Journal of
Chemical Information
and Modeling

Article

pubs.acs.org/jcim

Tertiary Structure Prediction of RNA–RNA Complexes Using a Secondary Structure and Fragment-Based Method

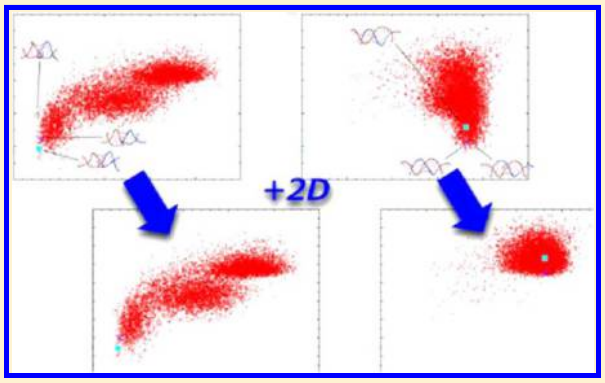
Satoshi Yamasaki,*[†] Takatsugu Hirokawa,[†] Kiyoshi Asai,[‡] and Kazuhiko Fukui[†]

[†]Molecular Profiling Research Center for Drug Discovery (molprof), National Institute of Advanced Industrial Science and Technology (AIST), 2-4-7 Aomi, Koto-ku, Tokyo 135-0064, Japan

[‡]Computational Biology Research Center (CBRC), National Institute of Advanced Industrial Science and Technology (AIST), 2-4-7 Aomi, Koto-ku, Tokyo 135-0064, Japan

S Supporting Information

ABSTRACT: A method has been developed for predicting the tertiary structures of RNA–RNA complex structures using secondary structure information and a fragment assembly algorithm. The linker base pair and secondary structure potential derived from the secondary structure information are particularly useful for prediction. Application of this method to several kinds of RNA–RNA complex structures, including kissing loops, hammerhead ribozymes, and other functional RNAs, produced promising results. Use of the secondary structure potential effectively restrained the conformational search space, leading to successful prediction of kissing loop structures, which mainly consist of common structural elements. The failure to predict more difficult targets had various causes but should be overcome through such measures as tuning the balance of the energy contributions from the Watson–Crick and non- Watson–Crick base pairs, by obtaining knowledge about a wider variety of RNA structures.



1. INTRODUCTION

Recent progress in molecular biology has led to the discovery of many types of small functional RNA molecules that control gene expression, translation, and replication. The RNA-induced silencing complex in eukaryotic organisms, which is a complex of micro-RNA (miRNA) and several types of proteins, suppresses the translation of mRNA targets by slicing them or by binding to them with their complementary sequences. We can artificially induce this mechanism by introducing small interfering RNA (siRNA). This technique, which is known as RNA interference (RNAi), has encouraged research on gene expression and gene regulation. Small nuclear RNAs (snRNAs) and small nucleolar RNAs (snoRNAs) are also functional RNA groups. snRNAs are found in the nuclei of eukaryotic cells. Small nuclear RNPs (snRNPs), which are complexes of snRNA and several proteins, control the RNA splicing function and are involved in important functions such as maintaining telomeres. snoRNAs are localized in the nucleolus of eukaryotic cells and build small nucleolar RNPs (snoRNPs) with several proteins. snoRNPs act as catalysts in RNA modification. Piwi-interacting RNAs (piRNAs) are localized in germ cells and suppress the activities of transposons. Many types of noncoding RNAs (ncRNAs) in eukaryotic cells bind to target RNA together with proteins and play important roles in gene regulation.

The functions of ncRNAs are also found in prokaryotic cells and virus particles. Antisense RNA in *E. coli* cells, which has a sequence complementary to the target RNA region, controls the frequency of plasmid replication.[1] Antisense RNA CopA binds to CopT, which is located in the leader region of RepA, and suppresses the expression of RepA proteins, which initiate plasmid replication.[2,3] In vitro studies revealed that CopA and CopT initially form a kissing-loop complex structure and then form a full duplex structure.[4−6] Several other sense−antisense pairs that regulate plasmid replication are also known and can form kissing-loop complexes.[7]

A similar mechanism for regulating replication has also been found in the human immunodeficiency virus type 1 (HIV-1). The dimerization of genomic RNA is an important process during the life cycle of HIV-1 and other retroviruses.[8] The dimerization initiation sites (DISs) of HIV-1, which have a stem-loop motif and complementary loop sequences, form the kissing-loop complex structure.[9] The formation of DIS-DIS kissing loops is considered to be important for processing group-specific antigen (gag) proteins, so such loops are a promising target for drug design.[10] Kissing-loop complex structures have been extensively studied, because they act as an important initiation process for many types of functions.[11,12] Although the tertiary structures of ncRNAs play important roles in various functions as described above, most of them have not yet been solved.

Received: September 9, 2013
Published: January 19, 2014

dx.doi.org/10.1021/ci400525t | *J. Chem. Inf. Model.* 2014, 54, 672−682

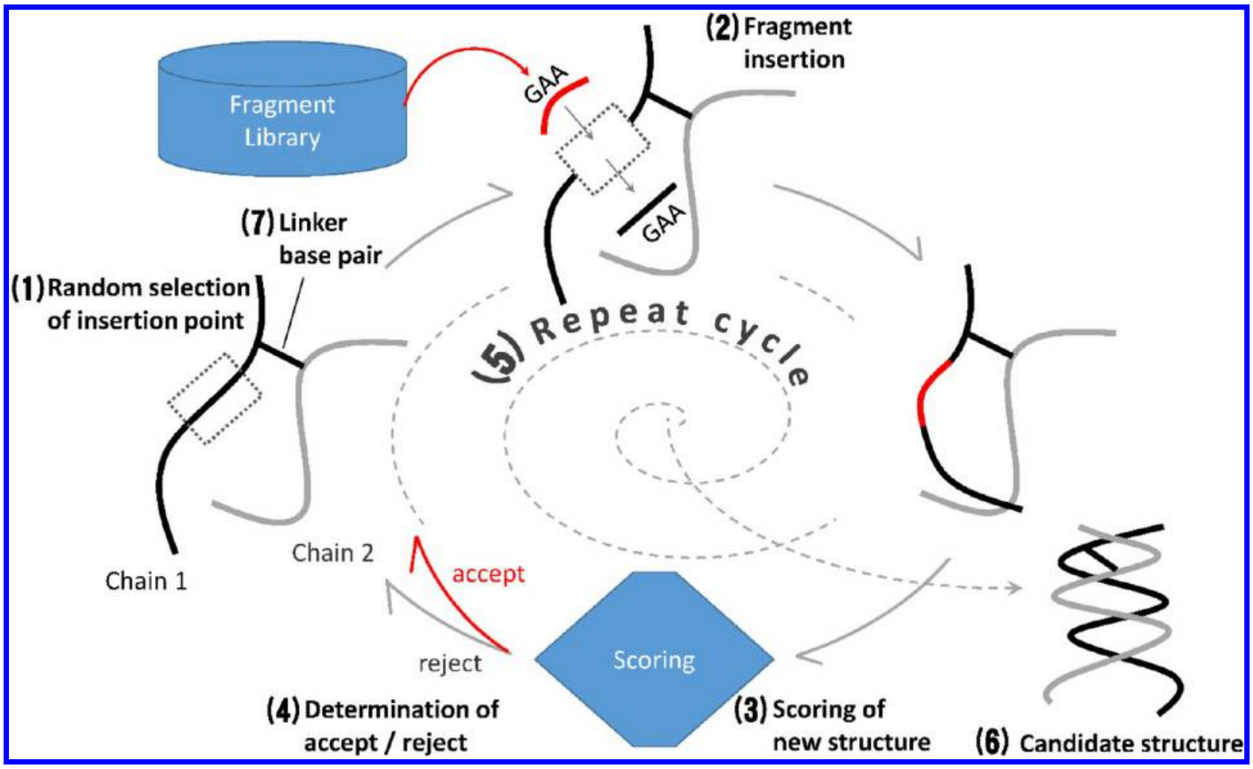


Figure 1. Fragment insertion cycle.

Computational studies on predicting the tertiary structure of RNA molecules have quickly progressed in recent years.[13−19] The tertiary structures of small and single-stranded RNA molecules can now be quickly predicted using various tools, such as FARNA,[13] RNA2D3D,[15] MCsym,[16] etc. The next step is predicting the tertiary structure of more difficult targets, such as RNA−RNA complexes. Simple duplex RNA−RNA complexes can be easily predicted using existing tools cited above. While difficult RNA−RNA complexes can be predicted using interactive modeling tools such as RNA2D3D,[15] prediction using such tools requires several interactive manipulations and a user highly knowledgeable about the RNA tertiary structure. A method that is fully automated and, therefore, is not dependent on the user is required.

Docking methods are well-suited to automatically deriving a bound structure when an unbound structure is available. There are two basic types of docking methods: rigid body docking and flexible docking. In rigid body docking, two unbound structures are docked into a complex structure without any structural change. Targets that require a large structural change before and after docking, such as RNA−RNA targets, are problematic for a rigid body docking method. In the kissing-loop complexes of HIV-1 DISs (PDBID:2F4X,[20] 2D19,[21] and others), for example, the loop nucleotides that form intermolecular interactions flip out toward those of another chain. The unbound state of such loop nucleotides must be a “closed” structure that includes many intramolecular base pairings and intramolecular stacking interactions like those of other monomeric stem loop structures. In fact, the predicted structure of the unbound state of DIS sequences results in such closed structures. Thus, rigid body docking would be unhelpful for RNA−RNA complexes.

Flexible docking, on the other hand, is applicable to targets requiring a large structural change. The structure of the target molecules may partly or wholly change from an unbound state to a bound one during flexible docking. Superior techniques and more computational power are required to perform the required computation. No outstanding flexible docking methods for RNA−RNA complexes are yet available.

The most difficult and critical point in docking is finding intermolecular interaction pairs. In this procedure, a great number of docking poses must be tried to determine the optimal complex structure. In the prediction of RNA−RNA complexes, this procedure can be omitted because the intermolecular pairs can be determined by using tools, such as the RactIP Web server,[22] to predict the secondary structure of RNA−RNA complexes. Therefore, RNA−RNA complexes can be predicted using a nondocking method.

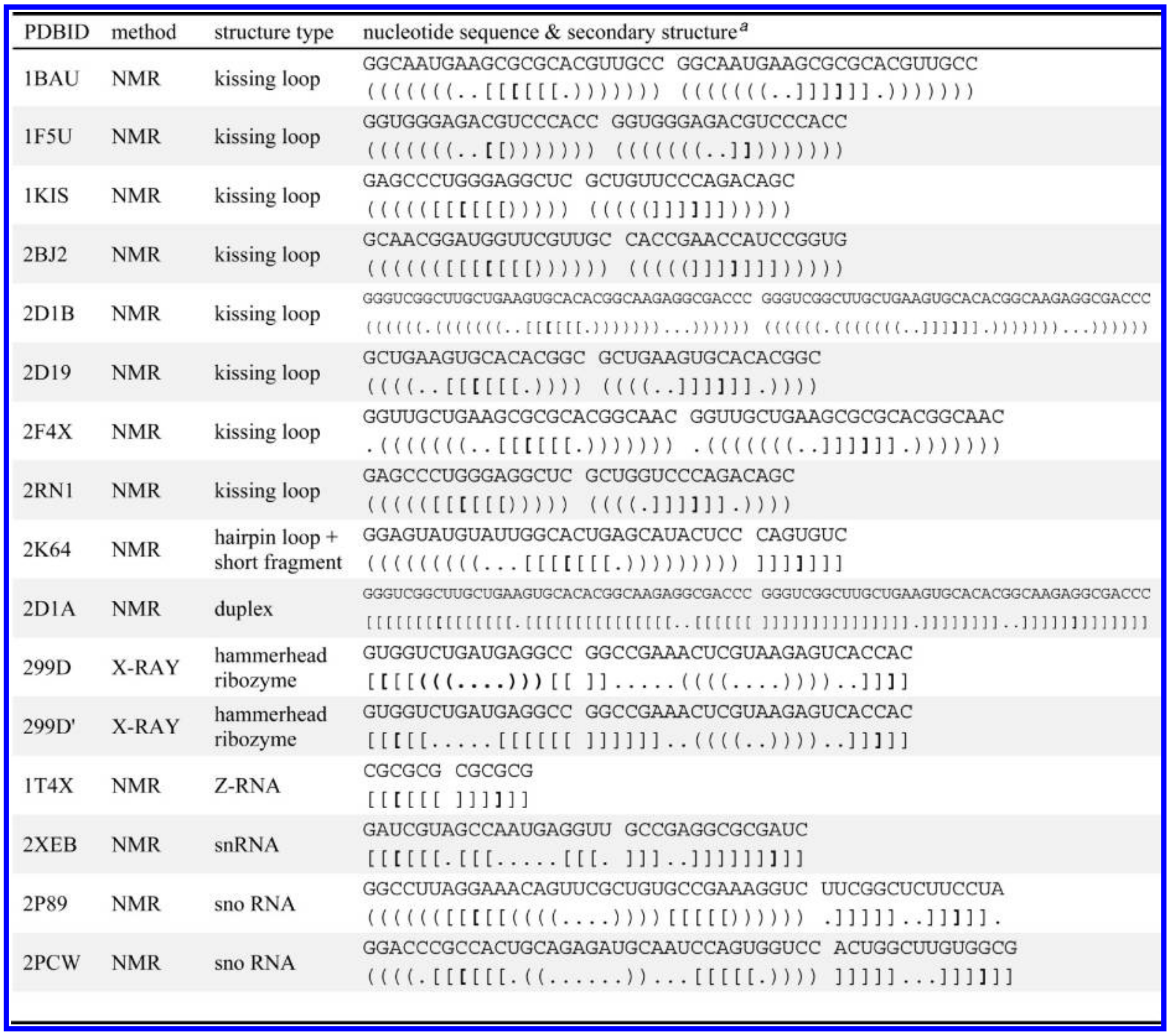
Here, we introduce our new method of fully automated predicting the tertiary structures of RNA−RNA complexes, called “Rascal”, which is based on fragment assembly method and can directly use given information on secondary structures. Fragment assembly method is well-balanced ab initio method, in terms of computational cost and assurance of deriving dataset at present. Using Rascal, we successfully predicted the tertiary structure of many kissing loops and many duplex targets from nucleotide sequences and secondary structures. Prediction of more difficult targets, like hammerhead ribozymes and snoRNAs, was partially successful. These results provide a stepping stone toward predicting complicated RNA−RNA structures.

2. MATERIALS AND METHODS

2.1. General Features of Fragment Assembly Method.

First, we will describe the general features of the fragment assembly method used for predicting the tertiary structures of RNA. A detailed description of the enhancements made to enable prediction of RNA−RNA complex structures is given in section 2.2.

Table 1. Properties of Target Sequences



PDBID	method	structure type	nucleotide sequence & secondary structure[a]
1BAU	NMR	kissing loop	GGCAAUGAAGCGCGCACGUUGCC GGCAAUGAAGCGCGCACGUUGCC (((((((..[[**[**[[[.))))))) (((((((..]]]**]**]].)))))))
1F5U	NMR	kissing loop	GGUGGGAGACGUCCCACC GGUGGGAGACGUCCCACC (((((((..**[**[))))))) (((((((..]**]**)))))))
1KIS	NMR	kissing loop	GAGCCCUGGGAGGCUC GCUGUUCCCAGACAGC ((((([[**[**[[[))))) (((((]]]**]**]])))))
2BJ2	NMR	kissing loop	GCAACGGAUGGUUCGUUGC CACCGAACCAUCCGGUG (((((([[[**[**[[[)))))) (((((]]]**]**]]])))))
2D1B	NMR	kissing loop	GGGUCGGCUUGCUGAAGUGCACACGGCAAGAGGCGACCC GGGUCGGCUUGCUGAAGUGCACACGGCAAGAGGCGACCC ((((((.(((((((..[[**[**[[[.)))))))...)))))) ((((((.(((((((..]]]**]**]].)))))))...))))))
2D19	NMR	kissing loop	GCUGAAGUGCACACGGC GCUGAAGUGCACACGGC ((((..[[**[**[[[.)))) ((((..]]]**]**]].))))
2F4X	NMR	kissing loop	GGUUGCUGAAGCGCGCACGGCAAC GGUUGCUGAAGCGCGCACGGCAAC .(((((((..[[**[**[[[.))))))) .(((((((..]]]**]**]].)))))))
2RN1	NMR	kissing loop	GAGCCCUGGGAGGCUC GCUGGUCCCAGACAGC ((((([[**[**[[[))))) ((((.]]]**]**]].))))
2K64	NMR	hairpin loop + short fragment	GGAGUAUGUAUUGGCACUGAGCAUACUCC CAGUGUC (((((((((...[[[**[**[[[.)))))))))]]]**]**]]]
2D1A	NMR	duplex	GGGUCGGCUUGCUGAAGUGCACACGGCAAGAGGCGACCC GGGUCGGCUUGCUGAAGUGCACACGGCAAGAGGCGACCC [[[[[[[**[**[[[[[[[.[[[[[[[[[[[[[[[..[[[[[[ ]]]]]]]]]]]]]]].]]]]]]]]..]]]]]**]**]]]]]]]
299D	X-RAY	hammerhead ribozyme	GUGGUCUGAUGAGGCC GGCCGAAACUCGUAAGAGUCACCAC [**[**[[**(((....)))**[[ ]].....((((....))))..]]**]**]
299D'	X-RAY	hammerhead ribozyme	GUGGUCUGAUGAGGCC GGCCGAAACUCGUAAGAGUCACCAC [[**[**[[.....[[[[[[ ]]]]]]..((((..))))..]]]**]**]
1T4X	NMR	Z-RNA	CGCGCG CGCGCG [[**[**[[[ ]]]**]**]]
2XEB	NMR	snRNA	GAUCGUAGCCAAUGAGGUU GCCGAGGCGCGAUC [[**[**[[[.[[[.....[[[.]]]..]]]]]]**]**]]
2P89	NMR	sno RNA	GGCCUUAGGAAACAGUUCGCUGUGCCGAAAGGUC UUCGGCUCUUCCUA (((((([[**[**[[((((....))))[[[[[)))))) .]]]]]..]]**]**]].
2PCW	NMR	sno RNA	GGACCCGCCACUGCAGAGAUGCAAUCCAGUGGUCC ACUGGCUUGUGGCG ((((.[[**[**[[[.((......))...[[[[[.))))]]]]]...]]]**]**]]

[a] Linker base pairs shown by bold square brackets.

The first report on predicting the tertiary structure of RNA using fragment assembly method was by Das and Baker,[13] which we followed in developing our Rascal method. While several potential energy functions were derived from the RASSIE method,[19] most of the source code for Rascal was written from scratch.

2.1.1. Fragment Insertion Cycle. The fragment insertion cycle (Figure 1) is an essential part of the fragment assembly method. It starts from an extended structure. In our work, the first 1000 insertion steps were used for randomizing the starting structure, and all 1000 insertions were accepted. At each insertion step, a trinucleotide insertion point was randomly selected from the target sequence (1), and the structure of that region was replaced by one of the fragment structures of the same trinucleotide sequence in the fragment library (2), and a new structure was derived. After insertion, the score calculated for the new structure using the potential energy function was compared with those of previous structures. The new structure was determined to be accepted or rejected according to the Metropolis Monte Carlo criterion (3) and (4). Repetition of this step for 50 000–100 000 (monomeric structure) or 100 000–200 000 (complex structure) cycles (5) resulted in the derivation of one candidate structure (6). Eight thousand (8000) runs of these insertion cycles were performed, and 8000 candidate structures for each target sequence were derived. The 8000 candidate structures were clustered with the root-mean-square deviation (RMSD) threshold chosen such that 20 structures were in the largest cluster. This clustering was performed based on the method used in Simons et al.[23] We used the center of the largest cluster as the predicted structure.

The fragment library and potential energy functions for scoring were necessary to compute this cycle. An RNA dataset was derived and used to build or estimate them. The details are described in the following subsection.

2.1.2. RNA Dataset. First, a list of 1153 nonredundant RNA sets was derived from the RNA 3D Hub website (http://rna.bgsu.edu/rna3dhub/nrlist/, release id 0.9-all).[24] This set included many types of RNAs, RNA–ligand complexes, RNA–RNA complexes, and protein–RNA complexes, which are immersed in varieties of solvent environments. Only the first models were used for the NMR models. The nucleotides, which included missing atoms and modified functional groups, were ignored and treated as missing residues. The chains, which included such noncanonical nucleotides, were divided into subchains, and chains less than five nucleotides in length were removed. Finally, a dataset of 1913 subchains without any nicks or noncanonical nucleotides was derived.

2.1.3. Fragment Library. The fragment library was derived from the RNA dataset described above. Fragments of three nucleotides in length were derived from the chains in the dataset by shifting the starting nucleotides one by one. Nucleotides at the 5′ and 3′ ends of each chain were ignored. After illegal fragments were removed, 80 787 fragment structures of three nucleotides in length were derived. These fragment structures were classified into 64 types of trinucleotide sequences. The trinucleotide sequences and internal coordinates (bonds, angles, and torsions) of the heavy atoms for all fragment structures were stored in the fragment library.

Since structured RNA consists mainly of helical stem structures, most of the fragment structures that we derived should have been helix-like structures. To confirm this, a reference structure for the helical structure, which was derived from an A-RNA model generated by nucleic acid builder (NAB) program of AmberTools,[25] was prepared, and RMSDs between the reference structure and fragment structures in our fragment library were calculated. These RMSDs were <1.0 Å for ~50% of our fragment structures. This bias reflects the nature of native structured RNAs, which mainly consist of helical stems. These helix-like fragments were extracted from the fragment library to build a "helix-like library." The fragment structures in the helix-like library were used to insert fragments into the base-paired regions when the secondary structure was given as input. Both helix-like and non-helix-like fragments were used to insert fragments into the non-base-paired regions.

2.1.4. Potential Energy Functions. The scores used to evaluate the structures were the weighted sum of the results for three potential energy functions: base-pairing potential, base-stacking potential, and steric potential. The weight factor was fitted for several small stem-loop structures. (w_{bp}, w_{coplanar}, w_{stacking}, w_{steric}) was (0.091, 0.60, 0.20, 0.070). Such statistical potential energy functions have worked well for functional analyses and structure predictions of nucleic acids.[13,26]

2.1.4.1. Base-Pairing Potential. The base-pairing potential was based on the spatial distribution of other bases around a nucleotide. A similar function was used in a study that evaluated the direct interaction energy between DNA and binding proteins.[26] The definition of a coordinate system for calculating this spatial distribution was as follows. The origin of this coordinate system was the geometric center of the heavy atoms of the base. The x-axis was parallel to the N9−C4 (for purines) or N1−C2 (for pyrimidines) vectors and started from the origin. The y-axis was perpendicular to the x-axis and antiparallel to the vertical line from N7 (for purines) or C5 (for pyrimidines) to the x-axis. The z-axis, in the right-hand reference frame, was the vector normal to the XY-plane. An 18 Å × 18 Å × 3 Å rectangular box around the origin was defined, and that box was divided into grid cells with a grid interval of 2 Å × 2 Å × 1 Å. When the base origin of another base was within the grid cells, the energy, which is defined as

$$\text{energy} = -\,kT\ \ln P$$

where P is the probability distribution of A, C, G, and U for that grid cell), was given. The probability distribution for each grid cell was calculated from the structures in our nonredundant dataset. Around the ith a nucleotide, the jth b nucleotide in grid cell g gave the base-pairing potential (a and b is one of A, C, G, or U):

$$E_{\text{bp}}(i, j) = -kT\ \ln P_{ab}(g) \tag{1}$$

Furthermore, the base coplanarity potential was based on the planarity of the base pairs. The coplanarity potential of the jth nucleotide, which existed in the rectangular box around the ith nucleotide, was given as

$$\begin{aligned} E_{\text{coplanar}}(i, j) &= E_{\text{angle}}(i, j) + E_{\Delta z}(i, j) \\ E_{\text{angle}}(i, j) &= -kT\ \cos \theta_{i,j} \\ E_{\Delta z}(i, j) &= -(1.0 - \Delta z_{i,j})kT \end{aligned} \tag{2}$$

The potential energy contributed from the base pairs was derived by summing up the contribution from all observed base pairs:

$$E_{\text{bp\_all}} = w_{\text{bp}} \sum_{\substack{\text{observed}\\ \text{pairs } (i,j)}} E_{\text{bp}}(i, j) + w_{\text{coplanar}} \sum_{\substack{\text{observed}\\ \text{pairs } (i,j)}} E_{\text{coplanar}}(i, j) \tag{3}$$

2.1.4.2. Base-Stacking Potential. The base-stacking potential was based on the relative conformation between two base origins ($E_{\text{stack\_position}}$) and the angle between two base planes ($E_{\text{stack\_norm}}$). The coordinate system used to calculate this potential was the same as that used for the base-pairing potential. The base-stacking potential was given when the origin of the jth base was in the range of $\{(x_i - x_j)^2 + (y_i - y_j)^2\}^{1/2} < 5$ and $2.5 < |z_i - z_j| < 4.16$. $E_{\text{stack\_position}}$ was given as $-kT\ \ln\ P\Delta z$, and $P\Delta z$, which is defined as a normal distribution, was derived by approximating the distribution of Δz between stacked nucleotides in the nonredundant dataset. If the angle between the ith and jth base was <30°, $-kT$ was given as $E_{\text{stack\_norm}}$.

$$\begin{aligned} E_{\text{stacking}} &= w_{\text{stacking}}(E_{\text{stack\_position}} + E_{\text{stack\_norm}}) \\ E_{\text{stack\_position}} &= \sum_{\substack{\text{observed}\\ \text{pairs } (a,b)}} \begin{cases} -kT\ \ln P_{\Delta z} & \text{if } \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} < 5, \\ & \quad 2.5 < |z_1 - z_2| < 4.16 \\ 0 & \text{otherwise} \end{cases} \\ P_{\Delta z} &= 0.120589 \exp\left[-\frac{(\Delta z - 3.32968)^2}{2 \times (0.270366)^2}\right] \\ E_{\text{stack\_norm}} &= \sum_{\substack{\text{observed}\\ \text{pairs } (a,b)}} \begin{cases} -kT & \text{if } \cos(\theta_{\vec{z}_1\vec{z}_2}) \geq \cos(30°) \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{4}$$

2.1.4.3. Steric Potential. The steric potential function penalized steric clashes between representative atoms on each nucleotide. Five sugar−phosphate backbone atoms (C1′, C2′, C3′, C5′, and P) and four base atoms (N6, N7, N9, and C2 for adenine; O6, N7, N9, and N2 for guanine; N4, O2, N1, and C5 for cytosine; and O4, O2, N1, and C5 for uracil) were selected as representative atoms. The values of the "steric radii" of these representative atoms were inferred from the third-shortest distance between two representative atoms observed in the nonredundant dataset. If the distance between two representative atoms with a candidate structure was less than their steric radii, an energy penalty of kT was imposed. The idea behind this potential energy function can also be found in the literature on FARNA.[13]

$$E_{\text{steric}} = w_{\text{steric}} \times \sum_{\substack{\text{crashed}\\ \text{atoms}}} kT \tag{5}$$

Table 2. Summary of Results

		Predicted Structure Using Rascal							
				Intra Base Pairs[b]		Inter Base Pairs[b]			
PDB ID	threshold	RMSD	iRMSD[a]	WC	nonWC	WC	nonWC	score[c]	2D score[d]
1BAU	3.38	6.90	2.41	8/10	0/4	4/4	0/1	−211.33	−99.68
1F5U	2.42	3.08	1.05	7/14	0/0	2/2	0/0	−146.86	−67.21
1KIS	2.75	4.05	2.57	5/8	0/2	1/1	0/5	−156.33	−73.06
2BJ2	2.88	4.32	2.13	3/11	0/0	7/7	0/0	−168.73	−78.91
2D1B	8.55	4.92	2.50	12/22	2/8	6/6	0/0	−346.18	−161.36
2D19	2.37	2.95	1.05	5/6	1/3	6/6	0/2	−160.27	−75.59
2F4X	3.51	5.31	2.54	10/12	0/2	6/6	0/0	−195.11	−92.11
2RN1	2.66	3.15	2.45	4/9	0/1	6/6	0/0	−141.39	−68.25
2K64	1.93	6.26	1.97	7/8	2/2	6/6	1/2	−203.23	−96.57
2D1A	8.52	8.73		0/0	0/0	19/28	3/9	−309.58	−142.90
299D	4.83	14.74	12.02	0/3	0/1	4/9	0/5	−114.77	−51.81
299D′	4.83	12.41	9.43	3/3	0/1	8/9	1/5	−160.30	−70.82
1T4X	0.48	8.79		0/0	0/0	6/6	0/0	−74.50	−35.04
2XEB	5.65	11.30		0/0	0/0	9/10	1/3	−149.44	−69.60
2P89	5.10	22.42		3/8	1/2	6/12	0/0	−128.17	−55.12
2PCW	2.33	21.10		0/6	0/4	7/7	0/2	−158.28	−72.27

[a]RMSD of interface nucleotides. [b]Number of base pairs found in predicted structure/native structure. [c]Calculated score for predicted structure. [d]Score contributed by secondary structure potential.

2.2. Special Treatment for RNA–RNA Complexes in the Fragment Assembly Method.

2.2.1. Introduction of "Linker Base Pairs". As described in the Introduction, intermolecular base pairs can be predicted by using secondary structure prediction tools in RNA–RNA problems. We provide a new restraint on the relative conformation of two chains through one of the predicted intermolecular base pairs, which we called a "linker base pair" (labeled "(7)" in Figure 1). For systems that include one intermolecular region, such as kissing loops and most simple targets, the central base pair of the intermolecular region is defined as a linker base pair. For systems that include several intermolecular regions such as hammerhead ribozymes, the central base pair of the first intermolecular region of the first chain is defined as a linker base pair. The relative conformation of the base atoms of linker base pairs was fixed to a given base–base conformation. The base–base conformation extracted from an A-RNA structure generated using the NAB program of AmberTools[25] was used as the reference relative conformation for the predictions discussed in this paper.

The conformation of atoms out of glycoside bonds changed, depending on the inserted fragments. Fragments were inserted alternately into the first chain (even steps) and the second chain (odd steps). Potential energy calculations and determination of accept or reject in the Metropolis Monte Carlo cycles were performed at every step. This cooperative fragment assembly method worked well for simple targets.

2.2.2. Secondary Structure Potential. Since an RNA–RNA complex structure is large and complicated, placing a restraint that restricts the conformational search space would be useful. Information on secondary structures is one of the best sources of such a restraint. A potential energy function that restrains its secondary structure can be used when information on the secondary structure is considered to be important. The secondary structure is basically a base-pairing pattern. Therefore, the potential energy function of a secondary structure can be defined using a base-paring potential energy function. The secondary structure of n-length RNA can be described using an $n \times n$ symmetric matrix, **M**. Most simply, if bases i and j make a base pair, then $M_{ij} = 1$; otherwise $M_{ij} = 0$. Using M_{ij}, the base pairing, and the base coplanarity potential (eq 3), we can express the secondary structure potential as

$$E_{\text{ss}} = w_{\text{bp\_ss}} \sum_{\substack{\text{observed}\\ \text{pairs } (i,j)}} M_{ij}\{E_{\text{bp}}(i,j)\} + w_{\text{coplanar\_ss}} \sum_{\substack{\text{observed}\\ \text{pairs } (i,j)}} M_{ij}\{E_{\text{coplanar}}(i,j)\} \quad (6)$$

In brief, the base pairing potential is doubled for the correct base pairs. Furthermore, M_{ij} should be −1 if bases i and j should not create a base pair. This approach can be effective for RNA–RNA complex structures. We set $M_{ij} = -1$ for bases i and j if they belonged to different chains and should not create a base pair. For example, elements in matrix **M** for a typical kissing loop are 1 (i and j are base-paired), −1 (i and j are not on the same chain and are not in the same base paired region), or 0 (other). In addition, the elements in matrix **M** can be a real number. If the probability of making base pairs has been derived, that probability can be reflected in matrix **M**. The values of ($w_{\text{bp\_ss}}$, $w_{\text{coplanar\_ss}}$) were the same as those of (w_{bp}, w_{coplanar}).

3. Target Selection.

The Nucleic Acid Database (NDB)[27] was used to select the test targets because we wanted to select RNA–RNA complex structures without any modified bases. We derived 464 entries for the RNA structure from the NDB by using search option "Structural Content; RNA=Y and others=N" and "Nucleic Acid Modifications; Base=N and Sugar=N and Phosphate=N." From these entries, 94 entries which contained two chains and less than 100 bases were selected. Helical duplex structures (72 of 94 entries) were ignored, except for one target (PDB ID: 2D1A[21]) because helical duplex structures are easy to predict. [Note: PDB stands for Protein Database.] Quadruplex structures (3 of 94) and helical-packing structures (4 of 94) were also excluded, because of the difficulty of predicting their secondary structures. One of the kissing loop structures (1BJ2) was excluded because the nucleotide sequence and secondary structure were the same as

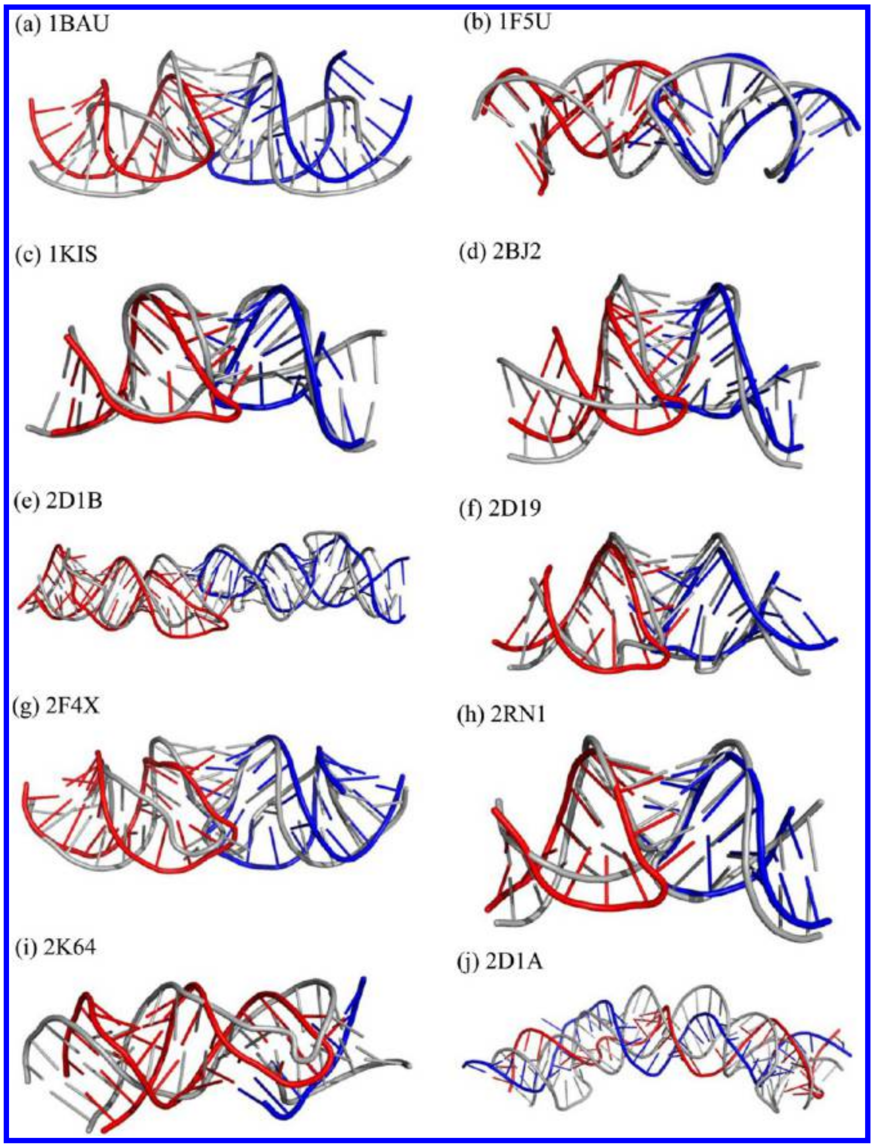


Figure 2. Predicted structures for (a−h) eight kissing-loop targets, (i) one hairpin loop − short fragment complex, and (j) one duplex structure. (Colored elements represent the predicted structures: red, chain A; blue, chain B; and gray, native structure.)

those of another (2BJ2).[28] The remaining 15 entries, comprising 8 kissing loops,[11,12,20,21,28−30] 1 hairpin loop−short RNA fragment complex, 1 duplex structure,[21] 1 hammerhead ribozyme,[31] 1 Z-RNA,[32] 1 snRNA,[33] and 2 snoRNAs,[34,35] were used as test targets (see Table 1).

The secondary structures of the 15 targets were predicted using the RactIP server. Since the predicted secondary structure of 299D was incorrect, we tried another prediction for 299D, 299D′, with a more native-like secondary structure. The predicted secondary structure of 2D1A was that of a kissing loop, because its sequence was the same as that of 2D1B, and then a native secondary structure (duplex) was used to predict the tertiary structure of 2D1A. The predicted secondary structure of 2D19 was duplex because it can form both duplex (2D18) and kissing-loop structures (2D19),[21] and then a native secondary structure (kissing loop) was used to predict the tertiary structure of 2D19. The predicted secondary structures of the other 12 targets were the same as the native secondary structures. The fragment library and potential energy functions were recalculated from our dataset without using the redundant entry for each target.

The predicted structures were evaluated using three criteria: RMSD, interface RMSD (iRMSD), and base pairs. The criteria for the predicted structures were compared with those for the native structures. RMSD was calculated using several representative atoms of all nucleotides, P, C3′, C4′, C5′, O3′, O5′, O4′, C1′, C2′, N1, N4, and C5. iRMSD was calculated using the same representative atoms of interface nucleotides, which are shown by the square brackets in Table 1. The base pairs were annotated with the RNAVIEW program.[36] Incomplete base pairs that had only single hydrogen bonds were excluded from the annotated base pairs. The base pairs were classified in accordance with two attributes: Watson−Crick (WC) or non-Watson−Crick (nonWC) base pair, and intramolecular or intermolecular base pair.

3. RESULTS AND DISCUSSION

3.1. Summary of Predictions. We predicted the tertiary structures of the 15 targeted RNA−RNA complexes. The secondary structures of all targets were predicted using the RactIP server. The predicted secondary structure of 299D included the non-native stem loop in chain A, so another

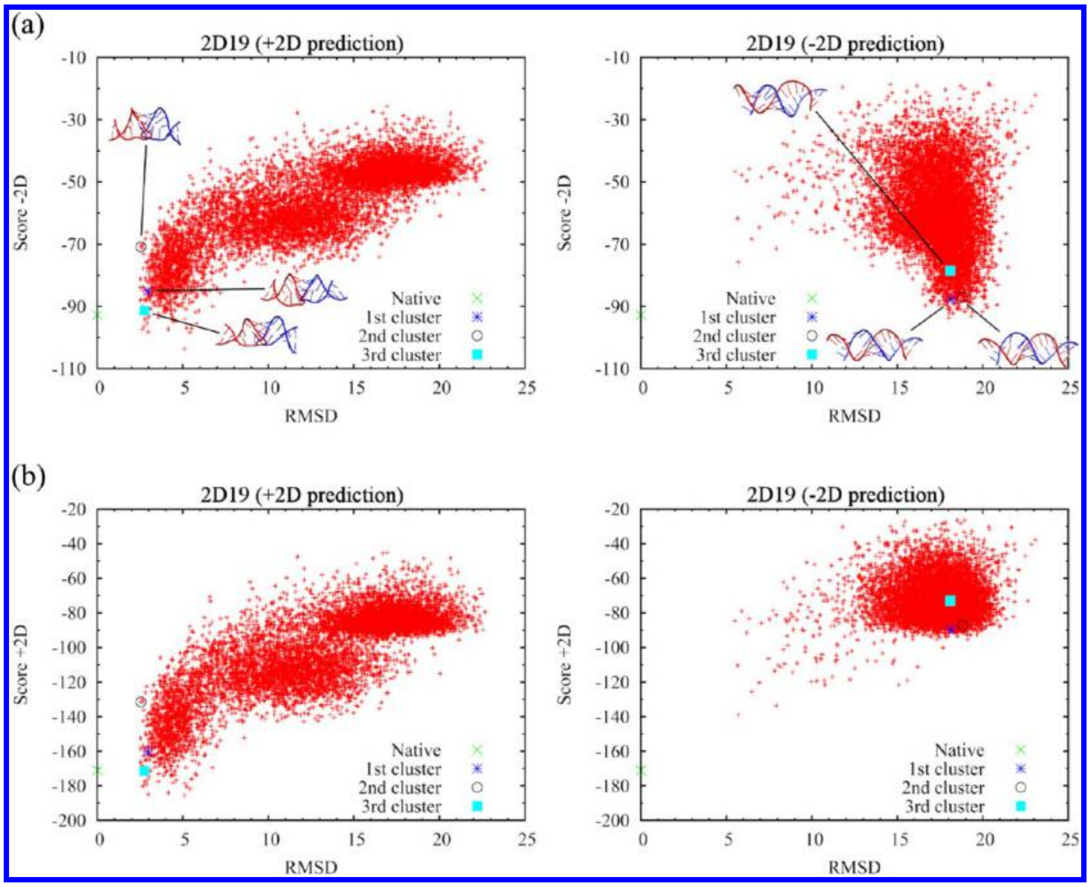


Figure 3. Scatter plots of RMSD between candidate structures and native structure (x-axis shown in Å units) versus score of candidate structures calculated by our scoring function (y-axis in kT units) for 2D19 +2D prediction (left) and −2D prediction (right): (a) score without secondary structure potential (Score −2D) and (b) score with secondary structure potential (Score +2D).

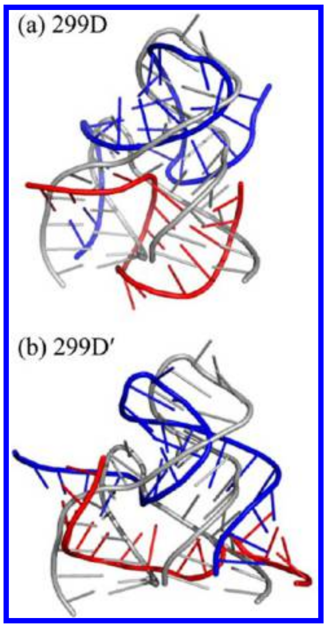


Figure 4. Predicted structures for (a) 299D and (b) 299D′. (Colored features represent predicted structures: red, chain A; blue, chain B; and gray, native structure.)

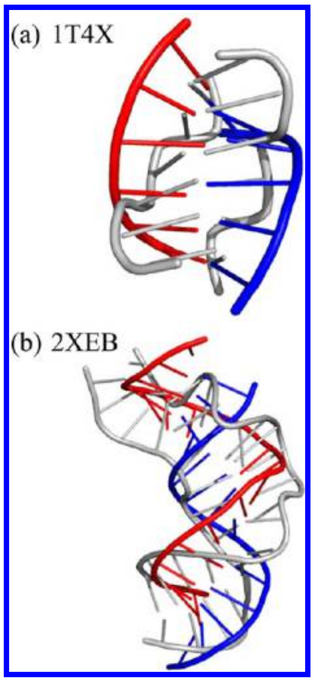


Figure 5. Predicted structures for (a) 1T4X and (b) 2XEB .(Colored features represent predicted structures: red, chain A; blue, chain B; and gray, native structure.)

prediction was added using a more native-like secondary structure (299D′). The predicted secondary structures of 2D1A and 2D19 differed from those of the native structure because these molecules can form both duplex (2D1A, 2D18) and kissing loop structures (2D1B, 2D19).[21] Therefore, native secondary structures were used for these targets.

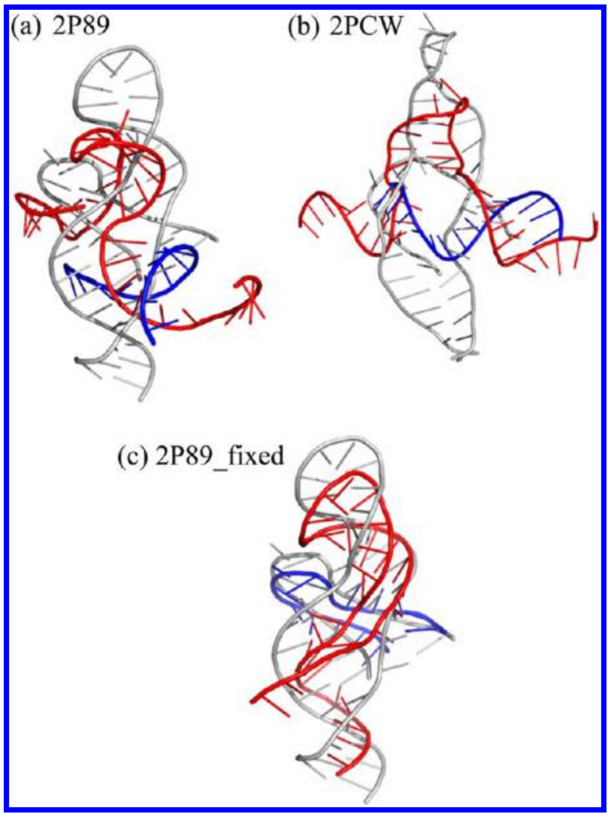


Figure 6. Predicted structures for (a) 2P89, (b) 2PCW, and (c) 2P89_fixed. (Colored features represent predicted structures: red, chain A; blue, chain B; and gray, native structure.)

The properties of the target sequences are summarized in Table 1. All predictions were made using the nucleotide sequences and the secondary structures. For each target, we performed 200 000 insertion cycles × 8000 runs and then derived 8000 candidate structures. These candidate structures were clustered on the basis of pairwise RMSD. The clustering thresholds in Table 2 were determined so that the size of the largest cluster was not more than 20. The center of the first cluster was used as the predicted structure. Table 2 summarizes the heavy atom RMSDs and the iRMSDs between the predicted structure and the native structure, the rate at which the native base pairs were reproduced, and the scores for the predicted structures.

We derived successful results for the values of RMSDs, especially for the eight kissing-oop targets and the 2K64 target. For these nine targets, the heavy-atom RMSDs were ~3−6 Å, and the iRMSDs were <3 Å. These predicted structures are shown in Figure 2. Therefore, our method was effective, especially for predicting kissing-loop targets.

3.2. Advantages of Using Secondary Structure Information. To determine whether the prediction of these targets without the secondary structure information would be successful, we performed control experiments to determine the effect of using secondary structure information. We used the same nine targets for which prediction was successful. We call these predictions "−2D predictions", in contrast to predictions with secondary structure information, which we call "+2D predictions." We must select linker base pairs without information on secondary structure information to make the −2D predictions. We defined two rules for selecting them: (1) the linker base pair must be the central base of the chain, and (2) the linker base pair must be a WC base pair. We were able to select correct linker base pairs for six of the nine targets (1F5U, 1KIS, 2BJ2, 2D1B, 2D19, and 2RN1) with these rules. Since we could not select correct linker base pairs with these rules for the other three targets, we tried to use the correct linker base pairs for these three targets. The parameters used for the −2D predictions were the same as those used for the +2D prediction, except for using the secondary structure potential energy function.

The −2D predictions were successful for only one target, 2K64, which is a native-like complex structure (see Figure S1(h) in the Supporting Information) containing a hairpin loop and a short RNA fragment. It is easy to predict the tertiary structures of short single-chain hairpin loops, so predicting 2K64 was as easy as predicting single hairpin loops. Therefore, successful −2D prediction of 2K64 was a matter of course. The other −2D predictions resulted in failure, and native-like candidate structures have not been derived. Most of the candidate structures derived by −2D prediction were duplex, because the targets contained two highly complementary chains. These results demonstrate the necessity of using the secondary structure potential for prediction.

Figure 3a shows scatter plots for the RMSDs between the candidate structures and the native structures (*x*-axis) versus the scores for the candidate structures (*y*-axis) for the 2D19 +2D prediction (left) and the 2D19 −2D prediction (right). The scores used in these plots were recalculated using 8000 candidate structures for each +2D prediction and −2D prediction without the secondary structure potential ("Score −2D"). The plots of "Score −2D" for the other eight targets are shown in Figure S1 in the Supporting Information.

These plots indicate that the "Score −2D" values for the centers of the first−third clusters for the +2D predictions were nearly equal to those for the −2D predictions. However, most structures in the −2D prediction results were deformed structure or duplex structure. Candidates with a kissing-loop structure rarely appeared. This suggests that the energy barriers between the deformed and kissing-loop structures were larger than those between the deformed and duplex structures.

Figure 3b shows scatter plots of the RMSDs versus the scores with the secondary structure potential for a native kissing-loop structure (Score +2D) for the +2D prediction (left) and the −2D prediction (right). The scores for the duplex structures increased greatly when the secondary structure potential was used (righthand side of Figure 3b) while those for kissing-loop structures decreased greatly (lefthand side of Figure 3b). Using the secondary structure potential energy efficiently eliminated the possibility of duplex structure folding and increased the possibility of kissing-loop structure folding. The use of the secondary structure potential energy in +2D prediction modifies the energy surface, which enhances the sampling of kissing-loop structures and prevents duplex structures from being sampled.

3.3. Other Difficult Targets. We tried to predict more difficult targets containing complicated secondary structures. The hammerhead ribozyme (299D) is one of the most significant targets, but the secondary structure predicted by RactIP differed from the native secondary structure. The nucleotides in the stem loop region of chain A in the predicted secondary structure (bold features shown in Table 1) are involved in the intermolecular non-WC base pairs in the native structure.

Table 3. RMSD and Number of Near-Native Fragments in the Fragment Library

2D19					2P89				
start[a]	sequence[b]	lowest RMSD[c]	<0.5[d]	<1.0[d]	start[a]	sequence[b]	lowest RMSD[c]	<0.5[d]	<1.0[d]
A1	GCU	0.120	625	1081	A1	GGC	0.453	4	728
A2	CUG	0.108	489	1082	A2	GCC	0.315	279	1244
A3	UGA	0.268	1	13	A3	CCU	0.302	15	778
A4	GAA	0.361	1	1	A4	CUU	0.928	0	1
A5	AAG	0.362	1	1	A5	UUA	0.978	0	1
A6	AGU	0.358	1	12	A6	UAG	0.415	2	343
A7	GUG	0.204	181	1023	A7	AGG	0.439	7	798
A8	UGC	0.131	487	759	A8	GGA	0.301	155	1050
A9	GCA	0.181	361	761	A9	GAA	0.283	133	620
A10	CAC	0.192	323	638	A10	AAA	0.297	145	908
A11	ACA	0.136	246	493	A11	AAC	0.307	63	581
A12	CAC	0.131	377	636	A12	ACA	0.307	61	439
A13	ACG	0.334	43	488	A13	CAG	0.233	418	839
A14	CGG	0.205	550	1325	A14	AGU	0.283	99	607
A15	GGC	0.127	914	1411	A15	GUU	0.620	0	42
B1	GCU	0.139	575	1073	A16	UUC	0.646	0	56
B2	CUG	0.129	488	1079	A17	UCG	0.583	0	45
B3	UGA	0.133	1	25	A18	CGC	0.780	0	3
B4	GAA	0.158	1	6	A19	GCU	0.127	3	895
B5	AAG	0.301	1	2	A20	CUG	0.205	200	1042
B6	AGU	0.229	1	2	A21	UGU	0.292	20	483
B7	GUG	0.179	239	1086	A22	GUG	0.655	0	6
B8	UGC	0.195	392	753	A23	UGC	0.898	0	3
B9	GCA	0.255	224	750	A24	GCC	0.264	143	1228
B10	CAC	0.191	283	633	A25	CCG	0.216	828	1420
B11	ACA	0.249	131	475	A26	CGA	0.273	173	562
B12	CAC	0.213	69	601	A27	GAA	0.381	21	556
B13	ACG	0.222	42	536	A28	AAA	0.330	52	767
B14	CGG	0.142	760	1363	A29	AAG	0.320	143	830
B15	GGC	0.105	916	1408	A30	AGG	0.335	122	1178
					A31	GGU	0.289	276	1278
					A32	GUC	0.320	248	883
					B35	UUC	0.300	51	413
					B36	UCG	0.381	23	592
					B37	CGG	0.349	55	1289
					B38	GGC	0.377	71	1368
					B39	GCU	0.640	0	6
					B40	CUC	1.126	0	0
					B41	UCU	1.148	0	0
					B42	CUU	1.184	0	0
					B43	UUC	0.343	14	387
					B44	UCC	0.304	226	751
					B45	CCU	0.361	11	745
					B46	CUA	0.314	22	437

[a] Chain ID and residue number for the first nucleotide of native trinucleotide fragment. [b] Nucleotide sequence of native trinucleotide fragment. [c] RMSD between a native fragment and the nearest native fragment in the fragment library. [d] Number of near native (RMSD < 0.5 Å/1.0 Å) fragment structure in the fragment library.

First, we tried to predict 299D using this non-native secondary structure. The results are summarized in Table 2 and Figure 4a. The predicted structure included the intermolecular interaction between 25–22 of chain A and 12–15 of chain B and the intramolecular interaction of the non-native stem loop of chain A. (Note that the index numbers for the nucleotides used here follow those used in the Protein Databank (PDB) file for 299D.[31]) A weak intramolecular interaction of the native stem loop of chain B was also found. One of the native intermolecular interaction regions was completely missing in the results. We concluded that conformation, which includes both intermolecular interaction regions and this non-native stem loop, was impossible.

We made another prediction of 299D, 299D′, using more near-native secondary structures (see Table 2 and Figure 4b). Although the value of RMSD was not good, the predicted structure was still a reasonable result, because most of the native WC base pairs were found in this predicted structure. These WC base pairs were in the stem loop region of chain B and in the intermolecular interaction regions of the 5′ and 3′ ends. However, the following non-WC base pairs were not found in the predicted structure: A:A90–B:G120, A:G80–

B:A130, A:U70−B:A140, and A:C30−B:C170. These non-WC base pairs elongated both helix stem regions at the 5′ and 3′ ends in the native structure. The lack of these non-WC base pairs induced large deformation around the nucleotides and resulted in large RMSDs. We tried another prediction, using even more native-like secondary structures,

[[[[[[...[[[[[[[ ]]]]]]].((((..)))).]]]]]]

which included A:A90−B:G120, A:G80−B:A130, A:U70−B:A140, and A:C30−B:C170 intermolecular pairs, but the results obtained were no better. In the end, the critical factor in this target was non-WC base pairs. Evaluation of non-WC base pairs is the key to improving the prediction of RNA structures.

Our method could provide only A-helix-based structures (see Table 2 and Figure 5) for Z-RNA (1T4X) and snRNA, which included a K-turn motif (2XEB). Since these targets were not A-helix structures, we made other predictions without using a "helix-like library," but the results did not improve. Because most known structures have A-helix or A-helix-like forms and Z-RNA types or other types of base pairing structures are rarely found, the fragment library and potential energy functions, which were derived from known structures, were not suitable for finding such unique structures.

Our method failed in the prediction of two snoRNA targets (2P89 and 2PCW) (see Table 2 and Figures 6a and 6b). Such unique tertiary structures, i.e., non-A-helix-based structures, are very difficult to predict with a method based on known structures because more than half the known structures consist of structures that are A-helix type. This can easily be seen by calculating the RMSD between an ideal A-helix model structure and the fragments derived from an RNA dataset. We calculated the RMSD between an A-helix model structure derived using the NAB program of AmberTools and the structures in our fragment library and found that more than half of our fragments consisted of helix-like structures. Of course, this is the nature of common RNA molecules, and they cannot be slightly modified to suit our purposes. For example, we could use a clustering method to modify our fragment library to reduce the population of helix-like fragment structures, but predictions with this modified fragment library would not be successful because the opportunity of going through helix-like structures would be decreased even for natural helix regions. This imbalance in the fragment library could have affected the performance of our method.

What was the main factor determining whether the target structures were successfully predicted? We found that the critical factor was the completeness of the fragment library; i.e., the correct structures of the failed targets were not in the fragment library. Table 3 summarizes the RMSDs and other statistical data for the native and fragment structures in the library for a successful target (2D19) and a failed target (2P89). We were able to find at least one fragment structure in the library for the successful target that was similar to the correct structure for all parts of the target. We were unable to find a fragment structure that was similar to the correct structure for the failed target for several parts of the target (hairpin loop of chain B and hairpin loop and internal loop of chain A). These parts never went through the structure near the correct one during the Monte Carlo simulation. It is no surprise that this target failed to be predicted.

We tried another prediction of 2P89 using additional restraints to fix the nucleotides (3−8 and 14−25 of the A chain and 39−44 of the B chain) to the correct structure and obtained greatly improved results (see Figure 6c, RMSD = 8.28 Å). This demonstrates that the critical factor in prediction failure was incompleteness of the fragment library. Therefore, the fragment library must be expanded to enable accurate prediction of such targets. Our method will likely become more widely used as more unknown structures of RNA molecules are derived and more varieties of RNA structures become known.

4. CONCLUSION

We have developed a method for predicting the tertiary structures of RNA−RNA complex structures from nucleotide sequences and secondary structures. The results demonstrated that our method can predict kissing-loop structures, which mainly consist of common structural elements. Use of the secondary structure potential derived from the given secondary structure information was the key to success—it provided effective restraints on the conformational search space. Comparison of the results for predictions made with and without the secondary structure information demonstrated that the secondary structure potential energy modifies the energy surface, thereby enhancing the sampling of correct structures and restraining the sampling of incorrect structures.

The failure to predict difficult targets had various causes. For targets with many non-WC base pairs such as a hammerhead ribozyme (299D), the failure to recapture the native non-WC base pairs was the main cause of the large deformation of the predicted structure. Tuning the balance of the energy contributions from the WC and non-WC base pairs should improve the results. For targets with a unique structural element such as snoRNA (2P89) and Z-RNA (1T4X), the lack of fragment structures in our fragment library similar to the native structures caused the misprediction. This problem should be solved when more varieties of unknown RNA structures become known. Structural sampling using molecular dynamics simulation should also be helpful. This method serves as a stepping stone in the prediction of RNA tertiary structures.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Scatter plots of RMSD between candidate structures and native structure (x-axis) versus "Score −2D" of candidate structures (y-axis) for +2D prediction (left) and −2D prediction (right) of 8 targets (Figure S1). This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: s.yamasaki@aist.go.jp.

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was partially supported by Platform for Drug Discovery, Informatics, and Structural Life Science from the Ministry of Education, Culture, Sports, Science and Technology, Japan.

■ ABBREVIATIONS

WC, Watson−Crick; nonWC, non-Watson−Crick; RMSD, root-mean-square deviation; iRMSD, interface root-mean-square deviation.

■ REFERENCES

(1) Nordström, K. Plasmid R1—Replication and its control. *Plasmid* **2006**, *55*, 1−26.

(2) Nordström, K.; Molin, S.; Light, J. Control of replication of bacterial plasmids: Genetics, molecular biology, and physiology of the plasmid R1 system. *Plasmid* **1984**, *12*, 71−90.

(3) Blomberg, P.; Wagner, E.; Nordström, K. Control of replication of plasmid R1: The duplex between the antisense RNA, CopA, and its target, CopT, is processed specifically in vivo and in vitro by RNase III. *EMBO J.* **1990**, *9*, 2331−2340.

(4) Persson, C.; Wagner, E. G.; Nordström, K. Control of replication of plasmid R1: Kinetics of in vitro interaction between the antisense RNA, CopA, and its target, CopT. *EMBO J.* **1988**, *7*, 3279−3288.

(5) Persson, C.; Wagner, E. G.; Nordström, K. Control of replication of plasmid R1: Structures and sequences of the antisense RNA, CopA, required for its binding to the target RNA, CopT. *EMBO J.* **1990**, *9*, 3767−3775.

(6) Persson, C.; Wagner, E. G.; Nordström, K. Control of replication of plasmid R1: Formation of an initial transient complex is rate-limiting for antisense RNA−target RNA pairing. *EMBO J.* **1990**, *9*, 3777−3785.

(7) Rist, M.; Marino, J. Association of an RNA kissing complex analyzed using 2-aminopurine fluorescence. *Nucleic Acids Res.* **2001**, *29*, 2401−2408.

(8) Skripkin, E.; Paillart, J. C.; Marquet, R.; Ehresmann, B.; Ehresmann, C. Identification of the primary site of the human immunodeficiency virus type 1 RNA dimerization in vitro. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 4945−4949.

(9) Ennifar, E.; Dumas, P. Polymorphism of bulged-out residues in HIV-1 RNA DIS kissing complex and structure comparison with solution studies. *J. Mol. Biol.* **2006**, *356*, 771−782.

(10) Paillart, J.; Skripkin, E. A loop−loop" kissing" complex is the essential part of the dimer linkage of genomic HIV-1 RNA. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 5572−5577.

(11) Chang, K. Y.; Tinoco, I. The structure of an RNA "kissing" hairpin complex of the HIV TAR hairpin loop and its complement. *J. Mol. Biol.* **1997**, *269*, 52−66.

(12) Mujeeb, A.; Clever, J.; Billeci, T. Structure of the dimer a initiation complex of HIV-1 genomic RNA. *Nat. Struct. Mol. Biol.* **1998**, *5*, 432−436.

(13) Das, R.; Baker, D. Automated de novo prediction of native-like RNA tertiary structures. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 14664−14669.

(14) Frellsen, J.; Moltke, I.; Thiim, M.; Mardia, K. V; Ferkinghoff-Borg, J.; Hamelryck, T. A probabilistic model of RNA conformational space. *PLoS Comput. Biol.* **2009**, *5*, e1000406.

(15) Martinez, H. M.; Maizel, J. V; Shapiro, B. a RNA2D3D: A program for generating, viewing, and comparing 3-dimensional models of RNA. *J. Biomol. Struct. Dyn.* **2008**, *25*, 669−683.

(16) Reinharz, V.; Major, F.; Waldispühl, J. Towards 3D structure prediction of large RNA molecules: An integer programming framework to insert local 3D motifs in RNA secondary structure. *Bioinformatics* **2012**, *28*, i207−i214.

(17) Shapiro, B. a; Yingling, Y. G.; Kasprzak, W.; Bindewald, E. Bridging the gap in RNA structure prediction. *Curr. Opin. Struct. Biol.* **2007**, *17*, 157−165.

(18) Sharma, S.; Ding, F.; Dokholyan, N. V iFoldRNA: Three-dimensional RNA structure prediction and folding. *Bioinformatics* **2008**, *24*, 1951−1952.

(19) Yamasaki, S.; Nakamura, S.; Fukui, K. Prospects for tertiary structure prediction of RNA based on secondary structure information. *J. Chem. Inf. Model.* **2012**, *52*, 557−567.

(20) Kieken, F.; Paquet, F.; Brulé, F.; Paoletti, J.; Lancelot, G. A new NMR solution structure of the SL1 HIV-1Lai loop−loop dimer. *Nucleic Acids Res.* **2006**, *34*, 343−352.

(21) Baba, S.; Takahashi, K.; Noguchi, S.; Takaku, H.; Koyanagi, Y.; Yamamoto, N.; Kawai, G. Solution RNA structures of the HIV-1 dimerization initiation site in the kissing-loop and extended-duplex dimers. *J. Biochem.* **2005**, *138*, 583−592.

(22) Kato, Y.; Sato, K.; Hamada, M.; Watanabe, Y.; Asai, K.; Akutsu, T. RactIP: Fast and accurate prediction of RNA−RNA interaction using integer programming. *Bioinformatics* **2010**, *26*, i460−i466.

(23) Simons, K. T.; Strauss, C.; Baker, D. Prospects for ab initio protein structural genomics. *J. Mol. Biol.* **2001**, *306*, 1191−1199.

(24) Leontis, N. B.; Zirbel, C. L. *Nonredundant 3D Structure Datasets for RNA Knowledge Extraction and Benchmarking*; Leontis, N., Westhof, E., Eds.; Springer: Berlin, Heidelberg, 2012; pp 281−298.

(25) Case, D. A.; Darden, T. A.; Cheatham III, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Götz, A. W.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wolf, R. M.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. *Amber13*; University of California, San Francisco, CA, 2012.

(26) Kono, H.; Sarai, A. Structure-Based Prediction of DNA Target Sites. *Proteins* **1999**, *131*, 114−131.

(27) Berman, H. M.; Olson, W. K.; Beveridge, D. L.; Westbrook, J.; Gelbin, A.; Demeny, T.; Hsieh, S.; Srinivasan, A. R.; Schneider, B. A comprehensive relational database of three-dimensional structures of nucleic acids. *Biophys. J.* **1992**, *63*, 751−759.

(28) Lee, a J.; Crothers, D. M. The solution structure of an RNA loop-loop complex: The ColE1 inverted loop sequence. *Structure* **1998**, *6*, 993−1005.

(29) Kim, C. H.; Tinoco, I. A retroviral RNA kissing complex containing only two G.C. base pairs. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 9396−9401.

(30) Van Melckebeke, H.; Devany, M.; Di Primo, C.; Beaurain, F.; Toulmé, J.-J.; Bryce, D. L.; Boisbouvier, J. Liquid-crystal NMR structure of HIV TAR RNA bound to its SELEX RNA aptamer reveals the origins of the high stability of the complex. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 9210−9215.

(31) Scott, W.; Murray, J.; Arnold, J.; Stoddard, B.; Klug, A. Capturing the structure of a catalytic RNA intermediate: The hammerhead ribozyme. *Science* **1996**, *274*, 2065−2069.

(32) Popenda, M.; Milecki, J.; Adamiak, R. W. High salt solution structure of a left-handed RNA double helix. *Nucleic Acids Res.* **2004**, *32*, 4044−4054.

(33) Falb, M.; Amata, I.; Gabel, F.; Simon, B.; Carlomagno, T. Structure of the K-turn U4 RNA: A combined NMR and SANS study. *Nucleic Acids Res.* **2010**, *38*, 6274−6285.

(34) Wu, H.; Feigon, J. H/ACA small nucleolar RNA pseudouridylation pockets bind substrate RNA to form three-way junctions that position the target U for modification. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 6655−6660.

(35) Jin, H.; Loria, J. P.; Moore, P. B. Solution structure of an rRNA substrate bound to the pseudouridylation pocket of a box H/ACA snoRNA. *Mol. Cell* **2007**, *26*, 205−215.

(36) Yang, H. Tools for the automatic identification and classification of RNA base pairs. *Nucleic Acids Res.* **2003**, *31*, 3450−3460.