Computational Biology and Chemistry 52 (2014) 9–17

Contents lists available at ScienceDirect

Computational Biology and Chemistry

journal homepage: www.elsevier.com/locate/compbiolchem

Circular code motifs in the ribosome decoding center

Karim El Soufi, Christian J. Michel *

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

ARTICLE INFO

Article history:
Received 8 June 2014
Received in revised form 1 August 2014
Accepted 1 August 2014
Available online 5 August 2014

Keywords:
Circular code motif
Trinucleotide
Ribosomal RNA
Translation code
Ribosome decoding center

ABSTRACT

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 rRNA of *Thermus thermophilus*, in particular in the ribosome decoding center. Three classes of *X* motifs are now identified in the rRNAs of bacteria *Escherichia coli* and *Thermus thermophilus*, archaea *Pyrococcus furiosus*, nuclear eukaryotes *Saccharomyces cerevisiae*, *Triticum aestivum* and *Homo sapiens*, and chloroplast *Spinacia oleracea*. The universally conserved nucleotides A1492 and A1493 in all studied rRNAs (bacteria, archaea, nuclear eukaryotes, and chloroplasts) belong to *X* motifs (called m_{AA}). The conserved nucleotide G530 in rRNAs of bacteria and archaea belongs to *X* motifs (called m_G). Furthermore, the *X* motif m_G is also found in rRNAs of nuclear eukaryotes and chloroplasts. Finally, a potentially important *X* motif, called *m*, is identified in all studied rRNAs. With the available crystallographic structures of the Protein Data Bank PDB, we also show that these *X* motifs m_{AA}, m_G, and *m* belong to the ribosome decoding center of all studied rRNAs with possible interaction with the mRNA *X* motifs and the tRNA *X* motifs. The three classes of *X* motifs identified here in rRNAs of several and different organisms strengthen the concept of translation code based on the circular code.

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 the translation process, the aminoacyl tRNA binds to the A-site where the decoding center containing the universally conserved dinucleotide AA (A1492 and A1493) is tasked with distinguishing 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.

The introduction of 3D crystallographic structure contributed significantly to the better understanding of the functionality of ribosomes. 3D structures and experimental biological results proved that the decoding center is located in the RNA region of the ribosome.

In 1996, a statistical analysis of occurrence frequencies of the 64 trinucleotides {AAA, . . . , TTT} in the three frames zero, one and two 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 zero is the reading frame in a gene, and the frames one and two are the reading frame zero 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 zero, one, and two, respectively, simultaneously of two large gene populations (protein coding regions): eukaryotes (26,757 sequences, 11,397,678 trinucleotides) and prokaryotes (13,686 sequences, 4,709,758 trinucleotides) (Arquès and Michel, 1996). This set *X*

* Corresponding author.
E-mail addresses: kelsoufi@unistra.fr (K. El Soufi), c.michel@unistra.fr (C.J. Michel).

http://dx.doi.org/10.1016/j.compbiolchem.2014.08.001
1476-9271/

contains the 20 following trinucleotides

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

The two sets X_1 and X_2, of 20 trinucleotides each, in the shifted frames one and two of genes can be deduced from X by the circular permutation map (see below). These three trinucleotide sets present several strong mathematical properties, particularly the fact that X is a 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 a few nucleotides positioned anywhere in a sequence generated with the circular code always retrieves the reading frame. For crossing the largest ambiguous words of the circular code X (words, not necessarily unique, in two or three frames), this window needs a length of 13 nucleotides with X (Fig. 3 in Michel, 2012). A window of 13 nucleotide length is the largest window of X to retrieve the reading frame for all the ambiguous words of X.

Recently, X circular code motifs (X motifs shortly), i.e., motifs generated from the circular code X, are identified in the 5′ and/or 3′ regions of 16 isoaccepting tRNAs of prokaryotes and eukaryotes (Michel, 2013). X motifs are also found in 16S rRNAs, in particular in the decoding center which recognizes the codon–anticodon helix in A-tRNA (Michel, 2012). A 3D visualization of X motifs in the ribosome shows several spatial configurations involving mRNA X motifs, A-tRNA and E-tRNA X motifs, and four 16S rRNA X motifs. These results led to the concept of a possible translation (framing) code based on the circular code (Michel, 2012).

By developing a search algorithm of X motifs in a DNA multiple sequence alignment, obtained here with a global multiple sequence alignment program, three classes of X motifs identified in multiple aligned rRNAs are involved in the ribosome decoding center of bacteria, archaea, nuclear eukaryotes, and chloroplasts.

2. Method

2.1. Definitions

A few basic properties of the trinucleotide circular code X (Eq. (1)) are presented here in order to recall to the reader that the X motifs can retrieve the reading frame (by definition), and thus, may be involved in gene translation.

Notation 1. The letters (or nucleotides or bases) define the genetic alphabet $A_4 = \{\mathrm{A}, \mathrm{C}, \mathrm{G}, \mathrm{T}\}$. The set of non-empty words (words resp.) over A_4 is denoted by A_4^+ (A_4^* resp.). The set of the 16 words of length 2 (dinucleotides or diletters) on A_4 is denoted by $A_4^2 = \{\mathrm{AA}, \ldots, \mathrm{TT}\}$. The set of the 64 words of length 3 (trinucleotides or triletters) on A_4 is denoted by $A_4^3 = \{\mathrm{AAA}, \ldots, \mathrm{TTT}\}$. Let $x_1 \cdots x_n$ be the concatenation of the words x_i for $i = 1, \ldots, n$, the symbol "·" being the concatenation operator.

There are two important biological maps involved in codes in genes on A_4.

Definition 1. The nucleotide complementarity map $\mathcal{C} : A_4 \to A_4$ is defined by $\mathcal{C}(\mathrm{A}) = \mathrm{T}$, $\mathcal{C}(\mathrm{C}) = \mathrm{G}$, $\mathcal{C}(\mathrm{G}) = \mathrm{C}$, $\mathcal{C}(\mathrm{T}) = \mathrm{A}$. According to the property of the complementary and antiparallel double helix, the trinucleotide complementarity map $\mathcal{C} : A_4^3 \to A_4^3$ is defined by $\mathcal{C}(l_0 \cdot l_1 \cdot l_2) = \mathcal{C}(l_2) \cdot \mathcal{C}(l_1) \cdot \mathcal{C}(l_0)$ for all $l_0, l_1, l_2 \in A_4$ e.g., $\mathcal{C}(\mathrm{ACG}) = \mathrm{CGT}$. By extension to a trinucleotide set S, the set complementarity map $\mathcal{C} : S \to S$ is defined by $\mathcal{C}(S) = \left\{v | u, v \in A_4^3, u \in S, v = \mathcal{C}(u)\right\}$ i.e., a complementary trinucleotide set $\mathcal{C}(S)$ is obtained by applying the complementarity map $\mathcal{C}$ to all its trinucleotides e.g., $\mathcal{C}(\{\mathrm{ACG}, \mathrm{AGT}\}) = \{\mathrm{ACT}, \mathrm{CGT}\}$.

Definition 2. The trinucleotide circular permutation map $\mathcal{P} : A_4^3 \to A_4^3$ is defined by $\mathcal{P}(l_0 \cdot l_1 \cdot l_2) = l_1 \cdot l_2 \cdot l_0$ for all $l_0, l_1, l_2 \in A_4$ e.g., $\mathcal{P}(\mathrm{ACG}) = \mathrm{CGA}$. The 2nd iterate of $\mathcal{P}$ is denoted $\mathcal{P}^2$ e.g., $\mathcal{P}^2(\mathrm{ACG}) = \mathrm{GAC}$. By extension to a trinucleotide set S, the set circular permutation map $\mathcal{P} : S \to S$ is defined by $\mathcal{P}(S) = \left\{v | u, v \in A_4^3, u \in S, v = \mathcal{P}(u)\right\}$ i.e., a permuted trinucleotide set $\mathcal{P}(S)$ is obtained by applying the circular permutation map $\mathcal{P}$ to all its trinucleotides e.g., $\mathcal{P}(\{\mathrm{ACG}, \mathrm{AGT}\}) = \{\mathrm{CGA}, \mathrm{GTA}\}$ and $\mathcal{P}^2(\{\mathrm{ACG}, \mathrm{AGT}\}) = \{\mathrm{GAC}, \mathrm{TAG}\}$.

Definition 3. A set $S \subset A_4^+$ of words is a code if, for each $x_1, \ldots, x_n, y_1, \ldots, y_m \in S$, $n, m \geq 1$, the condition $x_1 \cdots x_n = y_1 \cdots y_m$ implies $n = m$ and $x_i = y_i$ for $i = 1, \ldots, n$.

Definition 4. As the set $A_4^3 = \{\mathrm{AAA}, \ldots, \mathrm{TTT}\}$ is a code, its non-empty subsets are codes and called trinucleotide codes C.

Definition 5. A trinucleotide code $C \subset A_4^3$ is circular and called CC if, for each $x_1, \ldots, x_n, y_1, \ldots, y_m \in C, n, m \geq 1, r \in A_4^*, s \in A_4^+$, the conditions $sx_2 \cdots x_n r = y_1 \cdots y_m$ and $x_1 = rs$ imply $n = m$, $r = \varepsilon$ (empty word) and $x_i = y_i$ for $i = 1, \ldots, n$.

Remark 1. The fundamental property of a circular code is the ability to retrieve the reading (original or construction) 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

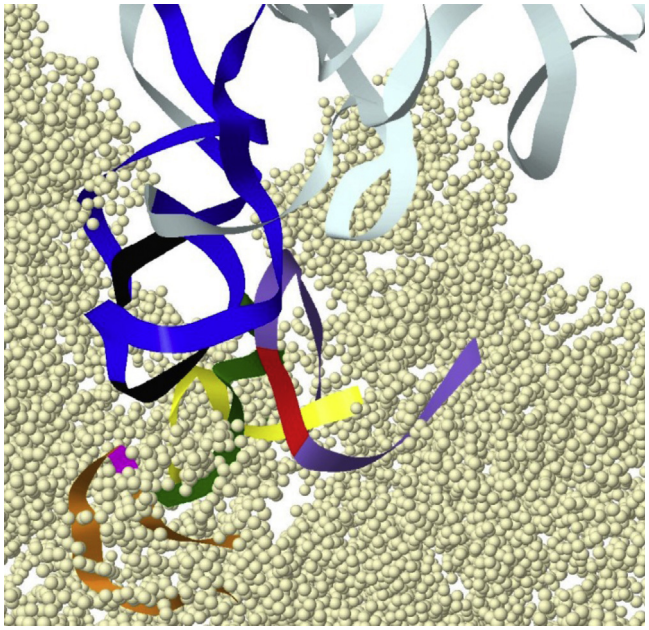

Fig. 1. X circular code motifs involved in the bacterial ribosome decoding center of *Escherichia coli* (crystallographic structure PDB 3J5T): the mRNA X motifs (green), the rRNA X motif m_{AA}(*E. coli*, 1487, 1501, 15) (purple with the conserved dinucleotide AA in red), the rRNA X motif m_{G}(*E. coli*, 527, 536, 10) (orange with the conserved nucleotide G in fuchsia), the rRNA X motif m(*E. coli*, 1396, 1404, 9) (yellow) and the tRNA X motifs (blue with the 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.)

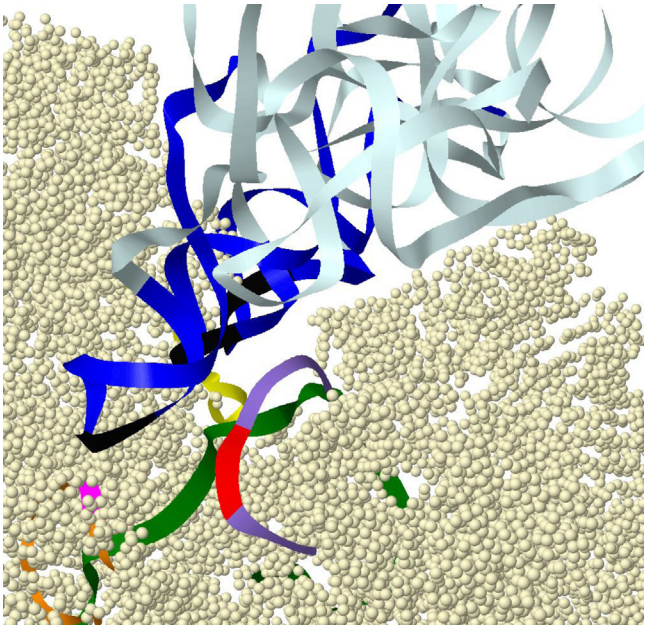

Fig. 2. *X* circular code motifs involved in the bacterial ribosome decoding center of *Thermus thermophilus* (crystallographic structure PDB 3I8G): the mRNA *X* motifs (green), the rRNA *X* motif m_{AA}(*T. thermophilus*, 1490, 1498, 9) (purple with the conserved dinucleotide AA in red), the rRNA *X* motif m_G(*T. thermophilus*, 528, 536, 9) (orange with the conserved nucleotide G in fuchsia), the rRNA *X* motif *m*(*T. thermophilus*, 1375, 1383, 9) (yellow) and the tRNA *X* motifs (blue with the 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.)

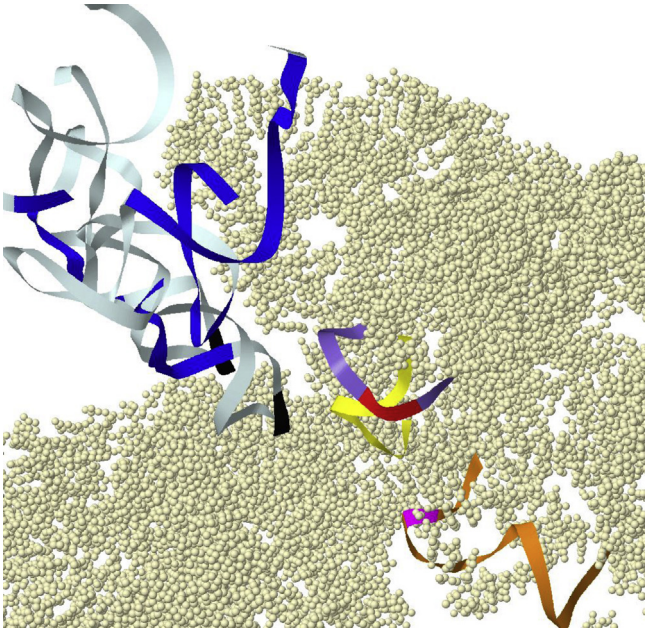

Fig. 3. *X* circular code motifs involved in the archaeal ribosome decoding center of *Pyrococcus furiosus* (crystallographic structure PDB 3J20): the rRNA *X* motif m_{AA}(*P. furiosus*, 1445, 1456, 12) (purple with the conserved dinucleotide AA in red), the rRNA *X* motif m_G(*P. furiosus*, 480, 497, 18) (orange with the conserved nucleotide G in fuchsia), the rRNA *X* motif *m*(*P. furiosus*, 1356, 1364, 9) (yellow) and the tRNA *X* motifs (blue with the 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.)

(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 word, not necessarily unique, which can be read in at least two frames, plus one letter (Fig. 3 in Michel, 2012, for an example).

Definition 6. A trinucleotide circular code $CC \subset A_4^3$ is self-complementary and called *SCC* if, for each $y \in CC$, $\mathcal{C}(y) \in CC$.

Definition 7. A trinucleotide circular code $CC \subset A_4^3$ is C^3 and called C^3CC if the two permuted trinucleotide sets $CC_1 = \mathcal{P}(CC)$ and $CC_2 = \mathcal{P}^2(CC)$ are trinucleotide circular codes.

Definition 8. A trinucleotide circular code $CC \subset A_4^3$ is C^3 self-complementary and called C^3SCC if $CC, CC_1 = \mathcal{P}(CC)$ and $CC_2 = \mathcal{P}^2(CC)$ are trinucleotide circular codes satisfying the following properties $CC = \mathcal{C}(CC)$ (self-complementary), $\mathcal{C}(CC_1) = CC_2$ and $\mathcal{C}(CC_2) = CC_1$ (CC_1 and CC_2 are complementary).

The trinucleotide set $X = X_0$ (Eq. (1)) coding the reading frames (frames 0) in eukaryotic and prokaryotic genes is a maximal (20 trinucleotides) C^3 self-complementary trinucleotide circular code C^3SCC with a window length equal to 13 nucleotides for biinfinite words (Arquès and Michel, 1996) and 12 nucleotides for right infinite words (Michel, 2012). Thus, $X, X_1 = \mathcal{P}(X)$ and $X_2 = \mathcal{P}^2(X)$ are trinucleotide circular codes verifying $X = \mathcal{C}(X), \mathcal{C}(X_1) = X_2$ and $\mathcal{C}(X_2) = X_1$.

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

We present here a search algorithm of *X* motifs of lengths greater than a given number of nucleotides in a DNA multiple sequence alignment (MSA) obtained here with the program ClustalW2 (global MSA). It will identify common *X* motifs in multiple aligned rRNAs.

Let a trinucleotide *t* of the circular code *X* defined in Eq. (1) be the three letters $t = l_1l_2l_3 \in A_4^3$. Let $\mathrm{Pref}_{\mathrm{let}}(X)$ be the set containing the letters $l_1 \in A_4$ of *X*, and $\mathrm{Pref}_{\mathrm{dilet}}(X)$ be the set containing the diletters $l_1l_2 \in A_4^2$ of *X*. Then, by inspection of *X*, we have

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

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

Remark 2. $\mathrm{Card}(\mathrm{Pref}_{\mathrm{let}}(X)) = 4$ and $\mathrm{Card}(\mathrm{Pref}_{\mathrm{dilet}}(X)) = 11$ (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 Xmotif searches a *X* motif at a given position startX (input parameter in integer) in a DNA sequence seq of length |seq| on A_4 or $A_4 \cup \{-\}$ (an aligned sequence with gaps) and returns its end position endX (output parameter in integer).

```
Xmotif[startX]
1.  endX = startX
2.  iX = 1 // index on X
3.  testX = true
4.  while testX
5.  {
8.    if {seq[endX] ⋂ A4} ≠ {} then
9.    {
6.      if iX = 1[3] then // Case 1 modulo 3: Pref1et
7.      {
8.        if endX ≤ |seq| and {seq[endX] ⋂ Pref1et} ≠ {} then
9.        {
10.         iX++
11.         endX++
12.       }
13.       else testX = false
14.     }
15.     if iX = 2[3] then // Case 2 modulo 3: Prefdilet
16.     {
17.       if endX ≤ |seq| and {seq[endX-1..endX] ⋂ Prefdilet} ≠ {} then
18.       {
19.         iX++
20.         endX++
21.       }
22.       else testX = false
23.     }
24.     if iX = 0[3] then // Case 0 modulo 3: X
25.     {
26.       if endX ≤ |seq| and {seq[endX-2..endX] ⋂ X} ≠ {} then
27.       {
28.         iX++
29.         endX++
30.       }
31.       else testX = false
32.     }
33.   }
33. }
34. return endX--
```

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

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

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 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_4 (here the wild character "*") and then applies the function Search_Xmotif_seq to the constructed seqCommon for identifying the common X motifs in MSA.

```
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]
```

An X motif returned by the Algorithm_Search_Xmotif_seqCommon beginning at position b and ending at position e is characterized by the triplet (b, e, l) where the nucleotide length $l = e - b + 1$, or by the quadruplet (organism, b, e, l).

2.3. Crystallographic data

In order to identify a general property of X circular code motifs in the ribosome decoding center, the Algorithm_Search_Xmotif_seqCommon is applied here to MSA 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, February 2014). 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 all their X motifs. PDB entries containing synthetic chains were excluded. The studied PDB crystallographic structures are for bacteria: *Escherichia coli* (Brilot et al., 2013) and *Thermus thermophilus* (Jenner et al., 2010); for archaea: *Pyrococcus furiosus* (Armache et al., 2013); for nuclear eukaryotes: *Saccharomyces cerevisiae* (Armache et al., 2010a); *Triticum aestivum* (Armache et al., 2010a,b; Gogala et al., 2014) and *Homo sapiens* (Anger et al.,

Table 1
X circular code motifs studied in seven crystallographic structures of the Protein Data Bank PDB. The main features of the studied crystallographic structures are given: PDB identification, kingdom, organism, type (16S for prokaryotes, 18S for eukaryotes) and base length (*b*) of rRNA, mRNA (Yes for available, No for unavailable), location 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	*Escherichia coli*	16S (1542 *b*)	Yes	Phe	Phe	No
3I8G	Bacteria	*Thermus thermophilus*	16S (1516 *b*)	Yes	Phe	Phe	Phe
3J20	Archaea	*Pyrococcus furiosus*	16S (1495 *b*)	No	No	Phe	Phe
3IZE	Eukaryote, nuclear	*Saccharomyces cerevisiae*	18S (1800 *b*)	Yes	No	Asp	No
3J5Z	Eukaryote, nuclear	*Triticum aestivum*	18S (1810 *b*)	Yes	No	Asp	No
3J3D	Eukaryote, nuclear	*Homo Sapiens*	18S (1869 *b*)	No	No	No	Met
3BBN	Eukaryote, chloroplast	*Spinacia oleracea*	16S (1491 *b*)	No	No	No	No

2013), and for chloroplasts (an organelle of eukaryotes): *Spinacia oleracea* (Sharma et al., 2007). Table 1 summarizes the main features of the crystallographic structures studied.

2.4. 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 X circular code motifs m_{AA} containing the universally conserved dinucleotide AA in rRNAs of bacteria, archaea, nuclear eukaryotes, and chloroplasts

The universally conserved dinucleotide AA (in position 1492 of the 16S rRNA in *E. coli*) has an experimentally proven biological function in the codon–anticodon binding (Moazed and Noller, 1990; Powers and Noller, 1994; Yoshizawa et al., 1999). Very unexpectedly, this universally conserved dinucleotide AA is found to be located in *X* circular code motifs (Table 2).

3.1.1. Bacteria

In rRNAs of both *E. coli* and *T. thermophilus*, the conserved dinucleotide AA occurs at the 2nd and 3rd sites of the trinucleotide GAA $\in X$ (Eq. (1)). In rRNA of *E. coli*, it belongs to the *X* motif m_{AA}(*E. coli*, 1487, 1501, 15) = G,GGT,G**AA**,GTC,GTA,AC of 15 nucleotide length starting with the nucleotide G suffix of CAG,CTG,GAG $\in X$ (Eq. (1)) followed by four trinucleotides GAA,GGT,GTA,GTC $\in X$ (Eq. (1)) (given in lexicographical order) and ending with the dinucleotide AC prefix of ACC $\in X$ (Eq. (1)). In rRNA of *T. thermophilus*, it belongs to the *X* motif m_{AA}(*T. thermophilus*, 1490, 1498, 9) = G,G**AA**,GGT,GC of nine nucleotide length starting with the nucleotide G, as in *E. coli*, followed by two trinucleotides GAA,GGT $\in X$ (Eq. (1)) and ending with the dinucleotide GC prefix of GCC $\in X$ (Eq. (1)). These two rRNA *X* motifs m_{AA}(*E. coli*) and m_{AA}(*T. thermophilus*) have completely different primary structures. Thus, the classical bioinformatics methods, such as sequence alignment or phylogenetic inference, were not able to identify these motifs which are only revealed by the circular code theory.

In rRNA of *T. thermophilus*, there is a dinucleotide AA in the *X* motif m_{AA^*}(*T. thermophilus*, 1461, 1475, 15) = G,GGC,GAA,GTC,GTA,AC of 15 nucleotide length aligned with the *X* motif m_{AA}(*E. coli*) with only one nucleotide difference (T in GGT replaced by C in *T. thermophilus*). However, the *X* motif m_{AA^*}(*T. thermophilus*) has a spatial structure far from the decoding center (Michel, 2012) and probably no function in modern rRNA of *T. thermophilus*.

3.1.2. Archaea

In rRNA of *P. furiosus*, the conserved dinucleotide AA occurs at the 2nd and 3rd sites of the trinucleotide GAA $\in X$ (Eq. (1)) and belongs to the *X* motif m_{AA}(*P. furiosus*, 1445, 1456, 12) = A,G**AA**,GTC,GTA,AC of 12 nucleotide length starting with the nucleotide A suffix of GAA,GTA $\in X$ (Eq. (1)) and then with a suffix of 11 nucleotides identical to the *X* motif m_{AA}(*E. coli*).

3.1.3. Nuclear eukaryotes

There are significant differences between prokaryotic and eukaryotic rRNAs, in particular eukaryotic 18S rRNAs are about 40% larger than prokaryotic 16S rRNAs. Nevertheless, some rRNA sites are conserved. In particular, the universally conserved dinucleotide AA in position 1492 of bacterial rRNAs occurs in eukaryotic rRNAs but at different positions: 1755 in *S. cerevisiae*, 1765 in *T. aestivum* (Fan-Minogue and Bedwell, 2008) and 1824 in *H. sapiens* (Bulygin et al., 2009).

In rRNA of *S. cerevisiae* and *H. sapiens*, the conserved dinucleotide AA is the prefix of the *X* motif m_{AA}(*S. cerevisiae*, 1755, 1764, 10) = m_{AA}(*H. sapiens*, 1824, 1833, 10) = **AA**,GTC,GTA,AC of 10 nucleotide length followed by two trinucleotides GTA,GTC $\in X$ (Eq. (1)) and ending with the dinucleotide AC prefix of ACC $\in X$ (Eq. (1)). In rRNA of *T. aestivum*, it occurs at the 2nd and 3rd sites of the trinucleotide

Table 2
Identification of *X* circular code motifs m_{AA} containing the universally conserved nucleotides A1492 and A1493 (in bold) in all studied rRNAs of bacteria, archaea, nuclear eukaryotes, and chloroplasts, and of *X* circular code motifs m_G containing the conserved nucleotide G530 (in bold) in rRNAs of bacteria and archaea.

PDB ID	Kingdom	Organism	*X* circular code motifs	Start	End	Length
3J5T	Bacteria	*E. coli*	m_{AA} = G,GGT,G**AA**,GTC,GTA,AC	1487	1501	15
			m_G = GC,G**G**T,AAT,AC	527	536	10
3I8G	Bacteria	*T. thermophilus*	m_{AA} = G,G**AA**,GGT,GC	1490	1498	9
			m_G = GC,**G**TT,ACC,C	528	536	9
3J20	Archaea	*P. furiosus*	m_{AA} = A,G**AA**,GTC,GTA,AC	1445	1456	12
			m_G = GC,G**G**T,AAT,ACC,GGC,GGC,C	480	497	18
3IZE	Eukaryote, nuclear	*S. cerevisiae*	m_{AA} = **AA**,GTC,GTA,AC	1755	1764	10
3J5Z	Eukaryote, nuclear	*T. aestivum*	m_{AA} = A,G**AA**,GTC,GTA,AC	1763	1774	12
3J3D	Eukaryote, nuclear	*H. sapiens*	m_{AA} = **AA**,GTC,GTA,AC	1824	1833	10
3BBN	Eukaryote, chloroplast	*S. oleracea*	m_{AA} = GT,G**AA**,GTC,GTA,AC	1438	1450	13

GAA ∈ *X* (Eq. (1)) and belongs to the *X* motif m_{AA}(*T. aestivum*, 1763, 1774, 12) = A,G**AA**,GTC,GTA,AC of 12 nucleotide length which is identical to the archaeal *X* motif m_{AA}(*P. furiosus*).

3.1.4. Chloroplasts

In rRNA of *S. oleracea*, the conserved dinucleotide AA occurs at the 2nd and 3rd sites of the trinucleotide GAA ∈ *X* (Eq. (1)) and belongs to the *X* motif m_{AA}(*S. oleracea*, 1438, 1450, 13) = GT,G**AA**, GTC,GTA,AC which is a suffix of 13 nucleotides of the bacterial *X* motif m_{AA}(*E. coli*).

3.1.5. X *circular code motif* m_{AA} *in the spatial structure of the ribosome decoding center of bacteria, archaea, nuclear eukaryotes, and chloroplasts*

In all studied rRNAs, the universally conserved dinucleotide AA precedes the trinucleotide GTC ∈ *X* (Eq. (1)), except in m_{AA}(*T. thermophilus*) where GTC is replaced by GGT. Thus, it always occurs at the 2nd and 3rd sites of a trinucleotide which is always GAA when the trinucleotide belongs to *X* (Eq. (1)). For one *X* motif m_{AA}(*S. cerevisiae*) = m_{AA}(*H. sapiens*), it does not belong to a trinucleotide of *X* (Eq. (1)). Figs. 1–7 show that the rRNA *X* motifs m_{AA}(*E. coli*), m_{AA}(*T. thermophilus*), m_{AA}(*P. furiosus*), m_{AA}(*S. cerevisiae*), m_{AA}(*T. aestivum*), m_{AA}(*H. sapiens*), and m_{AA}(*S. oleracea*) (purple with the conserved dinucleotide AA in red) of bacteria, archaea, nuclear eukaryotes and chloroplasts belong to the ribosome decoding center with spatial relations with mRNA *X* motifs (green) and tRNA *X* motifs (blue with the anticodon in black).

3.2. *Identification of* X *circular code motifs* m_G *containing the conserved nucleotide G in rRNAs of bacteria and archaea*

The bacterial conserved nucleotide G (in position 530 of the 16S rRNA in *E. coli*) has an experimentally proved biological function in

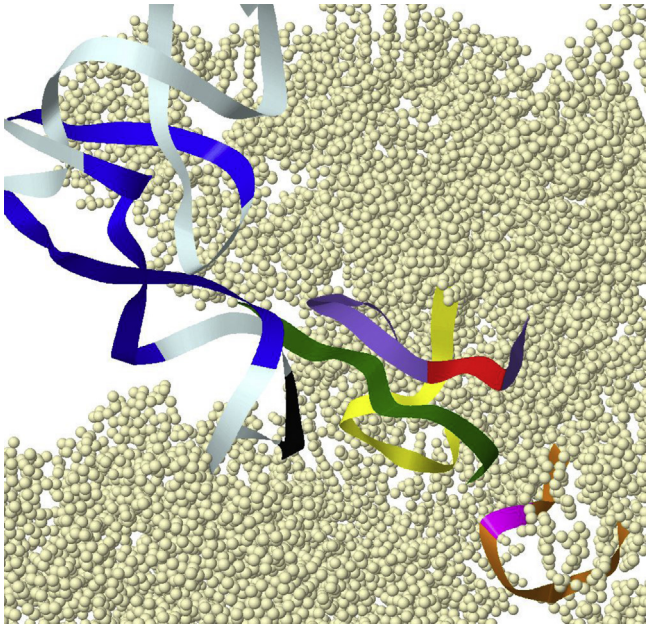

Fig. 5. *X* circular code motifs involved in the nuclear eukaryotic ribosome decoding center of *Triticum aestivum* (crystallographic structure PDB 3J5Z): the mRNA *X* motifs (green), the rRNA *X* motif m_{AA}(*T. aestivum*, 1763, 1774, 12) (purple with the conserved dinucleotide AA in red), the rRNA X motif m_G(*T. aestivum*, 578, 586, 9) (orange with the conserved nucleotide G in fuchsia), the rRNA *X* motif *m*(*T. aestivum*, 1641, 1649, 9) (yellow) and the tRNA *X* motifs (blue with the 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.)

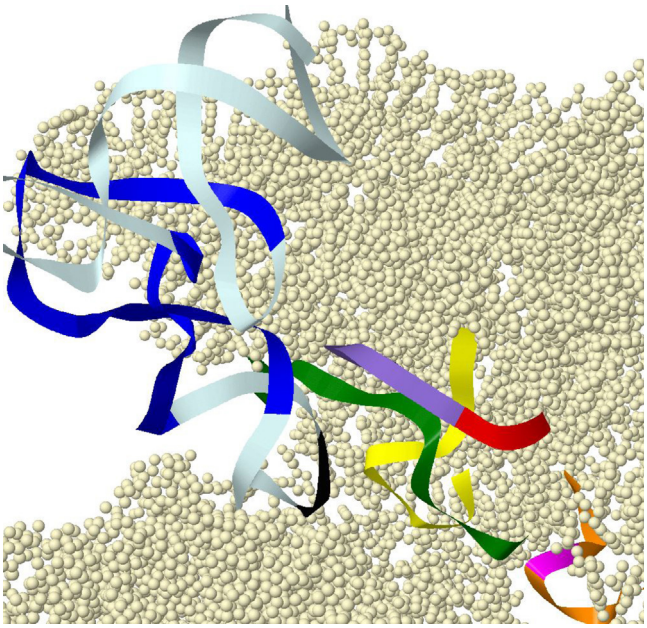

Fig. 4. *X* circular code motifs involved in the nuclear eukaryotic ribosome decoding center of *Saccharomyces cerevisiae* (crystallographic structure PDB 3IZE): the mRNA *X* motifs (green), the rRNA *X* motif m_{AA}(*S. cerevisiae*, 1755, 1764, 10) (purple with the conserved dinucleotide AA in red), the rRNA *X* motif m_G(*S. cerevisiae*, 574, 582, 9) (orange with the conserved nucleotide G in fuchsia), the rRNA *X* motif *m*(*S. cerevisiae*, 1633, 1641, 9) (yellow) and the tRNA *X* motifs (blue with the 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.)

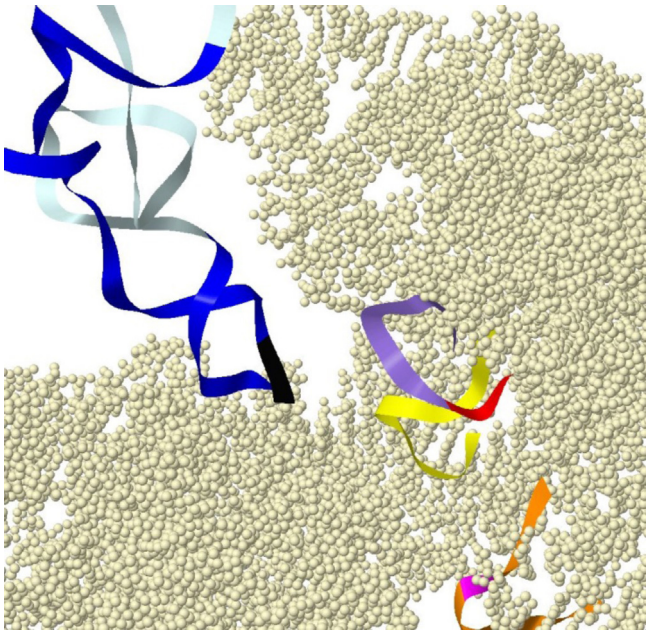

Fig. 6. *X* circular code motifs involved in the nuclear eukaryotic ribosome decoding center of *Homo sapiens* (crystallographic structure PDB 3J3D): the rRNA *X* motif m_{AA}(*H. sapiens*, 1824, 1833, 10) (purple with the conserved dinucleotide AA in red), the rRNA X motif m_G(*H. sapiens*, 623, 631, 9) (orange with the conserved nucleotide G in fuchsia), the rRNA *X* motif *m*(*H. sapiens*, 1697, 1705, 9) (yellow) and the tRNA *X* motifs (blue with the 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.)

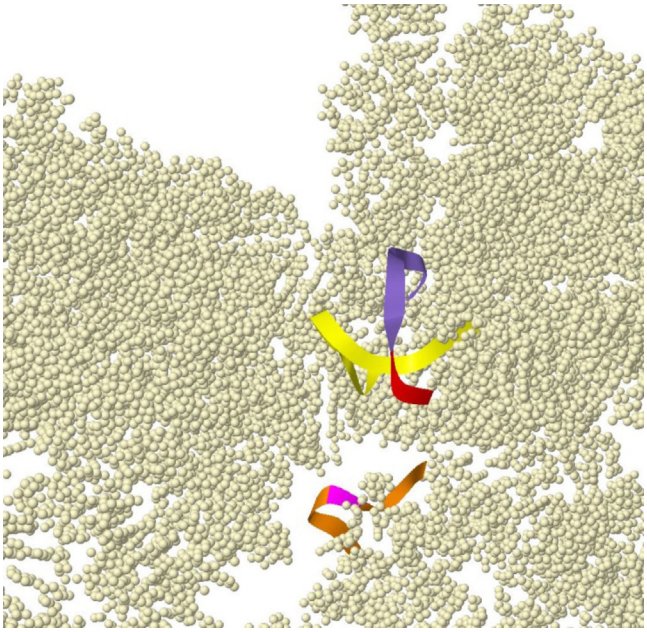

Fig. 7. *X* circular code motifs involved in the chloroplast eukaryotic ribosome decoding center of *Spinacia oleracea* (crystallographic structure PDB 3BBN): the rRNA *X* motif m_{AA}(*S. oleracea*, 1438, 1450, 13) (purple with the conserved dinucleotide AA in red), the rRNA X motif m_G(*S. oleracea*, 475, 481, 7) (orange with the conserved nucleotide G in fuchsia) and the rRNA *X* motif *m*(*S. oleracea*, 1345, 1353, 9) (yellow). The remaining rRNA (lemonchiffon) is outside the neighborhood of these *X* motifs, the mRNA and tRNA are 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.)

the codon–anticodon binding (Moazed and Noller, 1990; Powers and Noller, 1994; Yoshizawa et al., 1999). Very unexpectedly, this bacterial conserved G is also found to be located in *X* circular code motifs (Table 2).

3.2.1. Bacteria

In rRNA of *E. coli*, the conserved nucleotide G occurs at the 2nd site of the trinucleotide GGT $\in X$ (Eq. (1)) and belongs to the *X* motif m_G(*E. coli*, 527, 536, 10) = GC,G**G**T,AAT,AC of 10 nucleotide length starting with the dinucleotide GC suffix of GGC $\in X$ (Eq. (1)) followed by two trinucleotides AAT,GGT $\in X$ and ending with the dinucleotide AC prefix of ACC $\in X$ (Eq. (1)). In rRNA of *T. thermophilus*, the conserved nucleotide G occurs at the 1st site of GTT $\in X$ (Eq. (1)) and belongs to the *X* motif m_G(*T. thermophilus*, 528, 536, 9) = GC,**G**TT,ACC,C of nine nucleotide length starting with the dinucleotide GC, as in *E. coli*, followed by two trinucleotides ACC,GTT $\in X$ (Eq. (1)) and ending with the nucleotide C prefix of CAG,CTC,CTG $\in X$ (Eq. (1)). As with the conserved dinucleotide AA, the two rRNA *X* motifs m_G(*E. coli*) and m_G(*T. thermophilus*) have completely different primary structures and can only be revealed by the circular code theory.

3.2.2. Archaea

In rRNA of *P. furiosus*, the conserved nucleotide G occurs at the 2nd site of the trinucleotide GGT $\in X$ (Eq. (1)), as in m_G(*E. coli*), and belongs to the *X* motif m_G(*P. furiosus*, 480, 497, 18) = GC,G**G**T,AAT, ACC,GGC,GGC,C of 18 nucleotide length starting with the dinucleotide GC, as in m_G(*E. coli*) and m_G(*T. thermophilus*), followed by five trinucleotides AAT,ACC,GGC,GGT $\in X$ (Eq. (1)) and ending with the nucleotide C, as in m_G(*T. thermophilus*).

3.2.3. X *circular code motif* m_G *in the spatial structure of the ribosome decoding center of bacteria and archaea*

The bacterial *X* motif m_G(*E. coli*) is a prefix of 10 nucleotides of the archaeal *X* motif m_G(*P. furiosus*). The conserved nucleotide G occurs at the 2nd site of GGT $\in X$ (Eq. (1)) in m_G(*E. coli*) and m_G(*P. furiosus*), and at the 1st site of GTT $\in X$ (Eq. (1)) in m_G(*T. thermophilus*). Figs. 1–3 show that the rRNA *X* motifs m_G(*E. coli*), m_G(*T. thermophilus*) and m_G(*P. furiosus*) (orange with the conserved nucleotide G in fuchsia) of bacteria and archaea belong to the ribosome decoding center with spatial relations with mRNA *X* motifs (green) and tRNA *X* motifs (blue with the anticodon in black).

3.3. Identification of X *circular code motif* m_G *in the ribosome decoding center of nuclear eukaryotes and chloroplasts*

The search of the conserved nucleotide G in rRNAs of eukaryotes is impossible using this information alone. By using the circular code theory, the conserved nucleotide G in rRNAs of bacteria and archaea occurs in *X* motifs m_G. The rRNA *X* motifs m_G(*E. coli*) and m_G(*P. furiosus*) have the common *X* motif m_G({*E. coli*, *P. furiosus*}) = GC,G**G**T,AAT,AC, and the rRNA *X* motif m_G(*T. thermophilus*) is GC,**G**TT,ACC,C (Table 2). By using the Algorithm_Search_Xmotif_seqCommon (Section 2.2) with the global multiple sequence alignment ClustalW2 (gap open of 10 and gap extension of 0.2 and 0.1 in case of pairwise), these two *X* motifs m_G are searched in rRNAs of nuclear eukaryotes *S. cerevisiae*, *T. aestivum*, *H. sapiens* and chloroplasts *S. oleracea*. Very surprisingly, a common *X* motif m_G is identified in rRNAs of nuclear eukaryotes: m_G(*S. cerevisiae*, 574, 582, 9) = m_G(*T. aestivum*, 578, 586, 9) = m_G(*H. sapiens*, 623, 631, 9) = m_G(*NuclarEukaryotes*) = GC,GGT,AAT,T (Table 3). Furthermore, a *X* motif m_G is also identified in rRNA of chloroplasts: m_G(*S. oleracea*, 475, 481, 7) = m_G(*Chloroplasts*) = GC,GGT,AA (Table 3) which is a prefix of seven nucleotides of m_G(*NuclarEukaryotes*). As the common *X* motif m_G({*NuclarEukaryotes*, *Chloroplasts*}) = GC,GGT,AA is a prefix of the common *X* motif m_G({*E. coli*, *P. furiosus*}) = GC,G**G**T,AAT,AC,

Table 3
Identification of *X* circular code motifs m_G in rRNAs of nuclear eukaryotes and chloroplasts, and *m* in all studied rRNAs of bacteria, archaea, nuclear eukaryotes, and chloroplasts.

PDB ID	Kingdom	Organism	*X* circular code motifs	Start	End	Length
3J5T	Bacteria	*E. coli*	*m* = AC,ACC,GCC,C	1396	1404	9
3I8G	Bacteria	*T. thermophilus*	*m* = AC,ACC,GCC,C	1375	1383	9
3J20	Archaea	*P. furiosus*	*m* = AC,ACC,GCC,C	1356	1364	9
3IZE	Eukaryote, nuclear	*S. cerevisiae*	m_G = GC,GGT,AAT,T	574	582	9
			m = AC,ACC,GCC,C	1633	1641	9
3J5Z	Eukaryote, nuclear	*T. aestivum*	m_G = GC,GGT,AAT,T	578	586	9
			m = AC,ACC,GCC,C	1641	1649	9
3J3D	Eukaryote, nuclear	*H. sapiens*	m_G = GC,GGT,AAT,T	623	631	9
			m = AC,ACC,GCC,C	1697	1705	9
3BBN	Eukaryote, chloroplast	*S. oleracea*	m_G = GC,GGT,AA	475	481	7
			m = AC,ACC,GCC,C	1345	1353	9

Table 4
X circular code motifs identified by multiple sequence alignment of the seven studied 16S rRNAs *E. coli*, *T. thermophilus*, *P. furiosus*, *S. cerevisiae*, *T. aestivum*, *H. sapiens*, and *S. oleracea*.

E. coli (3J5T)			*T. thermophilus* (318G)			*P. furiosus* (3J20)			*S. cerevisiae* (3IZE)			*T. aestivum* (3J5Z)			*H. sapiens* (3J3D)			*S. oleracea* (3BBN)			
X motif	Start	End	*X* motif	Start	End	*X* motif	Start	End	*X* motif	Start	End	*X* motif	Start	End	*X* motif	Start	End	*X* motif	Start	End	Length
TA,GTA,GGT,G	249	257	TA,GTT,GGT,G	241	259	TA,GTT,GGT,G	245	253	TC,GAT,GGT,A	320	328	TC,GAT,GGT,A	324	332	TC,GAT,GGT,A	368	376	TA,GTT,GGT,G	220	228	9
TA,GCC,GAC,GA	273	282	AA,GCC,GAC,GA	265	274	AA,GCC,GAA,GA	269	278	AT,GGT,TTC,AA	344	353	AT,GGT,GGT,GA	348	357	AT,GGT,GAC,CA	392	401	AA,GGC,GAT,GA	244	253	10
GAG,GGT,GC	538	545	GAG,GGC,GC	518	525	C,GGC,GGC,C	490	497	CAG,CTC,CA	584	591	CAG,CTC,CA	588	595	CAG,CTC,CA	633	640	GAG,GAT,GC	486	493	8
C,GGT,GAA,AT	689	697	C,GGT,GAA,AT	669	677	G,GGT,GAA,AT	643	351	A,GGT,GAA,AT	900	908	A,GGT,GAA,AT	905	913	A,GGT,GAA,AT	957	965	C,GGT,GAA,AT	637	645	9
GC,AAC,GAG,C	1099	1107	GC,AAC,GAG,C	1078	1086	GT,AAC,GAG,C	1051	1059	AT,AAC,GAA,C	1319	1327	TT,AAC,GAA,C	1323	1331	AT,AAC,GAA,C	1376	1384	GC,AAC,GAG,C	1048	1056	9
AC,GGT,GAA,TA	1368	1377	GC,GGT,GAA,TA	1347	1356	GC,GGC,GAA,TA	1328	1337	GC,GTT,GAT,TA	1605	1614	GC,GTT,GAC,TA	1613	1622	GC,GTT,GAT,TA	1669	1678	GC,GGT,GAA,TT	1317	1326	10
AC,ACC,GCC,C	1396	1404	AC,ACC,GCC,C	1375	1383	AC,ACC,GCC,C	1356	1364	AC,ACC,GCC,C	1633	1641	AC,ACC,GCC,C	1641	1649	AC,ACC,GCC,C	1697	1705	AC,ACC,GCC,C	1345	1353	9
AA,GTC,GTA,AC	1492	1501	AA,GTC,GTA,AC	1466	1475	AA,GTC,GTA,AC	1447	1456	AA,GTC,GTA,AC	1755	1764	AA,GTC,GTA,AC	1765	1774	AA,GTC,GTA,AC	1824	1833	AA,GTC,GTA,AC	1441	1450	10

we can make the realistic hypothesis that the conserved nucleotide G in nuclear and chloroplast rRNAs occurs at the 2nd site of the trinucleotide GGT $\in X$ (Eq. (1)).

Furthermore, Figs. 4–7 show that the rRNA *X* motifs m_G(*S. cerevisiae*), m_G(*T. aestivum*), m_G(*H. sapiens*) and m_G(*S. oleracea*) (orange with the conserved nucleotide G in fuchsia) of nuclear eukaryotes and chloroplasts belong to the ribosome decoding center with spatial relations with mRNA *X* motifs (green) and tRNA *X* motifs (blue with the anticodon in black).

3.4. Identification of a potentially important X circular code motif m in the ribosome decoding center of bacteria, archaea, nuclear eukaryotes, and chloroplasts

By using the Algorithm_Search_Xmotif_seqCommon (Section 2.2) with the global multiple sequence alignment ClustalW2, an *X* motif *m* is identified which is universally conserved in rRNAs of bacteria, archaea, nuclear eukaryotes, and chloroplasts (Table 3): *m* = AC,ACC,GCC,C of nine nucleotide length starting with the dinucleotide AC suffix of AAC,GAC, TAC $\in X$ (Eq. (1)) followed by two trinucleotides ACC,GCC $\in X$ and ending with the nucleotide C prefix of CAG,CTC,CTG $\in X$ (Eq. (1)). The start and end positions of the *X* motif *m* in the seven studied organisms are given in Table 3.

Very unexpectedly, Figs. 1–7 show that the universally conserved rRNA *X* motif *m* (yellow) of bacteria, archaea, nuclear eukaryotes, and chloroplasts belongs to the ribosome decoding center with spatial relations with mRNA *X* motifs (green) and tRNA *X* motifs (blue with the anticodon in black).

3.5. Identification of six X circular code motifs less conserved

Table 4 shows eight *X* circular code motifs identified by multiple sequence alignment of the seven studied 16S rRNAs *E. coli*, *T. thermophilus*, *P. furiosus*, *S. cerevisiae*, *T. aestivum*, *H. sapiens*, and *S. oleracea*. There are two strictly conserved *X* motifs:

(i) AA,GTC,GTA,AC of 10 nucleotide length which is a suffix of the *X* motifs m_{AA} for *E. coli*, *P. furiosus*, *S. cerevisiae*, *T. aestivum*, *H. sapiens* and *S. oleracea*; and m_{AA^*} for *T. thermophilus* (studied in Section 3.1 and Table 2).
(ii) AC,ACC,GCC,C of nine nucleotide length which is the *X* motif *m* (studied in Section 3.4 and Table 3).

The six other *X* circular code motifs occurring at the same position in the multiple sequence alignment are less conserved. However, they may also have a biological function in modern or primitive translation codes.

4. Conclusion

The results obtained here bring several new contributions to the ribosome decoding center, in particular to its primary structure which is related to the mathematical property of circular code. The universally conserved dinucleotide AA in all studied rRNAs of bacteria, archaea, nuclear eukaryotes, and chloroplasts belongs to *X* motifs (m_{AA}; Table 2). The conserved nucleotide G in rRNAs of bacteria and archaea also belongs to *X* motifs (m_G; Table 2). The development of the Algorithm_Search_Xmotif_seqCommon (Section 2.2) associated with the global multiple sequence alignment allows to identify the *X* motifs m_G in nuclear and chloroplast rRNAs (Table 3). Furthermore, it reveals a new *X* motif (*m*) which is universally conserved in the seven studied organisms (Table 3). Finally, the three *X* motifs m_{AA}, m_G, and *m* belong to the ribosome decoding center in all studied rRNAs of bacteria, archaea, nuclear eukaryotes, and chloroplasts (Figs. 1–7).

Several biological considerations can be stressed from these results. The function of the ribosome decoding center which has been attributed to a very few nucleotides only, precisely the dinucleotide AA and the nucleotide G, may be related to motifs containing at least two successive trinucleotides up to a maximum of five successive trinucleotides (Tables 2 and 3). Furthermore, all these motifs are *X* circular code motifs; thus, motifs with the circular code property *CC* (Definition 5) allowing retrieval of the reading frame, the property C^3 (Definition 7) allowing retrieval of the two shifted frames and the complementary property *SCC* (Definition 6) allowing pairing with the *X* circular code motifs of mRNAs and tRNAs. These results 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-A 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-A 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.

Bulygin, K., Baouz-Drahy, S., Graifer, D., Favre, A., Karpova, G., 2009. Sites of 18S rRNA contacting mRNA 3′ and 5′ of the P site codon in human ribosome: a cross-linking study with mRNAs carrying 4-thiouridines at specific positions. Biochim. Biophys. Acta 1789, 167–174.

Fan-Minogue, H., Bedwell, D.M., 2008. Eukaryotic ribosomal RNA determinants of aminoglycoside resistance and their role in translational fidelity. RNA 14, 148–157.

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.

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.

Moazed, D., Noller, H.F., 1990. Binding of tRNA to the ribosomal A and P sites protects two distinct sets of nucleotides in 16S rRNA. J. Mol. Biol. 211, 135–145.

Powers, T., Noller, H.F., 1994. Selective perturbation of G530 of 16 S rRNA by translational miscoding agents and a streptomycin-dependence mutation in protein S12. J. Mol. Biol. 235, 156–172.

Sharma, M.R., Wilson, D.N., Datta, P.P., Barat, C., Schluenzen, F., Fucini, P., Agrawal, R. K., 2007. Cryo-EM study of the spinach chloroplast ribosome reveals the structural and functional roles of plastid-specific ribosomal proteins. Proc. Natl. Acad. Sci. U. S. A. 104, 19315–19320.

Wilson, D.N., 2014. Ribosome-targeting antibiotics and mechanisms of bacterial resistance. Nat. Rev. Microbiol. 12, 35–48.

Yoshizawa, S., Fourmy, D., Puglisi, J.D., 1999. Recognition of the codon–anticodon helix by ribosomal RNA. Science 285, 1722–1725.