Identification of Structural Motifs Using Networks of Hydrogen-Bonded Base Interactions in RNA Crystallographic Structures

Hazrina Yusof Hamdani [1,*] and **Mohd Firdaus-Raih** [2,3,*]

1 Advanced Medical and Dental Institute, Universiti Sains Malaysia, Bertam, Kepala Batas 13200, Pulau Pinang, Malaysia
2 Faculty of Science and Technology, Universiti Kebangsaan Malaysia (UKM), Bangi 43600, Selangor, Malaysia
3 Institute of Systems Biology, Universiti Kebangsaan Malaysia (UKM), Bangi 43600, Selangor, Malaysia
* Correspondence: hazrina@usm.my (H.Y.H.); firdaus@mfrlab.org (M.F.-R.)

Abstract: RNA structural motifs can be identified using methods that analyze base–base interactions and the conformation of a structure's backbone; however, these approaches do not necessarily take into consideration the hydrogen bonds that connect the bases or the networks of inter-connected hydrogen-bonded bases that are found in RNA structures. Large clusters of RNA bases that are tightly inter-connected by a network of hydrogen bonds are expected to be stable and relatively rigid substructures. Such base arrangements could therefore be present as structural motifs in RNA structures, especially when there is a requirement for a highly stable support platform or substructure to ensure the correct folding and spatial maintenance of functional sites that partake in catalysis or binding interactions. In order to test this hypothesis, we conducted a search in available RNA crystallographic structures in the Protein Data Bank database using queries that searched for profiles of bases inter-connected by hydrogen bonds. This method of searching does not require to have prior knowledge of the arrangement being searched. Our search results identified two clusters of six bases that are inter-connected by a network of hydrogen bonds. These arrangements of base sextuples have never been previously reported, thus making this the first report that proposes them as novel RNA tertiary motifs.

Keywords: RNA structural motifs; base-base interactions; classification of base arrangement; RNA crystallographic structures

1. Introduction

Hydrogen bonds are crucial for stabilizing the complex structures of ribonucleic acids (RNA). Conformational changes in a particular RNA molecule can result from variations of the hydrogen bond interactions present in a structure. Clusters of unbroken networks of hydrogen-bonded base interactions have been reported previously in RNA structures [1–3]. In this work, a hydrogen-bonded base interaction network is defined as an unbroken connection of bases that are interacting with each other through at least one hydrogen bond. Even though such hydrogen-bonded base interaction networks have been reported before, there is a paucity of work discussing large inter-connected hydrogen-bonded clusters of bases in three-dimensional (3D) arrangements.

Various types of smaller base arrangements consisting of triples, quadruples, and quintuples that can form tertiary-level motifs (3D motifs) have been observed, reported, and archived [4–6]. However, to our knowledge, there has been no systematic study to identify larger arrangements composed of six base clusters and beyond, despite their possible importance as building-block modules

of RNA structure. Identifying the existence of such 3D modules may lead to a more accurate design of functionally relevant synthetic RNA molecules in addition to improving the capacity to model RNA 3D structures from sequence information.

An RNA 3D motif can be defined as a tertiary arrangement of nucleotides, nucleosides, or bases that are repeatedly found in different locations either in the same RNA structure or in different RNA molecules. The annotation of RNA 3D motifs can be divided into three main approaches that consider: (i) the conformation of the RNA backbone [7,8]; (ii) the base–base interactions [1,3,9]; and (iii) the alignment of RNA 3D structures to detect similarities in folding and sub-folding [10,11]. Many of the known RNA 3D motifs discovered to date have an architectural role in RNA folding, functioning as stabilizers of RNA 3D structure or as sites of ligand binding or catalytic activity [12–14] More recently, RNA 3D motifs have been reported to play a role in miRNA biogenesis by acting as a guide for the Dicer-like 1 (DCL1) enzyme to perform cleavage of miRNA–miRNA* duplexes [15]. In this paper, we only focused on motifs that result from base–base interactions.

Expert visual examination of RNA structures has been a crucial aspect in the discovery of many currently known 3D base motifs [4,16]. RNA 3D motifs that have been discovered and annotated are available in various databases such as NCIR [4], RNA 3D Motif Atlas [9], and INTERRNA [6]. Several computer programs are available for identifying motifs in the available dataset of RNA crystallographic structures [10,17–19], and these programs allow for an automated search capacity that overcomes the limitations of manual visual curation such as the approach used for the NCIR database.

However, many of these computer programs require prior knowledge of the motifs to be provided as search queries, thus making them useful for structural annotation purposes but of limited utility for the discovery of novel motifs. The computer program COGNAC (COnnection tables Graphs for Nucleic Acids) was reported to be able to search for unbroken networks of hydrogen-bonded base interactions in RNA crystallographic structures that are available in the Protein Data Bank PDB [1]. A COGNAC search relies only on the user defining the connectivity of the bases by hydrogen bonds without the need for specific prior knowledge of how the bases are arranged in 3D space. Due to this capability, the COGNAC program may retrieve arrangements that are potentially novel motifs.

COGNAC annotations are independent of base sequence and any specific spatial arrangement definitions, thus making it possible to retrieve similar base components that, when visually examined, are in fact different in terms of spatial base arrangements. For example, the annotations for a query composed of GGGG, that was annotated as a planar base quadruple in the spinach RNA aptamer (G72.G29.G25.G68) (PDB ID: 4TS0), differed significantly from the GGGG base quadruple identified in the *Kluyveromyces lactis* 80S ribosomal structure (G1433.G1278.G1273.G1277) (PDB ID: 4V92) that was retrieved by the same search (Figure 1) [20,21]. While it is clear that the COGNAC program could potentially identify novel motifs, a secondary classification technique would be needed to sift through the hydrogen bond-connected base clusters that it retrieved.

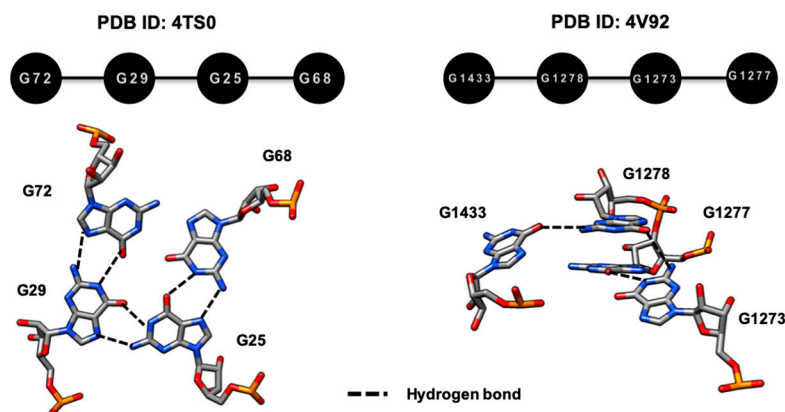| ID	Sequence	Bases in Interaction Chain_BaseResidueBaseNumber	Structure Description	PDBID
QUAD1_67767	GGGG	Y_G72: X_G29: X_G25: Y_G68	Crystal structure of the Spinach RNA aptamer in complex with DFHBI barium ions	4ts0
QUAD1_70860	GGGG	A_G1433: A_G1278: A_G1273: A_G1277	Kluyveromyces lactis 80S ribosome in complex with CrPV-IRES	4V92

Figure 1. An example of a GGGG arrangement found in the structure of a spinach RNA aptamer (PDB ID: 4TS0) and *Kluyveromyces lactis* 80S ribosome (PDB ID: 4V92) that have different tertiary base arrangements, despite being retrieved using the same graph representation query of the COGNAC computer program.

In this paper, we report the identification of 3D motifs in RNA structures by analyzing the data of clusters of base interactions involving six bases that are interconnected by hydrogen bonds. Our results revealed six base clusters that are repeated in different RNA molecules or can be found at different locations in the same molecules, which makes them possible tertiary motifs.
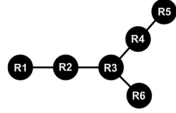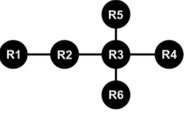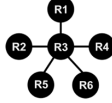
2. Materials and Methods

2.1. Dataset

In this study, 2158 structure coordinates containing RNA chains that were solved by X-ray diffraction of crystals were downloaded from the Protein Data Bank (PDB) [22]. This dataset includes coordinate files that contain RNA chains in the presence of other macromolecules or ligands. The downloaded structures represent a diverse repertoire of the available RNA molecules such as rRNA, ribozymes, riboswitches, mRNA, and tRNA. The resolution cutoff of the structures is ≤4 Å.

2.2. COGNAC Searches

The hydrogen bonding data for the 2158 structures were then generated by HBPRED, a program that was reported by Firdaus-Raih et al. [1]. The HBPRED program is based on the hydrogen bonding parameters used in the HBPLUS program but with specific modifications for use with RNA bases [23]. The hydrogen bonding information for the bases in all 2158 structures was then searched for specific arrangements of six base clusters (sextuples), where each base in a cluster is connected to another base by at least one hydrogen bond, using the COnnection tables Graphs for Nucleic ACids (COGNAC) computer program. The graph representations for the six possible arrangements of a base in a sextuple, each base being connected to at least one other base via at least one hydrogen bond (Table 1), were previously described by Firdaus-Raih et al. [1].

Table 1. Graph representations of six possible base connectivity patterns of a sextuple, where a base is connected to at least one other base by at least one hydrogen bond, and the number of occurrences found in the search dataset for each base sextuple type.
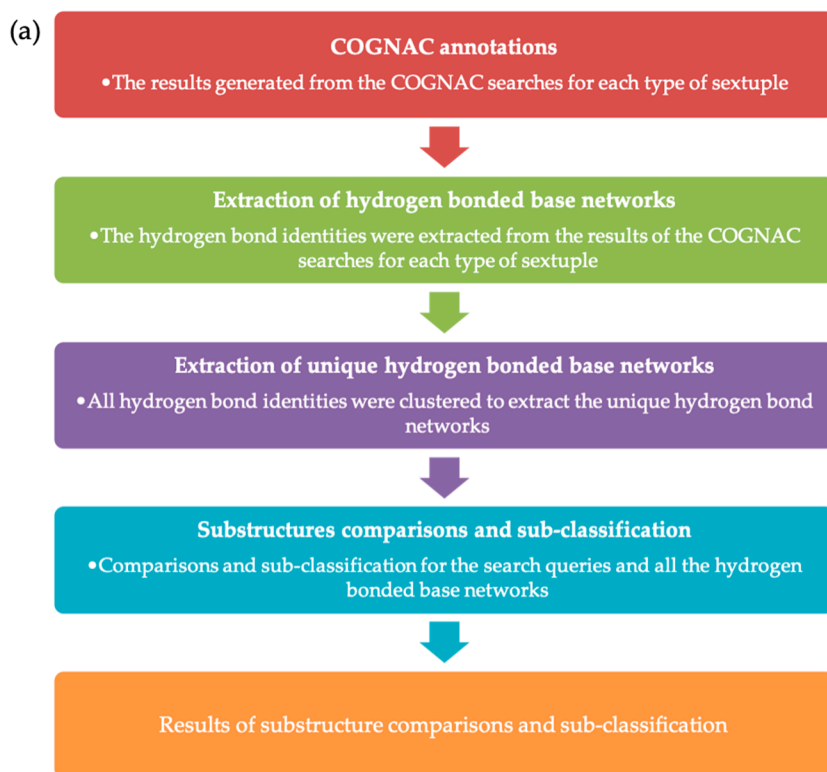
Pattern	Number of Annotations	Number of Unique Arrangements	Pattern	Number of Annotations	Number of Unique Arrangements
Type 1	10168	3572	Type 2	5448	2102
Type 3	4240	1862	Type 4	525	299
Type 5	8	6	Type 6	614	311

2.3. Sextuple Sub-Classification

The results generated from the COGNAC searches for each type of sextuple were deposited into the Interactions in RNA Structures Database (InterRNA) and can be accessed at http://mfrlab.org/interrna/. Each database entry is retrievable using an identification code referred to as the INTERRNA ID [6]. The flowchart for the data analysis and filtering process to sub-classify each type of sextuple is provided in Figure 2a.

The hydrogen bonding data for the COGNAC annotations were then parsed and extracted into matrices using a Perl program. The matrix for each hydrogen bond-connected base cluster contains information on the hydrogen bond donor base, the hydrogen bond donor atom, the hydrogen bond acceptor base, and the hydrogen bond acceptor atom. The matrices were compared to identify a unique set of hydrogen-bonded base networks. The extracted data for the hydrogen-bonded base networks served as the input for the next phase to sub-classify each sextuple.

Next, the search queries and all the hydrogen-bonded base clusters were compared, and each sextuple was further sub-classified using the PHP programming language. This phase of the data processing involved the extraction of unique base clusters from the COGNAC searches; the pseudocode for the process is provided in Figure 2b. Manual visual examinations of the structures were carried out using the UCSF Chimera molecular graphics suite [24]. These analyses included assessments of the different arrangements found superposed against each other using least-squares superpositions. Follow-up multiple-sequence alignments were carried out to detect the conservation of the observed sextuples at the sequence level.

(a)

COGNAC annotations
•The results generated from the COGNAC searches for each type of sextuple

Extraction of hydrogen bonded base networks
•The hydrogen bond identities were extracted from the results of the COGNAC searches for each type of sextuple

Extraction of unique hydrogen bonded base networks
•All hydrogen bond identities were clustered to extract the unique hydrogen bond networks

Substructures comparisons and sub-classification
•Comparisons and sub-classification for the search queries and all the hydrogen bonded base networks

Results of substructure comparisons and sub-classification

(b) **Input:** The COGNAC annotations of the sextuple search queries
Output: Classification of the hydrogen bonds networks

```
#to extract the hydrogen bonded base networks
for k ← 0 to n − 1
{
    hb_array[k]← hb_extraction(annotation[k])
}

#to extract the unique hydrogen bonded base networks
hb_array_unique ← array_unique(hb_array)

#to compare and sub-classify the search queries and hydrogen bonded base networks
for i ← 0 to n − 1
{
    key_hb_array_unique[i] ← array_search(hb_array[i], hb_array_unique)
}

for m ← 0 to n − 1
{
    x ← key_hb_array_unique[m]
    classification[x] ← classification[x].annotation[m]
}
```

Figure 2. (**a**) Flowchart of the sextuple sub-classification method based on the identity of hydrogen bonding; (**b**) pseudo-code of the classification technique based on the identity of hydrogen bonding.

3. Results and Discussion

3.1. Filtering of COGNAC Searches

The methodology that employs the COGNAC computer program used in this study has been proven to be able to identify hydrogen bond-connected bases from pairings to interactions involving

six bases [1,6]. Although the COGNAC approach was reported to be able to extract novel motifs without the need for prior knowledge of the tertiary base arrangements involved, it has the distinct limitation of being able to only search for arrangements in which the bases are connected by hydrogen bonds. That limitation, however, suited the ambit of this work in trying to find clusters of bases that are interconnected by hydrogen bonds.

This work further limited the scope of analysis and discussion to only sextuple patterns (Table 1) because smaller hydrogen bond-connected base arrangements such as triples, quadruples, and quintuples have been reported and are also likely to be constituents of a larger sextuple interaction. For example, the quadruple interaction (U438.A496.A498.G404) is a component of the sextuple interaction (U438.A496.A498.G404.A499.U439) annotated in the 16S rRNA of *Escherichia coli* (PDB ID: 5J7L) [25].

The COGNAC searches returned a total of are 3572 unique arrangements that were extracted from 10,168 annotations for the Type 1 sextuple pattern, 2102 pattern arrangements that were classified from 5448 annotations for the Type 2 sextuple pattern, and 1862 pattern arrangements that were classified from 4240 annotations for the Type 3 sextuple pattern. Furthermore, we obtained 299 pattern arrangements that were classified from 525 annotations for the Type 4 sextuple pattern, 6 pattern arrangements that were classified from 8 annotations for the Type 5 sextuple pattern, and 311 pattern arrangements that were classified from 614 annotations for the Type 6 sextuple pattern (Table 1).

Once the COGNAC search results for each sextuple type were sub-classified, the arrangements were analyzed to identify potentially novel structural motifs. In order to be considered as a potential motif, we required that the candidate arrangement fulfilled at least one of two criteria. The first criterion to be considered a motif was that the arrangement was repeatedly present in different RNA molecules, while the second criterion, especially if the first criterion was not met, was that an arrangement could be found in different locations of the same structure.

3.2. Hydrogen Bond-Connected Six-Base Interactions as Novel Structural Motifs

The results of the COGNAC searches were manually curated and were followed up by extensive visual examination to confirm the fitness of the patterns retrieved to our search criteria. Through this process, we were able to identify two sextuple base clusters that we propose as novel structural motifs, one meeting the first criterion, and the other fitting the second criterion [24].

3.2.1. A Base Sextuple Annotated in Different RNA Structures

One potentially novel motif that was uncovered by our visual examination process of the filtered COGNAC results is a GCA(A/U)(U/A)A Type 1 sextuple (Figure 3). The hydrogen bond donor–acceptor combinations (Figure 3a) clearly fitted that of a Type 1 sextuple pattern (Figure 3c). Our searches and analysis found this arrangement in four different RNA structures, i.e., 16S rRNA, 5S rRNA, a preQ1 riboswitch, and a S-adenosyl-(L)-homocysteine (SAH) riboswitch (Figure 3).

In this particular case, the base components at the R1, R2, R3, and R6 positions are conserved, with R1–R3 superposing very well onto each other (Figure 3c). However, in three structures, the R5 position is a uracil, while the R4 is an adenine, but in another structure the R4 is a uracil, while the R5 is an adenine. Despite this variation, it is clear that the 3D space taken up by both R4 and R5 are conserved due to the use of the same base components, despite them being on opposing sides of the pairing. This suggests that such a sextuple can be a structural module that can occur within specific spatial constraints and that variation in the sequence can exist, although the structures seem to be unrelated at sequence level (purine to pyrimidine). This is actually a practical mechanism for increasing sequence diversity that will at the same time maintain structural conservation, because the variations occur as a pair (AU to UA) that retains the interaction space of the pairing that was replaced.

Further scrutiny and visual examination of each occurrence of this GCA(A/U)(U/A)A Type 1 sextuple revealed that these motifs are found in parts of their respective RNA molecules that do not seem to share the same roles in different molecules. The GCA(A/U)(U/A)A Type 1 sextuple 3D motif,

found in the 16S rRNA (G113.C314.A51.U114.A313.A116), is situated in a location that does not appear to be directly involved in peptidyl-transferase activity. However, it is likely that this six-base cluster contributes to the structural stability of the ribosomal small subunit and, therefore, has an indirect role in protein synthesis. By visual examination of the cluster, we noted that it connects helix 5 and helix 7 to the body of the 16S rRNA structure, where helix 7 is one of three helices (helix 7, helix 44, and helix 16/17) that serve as the 'structural pole' of the 16S rRNA molecule [26]. We further observed that this potential 3D motif is located near and adjacent to bases that are involved in the interactions of the ribosomal proteins S12 and S16 [26,27].
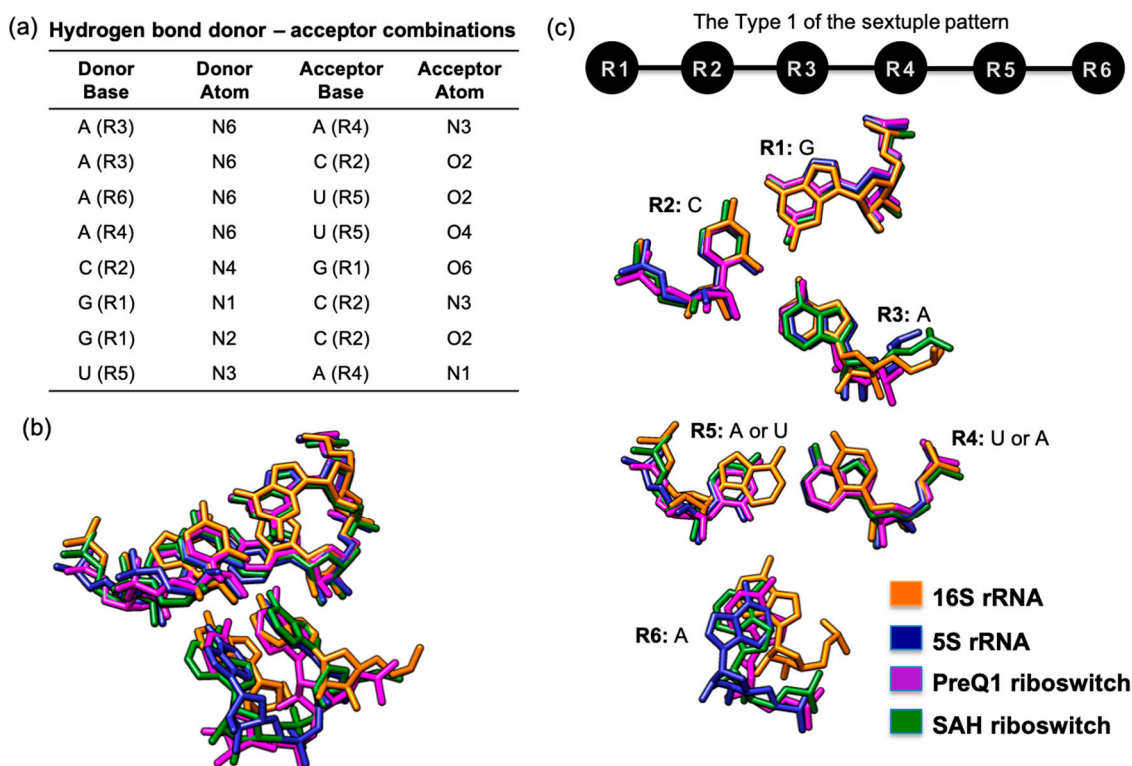

Figure 3. An example of a GCA(A/U)(U/A)A 3D arrangement that was annotated in four different RNA crystal structures. (**a**) Table of hydrogen bond donors and acceptors that define the hydrogen-bonded base network; (**b**) superposition of four similar base arrangements showing how they can be viewed in the four different RNA crystal structures; (**c**) graphical representation of the hydrogen bond network for the type 1 sextuple pattern (top) and corresponding base-by-base view of the superposed bases in the four different RNA crystal structures (bottom).

In the 5S rRNA structure, the G53.C29.A56.A54.U28.A57 cluster is located in loop B of domain I, which is between loop C of helix II and helix III that interact with Domain V of the 23S rRNA, with the presence of ribosomal proteins L5 and L18 (Figure 4b). The L5 ribosomal protein also interacts with the ribosomal protein subunit S13, and subsequently, the C terminal region of the ribosomal protein subunit S13 is located between the anticodon arms of the A- and P-tRNA sites, which functions as the nucleic acid decoding center; these sites are collectively known as the protuberance center [28]. This G53.C29.A56.A54.U28.A57 cluster does not appear to be directly involved in the protuberance center but more likely serves to reinforce the stability of the 5S rRNA (Domain 1).

In the preQ1 riboswitch structure, the G2.C21.A25.A3.U20.A26 cluster is located on the S1 stem and the L3 loop (Figure 4c). This particular Pre Q1 riboswitch is a 34-nucleotide H-type pseudoknot structure from *Bacillus subtilis* that serves as the preQ1 ligand recognition site [29]. In general, the H-type pseudoknot is folded via the interior of the hairpin loop, forming an intra-molecular interaction with the base of the exterior stem and resulting in a pseudoknot with two stems and two loops [30].
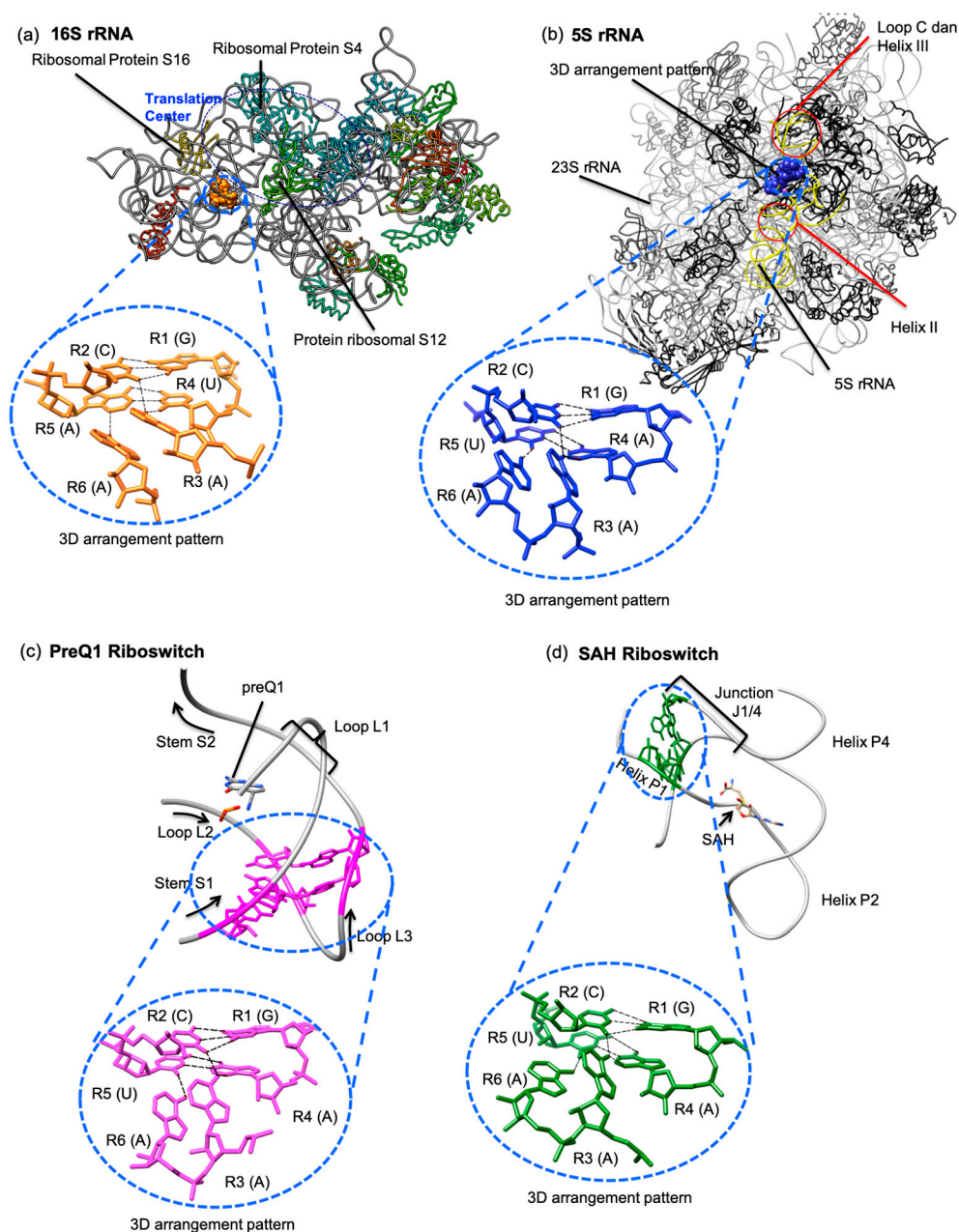
Figure 4. The GCA(A/U)(U/A)A motif (**a**) in the 16S rRNA crystal structure represented as orange spheres and magnified to show the base arrangements using orange stick representations (PDB ID: 1IBM) [27]; (**b**) in the 5S rRNA crystal structure represented as blue spheres and magnified to show the base arrangements using blue stick representations (PDB ID: 1JJ2) [31]; (**c**) in the preQ1 riboswitch crystal structure represented as purple sticks with the arrangement extracted and magnified (PDB ID: 3K1V) [29]; (**d**) in the SAH riboswitch crystal structure represented as green sticks with the arrangement extracted and magnified (PDB ID: 3NPQ) [32].

However, the H-type pseudoknot found in the 3K1V structure has two stems and three loops [29]. According to Aalberts and Hodas, the L3 loop where the G2.C21.A25.A3.U20.A26 cluster is annotated is a frequent location of many tertiary interactions [33]. There is also a quintuple base interaction, U6.A29.C18.G5.A28 (InterRNA ID: QUIN1_10653), very closely located at L1, end of S1, and end of L3, which has been identified as the ligand recognition site [29]. It is therefore highly possible that the G2.C21.A25.A3.U20.A26 cluster contributes to the stability of the ligand recognition site because it is located under the recognition site quintuple.

The fourth GCA(A/U)(U/A)A Type 1 sextuple can be found in the structure of a SAH riboswitch (G5.C34.A43.A6.U33.A44) (Figure 4d). The SAH riboswitches bind to the S-Adenosyl-l-homocysteine (SAH) molecule and can be found in many bacterial species [32]. The SAH riboswitch crystal structure consists of three helices, named P1, P2, and P3 [32]. The six-base cluster annotated in the SAH riboswitch is located at the P1 helix and the J1/4 junction [32], and our observations show that the sextuple does not interact with the SAH molecule.

However, two adenine bases (A43.A44) at the J1/4 junction form hydrogen bonds with the other four in the P1 helix, thus making the structure appear to be a type LL pseudoknot [32]. As with the previous observations regarding this sextuple, we also believe that it serves to increase the rigidity of the overall structure and is not directly associated with SAH binding that defines this riboswitch [32].

The structures in which the GCA(A/U)(U/A)A Type 1 sextuple could be found are not functionally related. Although the six nucleotides involved in this motif are closely situated in three-dimensional space, they span different sequence lengths and thus are not comparable using multiple-sequence alignments.

3.2.2. A Base Sextuple Annotated at Different Locations in the Same Structure

In addition to the six-base cluster that we found in four different RNA molecules, we were also able to identify a UAAGAC Type 2 sextuple that was annotated in three different locations in the same large ribosomal subunit structures, thus fitting our second criterion of base arrangements that could be potentially classified as tertiary motifs (Figure 5). This six-base cluster has a hydrogen bonded base connectivity pattern (Figure 6a) that fitted the Type 2 sextuple arrangement (Figure 6c).
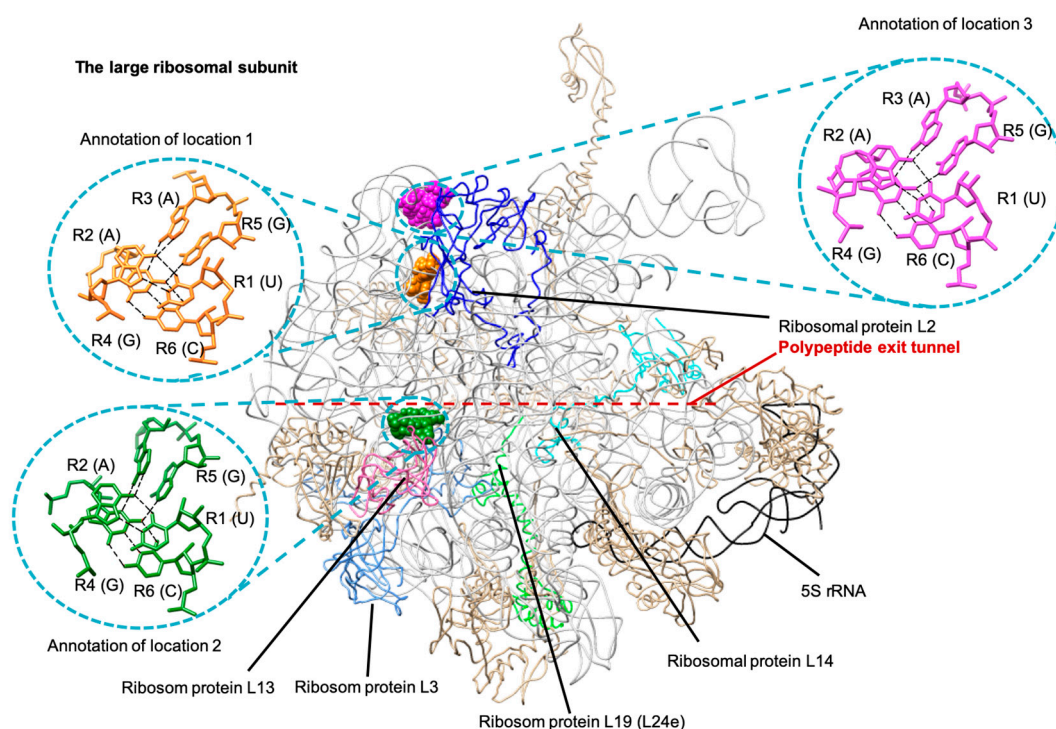

Figure 5. The UAAGAC base clusters that were annotated in three different locations in the large ribosomal subunit crystal structures are presented as colored spheres (PDB ID: 4LT8) [34].
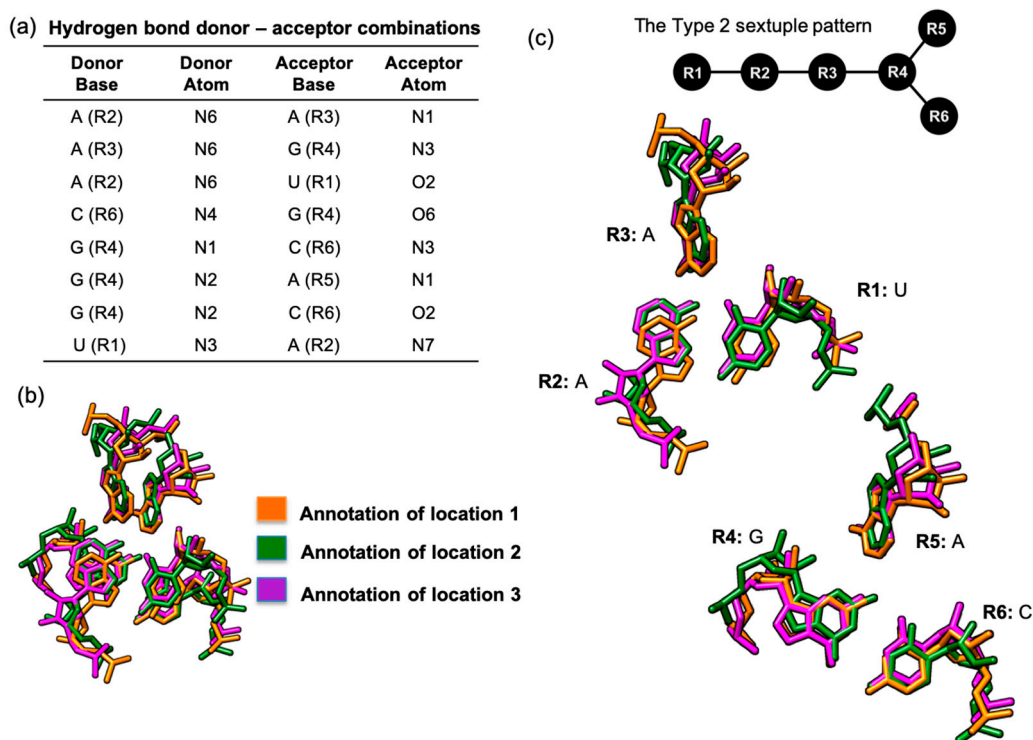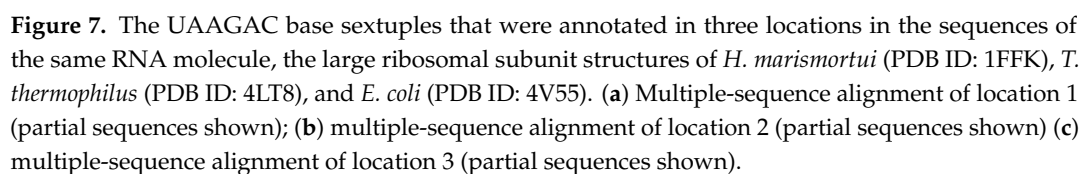
(a) **Hydrogen bond donor – acceptor combinations**

Donor Base	Donor Atom	Acceptor Base	Acceptor Atom
A (R2)	N6	A (R3)	N1
A (R3)	N6	G (R4)	N3
A (R2)	N6	U (R1)	O2
C (R6)	N4	G (R4)	O6
G (R4)	N1	C (R6)	N3
G (R4)	N2	A (R5)	N1
G (R4)	N2	C (R6)	O2
U (R1)	N3	A (R2)	N7

Figure 6. The UAAGAC base sextuple that was annotated in three locations in the same RNA molecule, a large ribosomal subunit structure (PDB ID: 4LT8). (**a**) Table of the hydrogen bond donors and acceptors that define the hydrogen-bonded base network; (**b**) superposition of three similar base arrangements as they can be viewed in the three different locations in same ribosomal subunit structure; (**c**) graphical representation of the hydrogen bond network for the Type 2 sextuple pattern and the corresponding base-by-base view of the superposed bases in the three different locations of the ribosomal subunit structure.

The first location of this UAAGAC sextuple is at helix 52 and helix 56 of Domain III in the large ribosomal subunit [35,36]. We also noted that this first site is located near the L2 ribosomal protein, which is one of the ribosomal proteins nearest to the peptidyl transferase center [37,38].

The second UAAGAC sextuple is located at Domain IV, helix 61, and helix 62 [35,36]. This second location is situated near two ribosomal proteins, L3 and L14 (Figure 5). The L3 ribosomal protein is one of the closest to the peptidyl transferase center [37,38]. In addition to the ribosomal protein L3, the ribosomal protein L14 is also involved indirectly in the binding site [39].

The third location for this UAAGAC sextuple, based on the large ribosomal subunit reference structures of *E. coli* and *Thermus thermophilus*, is at Domain III, helix 55, and helix 58 [35,36]. This UAAGAC sextuple is close to the ribosomal protein L2 [38]. However, we noted that this site could only be annotated in the large ribosomal subunit crystal structures of bacteria and was not found in the available examples for archaea.

Multiple-sequence alignments of the large ribosomal subunit of *Haloarcula marismortui*, *E. coli*, and *T. thermophiles* confirmed the results of the structural analysis that the third location for this UAAGAC Type 2 sextuple does not appear to be present in the archaea example (Figure 7). The third location appears to be in a more variable region in the structure of the large ribosomal subunit of *H. marismortui*. As with the previous GCA(A/U)(U/A)A Type 1 sextuple, the UAAGAC sextuple also appears to play a role in structural stabilization and does not appear to be directly involved in catalysis or as a binding site.

Figure 7. The UAAGAC base sextuples that were annotated in three locations in the sequences of the same RNA molecule, the large ribosomal subunit structures of *H. marismortui* (PDB ID: 1FFK), *T. thermophilus* (PDB ID: 4LT8), and *E. coli* (PDB ID: 4V55). (**a**) Multiple-sequence alignment of location 1 (partial sequences shown); (**b**) multiple-sequence alignment of location 2 (partial sequences shown) (**c**) multiple-sequence alignment of location 3 (partial sequences shown).

3.3. Presence of Known Motifs within the Sextuples

The sextuples that we report here are actually composed of elements that are already well known as tertiary motifs. For example, both sextuples contain A-minor motifs [16]. An AGC type I A-minor motif can be found in the GCA(A/U)(U/A)A Type 1 sextuple, while an AGC type II A-minor motif is a component of the UAAGAC Type 2 sextuple. Furthermore, the presence of other structural motifs, such as kink turns, can be seen and are expected to also be a feature of such larger motifs.

It is likely that other similarly large novel motifs discovered in the future will be composed of smaller known motifs. In this work, we demonstrate that, although the sextuples have known motifs as their constituent parts, mere visual examination was inadequate at identifying that they in fact partake in a larger conserved tertiary arrangement.

It is also worth noting that the COGNAC searches aimed at identifying novel motifs, as in this work, were intended for detecting tertiary motifs with the prerequisite of the component bases being connected by at least one hydrogen bond. It is therefore possible that similar arrangements may exist that were not retrieved by our search because they did not satisfy the hydrogen bonding criterion set. However, other programs that can be used to identify such motifs are already available and can be integrated in the process here reported, should they be required.

4. Conclusions

A cluster of six bases that are interconnected by a network of hydrogen bonds is expected to be a highly stable sub-structure. Due to this, it is unsurprising that our examination of the structures in which the GCA(A/U)(U/A)A Type 1 and the UAAGAC Type 2 sextuples can be found revealed a structural stabilization role that contributes to the correct folding and tertiary space maintenance of the associated functional sites. To our knowledge, this is the first such report that identifies these six bases as an associated cluster. Therefore, we propose that these base sextuples be classified as novel RNA tertiary motifs that may even have a wider role as RNA structural modules.

Author Contributions: Conceptualization, M.F.-R. Methodology and Software, M.F.-R. and H.Y.H.; Validation, H.Y.H.; Formal Analysis, H.Y.H.; Investigation, H.Y.H.; Resources, H.Y.H.; Writing—Original Draft Preparation, H.Y.H.; Writing—Reviewing & Editing, M.F.-R.; Visualization, H.Y.H.; Supervision, M.F.-R.; Project Administration, M.F.-R.; Funding Acquisition, H.Y.H. and M.F.-R.

References

1. Firdaus-Raih, M.; Hamdani, H.Y.; Nadzirin, N.; Ramlan, E.I.; Willett, P.; Artymiuk, P.J. COGNAC: A Web Server for Searching and Annotating Hydrogen-Bonded Base Interactions in RNA Three-Dimensional Structures. *Nucleic Acids Res.* **2014**, *42*, W382–W388. [CrossRef]

2. Hamdani, H.Y.; Artymiuk, P.J.; Firdaus-Raih, M. A Computational Approach for the Annotation of Hydrogen-Bonded Base Interactions in Crystallographic Structures of the Ribozymes. *AIP Conf. Proc.* **2015**, *1678*. [CrossRef]

3. Bhattacharya, S.; Jhunjhunwala, A.; Halder, A.; Bhattacharyya, D.; Mitra, A. Going beyond Base-Pairs: Topology-Based Characterization of Base-Multiplets in RNA. *RNA* **2019**, *25*, 573–589. [CrossRef]

4. Nagaswamy, U. NCIR: A Database of Non-Canonical Interactions in Known RNA Structures. *Nucleic Acids Res.* **2002**, *30*, 395–397. [CrossRef]

5. Firdaus-Raih, M.; Harrison, A.M.; Willett, P.; Artymiuk, P.J. Novel Base Triples in RNA Structures Revealed by Graph Theoretical Searching Methods. *BMC Bioinform.* **2011**, *12* (Suppl. 13). [CrossRef]

6. Appasamy, S.D.; Hamdani, H.Y.; Ramlan, E.I.; Firdaus-Raih, M. InterRNA: A Database of Base Interactions in RNA Structures. *Nucleic Acids Res.* **2016**, *44*, D266–D271. [CrossRef] [PubMed]

7. Ren, H.; Shen, Y.; Zhang, L. The λ-Turn: A New Structural Motif in Ribosomal RNA. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer: Berlin/Heidelberg, Germany, 2015; Volume 9226, pp. 456–466. [CrossRef]

8. Shen, Y.; Zhang, L. The Hasp Motif: A New Type of RNA Tertiary Interactions. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer: Berlin/Heidelberg, Germany, 2017; Volume 10362 LNCS, pp. 441–453. [CrossRef]

9. Parlea, L.G.; Sweeney, B.A.; Hosseini-Asanjan, M.; Zirbel, C.L.; Leontis, N.B. The RNA 3D Motif Atlas: Computational Methods for Extraction, Organization and Evaluation of RNA Motifs. *Methods* **2016**, *103*, 99–119. [CrossRef] [PubMed]

10. Zahran, M.; Sevim Bayrak, C.; Elmetwaly, S.; Schlick, T. RAG-3D: A Search Tool for RNA 3D Substructures. *Nucleic Acids Res.* **2015**, *43*, 9474–9488. [CrossRef] [PubMed]

11. Piątkowski, P.; Jabłónska, J.; Zyła, A.; Niedziałek, D.; Matelska, D.; Jankowska, E.; Waleń, T.; Dawson, W.K.; Bujnicki, J.M. SupeRNAlign: A New Tool for Flexible Superposition of Homologous RNA Structures and Inference of Accurate Structure-Based Sequence Alignments. *Nucleic Acids Res.* **2017**, *45*, e150. [CrossRef]

12. Hendrix, D.K.; Brenner, S.E.; Holbrook, S.R. RNA Structural Motifs: Building Blocks of a Modular Biomolecule. *Q. Rev. Biophys.* **2005**, *38*, 221–243. [CrossRef]

13. Schroeder, K.T.; Mcphee, S.A.; Ouellet, J.; Lilley, D.M.J. A Structural Database for K-Turn Motifs in RNA. *RNA* **2010**, *16*, 1463–1468. [CrossRef] [PubMed]

14. Lilley, D.M.J. The K-Turn Motif in Riboswitches and Other RNA Species. *Biochim. Biophys. Acta Gene Regul. Mech.* **2014**, *1839*, 995–1004. [CrossRef] [PubMed]

15. Miskiewicz, J.; Szachniuk, M. Discovering Structural Motifs in MiRNA Precursors from the Viridiplantae Kingdom. *Molecules* **2018**, *23*, 1367. [CrossRef] [PubMed]

16. Nissen, P.; Ippolito, J.A.; Ban, N.; Moore, P.B.; Steitz, T.A. RNA Tertiary Interactions in the Large Ribosomal Subunit: The A-Minor Motif. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 4899–4903. [CrossRef]

17. Lai, C.E.; Tsai, M.Y.; Liu, Y.C.; Wang, C.W.; Chen, K.T.; Lu, C.L. FASTR3D: A Fast and Accurate Search Tool for Similar RNA 3D Structures. *Nucleic Acids Res.* **2009**, *37* (Suppl. 2), W287–W295. [CrossRef]

18. Hamdani, H.Y.; Appasamy, S.D.; Willett, P.; Artymiuk, P.J.; Firdaus-Raih, M. NASSAM: A Server to Search for and Annotate Tertiary Interactions and Motifs in Three-Dimensional Structures of Complex RNA Molecules. *Nucleic Acids Res.* **2012**, *40*, W35–W41. [CrossRef]

19. Yen, C.Y.; Lin, J.C.; Chen, K.T.; Lu, C.L. R3D-BLAST2: An Improved Search Tool for Similar RNA 3D Substructures. *BMC Bioinform.* **2017**, *18*, 574. [CrossRef]

20. Warner, K.D.; Chen, M.C.; Song, W.; Strack, R.L.; Thorn, A.; Jaffrey, S.R.; Ferré-D'Amaré, A.R. Structural Basis for Activity of Highly Efficient RNA Mimics of Green Fluorescent Protein. *Nat. Struct. Mol. Biol.* **2014**, *21*, 658–663. [CrossRef]

21. Fernández, I.S.; Bai, X.C.; Murshudov, G.; Scheres, S.H.W.; Ramakrishnan, V. Initiation of Translation by Cricket Paralysis Virus IRES Requires Its Translocation in the Ribosome. *Cell* **2014**, *157*, 823–831. [CrossRef]

22. Burley, S.K.; Berman, H.M.; Bhikadiya, C.; Bi, C.; Chen, L.; Di Costanzo, L.; Christie, C.; Dalenberg, K.; Duarte, J.M.; Dutta, S.; et al. RCSB Protein Data Bank: Biological Macromolecular Structures Enabling Research and Education in Fundamental Biology, Biomedicine, Biotechnology and Energy. *Nucleic Acids Res.* **2019**, *47*, D464–D474. [CrossRef]

23. McDonald, I.K.; Thornton, J.M. Satisfying Hydrogen Bonding Potential in Proteins. *J. Mol. Biol.* **1994**, *238*, 777–793. [CrossRef] [PubMed]

24. Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Couch, G.S.; Greenblatt, D.M.; Meng, E.C.; Ferrin, T.E. UCSF Chimera—A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612. [CrossRef] [PubMed]

25. Cocozaki, A.I.; Altman, R.B.; Huang, J.; Buurman, E.T.; Kazmirski, S.L.; Doig, P.; Prince, D.B.; Blanchard, S.C.; Cate, J.H.D.; Ferguson, A.D. Resistance Mutations Generate Divergent Antibiotic Susceptibility Profiles against Translation Inhibitors. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 8188–8193. [CrossRef] [PubMed]

26. Schluenzen, F.; Tocilj, A.; Zarivach, R.; Harms, J.; Gluehmann, M.; Janell, D.; Bashan, A.; Bartels, H.; Agmon, I.; Franceschi, F.; et al. Structure of Functionally Activated Small Ribosomal Subunit at 3.3 Å Resolution. *Cell* **2000**, *102*, 615–623. [CrossRef]

27. Ogle, J.M.; Brodersen, D.E.; Clemons, W.M., Jr.; Tarry, M.J.; Carter, A.P.; Ramakrishnan, V. Recognition of Cognate Transfer RNA by the 30S Ribosomal Subunit. *Science* **2001**, *292*, 897–902. [CrossRef]

28. Dinman, J.D. 5S RRNA: Structure and Function from Head to Toe. *Int. J. Biomed. Sci.* **2005**, *1*, 2–7. [PubMed]

29. Klein, D.J.; Edwards, T.E.; Ferré-D'Amaré, A.R. Cocrystal Structure of a Class I PreQ1 Riboswitch Reveals a Pseudoknot Recognizing an Essential Hypermodified Nucleobase. *Nat. Struct. Mol. Biol.* **2009**, *16*, 343–344. [CrossRef]

30. Staple, D.W.; Butcher, S.E. Pseudoknots: RNA Structures with Diverse Functions. *PLoS Biol.* **2005**, *3*, e213. [CrossRef]

31. Klein, D.J.; Schmeing, T.M.; Moore, P.B.; Steitz, T.A. The Kink-Turn: A New RNA Secondary Structure Motif. *EMBO J.* **2001**, *20*, 4214–4221. [CrossRef]

32. Edwards, A.L.; Reyes, F.E.; Héroux, A.; Batey, R.T. Structural Basis for Recognition of S-Adenosylhomocysteine by Riboswitches. *RNA* **2010**, *16*, 2144–2155. [CrossRef]

33. Aalberts, D.P.; Hodas, N.O. Asymmetry in RNA Pseudoknots: Observation and Theory. *Nucleic Acids Res.* **2005**, *33*, 2210–2214. [CrossRef] [PubMed]

34. Maehigashi, T.; Dunkle, J.A.; Miles, S.J.; Dunham, C.M. Structural Insights into +1 Frameshifting Promoted by Expanded or Modification-Deficient Anticodon Stem Loops. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 12740–12745. [CrossRef] [PubMed]

35. Xu, W.; Wongsa, A.; Lee, J.; Shang, L.; Cannone, J.J.; Gutell, R.R. RNA2DMap: A Visual Exploration Tool of the Information in RNA's Higher-Order Structure. In Proceedings of the 2011 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2011, Atlanta, GA, USA, 12–15 November 2011; pp. 613–617. [CrossRef]

36. Borovinskaya, M.A.; Pai, R.D.; Zhang, W.; Schuwirth, B.S.; Holton, J.M.; Hirokawa, G.; Kaji, H.; Kaji, A.; Cate, J.H.D. Structural Basis for Aminoglycoside Inhibition of Bacterial Ribosome Recycling. *Nat. Struct. Mol. Biol.* **2007**, *14*, 727–732. [CrossRef] [PubMed]

37. Maguire, B.A.; Zimmermann, R.A. The Ribosome in Focus. *Cell* **2001**, *104*, 813–816. [CrossRef]

38. Jenner, L.; Starosta, A.L.; Terry, D.S.; Mikolajka, A.; Filonava, L.; Yusupov, M.; Blanchard, S.C.; Wilson, D.N.; Yusupova, G. Structural Basis for Potent Inhibitory Activity of the Antibiotic Tigecycline during Protein Synthesis. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 3812–3816. [CrossRef]

39. Ban, N.; Nissen, P.; Hansen, J.; Moore, P.B.; Steitz, T.A. The Complete Atomic Structure of the Large Ribosomal Subunit at 2.4 Å Resolution. *Science* **2000**, *289*, 905–920. [CrossRef]