ACS OMEGA

Open Access

http://pubs.acs.org/journal/acsodf

Article

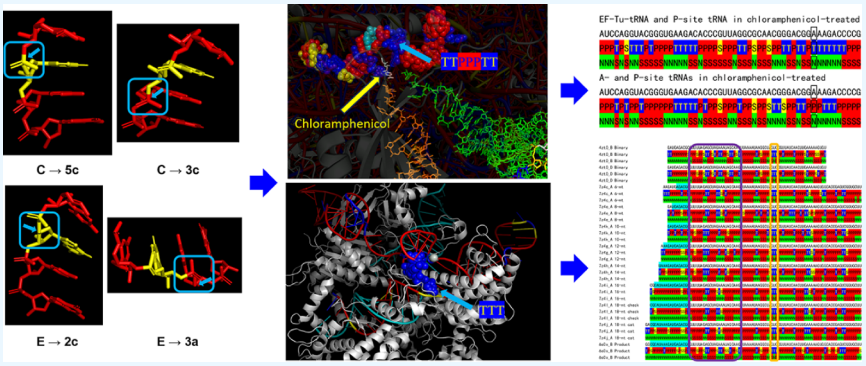
Supersecondary Structure Code for RNA: Trace of Conformational Change on the *Mycoplasma pneumoniae* Ribosome and the R-Loop Formation of Cas9

Hiroshi Izumi,* Laurence A. Nafie, and Rina K. Dukor

Cite This: *ACS Omega* 2025, 10, 4998−5005

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information



ABSTRACT: The shape of motifs is important for RNA functions and deeply reflects the structure of RNA at the supersecondary level, an intermediate level between secondary and spatial structure. However, there is currently no standardized classification system for the RNA supersecondary structures. Primary and secondary occupied conformations, accounting for 73% of the nucleic acid backbone units, were found by extending the concept of protein supersecondary structure code (SSSC) combined with the conformational code for organic molecules. The supersecondary structure code for RNA (SSCR) was introduced as a conformational term for each unit of the nucleic acid backbone using the letters P, S, T, and D, denoting respectively the primary occupied conformation (P), the secondary occupied conformation (S), the set of other conformations (T), and disordered residues (D). The alignment of SSCR sequences was used to compare with the different nucleic acid base sequences, depending on the species. SSCR can also trace the conformational change of motifs in RNA molecules such as ribosomal RNA (rRNA) and single-molecule guide RNA (sgRNA) in the R-loop formation process of Cas9. The assignment of supersecondary structure code T using the fuzzy search technique of structural code homology is an effective and quick detection method for motifs with conformational wobbling, such as the relatively rigid TTT motifs of sgRNA with Cas9, streptomycin-binding RNA aptamer, 23S rRNA, 2′-dG-II riboswitch, and human hepatitis B virus ε pregenomic RNA, which work as the scaffold for protein and RNA molecules or as the support stand for small external substrates.

INTRODUCTION

Secondary structures of RNA and protein molecules are defined by hydrogen-bonding interactions. RNA molecules first fold into secondary structures according to the canonical Watson−Crick (G:C and A:U),[1] wobble (G:U),[2] and other noncanonical base pairs (e.g., A:C and U:U), which form helices linked to each other by unpaired single-stranded nucleotides. In the diversity of these combinations in different structures, some recurring motifs have been found that are important for RNA function.[3,4] The description of these combinations and their characteristics considers the structure of RNA at the supersecondary level, an intermediate level between secondary and spatial structure. Such analysis allows, in many cases, the use of laborious "small-angle X-ray scattering" and, more importantly, to better understand the general principles of RNA folding.

RNA structures are highly variable and can form a wide range of motifs, such as helix−loop−helix, bulges with internal loops—regions where base pairing is disrupted; pseudo-knots—complex structures where a single-stranded region pairs with a region inside the helix, and so on.[5,6] There is

Received: November 24, 2024
Revised: December 23, 2024
Accepted: January 13, 2025
Published: January 27, 2025

currently no standardized classification system for RNA supersecondary structures.

Conformational change of ribosomal RNA (rRNA)[7] and long noncoding RNA (lncRNA)[8] is essential for the completion of functional expression. In particular, lncRNA collects RNA-binding proteins[9] as a scaffold and leads to liquid−liquid phase separation.[10] The characteristics of phase separation can be expected to be applied to a new field of organic synthesis. Further, R-loop formation and activation arise as a consequence of the complex conformational change of Cas9 with single-molecule guide RNA (sgRNA).[11,12] The description of the secondary structure for RNA is available to assess the effect of *the ssrA* leader sequence and T1 stem structure base pairing on transcriptional read-through by the introduction of two-point mutations.[13] However, the description is insufficient in relation to the conformational change of lncRNA with the interaction of RNA-binding proteins.

We have proposed the revised sequence rule in the IUPAC rule *p*-94.2 because it is necessary for the application of deep neural network for conformational comparison with different organic molecules.[14] The conformational code for organic molecules (Figure 1B) has also been adapted and applied[15,16] to the protein supersecondary structure code (SSSC),[17] which is represented as a conformation propensity using the letters

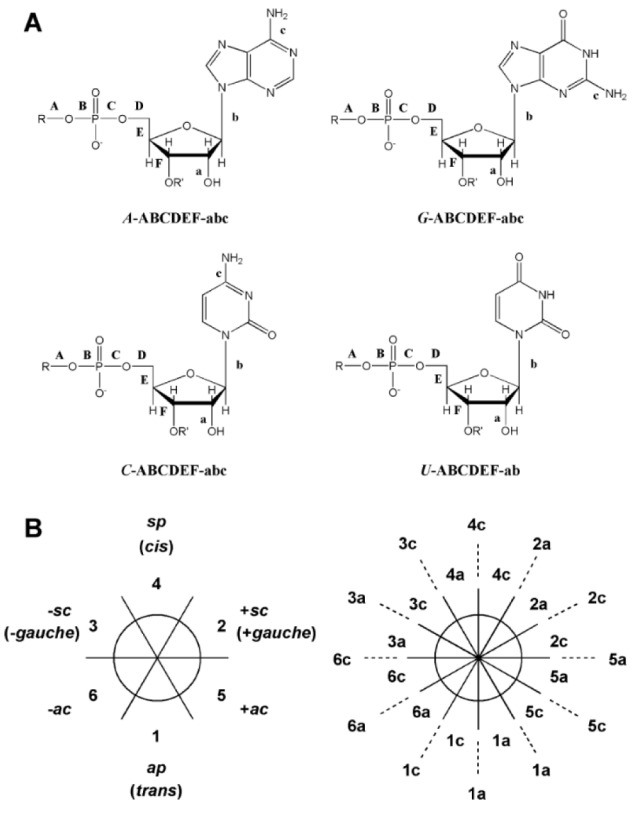


Figure 1. Definition of angle locations and conformational elements for the conformational code of main chains for RNA. (A) Angle locations consist of prefix of nucleic acid base (A, G, C, and U) and symbols indicating the bond locations (A,B,···etc.). (B) Conformational elements represent the classification of dihedral angles, and the elements **1**, **2**, **3**, **4**, **5**, and **6** correspond to conformational terms *ap* (antiperiplanar), + *sc* (+synclinal), − *sc* (−synclinal), *sp* (synperiplanar), + *ac* (+anticlinal), and − *ac* (−anticlinal), respectively. The terms **c** and **a**, in the conformational elements, mean clockwise and anticlockwise, respectively.

"H," "S," "T," and "D" for each amino acid peptide unit, referring to an α-helix-type conformation (H), a β-sheet-type conformation (S), a variety of other-type conformations (T), and disordered residues or the C-terminus (D), and this code has been approved as a protocol for a molecular biology database.[18] The deep neural network-based conformational variability prediction system of protein structures (SSSCPreds)[19] using SSSC shows a high level of prediction accuracy for the correlation between conformational variability and phenotypes by mutations of spike proteins for SARS-CoV-2 variants.[20,21] SSSCPreds has also been used to detect an uncompetitive inhibition mechanism related to drug resistance by amino acid mutations.[22] Furthermore, this method shows similar rigid patterns at the furin cleavage sites for the Omicron variant of SARS-CoV-2 and avian influenza A H5N1 virus with bovine and human infections since March 2024 in the United States and has also been approved as a protocol in a molecular biology database.[23] In this paper, we show that the concepts of the molecular conformational code and SSSC may be extended to the description of conformational changes of rRNA, lncRNA, and sgRNA. In particular, we report that a constructed supersecondary structure code for RNA (SSCR) can represent a key drug-bound RNA conformation on the rRNA site related to erythromycin resistance and can track the structural changes of rRNA and sgRNA in the R-loop formation process of Cas9.

RESULTS AND DISCUSSION

Derivation of Supersecondary Structure Code for RNA. The conformational code is composed of the combination of the codes of regional angle locations and the 12-divided segments (conformational elements) and is available for the description of the local structures of various organic molecules (Figure 1).[15] Then, unique motifs can be retrieved if maximal common substructures (MCS) of organic molecules are selected as the regional angle locations.[24] The conformational terms, synperiplanar, synclinal, anticlinal, and antiperiplanar, are defined as the classification of dihedral angles in the IUPAC rule *p*-94.2.[25] However, the fine 8-divided segments, + *ap* (+antiperiplanar), − *ap* (−antiperiplanar), + *sc* (+synclinal), − *sc* (−synclinal), + *sp* (+synperiplanar), − *sp* (−synperiplanar), + *ac* (+anticlinal), and − *ac* (−anticlinal) are not the equal division, so it is not suitable for mathematical procedures. The 12-divided segments of conformational elements are useful for the selection of fine and rough correspondence among the unique motifs, depending on the variable range setting, and then make the classification related to the energy barrier of the bond-rotation conformational change easy.

The conformational elements at each RNA angle location (Figure 1A) were calculated using protein data bank (PDB)[26] data of RNA molecules according to the classification of dihedral angles (Figure 1B). Specifically, four atomic coordinates of visible atoms at each RNA angle location for determination of dihedral angles from PDB data were extracted according to the following proposed rule for the IUPAC rule *p*-94.2: (a) If all the atoms or groups of a set are different, or if one atom or group is unique, select the atom or group that has priority by the sequence rule, except where one atom or group is unique between the rankings of each atom in the n^{th} sphere and in the $(n + 1)^{th}$ sphere (*p*-92.1.5 Exploration of a hierarchical digraph), in which case select that one; (b) if all the atoms or groups of a set are identical, select the atom or

https://doi.org/10.1021/acsomega.4c10681
ACS Omega 2025, 10, 4998−5005

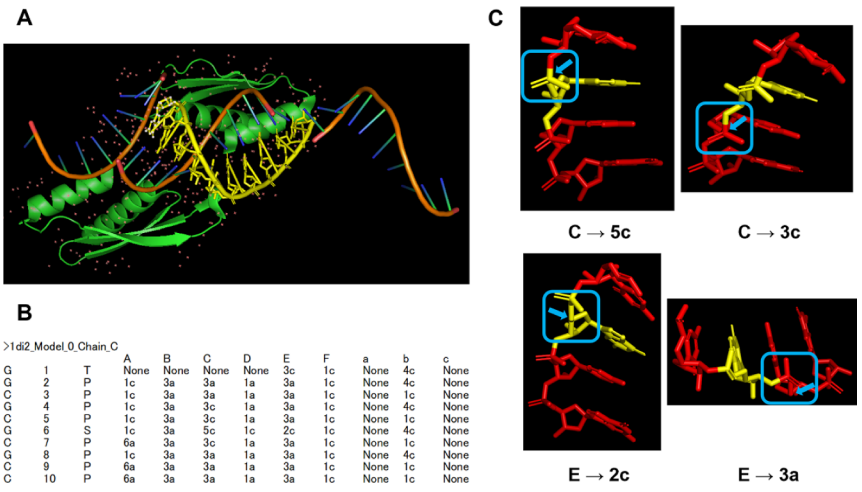


Figure 2. Primary and secondary occupied conformations for RNA molecules. (A) Molecular model of double stranded RNA of PDB ID 1di2 (yellow: chain C). (B) Conformational elements of chain C at each angle location (P, primary occupied conformation; S, secondary occupied conformation; T, other-type conformations). (C) Newman projections of conformational elements at RNA angle locations C and E for primary (right, C7) and secondary (left, G6) occupied conformations in chain C of 1di2.

group that provides the smallest torsion angle.[14] Then, the corresponding dihedral angles of the four atomic coordinates were converted to conformational codes. For example, as for C7 of chain C for PDB ID 1di2 (Figure 2), the four atomic coordinates of visible atoms, C_{OC}−C−O−P (−146.8°) for angle location A, C−O−P−O_C (−80.6°) for angle location B, O_C−P−O−C (−57.3°) for angle location C, P−O−C−C (166.8°) for angle location D, O−C−C−O (−64.3°) for angle location E, and O−C−C−O (−159.3°) for angle location F were selected, and then the pattern of conformational codes 6a3a3c1a3a1c for C7 of chain C was obtained. As shown in Figure 2, we found that unpublished different types of conformations exist in the double helix structure by using the conformational code. The completely different conformations at angle locations C and E complement each other and maintain the double helix (Figure 2C).

Next, the distribution of conformational elements at each angle location of main chain for 7,250,546 nucleic acid backbone units was tabulated. Major and minor peaks at angle locations C and E are also found in Table 1. Based on this distribution, the primary (P) and secondary (S) occupied conformation sequences were defined as the template patterns 6a3a3a1a3a6a and 6a3a5c1a2c6a, respectively. The assignment of the nucleic acid backbone conformations was carried out using the fuzzy search of structural code homology,[17] in which it was judged as the high homology if the conformational elements were included in a range of 90°, and the structural code homology of the main chain for each nucleic acid backbone unit was calculated as the logical conjunction of structural code homology at the angle locations A-F. Specifically, there were 12 sets, {4c, 2a, 2c}, {2a, 2c, 5a}, {2c, 5a, 5c}, {5a, 5c, 1a}, {5c, 1a, 1c}, {1a, 1c, 6a}, {1c, 6a, 6c}, {6a, 6c, 3a}, {6c, 3a, 3c}, {3a, 3c, 4a}, {3c, 4a, 4c}, and {4a, 4c, 2a}, in a range of 90°. It was supposed that a set of these sets was X = {p, q, r}, and a conformational element, which was compared at a specific angle location, was y. If all elements were satisfied with the equation, y ∈ X, it was judged that the structural code homology of the angle location was high.[17] From the nucleic acid backbone conformations of main chains in the PDB data of RNA molecules, 4,650,163 (64.1%) were assigned as P-type, 637,831 (8.8%) as S-type, and 1,962,552 (27.1%) as other-type (T). The T-type conformations further contained groups of 6a5c3a1a3a6a (K, 3.4%), 6a3a5c1a3a6a (L, 3.2%), 6a3a3a1a2c6a (M, 2.3%), 6a5c5c1a3a6a (G, 2.1%,), 6a5c5c1a2c6a (H, 1.2%), and 6a5c3a1a2c6a (I, 0.5%).

Table 1. Distribution of Conformational Elements at Each Angle Location of Main Chain for 7,250,546 Nucleic Acid Backbone Units[a]

	A	B	C	D	E	F
4a	8972	77864	100944	673	170188	0
4c	14475	53235	65876	693	160401	1
2a	30129	126551	214952	8225	561497	1
2c	28943	178048	295788	80521	441185	84
5a	12087	172762	199561	229661	108598	236
5c	16461	166727	288834	723717	67829	248
1a	123973	170645	484250	3715106	123543	1378
1c	2107295	179568	230324	1571402	74820	5923384
6a	3473627	271427	193816	584947	45024	305979
6c	1031716	768268	369919	212648	127375	733757
3a	287497	4222199	3097253	38764	3193849	215096
3c	25674	773394	1618861	4353	2105380	37
Other	89697	89858	90168	79836	70857	70345
Total	7250546	7250546	7250546	7250546	7250546	7250546

[a]The primary (red) and secondary (yellow) abundant conformational elements for the selection of template patterns are shown.

In a way analogous to SSSC, SSCR is represented as a conformational propensity for each unit of the nucleic acid backbone using the letters P, S, T, and D, denoting respectively the primary occupied conformation (P), the secondary occupied conformation (S), the set of other conformations (T), and disordered residues (D). The SSCR sequence only judges whether the nucleic acid backbone unit is close to P-type, S-type, T-type, or D-type conformations, but the shapes of motifs deeply reflect the structure of RNA at the supersecondary level, an intermediate level between secondary and spatial structure, so it can characterize particular RNA motifs which cannot be classified as secondary structures. The T-type conformations can also be assigned as the similar conformations for the other RNA molecules in the PDB data using the fuzzy search of structural code homology.[17] This

https://doi.org/10.1021/acsomega.4c10681
ACS Omega 2025, 10, 4998−5005

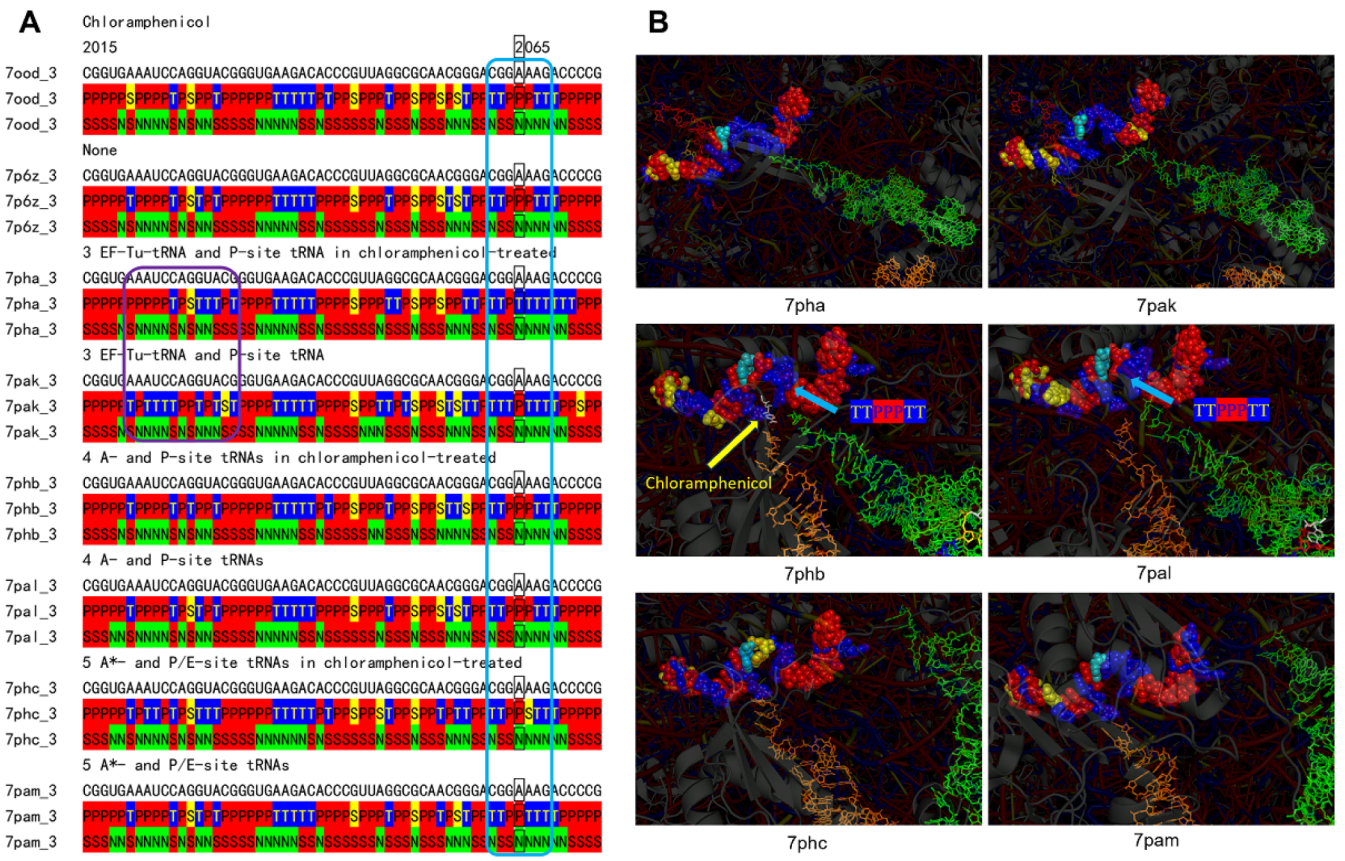


Figure 3. Conformational change of the key drug-bound rRNA site related to erythromycin resistance. (A) SSCR (middle) sequences (P: primary occupied conformation; S: secondary occupied conformation; T: other-type conformations) and SeqSSR (bottom) sequences (S: stem; N: nonstem) of the key drug-bound rRNA site related to erythromycin resistance (black rectangles: mutation sites of erythromycin resistance; cyan rectangular frame structure: sequence of drug-bound motif; purple rectangular frame structure: sequence of motif interacting with drug-bound motif). (B) Conformational change (red: primary occupied conformation; yellow: secondary occupied conformation; blue: other-type conformations; cyan: mutation site of erythromycin resistance; orange and green: tRNAs) of the key drug-bound rRNA site related to erythromycin resistance with chloramphenicol (left) and without (right).

simplification for the classification of SSCR makes structural changes easier to track, but of course, it risks overlooking finer details of RNA's structural dynamics, particularly complex interactions or subtle conformational shifts that may play critical roles in function. In that case, the fuzzy search of structural code homology with the narrow range setting is just used. However, such procedures are not effective empirically because RNA structures are highly variable.

In addition, the sequence of secondary structure for RNA (SeqSSR) was defined as stem (S) or nonstem (N), depending on whether the distance of paired nucleic acid bases is less than 3.5 Å or not.

tRNA Interactions in Chloramphenicol-Treated Cells on the *Mycoplasma pneumoniae* Ribosome. SSCR was applied to the translation dynamics at atomic detail inside a bacterial cell (*Mycoplasma pneumoniae*).[7] As shown in Figures 3A and S1, the SeqSSR sequences of a series of tRNA interactions in chloramphenicol-treated (*Cm*-treated) or non-treated samples were almost identical. On the other hand, the corresponding SSCR sequences were largely changed. It is suggested that the increase of continuous T-type conformations reflects an increase in strain by external interactions such as tRNA.

In many stages of tRNA interactions in *Cm*-treated cells, *Cm* was not visible in the cryo-electron microscopy (cryo-EM) structures (Figure 3B). However, the interaction between *Cm* and rRNA with EF-Tu-tRNA and P-site tRNA disconnected the close motif (purple rectangular frame structure in Figure 3A) interacting with the key drug-bound motif (cyan rectangular frame structure in Figure 3A) and eliminated the distortion from TPTTTT to PPPPPT.

Further, the key drug-bound motif corresponded to the mutation site of erythromycin resistance (A2065G). The nucleic acid base sequences of the key drug-bound motif were different depending on the species (*Mycoplasma pneumonia*, CGGAAAG; *Haloarcula marismortui*, AGCAAAG; *Deinococcus radiodurans*, CGAAAAG). However, the SSCR sequences of the tRNA-free motif (TTPPPTT) were almost identical (Figure S2). The tRNA-free motif (TTPSPST) of *Staphylococcus aureus* showed a different sequence pattern (Figure S2), but the deletion mutation in ribosomal protein uL22, associated with erythromycin resistance, forced a change in sequence.[27] *Cm* was visible only in the cryo-EM structures with the SSCR sequence (TTPPPTT), and it corresponded to the most abundant stage of the tRNA interaction. The SSCR sequence can explain rationally why *Cm* is not visible.

R-Loop Formation of Cas9. Conformational change of sgRNA in the R-loop formation process of Cas9[11] has not been fully figured out. The SeqSSR sequences of a series of sgRNA interactions in the process were also almost identical (Figure 4A). On the other hand, the binary structure (Figure 4B) without DNA showed the complex SSCR sequence pattern

https://doi.org/10.1021/acsomega.4c10681
ACS Omega 2025, 10, 4998−5005

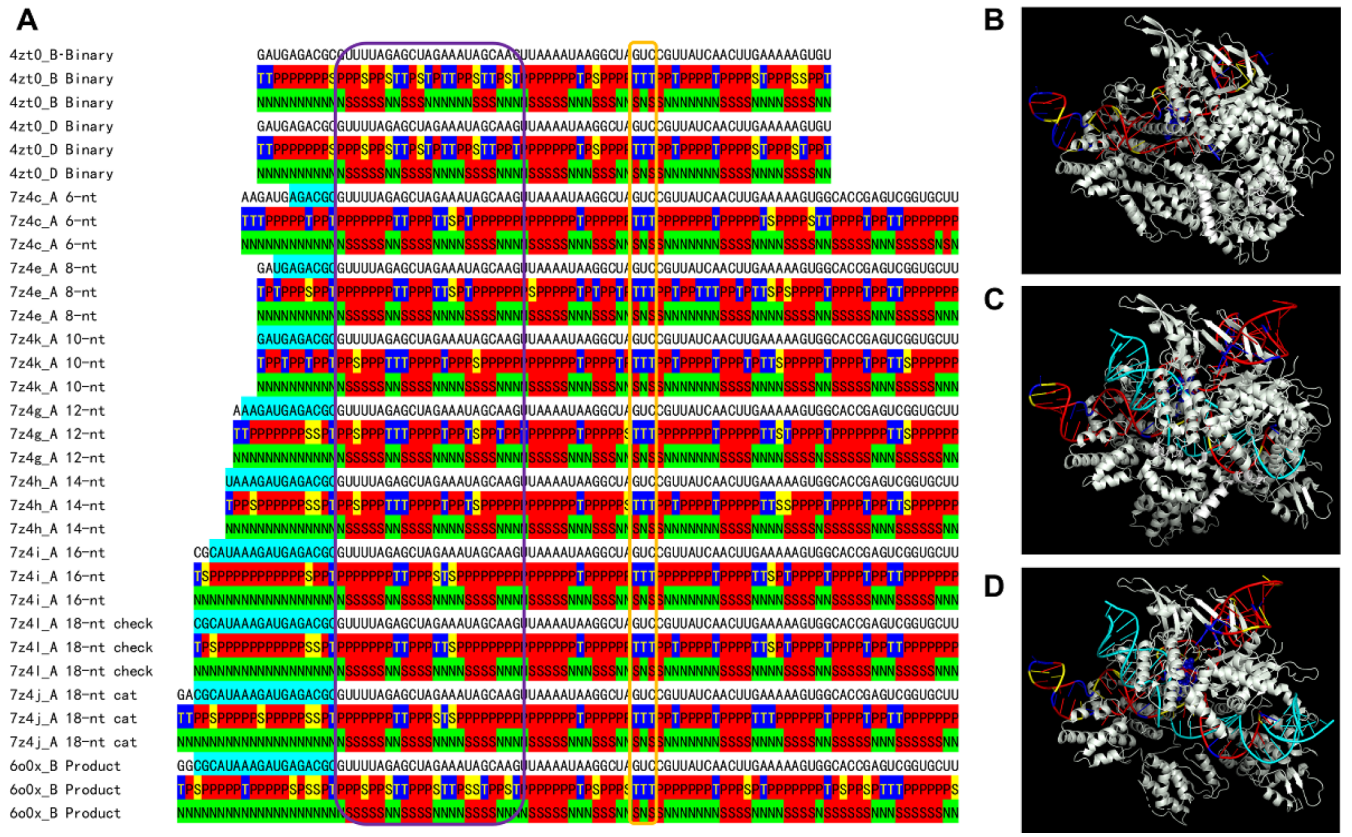


Figure 4. Conformational change of sgRNA in the R-loop formation process of Cas9. (A) SSCR (middle) sequences (P: primary occupied conformation; S: secondary occupied conformation; T: other-type conformations) and SeqSSR (bottom) sequences (S: stem; N: nonstem) of sgRNA in the R-loop formation process (cyan: DNA binding motif; purple rectangular frame structure: variable motif tied to DNA binding motif; orange rectangular frame structure: TTT motif). (B–D) Conformational change (red: primary occupied conformation; yellow: secondary occupied conformation; blue: other-type conformations; cyan: DNA) of sgRNA for the binary structure (B) on PDB ID 4zt0, the catalytic stage (C) on PDB ID 7z4j, and the product formation stage (D) on PDB ID 6o0x.

PPPSPPSTTPSTPTTPPSTTPST for the motif of GUUUUAGAGCUAGAAAUAGCAAG tied to the DNA binding motif. It suggested that the RNA-protein interaction between sgRNA and Cas9 increased in strain, and the quick conformational recovery from PPPPPPPTTPPPSTSPPPPPPPP (Figure 4C) of the catalytic stage to PPPSPPSTTPPPSTTPSSTPPST (Figure 4D) of the product formation stage supported the easy release of the cleaved DNA product.

In contrast with the variable motif tied to the DNA binding motif, the characteristic TTT motif was consistently rigid (Figures 4A and 5A). As shown in Figures 5B and S3, this TTT motif was bound to Cas9 tightly as a scaffold. To characterize the TTT motif, the conformational elements at each angle location for the nucleic acid backbone conformations of sgRNA were extracted, and then the assignment of supersecondary structure code T was carried out using the fuzzy search of structural code homology[17] in a similar way to proteins (Figure 5A).[18] crRNA, apo SAM-IV riboswitch models, *E. coli* tRNA-Arg2 (acg) anticodon stem and loop, transferrin receptor iron regulatory element B RNA, P-site tRNA, 2′-dG-II class of riboswitch, 61 nt human hepatitis B virus ϵ pregenomic RNA, mRNA, 25S rRNA, streptomycin RNA-aptamer, hc16 ligase product models, tmRNA, 23S rRNA, and 16S rRNA as a group of the TTT motifs were retrieved from 28,272 PDB subunit data for RNA molecules (Figure S3).

Antibiotics RNA aptamers are obtained by an approach based on the in vitro selection of small RNA molecules (30 to 60 nucleotides) that bind antibiotics with high selectivity and affinity.[28] The TTT motif of the streptomycin-binding RNA aptamer formed the binding pocket and worked as a relatively rigid support stand (Figure 5C). The other TTT motif of 23S rRNA contacted the 50S ribosomal protein L11. The conformational wobbling of TTT motifs for 23S rRNA within the determined threshold for the fuzzy search of structural code homology[17] (Figure 5D) was observed with the large conformational change of 23S rRNA between PDB ID 5mdz and 4v6t (Figure S3). Although the TTT motifs of 23S rRNA were identical, the conformational element at angle location D of G1087 showed a large wobbling, especially (Figure 5A). RNA structures are highly variable, so the fuzzy search technique of structural code homology[17] is an effective and quick detection method for such flexible motifs.

Bioinformatically defined classes of riboswitches have proved that there exists variation in both the cognate metabolite that regulates gene expression and the ability to discriminate between chemically related metabolites,[29] and 2′-dG-II riboswitch is one of them. The TTT motif was also found in this riboswitch and formed the support stand, which was very close to the external 2′-deoxyguanosine (Figure 5E). Further, the pseudotriloop of human hepatitis B virus ϵ pregenomic RNA corresponded to the TTT motif, and the hydroxyethylpiperidine tail for the predicted docking pose of Raloxifene was

https://doi.org/10.1021/acsomega.4c10681
ACS Omega 2025, 10, 4998−5005

A

sgRNA								
4zt0_B Binary	G58	T	6c	2c	6c	1a	3a	1c
4zt0_D Binary	G58	T	6c	2c	6c	1a	3a	1c
7z4j_A 18-nt cat	G58	T	6c	5a	6c	1a	3a	1c
7z4l_A 18-nt check	G58	T	6a	5a	3a	1a	3a	1c
6o0x_B Product	G58	T	6a	5c	3a	5c	3a	1c
4zt0_B Binary	U59	T	6c	3a	3c	1a	2c	1c
4zt0_D Binary	U59	T	6c	3a	3c	1a	2c	1c
7z4j_A 18-nt cat	U59	T	6c	3a	3a	5c	2c	6c
7z4l_A 18-nt check	U59	T	6c	3a	3a	1a	2c	6c
6o0x_B Product	U59	T	6c	3a	3c	1a	2c	1c
4zt0_B Binary	C60	T	1c	1a	3a	1a	3a	1c
4zt0_D Binary	C60	T	1c	1a	3a	1a	3a	1c
7z4j_A 18-nt cat	C60	T	1c	5a	3a	1c	3a	1c
7z4l_A 18-nt check	C60	T	1c	5a	3a	1a	3a	1c
6o0x_B Product	C60	T	1c	1a	3a	1a	3a	1c
Streptomycin RNA-aptamer								
1nta_B	A108	T	1c	1a	3a	5a	3c	6a
1nta_B	C109	T	6c	3a	3a	5c	2a	1c
1nta_B	G110	T	1a	1a	3a	1c	3c	1c
23S ribosomal RNA								
5mdz_1	G1087	T	6c	5c	1c	6a	3c	6a
4v6t_BA	G1087	T	3c	1a	1c	5a	3c	1c
5mdz_1	A1088	T	6a	3a	3a	1a	2c	6c
4v6t_BA	A1088	T	3a	6a	4a	5c	4c	1c
5mdz_1	A1089	T	1c	5a	3a	1a	3a	6c
4v6t_BA	A1089	T	6c	5c	6c	5c	3c	6a
2'-dG-II riboswitch								
6p2h_A	A23	T	3a	2c	1c	1a	3a	1c
6p2h_A	A24	T	6c	3c	3a	1c	2c	1c
6p2h_A	U25	T	6a	1a	3a	5c	3a	1c
Human Hepatitis B virus epsilon pre-genomic RNA								
6var_0_A	G35	T	6a	5c	3a	5c	3a	1c
6var_0_A	C36	T	6c	3c	6c	1a	2c	6c
6var_0_A	C37	T	6a	5a	3a	1c	3a	1c

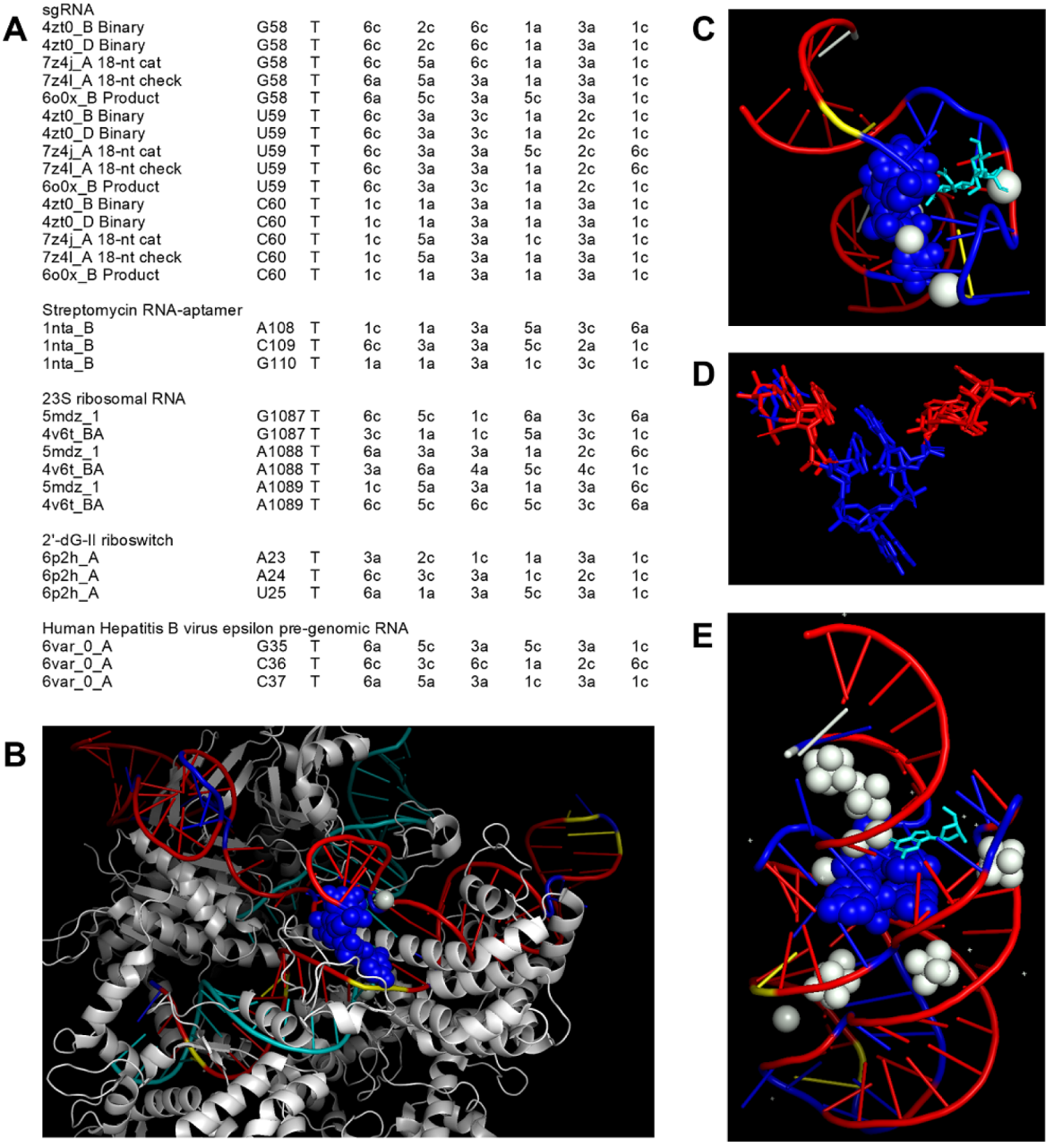


Figure 5. Characterization of TTT motifs for RNA molecules. (A) Conformational elements of TTT motifs for RNA molecules at each angle location. (B) Molecular model of TTT motif (sphere) for sgRNA of PDB ID 7z4j (red: primary occupied conformation; yellow: secondary occupied conformation; blue: other-type conformations; cyan: DNA). (C) Molecular model of TTT motif (sphere) for streptomycin-binding RNA aptamer of PDB ID 1nta (red: primary occupied conformation; yellow: secondary occupied conformation; blue: other-type conformations; cyan: streptomycin). (D) Conformational wobbling of TTT motifs for 23S rRNA between PDB ID 5mdz and 4v6t (red: primary occupied conformation; blue: other-type conformations). (E) Molecular model of TTT motif (sphere) for 2′-dG-II riboswitch of PDB ID 6p2h (red: primary occupied conformation; yellow: secondary occupied conformation; blue: other-type conformations; cyan: 2′-deoxyguanosine).

also close to the support stand formed by the TTT motif (Figure S3).[30]

CONCLUSIONS

In conclusion, we found that the P- and S-type conformations accounted for 73% of the nucleic acid backbone units. SSCR using these strings can trace the conformational change of motifs in RNA molecules such as rRNA and sgRNA. The conformational and functional analysis tools of lncRNA such as the liquid−liquid phase separation have been sought, but it is difficult to figure out unique motifs only from the three-dimensional structural data. We also detected the key drug-bound motif with the SSCR sequence TTPPPTT, and the conformational variability of this motif rationalized the invisible cause of *Cm* in many stages of the tRNA interactions in *Cm*-treated cells. The alignment of SSCR sequences was also especially effective in comparison with the different nucleic acid base sequences depending on the species.

In contrast with the variable motif of sgRNA tied to the DNA binding motif in the R-loop formation process of Cas9, the characteristic TTT motif was consistently rigid. crRNA, apo SAM-IV riboswitch models, *E. coli* tRNA-Arg2 (acg)

https://doi.org/10.1021/acsomega.4c10681
ACS Omega 2025, 10, 4998−5005

anticodon stem and loop, transferrin receptor iron regulatory element B RNA, P-site tRNA, 2′-dG-II class of riboswitch, 61 nt human hepatitis B virus ε pregenomic RNA, mRNA, 25S rRNA, streptomycin RNA-aptamer, hc16 ligase product models, tmRNA, 23S rRNA, and 16S rRNA contained the similar TTT motifs. It is suggested that the relatively rigid TTT motif is one of the important motifs to work as the scaffold for protein and RNA molecules or as the support stand for small external substrates. Here, we have constructed a deep neural network-based prediction system using SSCR and SeqSSR sequences for RNA molecules such as lncRNA, similar to the previous, successful protein analog SSSCPreds. The limitations for the conformational prediction are that the number of registrations for the PDB data of RNA molecules is very few against that of protein molecules (3.7%), and the data of rRNA account for the majority of data set.

COMPUTATIONAL METHODS

SSCRview. The conformational elements at each angle location were calculated using PDB data[26] of RNA molecules according to the classification of dihedral angles in the conformational code for organic molecules.[15] SSCR was represented as a conformation term for each nucleic acid backbone unit using the letters P, S, T, and D referring, respectively, to a primary occupied conformation (P), a secondary occupied conformation (S), a variety of other-type conformations (T), and disordered residues (D), which was derived from the template patterns, characterized as conformational codes, such as 6a3a3a1a3a6a (P-type conformation) and 6a3a5c1a2c6a (S-type conformation) in a similar way for protein supersecondary structure code (SSSC).[17,18]

The autoconversion conformation information needed for SSCR was carried out by the following procedures: 1) extraction process of four atomic coordinates for determination of dihedral angles from PDB data, 2) conversion process from dihedral angles to conformational codes, and 3) strict and/or fuzzy search processes of structural code homology. In the fuzzy search of structural code homology (process 3), it was judged that the homology was high if the conformational elements were included in the range of 90°. Specifically, there were 12 sets, {4c, 2a, 2c}, {2a, 2c, 5a}, {2c, 5a, 5c}, {5a, 5c, 1a}, {5c, 1a, 1c}, {1a, 1c, 6a}, {1c, 6a, 6c}, {6a, 6c, 3a}, {6c, 3a, 3c}, {3a, 3c, 4a}, {3c, 4a, 4c}, and {4a, 4c, 2a}, in a range of 90°. It was supposed that a set of these sets was X = {p, q, r}, and a conformational element, which was compared at a specific angle location, was y. If all elements were satisfied with the equation, $y \in X$, it was judged that the structural code homology of the angle location was high.[17] Finally, the structural code homology of the main chain for each nucleic acid backbone unit was calculated as the logical conjunction of structural code homology at angle locations A-F.

Simultaneously, the sequence of secondary structure for RNA (SeqSSR) was judged as stem (S) or nonstem (N) depending on whether the distance of paired nucleic acid bases is less than 3.5 Å or not.

Assignment of supersecondary structure code T was carried out using the fuzzy search of structural code homology with template patterns such as 6c3a3a5c2c6c (U59 of PDB ID 7z4j) in a similar way to proteins,[18] and the subunit file list of RNA molecules containing the T-type conformation within the determined threshold for the fuzzy search of structural code homology was obtained.

The observed PDB structure files were converted to FASTA-format files containing the nucleic acid base sequences with SSCR and SeqSSR sequences and Word files with the colored sequences using SSCRview. The subunit file list for the assignment of T conformations was also obtained by using SSCRview. SSCRview cannot convert chemically modified nucleic acid base sequences to conformational codes. Rename of chain names for PDB data may also be necessary because of the uppercase/lowercase error for the Windows operation system. Software SSCRview has been deposited on the H.I. website (https://staff.aist.go.jp/izumi.h/SSSCPreds/index-e.html) and is freely available.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.4c10681.

SSCR and SeqSSR sequences of the key drug-bound rRNA sites at tRNA interaction stages and depending on the species, retrieval of TTT motifs for RNA molecules, and SSCR and SeqSSR data of PDB structures for rRNA (PDF)

Biopython License Agreement; user's guide of SSCRview (ZIP)

AUTHOR INFORMATION

Corresponding Author

Hiroshi Izumi − *National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibaraki 305-8569, Japan;* orcid.org/0000-0002-8369-3038; Email: izumi.h@aist.go.jp

Authors

Laurence A. Nafie − *Department of Chemistry, Syracuse University, Syracuse, New York 13244-4100, United States; BioTools Inc., Jupiter, Florida 33458, United States;* orcid.org/0000-0002-3221-4332

Rina K. Dukor − *BioTools Inc., Jupiter, Florida 33458, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.4c10681

Author Contributions

H.I. carried out the conceptualization, investigation, and writing of the manuscript and provided the methodology and software. L.A.N. and R.K.D. performed research organization and writing of the manuscript.

Notes

The authors declare the following competing financial interest(s): H.I. listed as an inventor on patents US11119033B2 and JP4691728B2 related to software.

ACKNOWLEDGMENTS

This work was supported by JSPS KAKENHI Grant Number JP22K05073.

REFERENCES

(1) Xia, T.; SantaLucia, J. J.; Burkard, M. E.; Kierzek, R.; Schroeder, S. J.; Jiao, X.; Cox, C.; Turner, D. H. Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry* **1998**, *37*, 14719−14735.

(2) Dirks, R. M.; Lin, M.; Winfree, E.; Pierce, N. A. Paradigms for computational nucleic acid design. *Nucleic Acids Res* **2004**, *32*, 1392−1403.
(3) Mathews, D. H.; Disney, M. D.; Childs, J. L.; Schroeder, S. J.; Zuker, M.; Turner, D. H. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl. Acad. Sci. U.S.A* **2004**, *101*, 7287−7292.
(4) Wayment-Steele, H. K.; Kladwang, W.; Strom, A. I.; Lee, J.; Treuille, A.; Becka, A.; Das, R. RNA secondary structure packages evaluated and improved by high-throughput experiments. *Nat. Methods* **2022**, *19* (10), 1234−1242.
(5) Chen, J.-L.; Greider, C. W. Functional analysis of the pseudoknot structure in human telomerase RNA. *Proc. Natl. Acad. Sci. U.S.A* **2005**, *102*, 8080−8085.
(6) Ferré-D'Amaré, A. R.; Zhou, K.; Doudna, J. A. Crystal structure of a hepatitis delta virus ribozyme. *Nature* **1998**, *395*, 567−574.
(7) Xue, L.; Lenz, S.; Zimmermann-Kogadeeva, M.; Tegunov, D.; Cramer, P.; Bork, P.; Rappsilber, J.; Mahamid, J. Visualizing translation dynamics at atomic detail inside a bacterial cell. *Nature* **2022**, *610*, 205−211.
(8) Smola, M. J.; Christy, T. W.; Inoue, K.; Nicholson, C. O.; Friedersdorf, M.; Keene, J. D.; Lee, D. M.; Calabrese, J. M.; Weeks, K. M. SHAPE reveals transcript-wide interactions, complex structural domains, and protein interactions across the *Xist* lncRNA in living cells. *Proc. Natl. Acad. Sci. U.S.A* **2016**, *113*, 10322−10327.
(9) Loughlin, F. E.; Lukavsky, P. J.; Kazeeva, T.; Reber, S.; Hock, E.-M.; Colombo, M.; Von Schroetter, C.; Pauli, P.; Cléry, A.; Mühlemann, O.; Polymenidou, M.; Ruepp, M.-D.; Allain, F. H.-T. The solution structure of FUS bound to RNA reveals a bipartite mode of RNA recognition with both sequence and shape specificity. *Mol. Cell* **2019**, *73* (3), 490−504.e6.
(10) Yasuda, S.; Tsuchiya, H.; Kaiho, A.; Guo, Q.; Ikeuchi, K.; Endo, A.; Arai, N.; Ohtake, F.; Murata, S.; Inada, T.; Baumeister, W.; Fernández-Busnadiego, R.; Tanaka, K.; Saeki, Y. Stress- and ubiquitylation-dependent phase separation of the proteasome. *Nature* **2020**, *578*, 296−300.
(11) Pacesa, M.; Loeff, L.; Querques, I.; Muckenfuss, L. M.; Sawicka, M.; Jinek, M. R-loop formation and conformational activation mechanisms of Cas9. *Nature* **2022**, *609*, 191−196.
(12) Zhu, X.; Clarke, R.; Puppala, A. K.; Chittori, S.; Merk, A.; Merrill, B. J.; Simonović, M.; Subramaniam, S. Cryo-EM structures reveal coordinated domain motions that govern DNA cleavage by Cas9. *Nat. Struct. Mol. Biol* **2019**, *26*, 679−685.
(13) Davies, M. R.; Keller, N.; Brouwer, S.; Jespersen, M. G.; Cork, A. J.; Hayes, A. J.; Pitt, M. E.; De Oliveira, D. M. P.; Harbison-Price, N.; Bertolla, O. M.; Mediati, D. G.; Curren, B. F.; Taiaroa, G.; Lacey, J. A.; Smith, H. V.; Fang, N.-X.; Coin, L. J. M.; Stevens, K.; Tong, S. Y. C.; Sanderson-Smith, M.; Tree, J. J.; Irwin, A. D.; Grimwood, K.; Howden, B. P.; Jennison, A. V.; Walker, M. J. Detection of *Streptococcus pyogenes* $M1_{UK}$ in Australia and characterization of the mutation driving enhanced expression of superantigen SpeA. *Nat. Commun* **2023**, *14* (1), 1051.
(14) Izumi, H. Consideration of the sequence rule in rule P-94.2. *Chem. Int* **2018**, *40*, 36−37.
(15) Izumi, H.; Ogata, A.; Nafie, L. A.; Dukor, R. K. A revised conformational code for the exhaustive analysis of conformers with one-to-one corre-spondence between conformation and code: Application to the VCD analysis of (*S*)-ibuprofen. *J. Org. Chem* **2009**, *74*, 1231−1236.
(16) Izumi, H.; Yamagami, S.; Futamura, S.; Nafie, L. A.; Dukor, R. K. Direct observation of odd-even effect for chiral alkyl alcohols in solution using vibrational circular dichroism spectroscopy. *J. Am. Chem. Soc* **2004**, *126*, 194−198.
(17) Izumi, H.; Wakisaka, A.; Nafie, L. A.; Dukor, R. K. Data mining of supersecondary structure homology between light chains of immunogloblins and MHC molecules: absence of the common conformational fragment in the human IgM rheumatoid factor. *J. Chem. Inf. Model* **2013**, *53*, 584−591.
(18) Izumi, H. Homology searches using supersecondary structure code. *Methods Mol. Biol* **2019**, *1958*, 329−340.
(19) Izumi, H.; Nafie, L. A.; Dukor, R. K. SSSCPreds: Deep neural network-based software for the prediction of conformational variability and application to SARS-CoV-2. *ACS Omega* **2020**, *5*, 30556−30567.
(20) Izumi, H.; Nafie, L. A.; Dukor, R. K. Conformational variability correlation prediction of transmissibility and neutralization escape ability for multiple mutation SARS-CoV-2 strains using SSSCPreds. *ACS Omega* **2021**, *6*, 19323−19329.
(21) Izumi, H.; Aoki, H.; Nafie, L. A.; Dukor, R. K. Effect of conformational variability on seasonable thermal stability and cell entry of Omicron variants. *ACS Omega* **2023**, *8*, 7111−7118.
(22) Izumi, H.; Nafie, L. A.; Dukor, R. K. Effect of conformational variability on the drug resistance of *Candida auris* ERG11p and FKS1. *ACS Omega* **2024**, *9*, 19816−19823.
(23) Izumi, H. Conformational variability prediction of influenza virus hemagglutinins with amino acid mutations using supersecondary structure code. *Methods Mol. Biol* **2025**, *2870*, 63−78.
(24) Izumi, H.; Nafie, L. A.; Dukor, R. K. Three-dimensional chemical structure search using the conformational code for organic molecules (CCOM) program. *Chirality* **2016**, *28*, 370−375.
(25) Favre, H. A.; Powell, W. H. *IUPAC Recommendations and Preferred Name 2013. Nomenclature of Organic Chemistry (the 'Blue Book')*; The Royal Society of Chemistry: Cambridge, 2013; pp 1286−1288.
(26) Bekker, G.-J.; Yokochi, M.; Suzuki, H.; Ikegawa, Y.; Iwata, T.; Kudou, T.; Yura, K.; Fujiwara, T.; Kawabata, T.; Kurisu, G. Protein Data Bank Japan: Celebrating our 20th anniversary during a global pandemic as the Asian hub of three dimensional macromolecular structural data. *Protein Sci* **2022**, *31*, 173−186.
(27) Halfon, Y.; Matzov, D.; Eyal, Z.; Bashan, A.; Zimmerman, E.; Kjeldgaard, J.; Ingmer, H.; Yonath, A. Exit tunnel modulation as resistance mechanism of *S. aureus* erythromycin resistant mutant. *Sci. Rep* **2019**, *9* (1), 11460.
(28) Tereshko, V.; Skripkin, E.; Patel, D. J. Encapsulating streptomycin within a small 40-mer RNA. *Chem. Biol* **2003**, *10*, 175−187.
(29) Matyjasik, M. M.; Batey, R. T. Structural basis for 2'-deoxyguanosine recognition by the 2'-dG-II class of riboswitches. *Nucleic Acids Res* **2019**, *47*, 10931−10941.
(30) LeBlanc, R. M.; Kasprzak, W. K.; Longhini, A. P.; Olenginski, L. T.; Abulwerdi, F.; Ginocchio, S.; Shields, B.; Nyman, J.; Svirydava, M.; Vecchio, C. D.; Ivanic, J.; Schneekloth, J. S., Jr; Shapiro, B. A.; Dayie, T. K.; Grice, S. F. J. Structural insights of the conserved "priming loop" of hepatitis B virus pre-genomic RNA. *J. Biomol. Struct. Dyn* **2022**, *40* (20), 9761−9773.

https://doi.org/10.1021/acsomega.4c10681
ACS Omega 2025, 10, 4998−5005