Available online at www.sciencedirect.com

ScienceDirect

Theoretical Computer Science 395 (2008) 47–56

Theoretical Computer Science

www.elsevier.com/locate/tcs

A graphical criterion of planarity for RNA secondary structures with pseudoknots in Rivas–Eddy class

Shile Gao*, Kequan Ding

School of Electronic and Information Engineering, Dalian University of Technology, Dalian 116024, PR China

Received 27 August 2006; received in revised form 19 August 2007; accepted 11 September 2007

Communicated by A. Condon

Abstract

An RNA secondary structure is considered to be planar if its arc graph can be embedded into a plane without edge crossing. In this paper, a graphical criterion of planarity is presented based on graphical composition for RNA secondary structures with pseudoknots in Rivas–Eddy Class. Effective planar testing algorithms are introduced based on our graphical criterion.

Keywords: Pseudoknots; RNA secondary structures; Planar graph; Generator; Graph composition

---

1. Introduction

RNA molecules are sequences of nucleotides. They play diverse roles in living cells: as carriers of information, catalysts in cellular processes, and mediators in determining the expression level of genes [1]. Three-dimensional structures of RNA molecules are often the key to their functions [2]. The structures are strongly influenced by interactions between pairs of nucleotides, so-called base pairings. *RNA secondary structure* is the set of base pairings in the three-dimensional structure of a molecule.

In RNA secondary structures without pseudoknots, there are six basic structural elements: stacked pairings (stems), hairpin loops, bulge loops, internal loops, multiple loops, and external bases. A common structure is a hairpin and stem (see Fig. 1). The pairings of pseudoknot free structures lie in stems and they stacked each other. They can be understood easily and embedded into a plane without crossing edges [3]. When one loop pairs with another loop or some external bases, a new stem will be formed, which is named a pseudoknot (see Section 2 for an exact definition). RNA secondary structures with pseudoknots are complex due to the interactions among the overlapping base pairs (see Fig. 1). Some overlapping base pairs interact in “long distance”; i.e., between the two basic structural elements which form a pseudoknot there are many other structural elements; some pseudoknots exist in larger ones, which is called a recurrence of pseudoknots.

---

* Corresponding author. Tel.: +86 10 51057107.
E-mail address: gaoshile@sohu.com (S. Gao).

0304-3975/$ - see front matter
doi:10.1016/j.tcs.2007.09.026

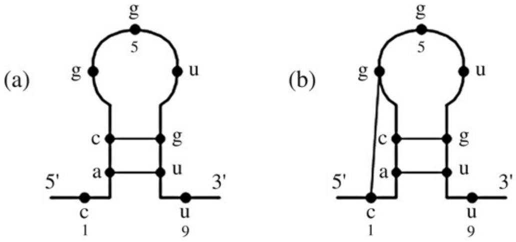

Fig. 1. Two examples of RNA secondary structures. (a) Nested structure: hairpin; (b) pseudoknotted structure: H-type pseudoknot.

Drawing an RNA molecule on a plane is often helpful to understanding its structure. Pseudoknot free RNA structures can be drawn in a plane as trees where the edges are the stems and the vertices are the rest of the basic structural elements. The pseudoknotted structures can be only drawn as graphs, even non-planar graphs, because they have additional interactions between different vertices of this tree [3]. It is important to draw pseudoknotted RNA secondary structures on a plane without edge crossing if possible since the graphs drawn without edge crossings have a natural advantage for visualization [4]. Therefore, the characterization problem of planarity is essential to understand RNA structures with pseudoknots.

The planar embedding of RNA secondary structures has been explored only in recent years. Witwer et al. [5] did case studies on some families of RNA pseudoknots that either can or cannot be drawn on a plane without intersecting lines. But, no simple criterion for recognizing planar pseudoknotted secondary structures has been given. Although Han et al. [3,6] have found that H-type pseudoknots, the simplest kind of pseudoknotted secondary structures, are all planar graphs and designed an embedding algorithm, other types of pseudoknots were not mentioned in their investigations.

Rivas and Eddy [7,8] have proposed a dynamic programming algorithm, R&E algorithm, for RNA secondary structure prediction including pseudoknots. The target of the R&E algorithm is called an R&E secondary structure class, or simply called R&E class [9]. R&E class is the largest one known by far. Condon et al. redefined R&E class using R&E patterns, and proved that the R&E class is exactly equivalent to R&E patterns. They pointed out a discriminance of R&E patterns and tested 486 structures from the PseudoBase and Gutell database, and there are only 3 exceptions. In this field, time complexity is important. To test a planar graph in general, complicated linear-time algorithms have been presented by Hopcroft et al. [10] and Booth et al. [11], and a simpler quadratic algorithm is further presented by Demoucron et al. [12]. A quadratic algorithm to test planarity is also given based on the proof of Kuratowski's theorem by Klotz [13]. In this paper, we will focus on the planarity of RNA secondary structures with pseudoknots in R&E class and give a planarity criterion via graph composition. A quadratic time testing algorithm and a linear time testing algorithm will also be introduced to recognize planar RNA secondary structures with pseudoknots in R&E class.

2. Graph composition of RNA secondary structures

There are many different ways to represent RNA secondary structures, such as conventional diagram representation [14–16], arc diagram representation [9,14,15], pattern representation [9,17], bracket representation [16], mountain representation [16], circle representation [16], and tree representation [18]. Generally, conventional diagram representation and arc diagram representation are more common in comparison to other representations. Arc diagram representation can be regarded as a special case of conventional diagram representation. The conventional diagram representation of RNA secondary structures is shown in Fig. 1. They are described once more by use of arc diagram representation, where base indices are shown as vertices on a straight line (backbone), ordered from the $5'$ end, and base pairs are indicated using arcs, as shown in Fig. 2(a). In an arc diagram, an arc from i to j $(i < j)$ can be written as (i, j).

As we know, inserting or deleting a node of degree 2 does not influence whether a graph is planar [19]. Arc diagrams can be simplified by deleting the unpaired bases since they have degree 2. We name the simplified arc diagram as *arc graph*. The arc diagrams in Fig. 2(a) can be simplified as arc graphs easily, as shown in Fig. 2(b). In the arc graph representation, there are $3n/2-1$ edges corresponding to n (even) nodes. n-1 of them are backbone edges and the others are arcs which represent pairings. An arc (i, j) is *crossed* if there exists another arc (i', j') satisfying

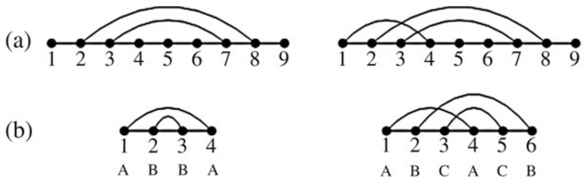

Fig. 2. Arc representations of RNA. (a) Arc diagrams of the nested and the pseudoknotted structures shown in Fig. 1; (b) arc graphs corresponding to the arc diagrams in (a).

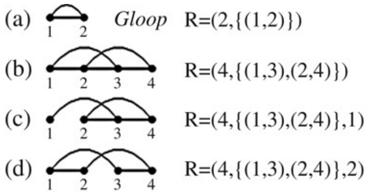

Fig. 3. (a)–(b) Examples of 0-structure graphs; (c)–(d) examples of 1-structure graphs.

$i < i' < j < j'$ or $i' < i < j' < j$. An arc is *nested* if it is not crossed. In the arc graphs, an RNA *pseudoknot* is defined as a structure containing at least one crossed arc [14]. A *region*, which is denoted as $[i; j]$ $(i < j)$, is defined as a contiguous set of base indices such that each base index in $[i; j]$ is paired with another base index in $[i; j]$ [17]. For example, the left example in Fig. 2(b) represents a pseudoknotted free RNA secondary structure. There are 4 vertices and 5 edges, including 3 backbone-edges and 2 nested arcs. Its regions are [2; 3] and [1; 4], respectively. The right example in Fig. 2(b) represents a pseudoknotted RNA secondary structure. There are 6 vertices and 8 edges, including 5 backbone-edges and 3 crossed arcs. In this case, there is only one region, [1; 6].

Now we will introduce a method of graph composition based on arc graph representation. In this method, the secondary structure of the RNA molecule is constructed from *generators* which form a set of basic structures to represent pairings (unpaired bases are deleted). We use 0-structures to represent the classical structures and 1-structures to represent the gapped ones used by Rivas and Eddy in [7,8]. 0-structures mean the usual foldings of RNA sequences with the unpaired bases deleted and 1-structures represent the generators of the composite method.

Definition 1 (*(0- and 1-Structures) [15]*). 0-structure $R = (n, P)$ consists of an even number n and a set $P \subset N_n^2$ of pairs (i, j) with $i < j$, such that for (i, j) and $(i', j') \in P$, the equivalence $i = i' \Leftrightarrow j = j'$ holds. Here P contains $n/2$ elements.

A 1-Structure $R = (n, P, k)$ consists of a 0-structure (n, P) and a natural number $0 < k < n$. (The backbone has a gap between k and $k + 1$.)

Definition 2 (*(0- and 1-Structure Graphs) [20]*). The structure graph $G(R) = (n, E)$ of a 0- or 1-structure R is a multi-graph, consisting of the vertex set N_n (represent bases) and two types of edges.

If $R = (n, P)$ is a 0-structure the set E of edges contains:

(1) Backbones: $\{i, i + 1\}$ for $1 \leqslant i < n$;
(2) Pairings: $\{i, j\}$ for $(i, j) \in P$.

If $R = (n, P, k)$ is a 1-structure the set E of edges contains:

(1) Backbones: $\{i, i + 1\}$ for $1 \leqslant i \leqslant k - 1$ and $k < i < n$;
(2) Pairings: $\{i, j\}$ for $(i, j) \in P$.

The structure graph of the 0-structure $R = (2, \{(1, 2)\})$ is shown in Fig. 3(a). It is the unique one-arc arc graph, the simplest one. We name it *Gloop* for convenience. The structure graph of the 0-structure $R = (4, \{(1, 3), (2, 4)\})$ is shown in Fig. 3(b). The structure graphs of the 1-structures $R = (4, \{(1, 3), (2, 4)\}, 1)$ and $R = (4, \{(1, 3), (2, 4)\}, 2)$ are shown in Figs. 3(c) and 3(d), respectively. One can see from the above examples that the structure graph of a 1-structure $R = (n, P, k)$ is the graph of the underlying 0-structure (n, P) without the backbone between k and $k + 1$.

In the following, we will define an operation, graph composition, to construct larger structure graphs using the generators.

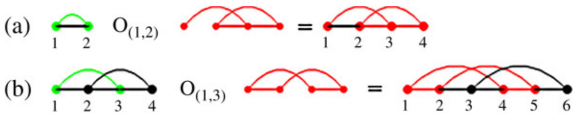

Fig. 4. Two examples of graph composition. (a) The composition upon arc (1, 2); (b) the composition upon arc (1, 3).

Fig. 5. A set of generators used for graph composition.

Definition 3 (*Graph Composition*). Let $R = (n, P)$ be a 0-structure and (i, j) be one of its pairings. For each 1-structure $R' = (m, Q, l)$, the structure $R\bigcirc_{(i,j)}R'$ is defined by

$$R\bigcirc_{(i,j)}R' = (n + m - 2, P\bigcirc_{(i,j)}Q).$$

$P\bigcirc_{(i,j)}Q$ contains the following elements:

$$(j_1', j_2') \text{ for all } (j_1, j_2) \in P \text{ with } j_x' = \begin{cases} j_x & \text{if } j_x < i; \\ j_x + l - 1 & \text{if } i < j_x < j; \\ j_x + m - 2 & \text{if } j < j_x. \end{cases}$$

$$(j_1', j_2') \text{ for all } (j_1, j_2) \in Q \text{ with } j_x' = \begin{cases} j_x + i - 1 & \text{if } j_x \leqslant l; \\ j_x + j - 2 & \text{if } l < j_x. \end{cases}$$

The operation of composition is to replace the arc by a generator and the composition upon the pairing (i, j) is denoted as $\bigcirc_{(i,j)}$. Two examples are taken in Fig. 4 to describe the operation of composition. In Fig. 4(a), the generator $(4, \{(1, 3), (2, 4)\}, 1)$ composes upon arc $(1, 2)$ in the 0-structure $(2, \{(1, 2)\})$ and the structure $(4, \{(1, 3), (2, 4)\})$ is obtained. The example shown in Fig. 4(b) is similar. If a 0-structure graph G is composed by a generator with 4-nodes and 2-arcs, the nodes of G increase 2, the edges increase 3, including 2 backbone-edges and an arc. The arc which was generated in the composition is called *new arc*; on the contrary, the arc upon which the generator composes is called *old arc*.

The above analysis shows that graph composition provides an efficient method to construct larger RNA structures from basic ones. In this method, a series of simple generators that can generate all the RNA secondary structure graphs in a certain class are essential to achieving graph composition. Fig. 5 shows a set of 1-structure generators named as *Glc* (Generator of left connect), *Grc* (Generator of right connect), *Gli* (Generator of left insert), *Gri* (Generator of right insert), *Glw* (Generator of left wrap), *Grw* (Generator of right wrap), *Gnest* (Generator of nest), and *Gcross* (Generator of cross), respectively. These 8 generators have elegant structure since all of them are made up of 4-node and 2-arc structures. The composition path is certain because only one backbone is missed. As a result, their composition function is intuitionistic and determined. In addition, *Gloop* can be considered as the only start arc graph of graph composition.

As the generators of the graph composition, the composition ability of this set of 8 generators is eagerly concerned. In what follows, we will exhibit that all of the RNA secondary structures in R&E class can be completely composed using them.

3. R&E graphs

Pattern representation has been introduced by Condon et al. [9,17], where the information about unpaired bases is lost but the pattern of nesting or overlaps among base pairs is preserved. A pattern p is a $2n$-length string in which there are n kinds of symbols, each of them appears exactly twice. For example, ABCACB is a pattern representation of the right example shown in Fig. 2(b). Condon et al. redefined R&E class using R&E patterns, and proved that R&E class is exactly equivalent to R&E patterns.

Theorem 1. *The arc graphs which compound through* $\{Gloop, \bigcirc_{(i,j)}, Gen\}$ *are exactly the structure graphs of R&E class, where Gen is the finite set of the generators in Fig. 5.*

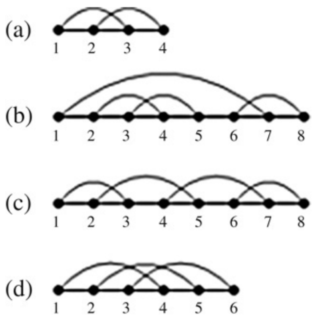

Fig. 6. Examples of R&E graphs. (a) H-type pseudoknot; (b) recurrence of pseudoknots; (c) chains of pseudoknots; (d) complex pseudoknot.

Proof. Since R&E class is exactly equivalent to R&E patterns, we only need prove the structure graphs which compound through $\{Gloop, \bigcirc_{(i,j)}, Gen\}$ are exactly the structures corresponding to R&E patterns.

Pattern p is R&E if $p \xrightarrow[R\&E]{*} \varepsilon$, where ε is an empty string [9]. The derivative rules of R&E pattern is defined implicitly by the definition of R&E pattern p; these rules are shown as the left part of formula (1). The symbols that satisfy the structural character like σ can be removed from p (τ represents any other symbol). The right part of formula (1) is the generators corresponding to the derivative rules. In the following, arc graphs refer to that generated by $\{Gloop, \bigcirc_{(i,j)}, Gen\}$ and patterns refer to R&E patterns. The relationship between them is expressed as

$$\begin{cases} \langle 1 \rangle \cdots \sigma\sigma \cdots & \leftrightarrow \quad \begin{Bmatrix} Glc \\ Grc \\ Gli \\ Gri \end{Bmatrix} \\ \langle 2 \rangle \cdots \tau \cdots \sigma\tau\sigma \cdots & \leftrightarrow \quad Grw \\ \langle 3 \rangle \cdots \sigma\tau\sigma \cdots \tau \cdots & \leftrightarrow \quad Glw \\ \left.\begin{matrix} \langle 4 \rangle \cdots \tau\sigma \cdots \sigma\tau \cdots \\ \langle 5 \rangle \cdots \sigma\tau \cdots \tau\sigma \cdots \end{matrix}\right\} & \leftrightarrow \quad Gnest \\ \left.\begin{matrix} \langle 6 \rangle \cdots \tau\sigma \cdots \tau\sigma \cdots \\ \langle 7 \rangle \cdots \sigma\tau \cdots \sigma\tau \cdots \end{matrix}\right\} & \leftrightarrow \quad Gcross. \end{cases} \tag{1}$$

The proof is performed by induction on the length of p. The shortest pattern has the same structural character as $\sigma\sigma$, which corresponds to the start arc graph *Gloop*. For the inductive step, supposing every $2k$-length ($k > 1$) pattern p corresponds to a $2k$-node arc graph G, we discuss any $2k+2$-node arc graphs G'. p must be derived from a $2k+2$-length pattern p' according to a certain derivative rule. Choose the generator corresponding to the derivative rule and compose it on the arc corresponding to symbol τ. One can get a $2k+2$-nodes R&E graph. This graph corresponds to p'.

However, the derivative rule $\langle 1 \rangle$ is an exception because there is no symbol τ in it. In this case, we can select the generators from the arcs neighboring $\sigma\sigma$. Suppose the indices of $\sigma\sigma$ are i and $i+1$, we can select the generators as follows:

$$\begin{cases} Glc & \text{if the right arc of } \sigma\sigma \text{ is } (i+2, j) \\ Grc & \text{if the left arc of } \sigma\sigma \text{ is } (k, i-1) \\ Gli & \text{if the left arc of } \sigma\sigma \text{ is } (i-1, j) \\ Gri & \text{if the right arc of } \sigma\sigma \text{ is } (l, i+2). \end{cases} \tag{2}$$

If $\sigma\sigma$ has left and right neighbors, two generators are suitable. We can select any one of them at will, and then the selected generator can be composed upon the arc which determines the selecting of the generators. □

Due to Theorem 1, it is reasonable to define the structure graphs compounded through $\{Gloop, \bigcirc_{(i,j)}, Gen\}$ as *R&E Graphs* visually since they are exactly the structure graphs of R&E class. All R&E graphs can be composed through these generators. Some typical R&E graphs are shown in Fig. 6, such as H-type pseudoknot, recurrence of pseudoknots, chains of pseudoknots, and complex pseudoknot.

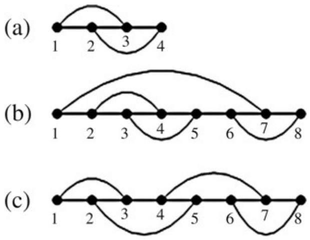

Fig. 7. Planar embeddings of (a) H-type pseudoknot; (b) recurrence of pseudoknots; (c) chains of pseudoknots.

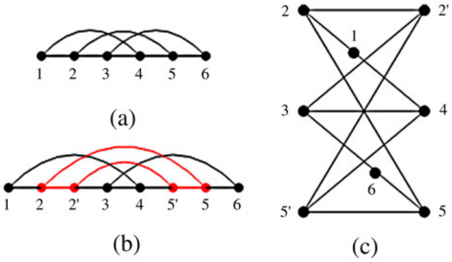

Fig. 8. (a) A typical example of non-planar R&E graph; (b) the graph obtained by replacing arc (2, 5) in (a) with *Gnest*; (c) an identical graph of (b) that contains a homeomorphic copy of $K_{3,3}$.

4. Planarity of R&E graphs

Some R&E graphs are planar, such as H-type pseudoknot, recurrence of pseudoknots, chains of pseudoknots shown in Figs. 6(a)–6(c), respectively. Fig. 7 shows a planar embedding for each of these types. However, some R&E graphs for RNA with complex pseudoknot are non-planar. An example that belongs to R&E graphs is shown in Fig. 8(a) [21]. Fig. 8(b) is obtained by adding an arc that is parallel to arc (2, 5). Figs. 8(a) and 8(b) are both planar or not simultaneously because arc $(2', 5')$ in Fig. 8(b) can be embedded into a plane at the same side of arc (2, 5) and without any intersections on it. Since Fig. 8(b) contains a homeomorphic copy of $K_{3,3}$ (see Fig. 8(c)), Fig. 8(a) is non-planar.

We are interested in the planarity of the R&E graphs since it is significant to understand RNA structures. In this section, we will try to find out the criterion to distinguish the planar R&E graphs from the non-planar ones.

Definition 4 (*P-Graphs*). The arc graphs which compound through $\{Gloop, \bigcirc_{(i,j)}, Gen'\}$ are called *P-graphs*, where Gen' is the finite set of the generators in Fig. 5 except *Gcross*.

Theorem 2. *Let G be an n nodes P-graph,* $[i; j]$ *is a region in G, then* $[1; i-1] \cup [j+1; n]$ *is a P-graph.*

Proof. The decomposition of G is defined to replace a generator in G using an arc or delete *Gloop* in G. A *P-graph* can be decomposed to empty graph in $n/2$ steps. For generators *Glc*, *Grc*, *Gli* and *Gri*, replacing them with an arc is equivalent to deleting *Gloop* in them, so the decomposition of G is deleting any *Gloop* in G or replacing *Glw*, *Grw* and *Gnest* with an arc.

The rest of the graph is denoted as G' after G is decomposed once. The nodes of G' are 2 less than those of G, the arcs of G' are 1 less than those of G. Supposing the deleted arc is α, we just use α to mark this decomposition. Because G contains $n/2$ arcs, we mark the decomposition sequence using the $n/2$ length sequence of arcs deleted in each step. The sequence is a permutation of whole arcs in G; we denote this sequence as $\alpha_1\alpha_2\cdots\alpha_{n/2}$.

Suppose $[i; j]$ is a region in G, which contains $(j-i+1)/2$ arcs. The arcs in $[i; j]$ must be in sequence $\alpha_1\alpha_2\cdots\alpha_{n/2}$. We select the arcs belonging to region $[i; j]$ and underline them in $\alpha_1\alpha_2\cdots\alpha_{n/2}$. The selected arcs are denoted as $\beta_1\beta_2\cdots\beta_{(j-i+1)/2}$. Now we will prove that the underlined symbols can be centralized to the left of $\alpha_1\alpha_2\cdots\alpha_{n/2}$.

Let α_k be an arc in $[i; j]$. If the decomposition of α_k accords to one of *Gloop*, *Glw* or *Grw* in G, the chosen one is included in $[i; j]$. If $\alpha_k \neq (i, j)$ and the decomposition of α_k accords to *Gnest*, *Gnest* is included in $[i; j]$. If $\alpha_k = (i, j)$, the decomposition of α_k accords to *Gnest*, and the arc parallel with α_k is $(i+1, j-1)$, *Gnest* is also included in $[i; j]$. In the above cases, the decomposition of α_k can be done in $[i; j]$. However, when the following three events exist simultaneously: $\alpha_k = (i, j)$, the decomposition of α_k accords to *Gnest*, and the arc parallel with α_k is $(i-1, j+1)$, *Gnest* will not be included in $[i; j]$. At this time, supposing α_k is β_s in the arc sequence

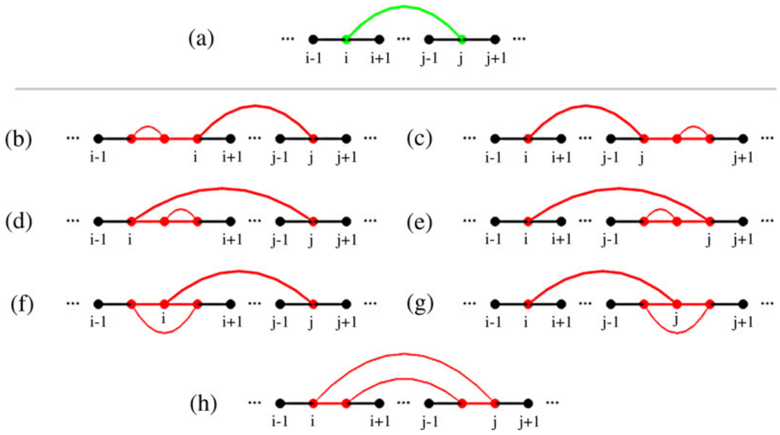

Fig. 9. (a) G': a planar embedding of G; (b)–(h) G_1': a planar embedding of $G\bigcirc_{(i,j)}$ *Gener* (one of *Glc* (b), *Grc* (c), *Gli* (d), *Gri* (e), *Glw* (f), *Grw* (g), and *Gnest* (h)).

of $\beta_1\beta_2\cdots\beta_{(j-i+1)/2}$, which can be rearranged as $\beta_1\beta_2\cdots\beta_{s-1}\beta_{s+1}\cdots\beta_{(j-i+1)/2}\beta_s$. According to the rearranged sequence, β_s, i.e., α_k, can be decomposed as a *Gloop*.

Since any arc in $[i; j]$ can be decomposed in itself, $[i; j]$ is a P-graph. $[1; i-1] \cup [j+1; n]$ is also a P-graph because G is a P-graph and $[i; j]$ is a P-graph. □

Theorem 3. *The P-graphs are exactly the whole planar R&E graphs.*

Proof. We will prove this theorem in two steps:

Step 1. P-graphs are planar.

The start arc graph of the composition of P-graphs is *Gloop*, which is planar. We suppose that the graph after n-step compositions on *Gloop*, which is denoted as G, is a planar. G' is a planar embedding of G, as shown in Fig. 9(a). $G_1 = G\bigcirc_{(i,j)}$ *Gener* (one of *Gen*$'$). A kind of planar embedding of G_1 is shown in Figs. 9(b)–9(h), respectively. In these figures, different G_1 correspond to different generators. Therefore, P-graphs are all planar.

Step 2. Each arc graph in R&E graphs but not in P-graphs is non-planar.

Based on the method of graph composition, the difference between the R&E graphs and the P-graphs is *Gcross*. Before the first use of *Gcross*, arc graph G is a P-graph. Now we consider the first composition of *Gcross*:

(i) Compose upon a crossed arc (i, j): $G\bigcirc_{(i,j)}$ *Gcross* is non-planar.

Arc (k, l) is supposed to be crossed with (i, j), as shown in the left arc graph in Fig. 10(a). $G\bigcirc_{(i,j)}$ *Gcross* contains a subgraph which consists of arcs $(k, l+1)$, $(i, j+1)$, and $(i+1, j+2)$ and the backbone edges from k to $j+2$ (see the right arc graph in Fig. 10(a)). This subgraph is non-planar since it contains a non-planar graph shown in Fig. 8(a). Therefore, $G\bigcirc_{(i,j)}$ *Gcross* is non-planar if (i, j) is a crossed arc. Because P-graphs are all planar, $G\bigcirc_{(i,j)}$ *Gcross* is not a P-graph but an non-planar R&E graph.

(ii) Compose upon a nested arc (i, j): $G\bigcirc_{(i,j)}$ *Gcross* is a P-graph.

$G\bigcirc_{(i,j)}$ *Gcross* is shown in Fig. 10(b). We affirm that $G\bigcirc_{(i,j)}$ *Gcross* can be composed without *Gcross*. $[i+1; j-1]$ is a region because (i, j) is a nested arc. $[1; i] \cup [j; n]$ is a P-graph because G is a P-graph (due to Theorem 2). $[1; i] \cup [j; n]$ is shown in Fig. 10(c).

The composition steps to construct the graph at the right hand of Fig. 10(b) without *Gcross* are described as follows (bracketed indices in Figs. 10(c)–10(e) are the final indices of the nodes in Fig. 10(f)):

(1) Embed $[1; i] \cup [j; n]$ in a plane using arc graphs, as shown in Fig. 10(c).
(2) Compose *Grw* on $(i, i+1)$, as shown in Fig. 10(d).
(3) Compose *Gri* on $(i, i+2)$, as shown in Fig. 10(e).
(4) Use arc $(i+2, i+3)$ as the basis to compose region $[i+2; j]$, as shown in Fig. 10(f).

Fig. 10(f) is the same as the right hand of Fig. 10(b). It means that $G\bigcirc_{(i,j)}$ *Gcross* can be composed without *Gcross*. The analysis is similar when the generator *Gcross* is used once more. □

Due to Theorem 3, the P-graphs are demonstrated to be exactly the total planar graphs in R&E class. Therefore, they can be called as *planar R&E Graphs* visually. The above analysis shows that the generator *Gcross* is the key of

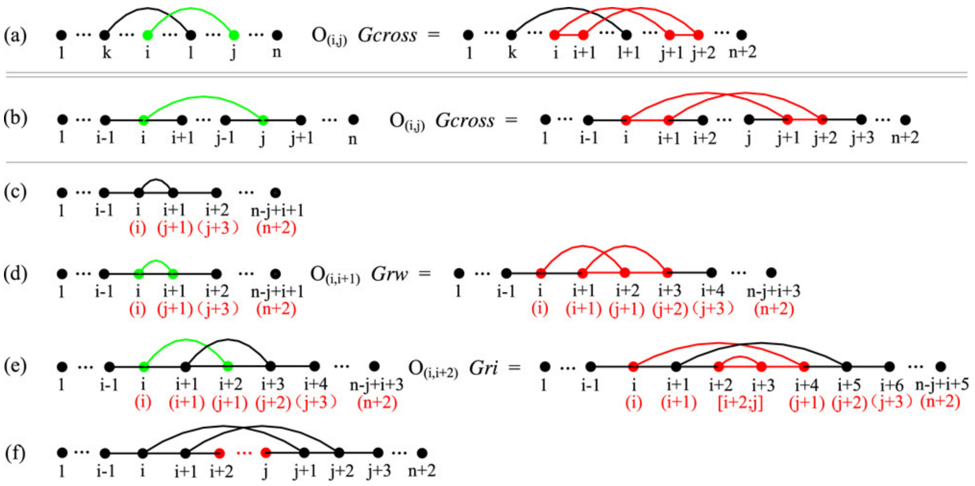

Fig. 10. (a) The composition of *Gcross* upon a crossed arc. (b) the composition of *Gcross* upon a nested arc; (c)–(f) the composition processes of the graph at the right hand of (b) without *Gcross* (bracketed indices in Fig. 10 (c)–(e) are the final indices of the nodes in Fig. 10(f)).

the planarity of RNA secondary structures in R&E class. When *Gcross* is composed upon a crossed arc the R&E graph will be non-planar. Otherwise it will be planar. This principle is consistent with the result obtained by other investigators previously [21]. All planar R&E graphs can be composed using the generators presented in Section 2 except for *Gcross*.

5. Planarity testing algorithms

Based on the graphical criterion, *Gcross*, we modified the conditions in the repeat section of the *R&E-pattern-test* algorithm [9]. The "directly adjacent" in *R&E-pattern-test* is changed to "stacked adjacent" or "wrapped adjacent". For an arc (i, j) in the arc graph G, if there exists another arc (k, l) satisfies $k = i - 1$ and $l = j + 1$, or $k = i + 1$ and $l = j - 1$, we say (k, l) is stacked adjacent to (i, j). Similarly, for the arc (i, j) in the arc graph G, if there exists an arc (k, l) satisfies $k = i - 1$ and $l = i + 1$, or $k = j - 1$ and $l = j + 1$, we say (k, l) is wrapped adjacent to (i, j). With these assumptions, one can get an algorithm of *planar-R&E-graph-test*. It is also a linear time algorithm, the same as the *R&E-pattern-test* algorithm, since both of them use the same physical structure.

Besides determining an arc graph is planar or not, another advantage of the graphical criterion lies in determining the planarity change before and after one step composition just in a constant time. The arc graph is supposed to be stored as a doubly linked list, with a pointer to the other node of the same arc. A Boolean field is used to record that an arc is crossed or nested and another Boolean field to record that a node is the left or right of an arc. The test algorithm is shown as follows:

Algorithm 1. A test algorithm for the planar R&E arc graph on one step composition.

```
Algorithm planar-R&E-graph-test-one-composition
data structure:
```

left link	name of arc	left/right	crossed/nested	partner link	right link

```
input: arc graph $G$, arc $\alpha$ (in $G$), Gener, planar or non-planar (of $G$).
output: "yes" if $G'$ ($G' = G\bigcirc_\alpha$ Gener) is planar and "no" otherwise.
if $G$ is non-planar then return "no"
elseif $\alpha$ is nested, or $\alpha$ is crossed and Gener is not Gcross then return "yes";
maintain the data structure according to Gener (crossed/nested field is shown in Table 1)
else return "no".
```

Table 1
Variations of the crossed/nested field of arcs on one step composition

	Gener	*Glc*, *Grc* *Gli*, *Gri*	*Gnest*	*Glw*, *Grw*	*Gcross*
before	after				
old arc 0	old arc	0	0	1	1
	new arc	0	0	1	1
old arc 1	old arc	1	1	1	1
	new arc	0	1	1	1

(0: nested, 1: crossed).

For an arbitrary recognizing planar arc graphs (or diagrams), the complexity of time is no more than the planar testing in general. Besides the general planar graph testing algorithms mentioned above, we present another quadratic time testing algorithm. Firstly, the crossed relation graph G^*, which is tightly related to the arc graph G, is defined as follows: each arc of G is represented as a node in G^*, and each intersection of two crossed arcs in G is represented by an arc in G^* that connects the two nodes corresponding to the two arcs in G. In the following, it is required to determine the crossed relation graph G^* is a bipartite graph or not, which has been widely investigated in graph theory [12,19]. G will be planar if G^* is a bipartite graph or else G will be non-planar. It will cost a quadratic time to obtain G^* from G in the most unfortunate case and the determination of a bipartite graph will cost a linear time. Therefore, this testing algorithm is quadratic. Since this testing algorithm is presented for the planar arc graphs in general, it is also suitable to test planar R&E arc graphs, special examples of planar arc graphs.

6. Conclusion

A graphical criterion for planarity of RNA secondary structures in R&E class is investigated. All of the R&E graphs are constructed by the method of graph composition using a few generators. It is demonstrated that the planarity of R&E graphs is related to the generator *Gcross*. The R&E graph is planar if *Gcross* is never composed on a crossed arc, and, otherwise, it is non-planar. Especially, all the planar R&E graphs can be composed by use of the generators except for *Gcross*. Several planar testing algorithms for R&E arc graphs were introduced based on this graphical criterion. It may be of help in designing the visual software that describes RNA secondary structures. In this paper, the planarity graphical criterion of RNA secondary structures in R&E class is studied. Moreover, the planarity of an arbitrary recognizing RNA secondary structure, especially those outside R&E class, is an interesting topic for further research.

References

[1] C. Dennis, The brave new world of RNA, Nature 418 (2002) 122–124.
[2] D.H. Mathews, D.H. Turner, Prediction of RNA secondary structure by free energy minimization, Curr. Opin. Struct. Biol. 16 (2006) 270–278.
[3] K. Han, D. Kim, H.-J. Kim, A vector-based method for drawing RNA secondary structure, Bioinformatics 15 (1999) 286–297.
[4] A. Liebers, Planarizing graphs — a survey and annotated bibliography, J. Graph Algorithms Appl. 5 (2001) 1–74.
[5] C. Witwer, I.L. Hofacker, P.F. Stadler, Prediction of consensus RNA secondary structures including pseudoknots, IEEE Trans. Comput. Biol. Bioinformatics 1 (2004) 66–77.
[6] K. Han, Y. Lee, W. Kim, Pseudoviewer: Automatic visualization of RNA pseudoknots, Bioinformatics 18 (2002) S321–S328.
[7] E. Rivas, S.R. Eddy, A dynamic programming algorithm for RNA structure prediction including pseudoknots, J. Mol. Biol. 285 (1999) 2053–2068.
[8] E. Rivas, S.R. Eddy, The language of RNA: A formal grammar that includes pseudoknots, Bioinformatics 16 (2000) 334–340.
[9] A. Condon, B. Davy, B. Rastegari, S. Zhao, F. Tarrant, Classifying RNA pseudoknotted structures, Theor. Comput. Sci. 320 (2004) 35–50.
[10] J. Hopcroft, R.E. Tarjan, Efficient planarity testing, J. Assoc. Comp. Mach. 21 (1974) 549–568.
[11] K.S. Booth, G.S. Lueaker, Tsting for consecutive one property, interval graphs, and graph planarity using PQ-tree algorithms, J. Comp. Syst. Sci. 13 (1976) 335–379.
[12] D.B. West, Introduction to Graph Theory, second ed., Prentice-Hall, 2004.
[13] W. Klotz, A constructive proof of Kuratowski's theorem, Ars Combin. 28 (1989) 51–54.
[14] R.B. Lyngsø, C.S. Pedersen, RNA pseudoknot prediction in energy based models, J. Comput. Biol. 7 (2000) 409–427.
[15] M. Brinkmeier, Pseudoknotted secondary RNA structure, Scientific Literature Digital Library, available from: http://citeseer.ist.psu.edu.
[16] V. Moulton, M. Zuker, M. Steel, R. Pointon, D. Penny, Metrics on RNA secondary structures, J. Comput. Biol. 7 (2000) 277–292.

[17] B. Rastegari, Linear time algorithm for parsing RNA secondary structure, Thesis of University of British Columbia, 2004.

[18] Y. Uemura, A. Hasegawa, S. Kobayashi, T. Yokomori, Tree adjoining grammars for RNA structure prediction, Theor. Comput. Sci. 210 (1999) 277–303.

[19] J.A. Bondy, U.S.R. Murty, Graph Theory with Applications, Macmillan, London, 1976.

[20] M. Brinkmeier, Structural alignments of pseudo-knotted RAN-molecules in polynomial time, Scientific Literature Digital Library, available from: http://citeseer.ist.psu.edu.

[21] S. Ieong, M.-Y. Kao, T.-W. Lam, W.-K. Sung, S.-M. Yiu, Predicting RNA secondary structures with arbitrary pseudoknots by maximizing the number of stacking pairs, J. Comput. Biol. 10 (2003) 981–995.