Biochemical and Biophysical Research Communications 388 (2009) 127–130

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

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

Intramolecular triple helix as a model for regular polyribonucleotide $(CAA)_n$

Alexander V. Efimov *, Alexander S. Spirin

Institute of Protein Research, Russian Academy of Sciences, Pushchino, Moscow Region, Russia 142290, Russian Federation

ARTICLE INFO

Article history:
Received 23 July 2009
Available online 30 July 2009

Keywords:
Tobacco mosaic virus (TMV) RNA
Omega sequence
Regular poly$(CAA)_n$ sequence
Enhancing leaders for translation initiation
Non-canonical base pairing
Base triads
Stacking interactions

ABSTRACT

The regular $(CAA)_n$ polyribonucleotide, as well as the omega leader sequence containing (CAA)-rich core, have recently been shown to form cooperatively melted and compact structures. In this report, we propose a structural model for the $(CAA)_n$ sequence in which the polyribonucleotide chain is folded upon itself, so that it forms an intramolecular triple helix. The triple helix is stabilized by hydrogen bonding between bases thus forming coplanar triads, and by stacking interactions between the base triads. A distinctive feature of the proposed triple helix is that it does not contain the canonical double-helix elements. The difference from the known triple helices is that Watson–Crick hydrogen bond pairings do not take place in the interactions between the bases within the base triads.

© 2009 Elsevier Inc. All rights reserved.

Introduction

A number of 5′-untranslated regions of mRNAs, called also leader sequences, are known to enhance initiation of translation in eukaryotes or prokaryotes. The leader sequence of tobacco mosaic virus RNA, the so-called omega sequence, is one of the best studied translational enhancer for eukaryotic translation systems. It consists of about 70 nucleotides and contains from 10 to 13 CAA repeats [1,2]. It was shown that the $(CAA)_n$ region is the most critical part for enhancing translation [2]. Moreover, it was demonstrated that the synthetic regular $(CAA)_n$ sequence also can be an efficient leader for initiation of translation [3]. Although little or no secondary structure of the Watson–Crick type is predicted to exist within the omega sequence, both the omega sequence and the regular $(CAA)_n$ sequence have been recently shown to form cooperatively melted, compact structures [4]. In contrast, a statistical (C,A) copolymer of an average length of about 80 nucleotides with the C-to-A ratio of 1:2 does not form a compact and cooperatively melted structure [4].

In this report, we propose a structural model for the $(CAA)_n$ sequence in which the polyribonucleotide chain is folded upon itself so that it forms an intramolecular triple helix. The triple helix is formed by three strands adjacent along the chain and packed so that the first and third strands are parallel and the second strand is antiparallel to them. The triple helix is stabilized by hydrogen bonding between A and C, and A and A, thus forming A:C:A, C:A:A and A:A:C triads of roughly equivalent size and geometry, and by stacking interactions between the base triads formed.

The triple helices of DNA and RNA described so far consist of the canonical double helices and the third strands that occupy the major groove of the double-helix forming Hoogsteen or reverse Hoogsteen pairs with the purines of Watson–Crick base pairs; these are polyU·polyA·polyU, polyI·polyA·polyI, poly(dA)·[poly(dT)]$_2$, etc. [5–10]. A distinctive feature of our model is that Watson–Crick hydrogen bond pairings do not take place in the interactions between the strands and that no canonical double helix is present there.

Materials and methods

The atomic coordinates of known structures of DNA and RNA have been obtained from the Protein Data Bank (http://www.rcsb.org/pdb). The stereochemical analysis of hydrogen-bonding and stacking interactions in the known structures of nucleic acids was performed by visual inspection of the structures with the computer graphics program RasMol [11]. Possible base pairs and base triads were modeled using the CPK space filling atomic models.

Results

The fact that a statistical (C,A) copolymer with the C-to-A ratio of 1:2 does not form a compact stable structure [4] suggests that the regularity of the core region of the omega leader and the $(CAA)_n$ sequences is critical for the formation of the structure. The regular primary structure should result in a regular three-dimensional structure. We propose that $(CAA)_n$ polyribonucleotide folds upon itself to form a regular intramolecular triple helix (Fig. 1). In order

* Corresponding author. Fax: +7 495 632 7871.
E-mail address: efimov@protres.ru (A.V. Efimov).

0006-291X/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.bbrc.2009.07.133

Fig. 1. Schematic representation of polyribonucleotide $(CAA)_n$ folds in two variants of the triple helix. Arrows indicate chain polarity of the stands. Small filled circles show hydrogen bonding within the base triads.

to be regular, the triple helix should be formed by base triads C:A:A, A:C:A, and A:A:C, having the same or nearly the same C1′–C1′ distances and angles between the corresponding glycosidic bonds. In other words, the regular triple helix can be formed if the base triads are isomorphous (or close to being isomorphous).

Fig. 2 represents possible hydrogen-bond interactions between two A's (I, II, III), or A and C (IV, V), or C and C (VI, VII), provided they can form at least two hydrogen bonds (for further descriptions, see [8,12]). Pairs I, II, VI and VII have dyad axes oriented perpendicular to the figure plane and consequently can be incorporated into a parallel duplex, with all nucleotides in the *anti* glycoside torsional conformation, or all in the *syn* conformation. Similarly, pair V can also be included into a parallel duplex. In contrast, pairs III and IV can be included into an antiparallel duplex, if both nucleotides in each pair have identical χ angles. Pairs III and IV can be built into a parallel duplex if one nucleotide in each pair adopts the *anti* and the other the *syn* glycosidic conformation. For the construction of a parallel duplex between the first and third strands (Fig. 1), all purine nucleotides must be in the *syn* and all pyrimidine nucleotides in the *anti* conformation, by analogy with Z-DNA [13]. In order to construct a regular parallel duplex, we have selected only those pairs which possess the same or near the same C1′–C1′-distances and angles between glycosidic bonds. Our analysis have shown that pairs II (A:A), IV (A:C and C:A), and VI (C:C) are the best suited for each other.

In the triple helix (Fig. 1), the second strand is antiparallel to the first and third strands forming the parallel duplex. Taking into account this fact, possible base triads are obtained by addition of a third nucleotide having an appropriate orientation and conformation to pairs II, IV and VI (Fig. 2) which have been selected for construction of the regular parallel duplex. Fig. 3 represents the triads obtained in this way. The triads I, II and III are best suited for the triple helix (Fig. 1) as they represent all three possible permutations of two adenines and one cytosine, A:C:A, A:A:C, and C:A:A. On the other hand, the corresponding C1′–C1′ distances and angles between glycosidic bonds are similar in these triads and consequently they can be incorporated into a quite regular triple helix.

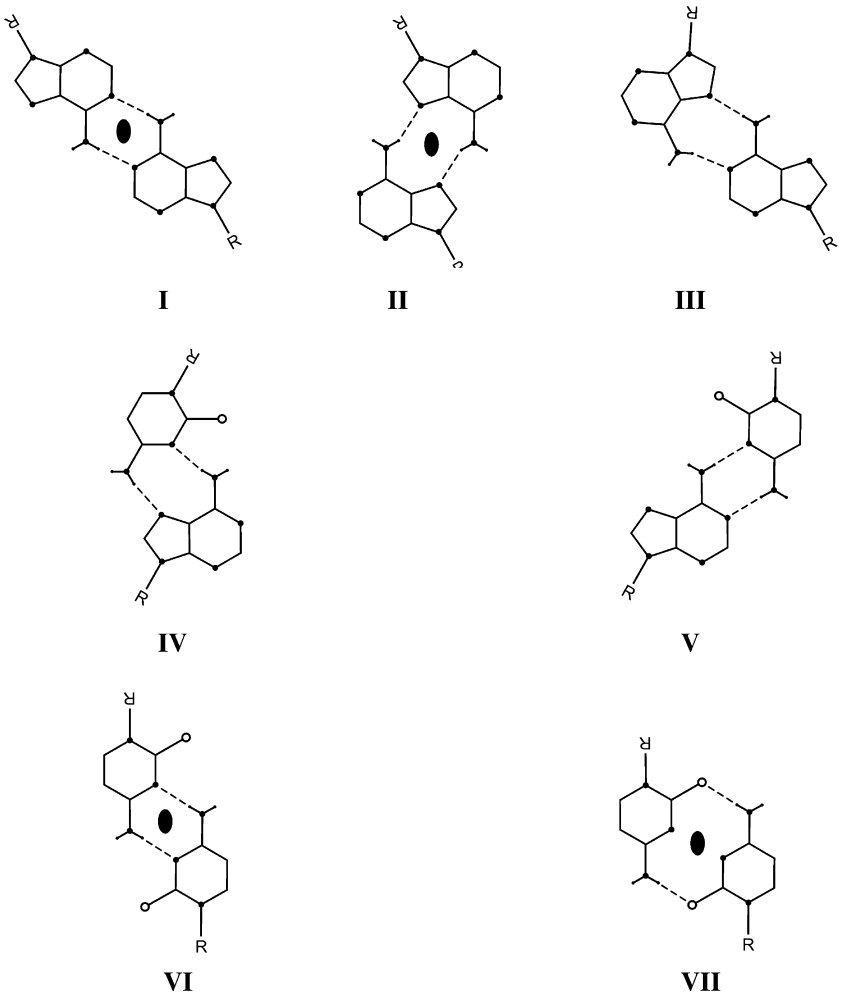

Fig. 2. Possible base pairs made of A and C connected by at least two hydrogen bonds (dashed lines). Filled ovals show dyad axes oriented perpendicular to the figure plane.

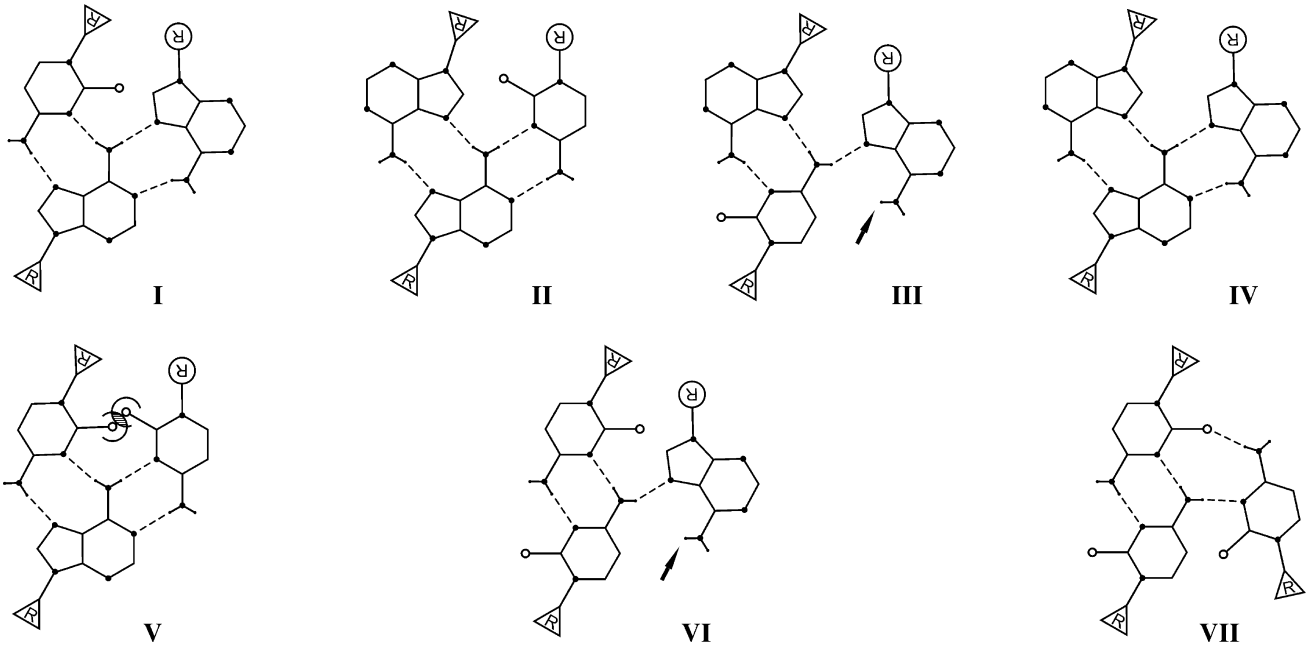

Fig. 3. Hydrogen bond and sterical interactions in the base triads described in the text. The orientation of the strands in the parallel duplex is indicated by triangles and that of the third strand (antiparallel to them) by circles at the positions of the C1′ atoms. In triad VII, all strands are parallel. All A nucleotides adopt the *syn* glycosidic conformation and all C nucleotides are in the *anti* conformation.

It should be noted that in triad III (Fig. 3) the N^6H group of adenine indicated by an arrowhead does not have a partner to form a hydrogen bond; this will somewhat destabilize the triple helix. Similar hydrogen-bond interactions take place in triad VI. However, the latter triad is not compatible with the regular triple helix formed by the $(CAA)_n$-sequence (Fig. 1) because it is made of two C and one A nucleotides. In triad VII, the added cytosine forms two hydrogen bonds with the C:C-pair, but this triad is not compatible with the triple helix because it is made of three C nucleotides and all its strands are parallel to each other. Triad VI can be transformed into a structure similar to triad VII by rotation of the A nucleotide; however, the obtained triad C:C:A is not compatible with the triple helix either. It should be noted that triad IV made of three A nucleotides has C1′–C1′ distances and the orientation of glycosidic bonds similar to those in triads I, II, and III. This implies that triad IV can be incorporated into the triple helix (indeed, the regular $(CAA)_n$ core of the natural omega sequence contains some extra A's). An interesting possibility is whether the poly(A) sequence, which has been shown to be also a translational enhancer [14], may be capable of forming a triple helix similar to that of the $(CAA)_n$ sequence.

There are two variants of the triple helix which have different hydrogen bond networks and arrangements of strands (Fig. 1). In one variant, the bases of the first strand form hydrogen bonds with the bases of both the second and the third strands and there are no hydrogen-bond interactions between the bases of the second and the third strands (see also triads I, II and III in Fig. 3). In the other variant, the bases of the third strand form hydrogen bonds with the bases of both the first and the second strands and there are no hydrogen-bond interactions between the bases of the first and the second strands. At the present stage of modeling, we are not able to decide which variant is better because the structure of the end loops is not modeled.

Discussion

The main result of this paper is that the intramolecular triple helix is demonstrated to be a plausible structure for the regular $(CAA)_n$ polyribonucleotide. At the present stage of modeling, we propose only the overall fold of the $(CAA)_n$ sequence without considering the end-loop regions and possible sequence-dependent defects of the triple helix. In any case, the proposed model is in an agreement with the known experimental data, first of all, with the temperature melting experiments showing a significant cooperative stability, and the sedimentation data demonstrating a high compactness of the three-dimensional structure [4]. It is noteworthy that the triple helix proposed can be formed only by the regular $(CAA)_n$ sequence, but not by a statistical (C,A) copolymer with the C-to-A ratio of 1:2. As it results from our stereochemical analysis, the base triads that include two or three C nucleotides, i.e., C:C:A, C:A:C, A:C:C and C:C:C, are unfavorable or prohibited (Fig. 3).

The next step in the study of the compact structure of $(CAA)_n$ polyribonucleotide can be the experimental testing of the model by means of chemical and enzymatic modifications. At the same time we understand that the final solution of the problem requires a direct approach to the atomic structure with the use of either NMR or X-ray diffraction technique. On the other hand, it seems to be important to check whether the compact structure – presumably the triple helix structure – of the $(CAA)_n$ sequence is necessary for the enhancing activity of the sequence in functional experiments, including its activity in binding of ribosomal particles and the factors required for efficient translation initiation.

Acknowledgment

The work was supported by the Program on Molecular and Cell Biology of the Russian Academy of Sciences.

References

[1] P. Coelet, G.P. Lomonossoff, P.G.J. Butler, M.E. Akam, M.J. Gait, J. Karn, Nucleotide sequence of tobacco mosaic virus RNA, Proc. Natl. Acad. Sci. USA 79 (1982) 5818–5822.
[2] D.R. Gallie, V. Walbot, Identification of the motifs within the tobacco mosaic virus 5′-leader responsible for enhancing translation, Nucleic Acid Res. 20 (1992) 4631–4638.

[3] T.M.A. Wilson, K. Saunders, M.J. Dowson-Day, D.E. Sleat, H. Trachsel, K.W. Mundry, Effects of the 5′-leader sequence of tobacco mosaic virus RNA, or derivatives thereof, on foreign mRNA and native viral gene expression, in: E.G. McCarthy, M.F. Tuite (Eds.), Post-transcriptional Control of Gene Expression, NATO ASI Series, vol. H49, Springer Verlag, Berlin/Heidelberg, 1990, pp. 261–275.
[4] A.A. Kovtun, N.E. Shirokikh, A.T. Gudkov, A.S. Spirin, The leader sequence of tobacco mosaic virus RNA devoid of Watson–Crick secondary structure possesses a cooperatively melted, compact conformation, Biochem. Biophys. Res. Commun. 358 (2007) 368–372.
[5] G. Felsenfeld, A. Rich, Studies on the formation of two- and, three-stranded polyribonucleotides, Biochim. Biophys. Acta 26 (1957) 457–468.
[6] J.-S. Sun, C. Helene, Oligonucleotide-directed triple-helix formation, Curr. Opin. Struct. Biol. 3 (1993) 345–356.
[7] M.D. Frank-Kamenetskii, S.M. Mirkin, Triplex DNA structures, Annu. Rev. Biochem. 64 (1995) 65–95.
[8] W. Saenger, Principles of Nucleic Acid Structure, Springer-Verlag, New York/Berlin/Heidelberg/Tokyo, 1984.
[9] V. Sklenar, J. Feigon, Formation of a stable triplex from a single DNA strand, Nature 345 (1990) 836–838.
[10] R. Haner, P.B. Dervan, Single-strand DNA triple-helix formation, Biochemistry 42 (1990) 9761–9765.
[11] R.A. Sayle, E.J. Milner-White, RASMOL – biomolecular graphics for all, Trends Biochem. Sci. 20 (1995) 374–376.
[12] M.E. Burkard, D.H. Turner, I. Tinoco Jr., Structures of base pairs involving at least two hydrogen bonds, in: R.F. Gesteland, T.R. Cech, J.F. Atkins (Eds.), The RNA World, second ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor/New York, 1999, pp. 675–680.
[13] A.H.-J. Wang, G.J. Quigley, F.J. Kolpak, G. van der Marel, J.H. van Boom, A. Rich, Left-handed double helical DNA: variations in the backbone conformation, Science 211 (1981) 171–176.
[14] A.T. Gudkov, M.V. Ozerova, V.M. Shiryaev, A.S. Spirin, 5′-poly(A) sequence as an effective leader for translation in eukaryotic cell-free systems, Biotech. Bioeng. 91 (2005) 468–473.