Review Article

A Review of RNA Motifs, Identification Algorithms and their Function on Plants

Davoud Naderi[1*], Raheba Jami[2], Fazal ur Rehman[3]

[1] *Agricultural Biotechnology Research Institute, University of Zabol, Zabol, Iran*
[2] *Department of Horticulture, Faculty of Agriculture, Herat University, Herat, Afghanistan*
[3] *Department of Plant Pathology, College of Agriculture, University of Sargodha, Punjab, Pakistan*

Use yor device to scan and read the article online

Citation Naderi D, Jami R, Rehman F U. (2021). A Review of RNA Motifs, Identification Algorithms, and their Function on Plants. *J. Agri. Sci. Eng.* 3(1): 28-40

https://doi.org/10.22034/jpbb.2021.271442.1001

Article info
Received: 02 February 2021
Accepted: 28 May 2021
Available Online: 28 May 2021
Checked for Plagiarism: Yes
Peer Reviewers Approved by:
Bahman Fazeli-Nasab
Editor who Approved Publication:
Dr. Behzad Ghareyazi

Keywords:
Amino acid, Algorithms, Functional motifs, ORRM family, Structural motifs

ABSTRACT

Recent Genomic research has significantly developed human knowledge on structural non-coding RNAs (ncRNAs) which folds into characteristic secondary structures and performs specific-structure dependent biological functions. Hence, RNA secondary structure prediction is among the most commonly evaluated issues in computational RNA biology. The aim of the present study is to introduce the key role of RNA motifs in biological processes. Such motifs are specifically effective in regulating gene expression, maintaining structure and strength of RNA molecule, splicing the early mRNA, and providing the most appropriate recognition site for protein binding. Doing research and analyzing RNA motifs requires using several methods and algorithms which can provide the structure and properties of these building blocks in organisms. Generally, the most successful computational methods used in organizing RNA include the QRNA, RNAz, and CMfinder algorithms, which correlate nucleotides and other features of RNA structure formation and maintenance. In plants, RNA Recognition Motif (RRM), an 80- amino acid protected motif, is one of the most abundant protein motifs in eukaryotes. This motif has various important roles like participating in growth processes and responding to stresses in plants, and accelerating different metabolic processes. In addition, further analysis of ORRM family members represented presence of ORRM2, ORRM3 and ORRM4 acting as RNA editing factors in mitochondria as well as ORRM6 which is a chloroplast RNA editing factor. Among the construction motifs, the Pseudoknot motif which contributes to several biological activities such as altering expression of pathogenic genes in some viruses and formation of telomerase and self-truncating introns is of great significance, since these are important breeding factors in biotechnology. Based on the results of the study, it can be proposed that further studies on bioinformatics analysis of plant motifs are required to be implemented to open new windows on controlling pathogens in plants.

1. Introduction

Intermediates between coding DNA sequences and proteins, RNAs have been ascribed a plethora of novel roles in the past decades. It is recognized as active regulatory molecules, influencing processes such as chromatin organization, genome stability or gene expression that affect all aspects of a plant's life [1]. As RNA are being synthesized, they fold into secondary structures, which can influence RNA metabolism at multiple levels, including

***Corresponding Author:** *Davoud Naderi (naderi.davoud02@gmail.com)*

transcript processing and stability, or protein translation [2]. After transcription, RNAs require proper maturation through a multi-step process integrating capping, splicing, and polyadenylation. Pieczynski, reported the importance of accurate splicing of the precursor mRNA (pre-mRNA) encoding the small subunit of the Arabidopsis nuclear cap-binding complex [3]. Within the RNA molecule, structural elements (motifs) are so organized that they can maintain the stability of RNA structure. Structural motifs also stabilize RNA strands through hydrophobic interactions just in the form of extensive networks in hydrogen bonds [4].

Motif is defined as a short and protected portion of amino acids or nucleotides (in DNA or RNA strand). Specifically in genetics, a sequence motif is considered as a pattern of nucleotides or amino acids with a great biological significance. For proteins, the sequence motif is distinct from the structural motif. Most structural motifs are also structural domains. The domain refers to part of protein that folds independently. Similarly, a great number of structural motifs are also secondary structures which fold separately. In a large biological molecule like a protein, a structural motif such as β-Hairpins, α-Helix, and β-α-β is adjacent to several secondary structures which are smaller in size than the second protein [5].

Motifs, present at the levels of DNA, RNA, and proteins, can be divided into different types in each of these molecules. For proteins, motif is defined as a special combination of secondary structures. A three-dimensional structure which is produced through a special sequence of amino acids can also lead to a specific function in that protein. However, for nucleic acids, motif is a short nucleotide sequence that creates a recognition site to which other proteins bind. The first DNA-binding protein motif, the Helix-turn-helix motif, which was identified in bacterial proteins included two alpha helixes linked by a short chain of amino acids active in binding proteins to DNA [6]. The basic structure of ribonucleic acid (RNA) is a nucleotide that contains ribose sugar, a base, and a phosphate group. These nucleotides bind in a chain of polynucleotides that defines the primary structure of RNA.

A number of proteins which can detect specific sequences in RNA molecules or computational models of RNA are called motifs and are commonly used in *in vitro* experiments like RNA Compete [7] Bind-n-Seq RNA [8] or Systematic Evolution of Ligands by Exponential Enrichment [9]. In order to study and interpret RNA motifs, several methods and algorithms are required to provide the structure and other properties of these building blocks in organisms [10]. Moreover, most of these algorithms have been successfully used to identify non-coding RNAs and to discover new classes of motifs in the genomes of organisms such as fungi [11]. Successful computational methods which are used to organize RNA include QRNA [12], RNAz [13], and CMfinder [14]. It is obvious that these algorithms correlate with nucleotides and other features of RNA structure formation and maintenance. In this study, the CMfinder method with the MEME was used to detect the secondary structure of RNA from non-homogeneous sequences by covariance models [15]. To improve the prediction accuracy of the agreement secondary structure, a thermodynamic model was included in the CMfinder algorithm to predict base pairs [16].

In practice, identifying a motif is more complicated than what it seems. Given the high significance of motifs in biological systems which led to the advent of new genes, further efforts in this field seem important. The aim of this study was accordingly to introduce RNA motifs which have valuable functions in biological processes. Such motifs are generally effective in regulating gene expression, maintaining the structure and strength of the RNA molecule, splicing the early mRNA, and providing the most suitable recognition site for protein binding. In this research, some of these motifs were expressed and their significance was explained separately.

2. Motif recognition algorithms

2. 1. QRNA algorithm

The QRNA algorithm (QRNA) is identifiable among other algorithms with the ID OMICS_15802. It is an algorithm which can provide a gene-finder prototype of non-coding RNA based on comparative analysis of genome sequences. The QRNA algorithm detects both second protected RNA structures including ncRNA genes and the regulatory structures of the cis RNA molecule. It also uses three different probabilistic models, namely, probabilistic RNA structure, probabilistic coding sequence, and position-based evaluation, to test the pattern of mutations in two sequences [12]. Under the GNU General Public License, this algorithm is available for free at (http://www.genetics.wustl.edu/eddy/software/).

2.2. RNAz algorithm

The RNAz algorithm is a program used for predicting the protected structure and the thermodynamically stable secondary structure of an RNA molecule in multiple sequence alignments. This program can also be useful in extensive genome representation to discover and track RNA functional structures among which the most important are non-coding RNAs and active Cis regulatory elements of mRNA molecule [17]. This algorithm is accessed by the link (http://rna.tbi.univie.ac.at/cgi-bin/RNAz/RNAz.cgi). The RNAz program creates a consensus secondary structure for an alignment using the RNAalifold method which acts just the same as the single sequence folding algorithm (RNAfold); however, the energy model is computed through correlation information [18].

2.3. CMfinder algorithm

CMfinder tool which is also directly synchronized with the genome homology search method can be used to refine and develop RNA families and it is a powerful algorithm for finding stable structured ncRNAs. In 2017, CMfinder was used to predict the RNA structure some mammalian genomes and 510,000 human structured ncRNAs were predicted. It is reported that CMfinder can generate useful alignments even with low sequence conservation [19].

2.4. RNAalifold algorithm

As the first important step, doing subsequent analysis is required to predict a consensus structure for a set of dependent RNAs. In this regard, the RNAalifold method is one of the oldest and most widely used tools which can prepare RNA structure through a set of protected sequences[20]. In order to use this algorithm, it is suggested to create an agreement building of the desired sequence by downloading the link (http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAalifold.cgi).

3. Case study of RNA motifs in plants

3.1. Organelle RNA recognition motif (ORRM)

The RNA recognition motif (RRM), considered as one of the most abundant protein motifs in eukaryotesm is a protected amino acid motif which includes 80 amino acids. Its main roles in plants are participation in growth processes, response to stresses in plants and various metabolic processes [21]. It is reported that RRM proteins targeted by plant organs can act on various RNA processes such as RNA binding, RNA editing, and RNA stability. They can also participate in plant growth and / or stress responses. Similarly, mutation in ORRM motif showed delayed flowering and growth retardation[22].

As reported by Sun et al. (2013), the ORRM1 motif is essential for RNA editing after transcription of cytidine (C) to uridine (U) in Arabidopsis chloroplasts. Besides, analysis of ORRM family members showed that ORRM2, ORRM3 and ORRM4 are RNA editing factors in mitochondria, whereas ORRM6 is just a chloroplast RNA editing factor [23].

3.2. Introduction and role of ORRM5 in plants

The function of members of the ORRM family was evaluated through the analysis of T-DNA mutations. The expression of a modified gene,

At4g13850, in the ORRM family delayed the plant growth and flowering. This protein known as ORRM5, which used to evaluate the possible causes of morphological defects in ORRM5 mutants, on the RNA splicing process, frequency of specific transcripts, and RNA editing. Results showed that ORRM5 mutations could reduce the cis-splicing function of the first intron of mitochondrial nad5 transcripts and might play additional roles in RNA metabolism. Mutations in ORRM5 also decreased editing performance in 18 mitochondrial C targets, while editing rates increased in 79 mitochondrial sites compared with wild-type editing levels. Therefore, the absence of ORRM5 leads to an increase in editing at 14% of the mitochondrial sites examined. The most important patterns described for the ORRM5 plant motif are [24, 25]:

a) ORRM5 mutations which lead to growth retardation and late flowering,
b) ORRM5 mutations which change the rate of mitochondrial RNA editing,
c) Stable expression of ORRM5 which completes the editing defect in orrm5 mutations,
d) Mutation in ORRM5 which has no effect on the steady state level of RNA transcript, and,
e) Stable expression of ORRM5 which complements the morphological defect of ORRM5 mutations.

3.3. Allelic RNA motif in potato

Plant intercellular RNA communication has a significant function in numerous aspects of biology. Viroids, which are circular non-coding RNAs causing plant diseases, are an autopsy model affecting the role of structural RNA motifs in regulating intercellular RNA communication in plants. Recent studies on potato wart spindle gland (PSTVd) showed the significant role of RNA motif Loop 19 in spreading PSTVd from secretory to spongy mesophyll in infected leaves[26]. The potato spindle viral RNA genome (PSTVd, viral species of potato spindle, Pospiviroidae family) consists of 359 nt in a secondary rod-like structure which by itself comprises 27 RNA loop motifs surrounded by short double helixes. The Loop 19 motif is also a critical link for the systemic spread of PSTVd in N. benthamiana. A recent report suggests that the possible importance of this loop for PESVd entry into spongy mesophiles from mesophiles which are secreted in inoculated leaves [27], and in terms of function is similar to the RNA 3D motif of the PSTVd 6 loop [26].

4. Functional Sequence Motifs

According to a report [16], several algorithms were proposed to identify the sequence motifs which have their own unique characteristics. These algorithms include Cmfinder, RNAz, QRNA and MEME analog algorithm. The CMfinder algorithm which is used for predicting motifs for RNA in heterogeneous sequences is the maximum prediction algorithm that uses covariance models to describe motifs, carefully crafted heuristics models for effective motif search, and the new Bayesian framework model for making prediction through combining folding energy and sequence covariance. Similarly, the RNAz algorithm which is a program for predicting the protected structure and thermodynamically stable secondary structure of an RNA molecule in multiple sequence alignments can be effective in extensive genome representation to discover and track RNA functional structures. Here, the most important structures are non-coding RNAs and active cis regulatory elements of the mRNA molecule [17]. The following are some motifs which are identified by these algorithms.

4.1. SDC motif

The SDC motif, which is reported in 26 species of fungi and can typically form a series of pins with an 11-bp stem (Figure 1A), represents some changes in stem intensity and length, which for the current purpose are referred to as the SDC motif in Neurospora crassa (Figure 1B). Each case upstream of the SAM (S-adenosyl methionine) decarboxylase gene indicates the significance of formation of hairpins and dependent conserved nucleotides for regulation of SDC gene [16]. The SDC enzyme catalyzes the synthesis of S-adenosyl methionine, which provides propylamine for polyamine production like spermidine and

spermine. The polyamine production is further regulated strongly by the cell at multiple levels including transcription, translation, enzyme activation, and protein depletion. At the transcriptional level, the expression of the SDC gene in mammals and plants is regulated by the expression of uORFs positioned at the levels of 5' UTRs [28]. In N. crassa, the two uORFs are short peptide coding regions which in some mRNAs are slightly above the original ORF (Figure 1C). Such uORF regions are commonly involved in controlling translational initiation in adjacent genes [29]. Two uORF sequences belonging to the SDC motif of the N. crassa fungus are located next to two 5 'agreed binding sites and one 3' site. In another experiment [16], the function of the SDC motif in N.crassa was tested and a luciferase reporter gene was added to the 5'UTR locus (Figure 1C). The compound was then transferred to the fungal cell and the results were analyzed by RT-PCR. The results showed that the new mRNA body was effectively found periodically in two major forms of Sp-I containing two uORFs and a form of Sp-II which had none of these two uORFs (Figure 1C).

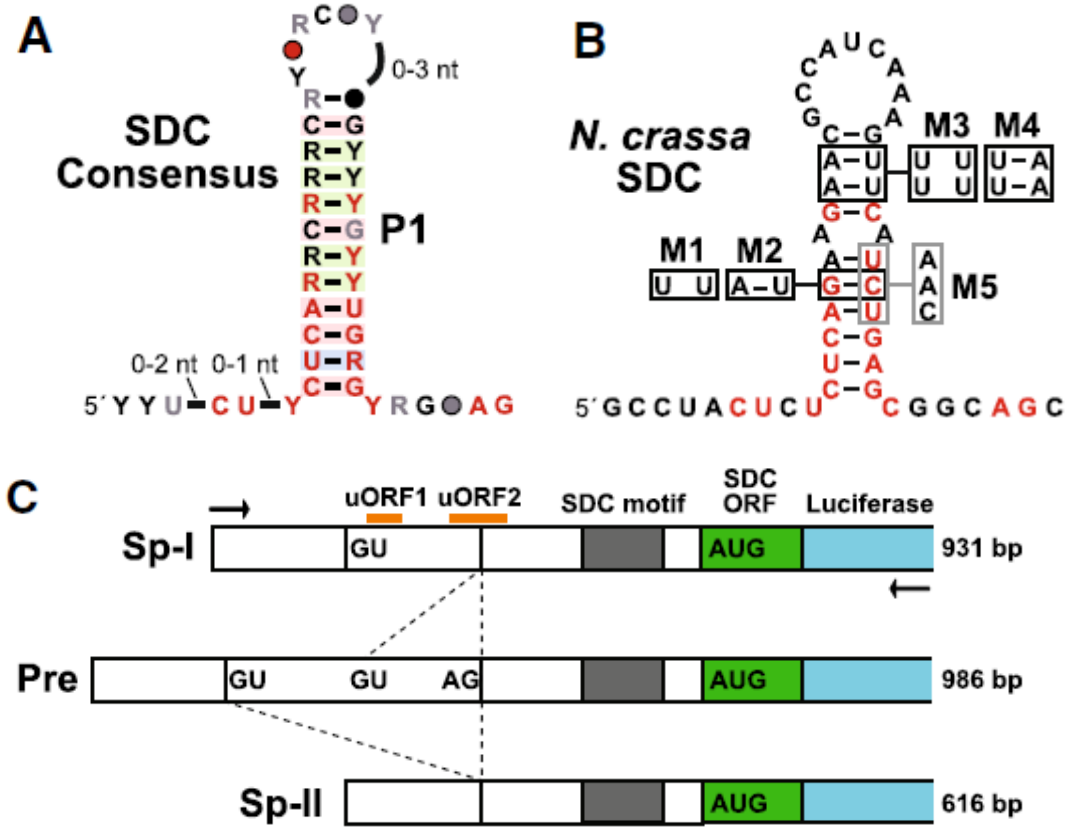

Figure 1. Structure and gene control function of SDC motif. A) Motif agreement model that describes the protected sequences of the secondary structure of the SDC RNA motif. B) Sequence and model show the secondary structure of the SDC motif from N.crassa. The nucleotides shown in red belong to the conserved nucleotides which are present in the consensus sequence. M1 to M5 which also indicate the nucleotide differences at the sites stand for mutation and are used to assess the importance of the P1 stem in gene expression. C) It shows the genetic elements present next to the SDC motif, which include the location of the luciferase reporter gene [16]

4.2. Amd motif

A total of 23 cases of amd motif were reported among 20 species of fungi. The sequence agreement and the second building model which was based on this sequence also proved that this motif probably adopts a 2-stranded bifurcation in its sequence, so that the protected nucleotides are located inside the

ring and the P1 and P2 sections next to the ring. In addition, nucleotides upstream of the P1 region and several other segments in the inner ring which are emerged to encode a short upstream ORF (Figure 2) suggest that ORF and the amn RNA overlap structure of the amd motif may work together to regulate downstream genes [16].

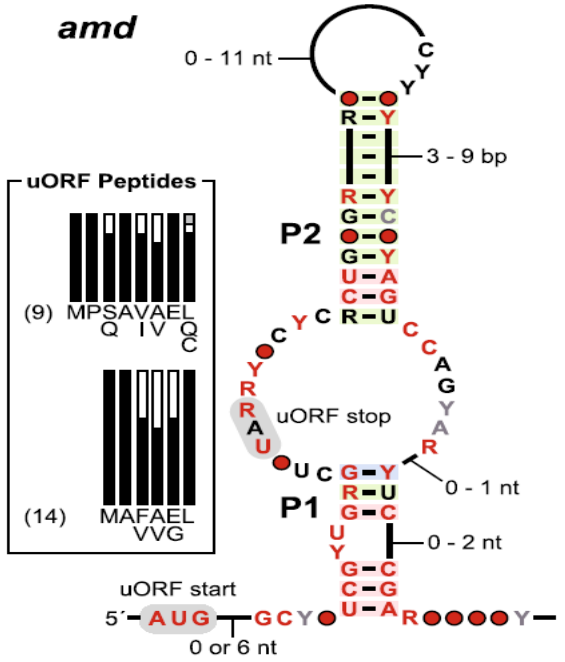

Figure 2. Sequence and agreement structure of uORF peptide products for amd motif. White, black, and gray represent the highest to lowest common amino acids, respectively. The numbers in parentheses indicate the number of specific amino acids in the peptide

4.3. *Ies6* motif

This type of RNA motif forms a broad hairpin structure with a short loop and a long inner loop (Figure 3). Pin rings which also correspond to a GNRA tetralope are frequently found in structural RNAs. GNRA tetralope can cause tertiary interactions with tetralope receptor structures. These predicted RNA structures are approximately close to a protein encoding gene which is similar to the Ies6 subunit chromatin-modifying complex (Ino80 subunit 6) involved in chromatin splicing, chromosomal differentiation, and telomere length regulation[30, 31].

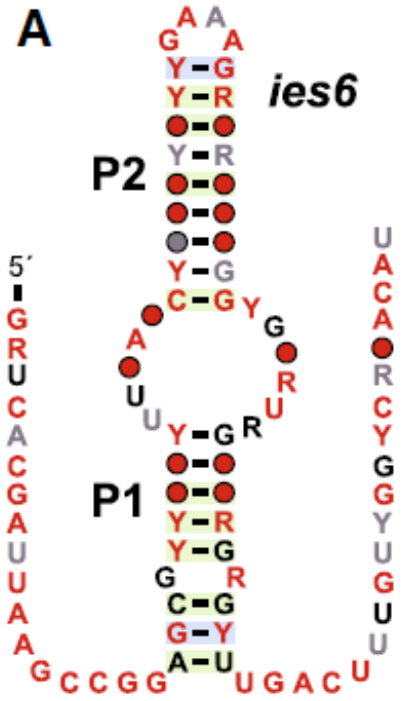

Figure 3. Consensus sequence and second structure of ies6 RNA motif [16]

4.4. *Hexa* motif

The second proposed building for this motif is a long series bracket interrupted by two inner rings. The hexA motif is found in the 5 'UTR region or intron of annotated genes and acts as the encoders of hypothetical proteins (40%) or the Woronin body subunit of the Hex protein (60%) (Figure 4a). Veronine bodies are fungal specific organs that during physical injury, are rapidly inserted into septal nuclei of cells to prevent cell volume loss [32]. Hex subunit proteins which are also interconnected with different forms of mRNA are among the main and essential components of the veronin body [32, 33]. On the other hand, a portion of the constituent sequence to the right of the HexA motif series pin corresponds closely with the 3 'agreement binding site. In addition, the 3 'long tail nucleotides show the ability of these nucleotides in protein encoding. In particular, the lowest conserved nucleotides in this region which are located at all three sites may be related to the wobble site of the codons. These features suggest the regulatory potential of hexA motif RNA structure in that part to

generate a modified ORF sequence for translation[16].

4.5. *SART-1* motif

The SART-1 motif model is based on 12 unique examples from 11 fungal species. The second structure of this motif potentially contains at least two Hairpins (Figure 4B). The hairpin RNA loops in this sequence are involved in stabilizing the pairing regions [34]. Similarly, the SART-1 motif is similar to the yeast Snu66 protein. In mammals, the SART-1 gene encodes separate ORF sequences translated by a (-1 framshifting) mechanism.

4.6. AU rich series brooch motif

Among the only 12 cases of this motif were reported from 3 species of fungi, 8 are from Rhizoctonia solani. Two other specimens which had a remarkable sequence and structural similarity and were found in the genome of the bacterium Orientia tsussugamushi were reported as the members of a mraW class which belonged to ncRNAs [35]. Only a small fraction of the agreed sequence for the mraW RNA motif is similar to the agreed sequence of the AU nucleotide-rich series motif derived from the fungal sample (Figure 4C). Yet, the use of mraW RNA motifs is not known, and on the other hand, the genetic connections of fungal specimens vary greatly, so no clues are found for the biological maps of the AU-rich motif [16].

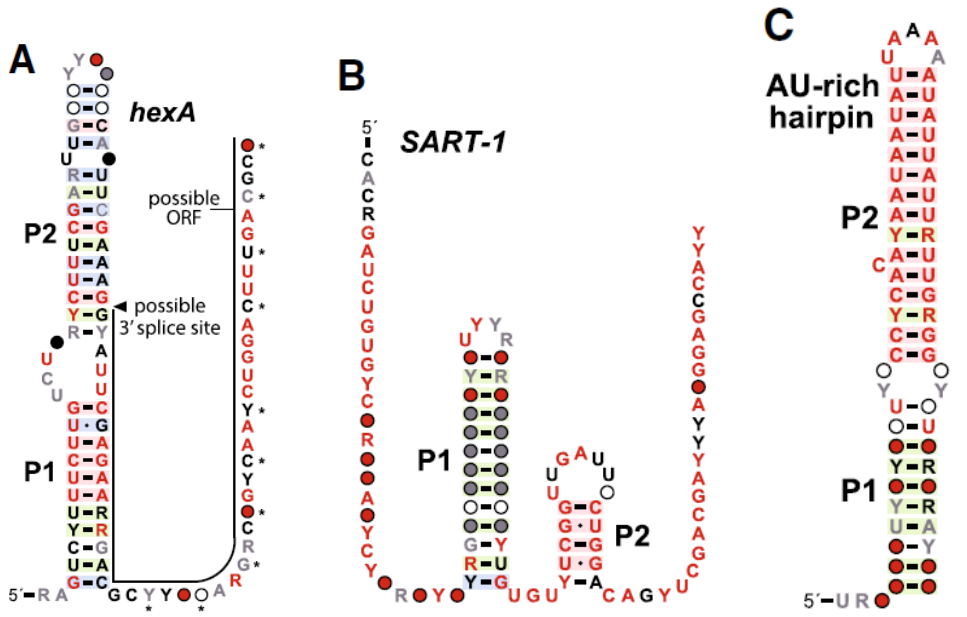

Figure 4. Agreed sequence and model of second structure for predicted RNA motif. A) hexA motif. B) SART-1 motif. C) AU rich series brooch motif. [16]

5. RNA structural motifs

The RNA molecule is described by its sequence of actions and structural constraints. This is a feature of both non-coding RNAs and other functional RNAs. In other words, all are characterized by ordered pairwise interactions that not only organize molecules into motifs and domains but also provide a framework for functional reactions. The RNAMotif algorithm is a program used to identify these motifs and detect the second and the third structures in the RNA sequence. These structural motifs include primary pairs, helix and unpaired nucleotides in hairpins, and multi-stem loops. Here, some related motifs are explained.

5.1. Pseudoknots motif

Pseudoknot is a motif in which nucleotides in a series pin loop along with a complementary sequence next to the loop can form open pairs and establish the Stem/Loop regions. These extra stems together with the hairpin, loop, and

stem, can also build a mass on top of each other and participate in the formation to stabilize the entire RNA structure. Similarly, the Pseudoknot motif participates in various biological activities. In some viruses, this motif alters gene expression through inducing a ribosomal framework which motivates ribosomal framing [36]. This type of motif can also be telomerase [37], self-splicing introns [38] and the catalytic nucleus of various ribosomes [39]. The structure (Figure 5) shows a Pseudoknot motif as a potent inhibitor of the reverse transcriptase enzyme which acts through detecting the binding site of this enzyme in humans and binding to this site (Figure 5). An example is the human immunodedi function virus (HIV-1) reverse transcriptase enzyme [40].

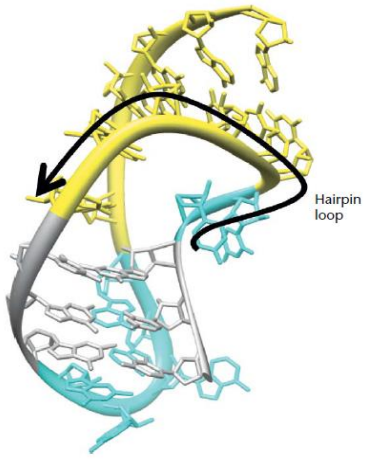

Figure5. Pseudoknot motif. The stem of the series pin is shown in gray. The pin of the series is marked with a curved arrow and an extra stem is shown in yellow [4]

5.2. Hairpins Motif

A series pin is formed as a double helix is covered by a loop. Pins which act as the most common secondary structure motif in the RNA molecule are essential compounds in the architecture of the third RNA structure and the formation of binding sites for other molecules. At least three nucleotides are required to form a thermodynamically stable series pin (Trillops). Moreover, 4-nucleotide (four-loop) series pins are the most abundant of these motifs which are observed in natural RNA molecules [41]. Stable three-ring or four-ring motifs are which are usually conserved during evolution can be easily identified by phylogenetic analysis.

Three-ring series pin motifs are reported for 7% and 16% of 16s-like rRNA rings in bacteria and eukaryotes, respectively [42]. In the stable (thermodynamically) three-loop series pins, abundant U-rich sequences were observed. In particular, the uracil U base is located at positions 1 and 2, and the adenine A base is located at position 3 [43]. The two interconnected U bases appear to provide a detection site for protein binding or further twisting of the RNA molecule. The three stable loops are often affected by open pairs which are in adjacent to the ring (Figure 6).

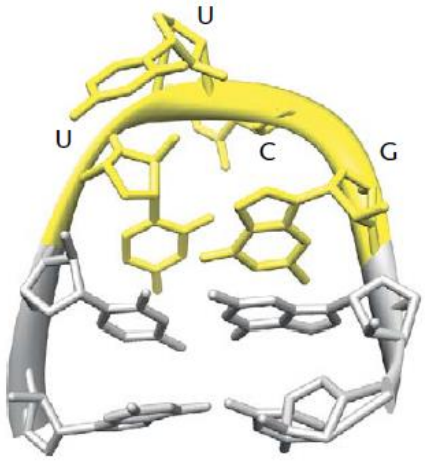

Figure 6. UUCG tetraloop Hairpin motif [4]

5.3. Bulge – helix – bulge motif

By using an all-protein system, ancient bacteria can bind some RNA transcripts to form a special motif called a bulge – helix – bulge. This motif, in essence, contains a seven-game internal loop motif (Figure 7). Four bases at the end of 5, both strings are motifs from a spiral A, four pairs of motifs, and three bases at the end of the three primers are all protrusions [44]. The Bulge – helix – bulge motif found at the intron-exon boundaries of the primary mRNA in archaea suggests that the primary mRNA splicing probably depends upon a splicing system that requires a Bulge – helix – bulge motif or motifs for the active site of special processing [45].

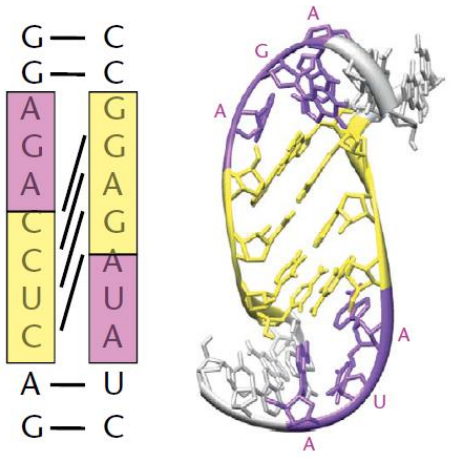

Figure 7. Sequence and structure of Bulge – helix – bulge motif. The two protrusions are on top of the two opposite strands (purple). The central helix is shown in yellow [4]

5.4. G-bulge motif

In this motif, there is a single-game protruding ring of guanine in double helix A. To illustrate the G-bulge motif, a compound found in the sarcin/ricin loop structure of the 28 s mouse rRNA molecule is used (Figure 8). The G base in the G-bulge motif which forms two hydrogen bonds with a phosphate group on the opposite strand can build a hydrogen bond with a three-uracil primer base. Studies have shown that the G-bulge motif has a highly variable conformation in the collective refining of structures [46], which may lead to greater adaptation to binding to target molecules. Accordingly, this motif is reported to be essential for packaging and infection of the human immunodeficiency virus (HIV) [47].

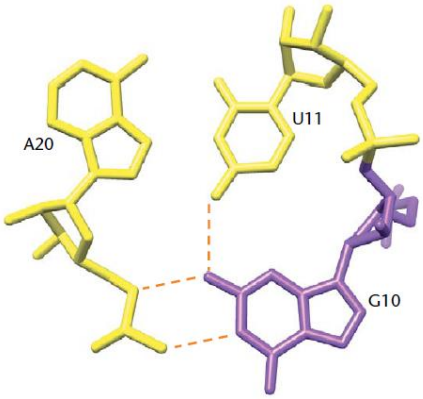

Figure 8. G-bulge motif building. The guanine nucleotide involved in the motif is shown in purple. The hydrogen bond between the guanine and the helix appears as dark lines [4]

5.5. Ribose zipper motif

When two coils are placed adjacent, ribose sugar in two opposite strands of two different coils can form a network of hydrogen bonds at the point of collision (contact) to establish and stabilize the order of the two strands [48]. This ribose sugar hydrogen bonding network is known as the "zipped ribose ribbon". Proteins involved in the early stages of ribosomal assembly and ribosome function are able to stabilize the interactions of the zipped ribose motif [49] which is the only structural motif present to essentially cause strength and skeleton of the RNA molecule. The structure of the Ribose zipper motif is shown in Figure 9.

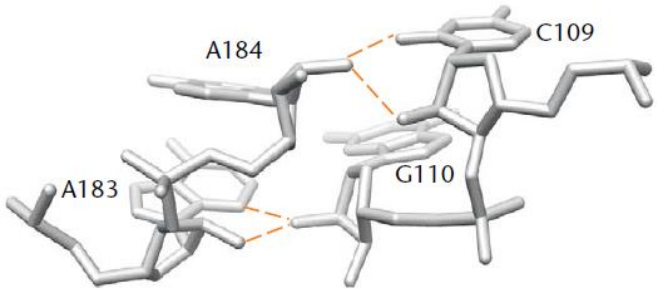

Figure 9. Ribose zipper motif structure. This motif allows two RNAs of two different strands to be adjacent [4]

6. Conclusion

In this study, a number of functional and structural RNA motifs were introduced using QRNA, RNAz, Cmfinder, and RNAalifold detection algorithms among which the most significant is Pseudoknot motif, mainly due to its ability to participate various biological activities such as variation in the expression of pathogenic genes in some viruses and formation of telomerase and self-truncated introns which are important breeding factors in biotechnology. Another important feature of this motif is inhibition of binding in enzyme Reverse transcriptase of the acquired immunodeficiency virus (HIV), which act effectively and efficiently in treatment of this deadly disease. The G-bulge motif is also another important motif, which due to its infectious nature against HIV, can be used in modern medicine in the future through genetic engineering methods. Given the useful and efficient functions provided in various fields by RNA motifs, it is suggested to conduct more extensive research on these motifs to pave the way for further success in the future, especially in medical science and other issues related to human health.

Abbreviations

CMfinder: A covariance model based RNA motif finding algorithm
MEME: Maximum expected analog algorithm
ncRNAs: Noncoding RNAs
PSTVd: potato wart spindle gland
QRNA: it is a powerful program and has been widely used as an efficient analysis tool to detect ncRNA gene at present
RRM: RNA recognition motif

Conflict of interest

None.

Consent for publications

The author read and proved the final manuscript for publication.

Availability of data and material

All data generated during this study are included in this published article

Authors' Contribution

All authors had equal role in study design, work, statistical analysis and manuscript writing.

Ethics approval and consent to participate

No human or animals were used in the present research.

Funding

No funding was received.

Ethics approval and consent to participate

No human or animals were used in the present research.

Orcid:

Davoud Naderi
https://orcid.org/0000-0003-1051-1771
Raheba Jami
https://orcid.org/0000-0002-5897-9341
Fazal ur Rehman
https://orcid.org/0000-0001-7581-6344

References

1. Szakonyi D, Confraria A, Valerio C, Duque P, Staiger D. (2019). Plant RNA Biology. *Frontiers in plant science*, 10: 887. https://doi.org/10.3389/fpls.2019.00887
2. Yang X, Yang M, Deng H, Ding Y. (2018). New era of studying RNA secondary structure and its influence on gene regulation in plants. *Frontiers in plant science*, 9: 671. https://doi.org/10.3389/fpls.2018.00671
3. Pieczynski M, Kruszka K, Bielewicz D, Dolata J, Szczesniak M, Karlowski W, Jarmolowski A, Szweykowska-Kulinska Z. (2018). A role of U12 intron in proper pre-mRNA splicing of plant cap binding protein 20 genes. *Frontiers in plant science*, 9: 475. https://doi.org/10.3389/fpls.2018.00475

4. Zhang M, Perelson A S, Tung C-S. (2011). RNA structural motifs. *eLS*, 2011(August): 1-10. https://doi.org/10.1002/9780470015902.a0003132.pub2
5. Chiang Y S, Gelfand T I, Kister A E, Gelfand I M. (2007). New classification of supersecondary structures of sandwich-like proteins uncovers strict patterns of strand assemblage. *Proteins: Structure, Function, and Bioinformatics*, 68(4): 915-921. https://doi.org/10.1002/prot.21473
6. Raff M, Alberts B, Lewis J, Johnson A, Roberts K. (2002). Molecular biology of the cell 4th edition: *National Center for Biotechnology InformationÕs Bookshelf*.
7. Ray D, Kazan H, Cook K B, Weirauch M T, Najafabadi H S, Li X, Gueroussov S, Albu M, Zheng H, Yang A. (2013). A compendium of RNA-binding motifs for decoding gene regulation. *Nature*, 499(7457): 172-177. https://doi.org/10.1038/nature12311
8. Lambert N, Robertson A, Jangi M, McGeary S, Sharp P A, Burge C B. (2014). RNA Bind-n-Seq: quantitative assessment of the sequence and structural binding specificity of RNA binding proteins. *Molecular cell*, 54(5): 887-900. https://doi.org/10.1016/j.molcel.2014.04.016
9. Shi X, Germain A, Hanson M R, Bentolila S. (2016). RNA recognition motif-containing protein ORRM4 broadly affects mitochondrial RNA editing and impacts plant development and flowering. *Plant Physiology*, 170(1): 294-309. https://doi.org/10.1104/pp.15.01280
10. Daub J, Eberhardt R Y, Tate J G, Burge S W. (2015). Rfam: annotating families of non-coding RNA sequences: *Springer*. https://doi.org/10.1007/978-1-4939-2291-8_22
11. Liu N, Xiao Z-D, Yu C-H, Shao P, Liang Y-T, Guan D-G, Yang J-H, Chen C-L, Qu L-H, Zhou H. (2009). SnoRNAs from the filamentous fungus Neurospora crassa: structural, functional and evolutionary insights. *BMC genomics*, 10(1): 1-13. https://doi.org/10.1186/1471-2164-10-515
12. Rivas E, Eddy S R. (2001). Noncoding RNA gene detection using comparative sequence analysis. *BMC bioinformatics*, 2(1): 1-19. https://doi.org/10.1186/1471-2105-2-8
13. Washietl S, Hofacker I L, Stadler P F. (2005). Fast and reliable prediction of noncoding RNAs. *Proceedings of the National Academy of Sciences*, 102(7): 2454-2459. https://doi.org/10.1073/pnas.0409169102
14. Yao Z, Weinberg Z, Ruzzo W L. (2006). CMfinder—a covariance model based RNA motif finding algorithm. *Bioinformatics*, 22(4): 445-452. https://doi.org/10.1093/bioinformatics/btk008
15. Eddy S R, Durbin R. (1994). RNA sequence analysis using covariance models. *Nucleic acids research*, 22(11): 2079-2088. https://doi.org/10.1093/nar/22.11.2079
16. Li S, Breaker R R. (2017). Identification of 15 candidate structured noncoding RNA motifs in fungi by comparative genomics. *BMC genomics*, 18(1): 1-17. https://doi.org/10.1186/s12864-017-4171-y
17. Gruber A R, Findeiß S, Washietl S, Hofacker I L, Stadler P F. (2010). RNAz 2.0: improved noncoding RNA detection Biocomputing 2010 (pp. 69-79): *World Scientific*. https://doi.org/10.1142/9789814295291_0009
18. Pettersen E F, Goddard T D, Huang C C, Couch G S, Greenblatt D M, Meng E C, Ferrin T E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry*, 25(13): 1605-1612. https://doi.org/10.1002/jcc.20084
19. Hou L, Xie J, Wu Y, Wang J, Duan A, Ao Y, Liu X, Yu X, Yan H, Perreault J. (2021). Identification of 11 candidate structured noncoding RNA motifs in humans by comparative genomics. *BMC genomics*, 22(1): 1-14. https://doi.org/10.1186/s12864-021-07474-9
20. Glisovic T, Bachorik J L, Yong J, Dreyfuss G. (2008). RNA-binding proteins and post-transcriptional gene regulation. *FEBS letters*, 582(14): 1977-1986. https://doi.org/10.1016/j.febslet.2008.03.004

21. Lorković Z J. (2009). Role of plant RNA-binding proteins in development, stress response and genome organization. *Trends in plant science*, 14(4): 229-236. https://doi.org/10.1016/j.tplants.2009.01.007
22. Shi X, Bentolila S, Hanson M R. (2016). Organelle RNA recognition motif-containing (ORRM) proteins are plastid and mitochondrial editing factors in Arabidopsis. *Plant signaling & behavior*, 11(5): 294-309. https://doi.org/10.1080/15592324.2016.1167299
23. Hackett J B, Shi X, Kobylarz A T, Lucas M K, Wessendorf R L, Hines K M, Bentolila S, Hanson M R, Lu Y. (2017). An organelle RNA recognition motif protein is required for photosystem II subunit psbF transcript editing. *Plant Physiology*, 173(4): 2278-2293. https://doi.org/10.1104/pp.16.01623
24. Shi X, Castandet B, Germain A, Hanson M R, Bentolila S. (2017). ORRM5, an RNA recognition motif-containing protein, has a unique effect on mitochondrial RNA editing. *Journal of experimental botany*, 68(11): 2833-2847. https://doi.org/10.1093/jxb/erx139
25. Castandet B, Hotto A M, Strickler S R, Stern D B. (2016). ChloroSeq, an optimized chloroplast RNA-Seq bioinformatic pipeline, reveals remodeling of the organellar transcriptome under heat stress. *G3: Genes, Genomes, Genetics*, 6(9): 2817-2827. https://doi.org/10.1534/g3.116.030783
26. Takeda R, Zirbel C L, Leontis N B, Wang Y, Ding B. (2018). Allelic RNA motifs in regulating systemic trafficking of Potato spindle tuber viroid. *Viruses*, 10(4): 160. https://doi.org/10.3390/v10040160
27. Jiang D, Wang M, Li S. (2017). Functional analysis of a viroid RNA motif mediating cell-to-cell movement in Nicotiana benthamiana. *Journal of General Virology*, 98(1): 121-125. https://doi.org/10.1099/jgv.0.000630
28. Hu W-W, Gong H, Pua E C. (2005). The pivotal roles of the plant S-adenosylmethionine decarboxylase 5′ untranslated leader sequence in regulation of gene expression at the transcriptional and posttranscriptional levels. *Plant Physiology*, 138(1): 276-286. https://doi.org/10.1104/pp.104.056770
29. Vilela C, McCarthy J E. (2003). Regulation of fungal gene expression via short open reading frames in the mRNA 5′ untranslated region. *Molecular microbiology*, 49(4): 859-867. https://doi.org/10.1046/j.1365-2958.2003.03622.x
30. Fiore J L, Nesbitt D J. (2013). An RNA folding motif: GNRA tetraloop–receptor interactions. *Quarterly reviews of biophysics*, 46(3): 223-264. https://doi.org/10.1017/S0033583513000048
31. Chambers A L, Ormerod G, Durley S C, Sing T L, Brown G W, Kent N A, Downs J A. (2012). The INO80 chromatin remodeling complex prevents polyploidy and maintains normal chromatin structure at centromeres. *Genes & development*, 26(23): 2590-2603. https://doi.org/10.1101/gad.199976.112
32. Beck J, Ebel F. (2013). Characterization of the major Woronin body protein HexA of the human pathogenic mold Aspergillus fumigatus. *International Journal of Medical Microbiology*, 303(2): 90-97. https://doi.org/10.1016/j.ijmm.2012.11.005
33. Yuan P, Jedd G, Kumaran D, Swaminathan S, Shio H, Hewitt D, Chua N-H, Swaminathan K. (2003). A HEX-1 crystal lattice required for Woronin body function in Neurospora crassa. *Nature Structural & Molecular Biology*, 10(4): 264-270. https://doi.org/10.1038/nsb910
34. Molinaro M, Tinoco Jr I. (1995). Use of ultra stable UNCG tetraloop hairpins to fold RNA structures: thermodynamic and spectroscopic applications. *Nucleic acids research*, 23(15): 3056-3063. https://doi.org/10.1093/nar/23.15.3056
35. Weinberg Z, Wang J X, Bogue J, Yang J, Corbino K, Moy R H, Breaker R R. (2010). Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea, and their metagenomes. *Genome biology*, 11(3): 1-17. https://doi.org/10.1186/gb-2010-11-3-r31
36. Staple D W, Butcher S E. (2005). Pseudoknots: RNA structures with diverse

functions. *PLoS Biol*, 3(6): e213. 10.1371/journal.pbio.0030213
37. Takeda R, Petrov A I, Leontis N B, Ding B. (2011). A three-dimensional RNA motif in Potato spindle tuber viroid mediates trafficking from palisade mesophyll to spongy mesophyll in Nicotiana benthamiana. *The Plant Cell*, 23(1): 258-272. PMID: 21258006; PMCID: PMC3051236; https://doi.org/10.1105/tpc.110.081414
38. Adams P L, Stahley M R, Kosek A B, Wang J, Strobel S A. (2004). Crystal structure of a self-splicing group I intron with both exons. *Nature*, 430(6995): 45-50. https://doi.org/10.1038/nature02642
39. Ke A, Zhou K, Ding F, Cate J H, Doudna J A. (2004). A conformational switch controls hepatitis delta virus ribozyme catalysis. *Nature*, 429(6988): 201-205. https://doi.org/10.1038/nature02522
40. Jaeger J, Restle T, Steitz T A. (1998). The structure of HIV-1 reverse transcriptase complexed with an RNA pseudoknot inhibitor. *The EMBO journal*, 17(15): 4535-4542. https://doi.org/10.1093/emboj/17.15.4535
41. Tung C-S. (1997). A computational approach to modeling nucleic acid hairpin structures. *Biophysical journal*, 72(2): 876-885. https://doi.org/10.1016/S0006-3495(97)78722-8
42. Wolters J. (1992). The nature of preferred hairpin structures in 16S-like rRNA variable regions. *Nucleic acids research*, 20(8): 1843-1850. https://doi.org/10.1093/nar/20.8.1843
43. Shu Z, Bevilacqua P C. (1999). Isolation and characterization of thermodynamically stable and unstable RNA hairpins from a triloop combinatorial library. *Biochemistry*, 38(46): 15369-15379. https://doi.org/10.1021/bi991774z
44. Diener J L, Moore P B. (1998). Solution structure of a substrate for the archaeal pre-tRNA splicing endonucleases: the bulge-helix-bulge motif. *Molecular cell*, 1(6): 883-894. https://doi.org/10.1016/S1097-2765(00)80087-8
45. Yoshinari S, Itoh T, Hallam S J, DeLong E F, Yokobori S-i, Yamagishi A, Oshima T, Kita K, Watanabe Y-i. (2006). Archaeal pre-mRNA splicing: a connection to hetero-oligomeric splicing endonuclease. *Biochemical and Biophysical Research Communications*, 346(3): 1024-1032. https://doi.org/10.1016/j.bbrc.2006.06.011
46. Ulyanov N B, Mujeeb A, Du Z, Tonelli M, Parslow T G, James T L. (2006). NMR structure of the full-length linear dimer of stem-loop-1 RNA in the HIV-1 dimer initiation site. *Journal of biological chemistry*, 281(23): 16168-16177. https://doi.org/10.1074/jbc.M601711200
47. Clever J L, Parslow T G. (1997). Mutant human immunodeficiency virus type 1 genomes with defects in RNA dimerization or encapsidation. *Journal of virology*, 71(5): 3407-3414. PMCID: PMC191485; PMID: 9094610
48. Cate J H, Gooding A R, Podell E, Zhou K, Golden B L, Kundrot C E, Cech T R, Doudna J A. (1996). Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science*, 273(5282): 1678-1685. https://doi.org/10.1126/science.273.5282.1678
49. Tamura M, Holbrook S R. (2002). Sequence and structural conservation in RNA ribose zippers. *Journal of molecular biology*, 320(3): 455-474. https://doi.org/10.1016/S0022-2836(02)00515-6