Vol.2009-BIO-17 No.13
2009/5/26

Prediction of RNA Secondary Structures with Binding Sites Using Dynamic Programming Algorithm

Unyanee Poolsap,[†1] Yuki Kato[†1]
and Tatsuya Akutsu[†1]

Noncoding antisense RNAs have recently occupied considerable attention and several computational studies have been made on RNA-RNA interaction prediction. In this technical report, we present novel dynamic programming algorithms for predicting the minimum energy secondary structure with binding sites of one of the two interacting RNAs. Experimental results on several known RNA-RNA interaction data show that our proposed method achieves good performance in accuracy.

1. Introduction

In recent years, analysis of noncoding RNAs has attained great importance. They play a crucial role in some biological processes including posttranscriptional regulation of gene expression. Some noncoding RNAs, called *antisense RNAs*, aim at inhibiting their target RNA function through base complementary binding. Some antisense RNAs use full complementarity to their target for binding, whereas a number of antisense RNAs use partial complementarity[4], and several *kissing hairpin* structures (Fig. 1) caused by loop-loop interaction have been reported[5].

To predict joint secondary structures of interacting RNAs (e.g., antisense-target RNA complexes), several dynamic programming (DP) algorithms have been proposed so far. Andronescu et al.[2] developed the PairFold algorithm for secondary structure prediction of two interacting RNAs of minimum free energy. Since this algorithm is based on the Zuker's algorithm[15] for predicting pseudoknot-free structure of a single RNA, its time complexity is $O(n+m)^3$ where n and m are respective lengths of two input sequences. The PairFold algorithm, however, cannot deal with any kissing hairpins, which are essentially equivalent to pseudoknotted structures when concatenating two interacting sequences. On the other hand, DP algorithms presented by Pervouchine[11], Alkan et al.[1] and Kato et al.[7] can predict joint secondary structures including kissing hairpins in $O(n^3m^3)$ time. However, the time complexity of these algorithms is prohibitive in case $n \simeq m$ (i.e., $O(n^6)$), which is the same complexity of a prediction algorithm for pseudoknots[13].

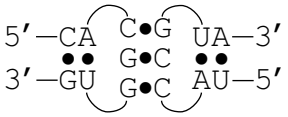

Fig. 1 A kissing hairpin.

Viewing RNA-RNA interaction prediction from a different angle inspires us to consider the situation where we aim at predicting the secondary structure with binding sites of one of the two interacting RNAs on condition that interacting sites of the other RNA are known. In fact, we assume that a "profile" of intermolecular binding is given in advance. This assumption could be helpful in target site prediction for antisense RNAs. In this technical report, we propose novel DP algorithms for predicting RNA secondary structures with binding site information. Notice that our formulation of the prediction problem requires that the order in which binding sites appear for antisense RNA should be the same as the order for its target RNA. To deal with base-paired structure as well as binding sites, we design an extension of the classical Nussinov's algorithm[9], which minimizes the sum of base pair energies. In addition, we develop a DP algorithm that can incorporate stacking energy, which is based on the Zuker's algorithm[15]. Both of our algorithms can run in $O(N^3n^3)$ time where N is the number of binding sites and n is an input length. Since N can be regarded as a constant in most cases, the time complexity of our algorithms can be evaluated as $O(n^3)$. We demonstrate the performance of our approach using the proposed algorithm for base pair energy.

The rest of this technical report is organized as follows. In Section 2, we provide

†1 Bioinformatics Center, Institute for Chemical Research, Kyoto University

formal descriptions of the prediction problem concerned and DP algorithms. We then show some experimental results of the prediction for real interacting RNAs in Section 3. Section 4 concludes the technical report.

2. Methods

In this section, we will present algorithms based on dynamic programming (DP) for predicting RNA secondary structures with binding sites. Given an RNA sequence (target sequence) with unknown structure and a profile of intermolecular binding, our algorithms return the optimum secondary structure and locations of binding sites in order from $5'$ to $3'$. Before going through the details of the algorithms, let us begin with definitions of RNA secondary structure and prediction problem considering binding sites.

2.1 Preliminaries

Definition 1 (RNA secondary structure)**.** An RNA sequence is represented by a string of n characters $s = s_1 s_2 \cdots s_n$ where $s_i \in \Sigma = \{\mathrm{A}, \mathrm{C}, \mathrm{G}, \mathrm{U}\}$. A secondary structure of the sequence s is a set of base pairs (s_i, s_j) such that the following conditions hold:

- $1 \leq i < j \leq n$;
- Each base can be paired with at most one base;
- (s_i, s_j) is a valid base pair, i.e., any of Watson-Crick pairs $\{\mathrm{A}, \mathrm{U}\}$ and $\{\mathrm{C}, \mathrm{G}\}$, and a wobble pair $\{\mathrm{G}, \mathrm{U}\}$;
- $j - i \geq t$, where t is a small positive number.

As stated earlier, our algorithms take a target RNA sequence and a binding site profile as inputs. Therefore, the binding site profile has to be prepared in advance. Informally, the steps of how to create the binding site profile are as follows:

(1) Given an antisense-target RNA complex with known structure, we search for all binding sites of the antisense RNA, written as $\bar{B}_1, \bar{B}_2, \ldots, \bar{B}_N$.

(2) Using the fact that the antisense binds its target via base complementarity, we compute the complementary subsequences to $\bar{B}_1, \bar{B}_2, \ldots, \bar{B}_N$, denoted by $B_1, B_2, \ldots, B_N$. For example, if $\bar{B}_1 = \mathrm{GGACU}$, $B_1 = \mathrm{CCUGA}$. Note that B_p $(1 \leq p \leq N)$ itself is not a subsequence of the target sequence.

(3) For each p $(1 \leq p \leq N)$, if a subsequence of the target that matches B_p can be found at the location that starts from i and ends at j, we assign a finite value to the profile, represented by $I_p(i, j)$, otherwise ∞ is assigned.

Table 1 Example of an energy function e

Base pair	Energy value
G-U	-1
A-U	-2
C-G	-3

Now, let us formally define the binding site profile.

Definition 2 (Binding site profile)**.** Let N be the number of binding sites and $\bar{B}_p = \bar{s}_{i_p} \bar{s}_{i_p+1} \cdots \bar{s}_{j_p} \in \Sigma^*$ $(1 \leq p \leq N)$ denote a binding site (subsequence) of an antisense RNA sequence $\bar{s} = \bar{s}_1 \bar{s}_2 \cdots \bar{s}_m \in \Sigma^*$, where i_p and j_p are fixed and satisfy $1 \leq i_p < j_p \leq m$ for each p. Let $s_i s_{i+1} \cdots s_j$ be a subsequence of a target RNA sequence $s = s_1 s_2 \cdots s_n \in \Sigma^*$. Then, for each p $(1 \leq p \leq N)$, a binding site profile $I_p(i, j)$ of $s_i s_{i+1} \cdots s_j$ is defined as follows:

$$I_p(i,j) = \begin{cases} c \sum_{k=i}^{j} e(s_k, \bar{s}_{k_p}) & (j - i = j_p - i_p, \\ & \quad \text{and } s_k \text{ is complementary to } \bar{s}_{k_p}), \\ \infty & (\text{otherwise}) \end{cases} \tag{1}$$

where c is a positive constant, and e is an energy function that maps from a valid base pair to the corresponding energy value (see Table 1).

With these definitions, we define the prediction problem of RNA secondary structure with binding sites.

Definition 3 (RNA secondary structure prediction with binding sites)**.**

Input: a target RNA sequence $s = s_1 s_2 \cdots s_n \in \Sigma^*$ and N binding site profiles $I_1, I_2, \ldots, I_N$ of s.

Output: the optimum secondary structure of s whose subsequences match the binding sites in the order from I_1 to I_N.

2.2 DP algorithms

We develop two prediction models based on DP. The first DP model is an extension of the Nussinov's algorithm[9] using a simple base pair energy function. For the second model, we extend the first model to utilize the stacking energy and loop energy functions instead of the simple energy function, which is based

on the Zuker's algorithm[15].

2.2.1 Base pair energy model

In the beginning, we define DP tables to predict secondary structure with biding sites. Let $s = s_1 s_2 \cdots s_n$ be an RNA sequence. As in the conventional case, we let $W(i,j)$ denote the minimum free energy of secondary structure formed from a subsequence $s_i s_{i+1} \cdots s_j$ of s. In addition, let $W_{pq}(i,j)$ be the minimum free energy of secondary structure for $s_i s_{i+1} \cdots s_j$ that contains binding sites corresponding to $I_p, I_{p+1}, \ldots, I_q$ $(1 \leq p \leq q \leq N)$.

These DP tables are initialized as follows:

$$W(i,i) = 0,\ W_{pq}(i,i) = \infty \quad (1 \leq \forall i \leq n;\ 1 \leq \forall p \leq \forall q \leq N).$$

The recursions are classified into three cases as shown below: In the first case, we use the simple Nussinov's algorithm to predict secondary structure that does not contain any binding sites. The second case is used for dealing with the structure with just one binding site. The third case is used for predicting the structure with two or more binding sites.

Case 1 (the Nussinov's algorithm):

$$W(i,j) = \min \begin{cases} W(i+1,j), \\ W(i,j-1), \\ W(i+1,j-1) + e(i,j), \\ \min\limits_{i \leq k < j} \{W(i,k) + W(k+1,j)\} \end{cases} \tag{2}$$

where $e(i,j)$ is the simple energy function of a base pair (s_i, s_j) (see Table 1). In the above DP recursion, the first and the second cases of minimization represent the cases where s_i and s_j do not form a base pair. The third case says that s_i and s_j form a base pair, and the resulting energy $e(s_i, s_j)$ is added to the present value of W. The forth formula represents the bifurcation structure. Note that k is the position at which the structure bifurcates in such a way that the sum of energies of two substructures is minimized.

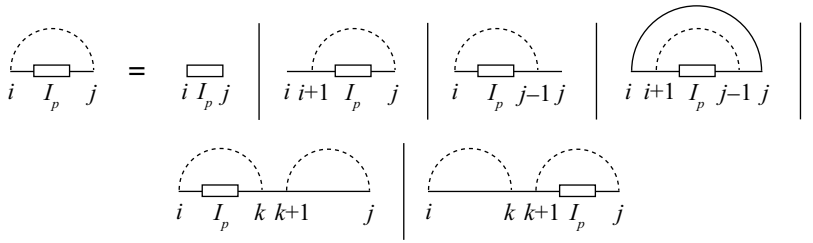

Fig. 2 Recursion for $W_{pp}(i,j)$. A dashed curve indicates that we do not know whether or not two bases connected by the curve form a base pair, and a solid curve shows that two bases connected by it definitely form a base pair.

Case 2 $(p = q)$:

$$W_{pp}(i,j) = \min \begin{cases} I_p(i,j), \\ W_{pp}(i+1,j), \\ W_{pp}(i,j-1), \\ W_{pp}(i+1,j-1) + e(i,j), \\ \min\limits_{i \leq k < j} \{W_{pp}(i,k) + W(k+1,j)\}, \\ \min\limits_{i \leq k < j} \{W(i,k) + W_{pp}(k+1,j)\}. \end{cases} \tag{3}$$

The first case means that $s_i s_{i+1} \cdots s_j$ is a binding site and we adopt the corresponding score $I_p(i,j)$ computed in the equation (1). The formulas from the second through the fourth are similar to the ones from the first through the third in the recursion (2). The fifth case represents the bifurcation structure where the binding site is contained in the former part of the bifurcation. Because the latter part of the bifurcation does not contain any binding sites, we use W computed in the recursion (2). The last case is a counterpart of the fifth case. Following a diagrammatic representation in[13], we provide a schematic representation of the recursion for $W_{pp}(i,j)$ in Fig. 2.

Vol.2009-BIO-17 No.13
2009/5/26

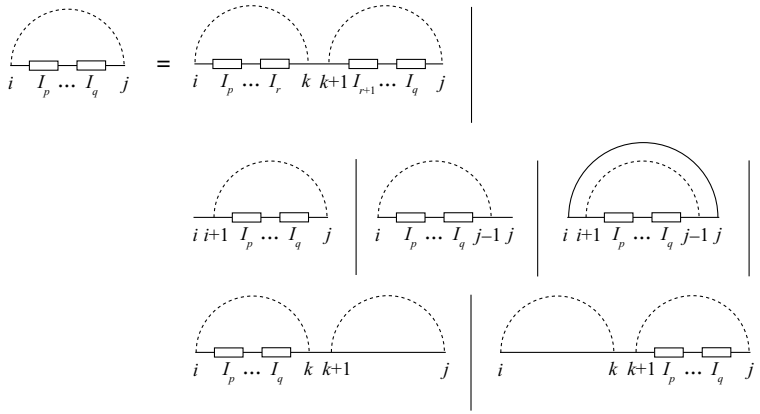

Fig. 3 Recursion for $W_{pq}(i,j)$.

Case 3 ($q \geq p+1$):

$$W_{pq}(i,j) = \min \begin{cases} \min\limits_{i \leq k < j} \min\limits_{p \leq r < q} \{W_{pr}(i,k) + W_{r+1,q}(k+1,j)\}, \\ W_{pq}(i+1,j), \\ W_{pq}(i,j-1), \\ W_{pq}(i+1,j-1) + e(i,j), \\ \min\limits_{i \leq k < j} \{W_{pq}(i,k) + W(k+1,j)\}, \\ \min\limits_{i \leq k < j} \{W(i,k) + W_{pq}(k+1,j)\}. \end{cases} \tag{4}$$

The first case is designed for computing the bifurcation of secondary substructures, each of which contains the binding sites. It should be noted that we have to find the position r at which a series of the binding sites is divided in such a way that the total energy of substructures is minimized. The other cases can be interpreted as in Case 2. Fig. 3 illustrates the above DP recursion.

We now evaluate the complexity of the above algorithm. Computing the equation (2) takes $O(n^3)$ time. The equations (3) and (4) can be computed in $O(Nn^3)$ and $O(N^3n^3)$ time, respectively. Therefore, the overall time complexity is evaluated as $O(N^3n^3)$. By similar evaluation, we can see that the space complexity is $O(N^2n^2)$.

The minimum energy of the secondary structure of the input sequence is equivalent to $W_{1,N}(1,n)$, and the optimal secondary structure can be retrieved by tracing back the DP table from $W_{1,N}(1,n)$.

2.2.2 Stacking energy model

Since the energy function used in the above DP algorithm is very simple, there is room for further improvement of our DP model. It is widely accepted that calculating contributions for stacking energy rather than individual contributions for each base pair yield better prediction. Hence, we extend the above DP algorithm based on this idea. In order to incorporate stacking energy into our previous DP model, we introduce additional DP tables. Let $V(i,j)$ be the minimum free energy of secondary structure formed from a subsequence $s_i s_{i+1} \cdots s_j$ such that s_i and s_j form a base pair. Let $V_{pq}(i,j)$ be the minimum free energy of secondary structure for $s_i s_{i+1} \cdots s_j$ that contains binding sites corresponding to $I_p, I_{p+1}, \ldots, I_q$ such that s_i and s_j form a base pair. Note that $W(i,j)$ and $W_{pq}(i,j)$ are defined in the same way as in the base pair energy model.

Initialization conditions for W and V are as follows:

$$W(i,i) = \infty,\ V(i,i) = \infty,\ W_{pq}(i,i) = \infty,\ V_{pq}(i,i) = \infty$$
$$(1 \leq \forall i \leq n;\ 1 \leq \forall p \leq \forall q \leq N).$$

The revised version of the DP recursions is as follows:

Case 1 (the Zuker's algorithm):

$$W(i,j) = \min \begin{cases} W(i+1,j), \\ W(i,j-1), \\ V(i,j), \\ \min\limits_{i \leq k < j} \{W(i,k) + W(k+1,j)\}, \end{cases} \tag{5}$$

$$V(i,j) = \min \begin{cases} eh(i,j), \\ V(i+1,j-1) + es(i,i+1,j-1,j), \\ \min\limits_{i < i' < j' < j} \{V(i',j') + ebi(i,i',j',j)\}, \\ \min\limits_{i < k < j-1} \{W(i+1,k) + W(k+1,j-1)\} + b \end{cases} \tag{6}$$

where $eh(i,j)$ is an energy of a hairpin loop closed by a pair of (s_i, s_j), $es(i,i+1,j-1,j)$ is a stacking energy of (s_i, s_j) and (s_{i+1}, s_{j-1}), $ebi(i,i',j',j)$ is an

energy of a bulge or an interior loop closed by (s_i, s_j) and $(s_{i'}, s_{j'})$, and b is a penalty for a bifurcation structure. Note that in the recursion formula (5), the case where s_i and s_j form a base pair is changed to $V(i,j)$. As can be seen in the equation (6), $V(i,j)$ is the minimization of four cases: The first case represents an energy of a hairpin loop closed by (s_i, s_j). The second formula adds the stacking energy of (s_i, s_j) and (s_{i+1}, s_{j-1}) to the present value of V. The third case represents a substructure where a bulge or an interior loop occurs in $s_i \cdots s_{i'}$ and $s_{j'} \cdots s_j$. The fourth formula is used for computing bifurcation.

Case 2 ($p = q$):

$$W_{pp}(i,j) = \min \begin{cases} W_{pp}(i+1,j), \\ W_{pp}(i,j-1), \\ V_{pp}(i,j), \\ \min\limits_{i \le k < j} \{W_{pp}(i,k) + W(k+1,j)\}, \\ \min\limits_{i \le k < j} \{W(i,k) + W_{pp}(k+1,j)\}, \end{cases} \tag{7}$$

$$V_{pp}(i,j) = \min \begin{cases} \min\limits_{i < i' < j' < j} \{I_p(i',j') + eh'(i,i',j',j)\}, \\ V_{pp}(i+1,j-1) + es(i,i+1,j-1,j), \\ \min\limits_{i < i' < j' < j} \{V_{pp}(i',j') + ebi(i,i',j',j)\}, \\ \min\limits_{i < k < j-1} \{W_{pp}(i+1,k) + W(k+1,j-1)\} + b, \\ \min\limits_{i < k < j-1} \{W(i+1,k) + W_{pp}(k+1,j-1)\} + b \end{cases} \tag{8}$$

where $eh'(i,i',j',j)$ is an energy of a hairpin closed by (s_i, s_j) that contains a binding site $s_{i'} s_{i'+1} \cdots s_{j'}$. $V_{pp}(i,j)$ is computed by minimizing among five choices: The first formula represents the case where I_p is contained in the hairpin loop closed by a base pair (s_i, s_j) and $s_{i'} s_{i'+2} \cdots s_{j'}$ is a binding site. The other cases are similar to those of $V(i,j)$ recursion. We show an illustration of the recursion $V_{pp}(i,j)$ in Fig. 4.

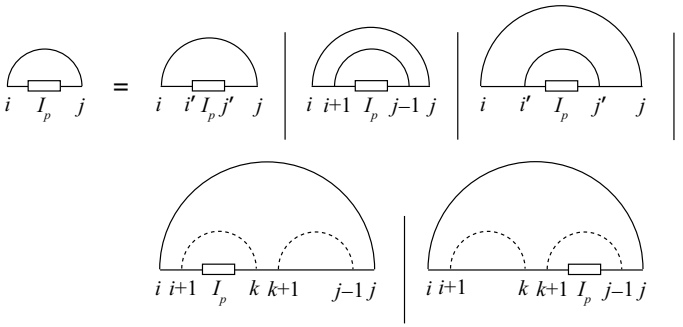

Fig. 4 Recursion for $V_{pp}(i,j)$.

Case 3 ($q \ge p+1$):

$$W_{pq}(i,j) = \min \begin{cases} \min\limits_{i \le k < j} \min\limits_{p \le r < q} \{W_{pr}(i,k) + W_{r+1,q}(k+1,j)\}, \\ W_{pq}(i+1,j), \\ W_{pq}(i,j-1), \\ V_{pq}(i,j), \\ \min\limits_{i \le k < j} \{W_{pq}(i,k) + W(k+1,j)\}, \\ \min\limits_{i \le k < j} \{W(i,k) + W_{pq}(k+1,j)\}, \end{cases} \tag{9}$$

$$V_{pq}(i,j) = \min \begin{cases} \min\limits_{i < k < j-1} \min\limits_{p \le r < q} \{W_{pr}(i+1,k) + W_{r+1,q}(k+1,j-1)\} + b, \\ \min\limits_{i < i' < j' < j} \{W_{pq}(i',j') + eh'(i,i',j',j)\}, \\ V_{pq}(i+1,j-1) + es(i,i+1,j-1,j), \\ \min\limits_{i < i' < j' < j} \{V_{pq}(i',j') + ebi(i,i',j',j)\}, \\ \min\limits_{i < k < j-1} \{W_{pq}(i+1,k) + W(k+1,j-1)\} + b, \\ \min\limits_{i < k < j-1} \{W(i+1,k) + W_{pq}(k+1,j-1)\} + b. \end{cases} \tag{10}$$

$V_{pq}(i,j)$ in Case 3 differs from $V_{pp}(i,j)$ in Case 2 in that $s_i s_{i+1} \cdots s_j$ contains at least two binding sites. The first case of minimization has the same meaning as that of $W_{pq}(i,j)$ except that (s_i, s_j) is already known to form a base pair. Fig. 5

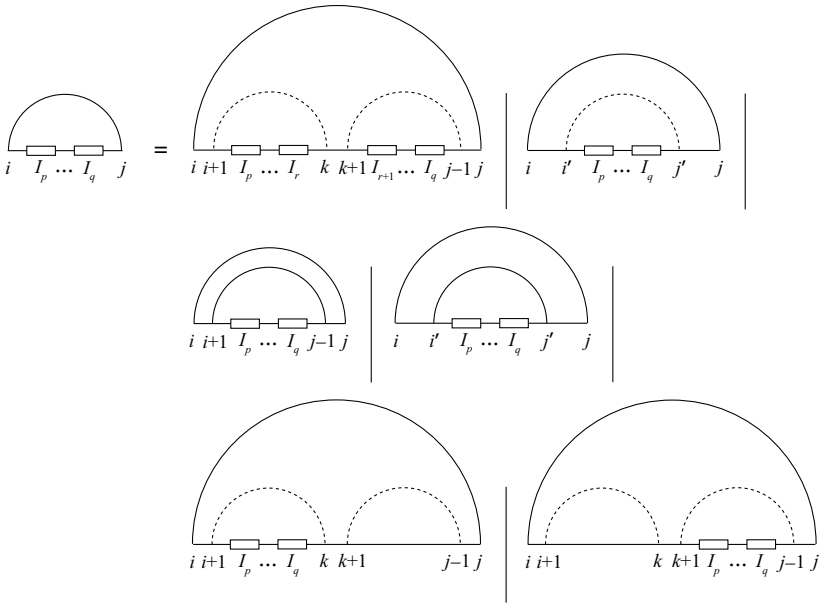

Fig. 5 Recursion for $V_{pq}(i,j)$.

shows the recursion for $V_{pq}(i,j)$.

Finally, we evaluate the complexity of this algorithm. Obviously, complexity for computing the equation (10) dominates the overall complexity of the algorithm. Exact analysis of the second and forth formulas of the equation (10) reveals time complexity of $O(N^2n^4)$. However, in actual case, the loop size is bounded by a constant, and thus the complexity can be reduced to $O(N^2n^2)$. Computing the first formula takes $O(N^3n^3)$ time. Therefore, the overall time complexity is evaluated as $O(N^3n^3)$. The space complexity is $O(N^2n^2)$.

3. Results

The DP for base pair energy model was tested by the data set consisting of six antisense-target complexes, taken from literatures (see Table 2). The secondary structures of antisense-target complexes in the data set are known, also each of them is known to contain at least one binding site. The length of target sequences in the data set ranges from 16–100 bases. The maximum number of binding sites is three (CopA-CopT).

Table 2 Results of DP algorithm for base pair energy, where n is the length of a target sequence and N is the number of binding sites.

Antisense-Target	n	N	Sensitivity (%)	Specificity (%)	F-measure (%)	CPU Time (sec)
Tar-Tar*[6]	16	1	100.00	90.00	94.74	0.23
R1inv-R2inv[12]	19	1	100.00	100.00	100.00	0.33
DIS-DIS[10]	35	1	82.35	73.68	77.78	1.07
CopA-CopT[14]	57	3	77.42	75.00	76.19	17.14
IncRNA54-RepZ[3]	61	2	72.97	72.97	72.97	9.98
OxyS-fhlA[14]	100	2	72.73	65.31	68.82	41.46

We measured the prediction accuracy using sensitivity, specificity and F-measure as defined below:

$$\text{sensitivity} = \frac{\sharp\text{correctly predicted base pairs} + \sharp\text{correctly predicted bases in the binding sites}}{\sharp\text{observed base pairs} + \sharp\text{observed bases in the binding sites}},$$

$$\text{specificity} = \frac{\sharp\text{correctly predicted base pairs} + \sharp\text{correctly predicted bases in the binding sites}}{\sharp\text{predicted base pairs} + \sharp\text{predicted bases in the binding sites}},$$

$$\text{F-measure} = \frac{2 \times \text{sensitivity} \times \text{specificity}}{\text{sensitivity} + \text{specificity}}.$$

Table 2 shows accuracy of the base pair energy model. As we can see from the table, the DP algorithm can predict the secondary structure with binding sites with at most 100.00% of sensitivity and specificity for the target sequence R2inv, and at least 72.73% of sensitivity and 65.31% of specificity for the target sequence fhlA despite use of only simple energy function. Fig. 6 depicts the secondary structure predicted by the algorithm in dot-parenthesis representation where the binding sites are indicated by the * symbol.

4. Conclusion

We proposed new dynamic programming algorithms for predicting RNA secondary structures with binding sites. Our approach is a novel method of RNA-RNA interaction prediction from a different point of view (i.e., use of profile of

Observed

```
5'-AAACCCCGAUAAUCUUCUUCAACUUUGGCGAGUACGAAAAGAUUACCGGGGCCCCAC-3'
   ...(((((.(((((((.******.....*********)))))))..)))))..****
```

Predicted

```
5'-AAACCCCGAUAAUCUUCUUCAACUUUGGCGAGUACGAAAAGAUUACCGGGGCCCCAC-3'
   ...((((((.((.(..)******)))((.*********((..))))))))..****
```

(a) CopT

Observed

```
5'-AUGACCUUUUGCACCGCUUUGCGGUGCUUUCCUGGAAGAACAAAAUGUCAUAUACACCGAUGAGUGAUCUCGGACAACAAGGGUUGUUCGACAUCACUCG-3'
   (((((.((((((((((...)))))))..*******.....)))).)))))........(((((((.((((....*********.)))).)))))))
```

Predicted

```
5'-AUGACCUUUUGCACCGCUUUGCGGUGCUUUCCUGGAAGAACAAAAUGUCAUAUACACCGAUGAGUGAUCUCGGACAACAAGGGUUGUUCGACAUCACUCG-3'
   ((((((((((.(..(..))((...))((*******).).))))))))))((.(...)))(((((((..)(((((*********)).)))))))))).
```

(b) fhlA

Fig. 6 Prediction results for the base pair energy model.

intermolecular interaction), and achieved lower time complexity compared with earlier methods. The performance of the algorithm for base pair energy was demonstrated for several known RNA-RNA interaction data. To improve the prediction accuracy, we are now trying to carry out prediction tests using the algorithm for stacking energy. Our method could be applicable to RNA-protein interaction if a profile of interacting protein is available, which is also left as our future work.

Acknowledgments The first author thanks the Japanese Government (Monbukagakusho:MEXT) Scholarship from the Ministry of Education, Culture, Sports, Science and Technology. This work was supported in part by Grant-in-Aid for Young Scientists (Start-up) (KAKENHI) #20800023 from Japan Society for the Promotion of Science (JSPS), and by Grant-in-Aid for Scientific Research (KAKENHI) #19200022 from MEXT, Japan.

References

1) C.Alkan, E.Karakoç, J.H.Nadeau, S.C.Şahinalp and K.Zhang: RNA-RNA interaction prediction and antisense RNA target search, *Journal of Computational Biology*, vol.13, no.2, pp.267–282, 2006.
2) M.Andronescu, Z.C.Zhang and A.Condon: Secondary structure prediction of interacting RNA molecules, *Journal of Molecular Biology*, vol.345, issue 5, pp.987–1001, 2005.
3) K. Asano and K. Mizobuchi: Structural analysis of late intermediate complex formed between plasmid ColIb-P9 Inc RNA and its target RNA, *The Journal of Biological Chemistry*, vol.275, no.2, pp.1269–1274, 2000.
4) S. Brantl: Antisense-RNA regulation and RNA interference, *Biochimica et Biophysica Acta*, vol.1575, issues 1–3, pp.15–25, 2002.
5) C.Brunel, R.Marquet, P.Romby and C.Ehresmann: RNA loop-loop interactions as dynamic functional motifs, *Biochimie*, vol.84, issue 9, pp.925–944, 2002.
6) K.-Y.Chang and I.Tinoco Jr: The structure of an RNA "kissing" hairpin complex of the HIV TAR hairpin loop and its complement, *Journal of Molecular Biology*, vol.269, issue 1, pp.52–66, 1997.
7) Y.Kato, T.Akutsu and H.Seki: A grammatical approach to RNA-RNA interaction

prediction, *Pattern Recognition*, vol.42, issue 4, pp.531–538, 2009.
8) F.A.Kolb, C.Malmgren, E.Westhof, C.Ehresmann, B.Ehresmann, E.G.H.Wagner and P.Romby: An unusual structure formed by antisense-target RNA binding involves an extended kissing complex with a four-way junction and a side-by-side helical alignment, *RNA*, vol.6, no.3, pp.311–324, 2000.
9) R. Nussinov, G. Pieczenik, J.R. Griggs and D.J. Kleitman: Algorithms for loop matchings, *SIAM Journal on Applied Mathematics*, vol.35, no.1, pp.68–82, 1978.
10) J.-C.Paillart, E.Skripkin, B.Ehresmann, C.Ehresmann and R.Marquet: A loop-loop "kissing" complex is the essential part of the dimer linkage of genomic HIV-1 RNA, *Proceedings of the National Academy of Sciences of the USA*, vol.93, no.11, pp.5572–5577, 1996.
11) D.D.Pervouchine: IRIS: intermolecular RNA interaction search, *Proceedings of the 15th International Conferences on Genome Informatics (GIW2004)*, pp. 92–101, 2004.
12) M.J.Rist and J.P.Marino: Association of an RNA kissing complex analyzed using 2-aminopurine fluorescence, *Nucleic Acids Research*, vol.29, no.11, pp.2401–2408, 2001.
13) E. Rivas and S.R. Eddy: A dynamic programming algorithm for RNA structure prediction including pseudoknots, *Journal of Molecular Biology*, vol. 285, issue 5, pp.2053–2068, 1999.
14) E.G.H.Wagner and K.Flardh: Antisense RNAs everywhere?, *TRENDS in Genetics*, vol.18, no.5, pp.223–226, 2002.
15) M.Zuker and P.Stiegler: Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information, *Nucleic Acids Research*, vol.9, no.1, pp. 133–148, 1981.