–177.
- Wilson, D.N., 2014. Ribosome-targeting antibiotics and mechanisms of bacterial resistance. *Nat. Rev. Microbiol.* 12, 35–48.