Journal of Mathematical Chemistry Vol. 37, No. 4, May 2005 (© 2005)
DOI: 10.1007/s10910-004-1099-7

The adjoints of DNA graphs*

Jianxiu Hao
Department of Mathematics, Zhejiang Normal University, Jinhua, Zhejiang 321004, P.R. China
sx35@mail.zjnu.net.cn

Received 13 February 2004; revised 29 June 2004

In order to read a DNA sequence, we propose a method which induces the concept of DNA graph. In this paper, by discussing the adjoints of DNA graphs, we obtain more DNA graphs from known DNA graphs.

KEY WORDS: DNA, DNA graph, recognition, adjoint

AMS subject classification: 92D20

1. Introduction

Since Watson and Crick [1] proposed the helical structure of DNA, many problems about this structure are posed. An important problem is how to read a DNA sequence, one method is hybridization and reconstruction. All short fragments of nucleic acids (oligonucleotides) of length l (a library composed of 4^l subchains) are used in the hybridization experiment and thus, the formation of the complex indicates the occurence of a sequence complementary to the oligonucleotides in the DNA chain. It is detected by a nuclear or spectroscopic detector. As a result of the experiment one gets a set (called spectrum) of all l-long oligonucleotides which are used to hybridize with the investigating DNA sequence of length n.

Now raising a new problem: in what order to reconstruct these fragments? Lysov et al. [2] proposed a method which is to formulate the problem of finding original DNA sequence as looking for a Hamiltonian path. Later, Pevzner et al. [3] refined this problem as looking for a Eulerian trail. The above approach raised some interesting questions in graph theory. They are concerned with the labeling graph which will be referred to as DNA graph. About previous results please see [2–5].

However, when we use obligonucleotides to hybridize with the investigating DNA sequence, we may make mistakes. How to check whether the operation of

*The Project Supported by Zhejiang Provincial Natural Science Foundation of China (102055) and The Foundation of Zhejiang Universities' Youth Teachers.

the experiment is proper? We transfer this problem to the conception of graph theory as follows: how to identify a graph being a DNA graph ? In this paper, by discussing the adjoint of DNA graph, we obtain more DNA graphs from known DNA graphs. In this way, we obtain more checking methods. The organization of the paper is as follows. In section 2, for convenience, we present some definitions. Section 3 is devoted to the main results.

2. Preliminaries

In this paper, when we say a directed graph, we means a directed 1-graph without directed loops. Because the purpose of our study is to look for a sequence of A, T, G, C, we consider directed loops and p-graphs ($p \geqslant 2$) is valueless for the arrangement of a DNA sequence.

Definition 2.1. [6] A directed graph is a p-graph if given any ordered pair x, y of vertices (x possibly equal to y), there are at most p parallel arcs from x to y.

Definition 2.2. [4] Let $k > 1$ and $\alpha > 0$ be two integers. We say that a directed graph $D = (V, A)$ can be (α, k) – labeled if it is possible to assign a label $(l_1(x), \ldots, l_k(x))$ of length k to each vertex x of D such that

1. $l_i(x) \in \{1, \ldots, \alpha\}, \forall i \forall x \in V$;
2. all labels are different, that is $(l_1(x), \ldots, l_k(x)) \neq (l_1(y), \ldots, l_k(y))$ if $x \neq y$;
3. If $x \neq y$, $(x, y) \in A \Leftrightarrow (l_2(x), \ldots, l_k(x)) = (l_1(y), \ldots, l_{k-1}(y))$.

Definition 2.3. [4]. Given two integers $k > 1$ and $\alpha > 0$, L_k^α is the class of directed graphs that can be (α, k) – labeled.

Definition 2.4. [4]. A directed graph D is a DNA graph if and only if $\exists k > 1$ such that $D \in L_k^4$.

Definition 2.5. [4]. The adjoint $D' = (V, U)$ of a graph $D = (X, V)$ is the directed graph with vertex set V and such that there is an arc from a vertex x to a vertex y in D' if and only if the terminal endpoint of the arc x in D is the initial endpoint of arc y in D.

A graph D' is an adjoint if there exists some graph D such that D' is the adjoint of D.

Definition 2.6. Let D' be the adjoint of D. If D is isomorphic to D', we call D self-adjoint.

Definition 2.7. Define $S_{v_j,m,i}$ as follows:

$$V(S_{v_j,m,i}) = \{v_j, x_1, x_2, \ldots, x_m, y_1, y_2, \ldots, y_i\},$$
$$A(S_{v_j,m,i}) = \{(x_1, v_j), (x_2, v_j), \ldots, (x_m, v_j), (v_j, y_1), (v_j, y_2), \ldots, (v_j, y_i)\},$$

where $m \geqslant 0, \ i \geqslant 0$.

Let C_n be a directed n-cycle:

$$V(C_n) = \{v_1, v_2, \ldots, v_n\},$$
$$A(C_n) = \{(v_1, v_2), (v_2, v_3), \ldots, (v_{n-1}, v_n), (v_n, v_1)\},$$

where $n \geqslant 2$. Then, we paste $S_{v_j,m,0}$ to C_n such that $V(S_{v_j,m,0}) \cap V(C_n) = \{v_j\}$, where $j = 1, 2, \ldots, n$, we obtain a directed graph, denoted $C(v_1, v_2, \ldots, v_n)$.

For example, one of $C(v_1, v_2, v_3)$ is as follows:

$$V(C(v_1, v_2, v_3)) = \{v_1, v_2, v_3, x_{11}, x_{12}, x_{21}, x_{22}, x_{31}\},$$
$$A(C(v_1, v_2, v_3)) = \{(v_1, v_2), (v_2, v_3), (v_3, v_1), (x_{11}, v_1), (x_{12}, v_1), (x_{21}, v_2), (x_{22}, v_2), (x_{31}, v_3)\}.$$

Define A_i as follows:

$A_0 = C_n$, a directed n-cycle; $A_1 = C(v_1, v_2, \ldots, v_n)$; defined above. Suppose A_i is constructed, where $i \geqslant 1$, we construct A_{i+1} as follows:

For every vertex $u \in V(A_i)$ with $d^+_{A_i}(u)=1$ and $d^-_{A_i}(u) = 0$, we paste $S_{u,m,0}$ at vertex u, $\forall u \in V(A_i)$, $m \geqslant 0$, for different $u \in A_i$, m may be different, $V(S_{u,m,0}) \cap V(A_i) = \{u\}$.

Definition 2.8. [7]. The converse $\overleftarrow{D}$ of D is the directed graph obtained from D by reversing the orientation of each arc.

The main results of this paper are as follows:

Theorem 3.1. Let $\{D\}$ denote the set of DNA graphs and $\{D'\}$ denote the set of adjoints of DNA graphs in $\{D\}$, then $\{D'\}$ is a proper subset of $\{D\}$.

Theorem 3.7. The connected self–adjoints are A_n and $\overleftarrow{A}_n$, where A_n is defined in definition 2.7, $\overleftarrow{A}_n$ is defined in definition 2.8.

3. Results

Theorem 3.1. Let $\{D\}$ denote the set of DNA graphs and $\{D'\}$ denote the set of adjoints of DNA graphs in $\{D\}$, then $\{D'\}$ is a proper subset of $\{D\}$.

Proof. Let D be a DNA graph and $D^{'}$ its adjoint. Suppose $u, v \in V(D)$ and $a = (u, v) \in A(D)$. Since D is a DNA graph, by definitions 2.2 and 2.4, we have

$$l(u) = (l_1(u), l_2(u), \dots, l_k(u)),$$
$$l(v) = (l_2(u), \dots, l_k(u), l_k(v)),$$

we define the label of $a \in V(D')$ as follows:

$$l(a) = (l_1(u), l_2(u), \dots, l_k(u), l_k(v)).(*).$$

In the following, we prove that by (*) we give D' a proper $(4, k+1)$-labeling, hence, D' is also a DNA graph.

Let $x, y, u, v \in V(D)$, $x \neq v$, $a_1 = (x, y) \in A(D)$ and $a_2 = (u, v) \in A(D)$. Since $x \neq v$, we have $l(a_1) \neq l(a_2)$.

Case 1. $y = u$. Let $l(x) = (l_1(x), l_2(x), \dots, l_k(x))$, by definition 2.2 we have

$$l(y) = (l_2(x), \dots, l_k(x), l_k(y)),$$
$$l(v) = (l_3(x), \dots, l_k(x), l_k(y), l_k(v)).$$

By (*), we obtain the labels of $a_1 \in V(D')$ and $a_2 \in V(D')$ as follows:

$$l(a_1) = (l_1(x), l_2(x), \dots, l_k(x), l_k(y)),$$
$$l(a_2) = (l_2(x), \dots, l_k(x), l_k(y), l_k(v)).$$

By Definition 2.2 we have $(a_1, a_2) \in A(D')$, which is what definition 2.5 requested.

If $(a_2, a_1) \in A(D')$, by definition 2.2 we have

$$(l_3(x), \dots, l_k(x), l_k(y), l_k(v)) = (l_1(x), l_2(x), \dots, l_k(x)),$$

thus, $l(v) = l(x)$, which contradicts with definition 2.2. Thus, (a_2, a_1) is not an arc of D', which is what definition 2.5 requested.

Case 2. $y \neq u$.

Similar as case 1, we obtain

$$l(x) = (l_1(x), l_2(x), \dots, l_k(x)),$$
$$l(y) = (l_2(x), \dots, l_k(x), l_k(y)),$$
$$l(u) = (l_1(u), l_2(u), \dots, l_k(u)),$$
$$l(v) = (l_2(u), \dots, l_k(u), l_k(v)),$$
$$l(a_1) = (l_1(x), l_2(x), \dots, l_k(x), l_k(y)),$$
$$l(a_2) = (l_1(u), l_2(u), \dots, l_k(u), l_k(v)),$$

where $y, u \in V(D)$, $a_1, a_2 \in V(D')$.

Since $y \neq u$, by definition 2.2 we have

$$(l_2(x), \dots, l_k(x), l_k(y)) \neq (l_1(u), l_2(u), \dots, l_k(u)).$$

By definition 2.2 there is no arc from a_1 to a_2 in D', which is what definition 2.5 requested. Thus, D' is a DNA graph.

Define W' as follows:

$$V(W') = \{x, y, z, u\},$$
$$A(W') = \{(x, y), (y, z), (z, u), (u, x), (x, z)\}.$$

We label the vertices of W' as follows:

$$l(x) = (2, 1, 1), l(y) = (1, 1, 1),$$
$$l(z) = (1, 1, 2), l(u) = (1, 2, 1).$$

Thus, W' is a DNA graph.

Claim. There is no DNA graph whose adjoint is W'.

By contradiction. Suppose W' is the adjoint of DNA graph W. Because there are 4 vertices in W', by definition 2.5 there are exactly 4 arcs in W. Let $A(W) = \{a_1, a_2, a_3, a_4\}$. Because the directed triangle xzu is contained in W', by definition 2.5 we have

$$a_1 = (v_1, v_2),\ a_2 = (v_2, v_3),\ a_3 = (v_3, v_1).$$

By the symmetry of v_1, v_2, v_3, we have four cases to consider:

Case 3. $a_4 = (v_4, v_5)$.

By definition 2.5 a_4 corresponds to an isolated vertex in W', which is a contradiction.

Similarly, we can prove that cases 4–6 are impossible.

Case 4. $a_4 = (v_4, v_1)$.

Case 5. $a_4 = (v_1, v_4)$.

Case 6. $a_4 = (v_2, v_1)$.

The theorem follows.

Remark. Theorem 3.1 provides a general method, using it we obtain more DNA graphs from known results, as demonstrated in theorems 3.3 and 3.4.

By theorem 3.1 the following corollary is obvious.

Corollary 3.2. Let D_{i+1} be the adjoint of D_i, where $i = 0, 1, 2, \ldots, n, n$ is an arbitrary integer number. If D_0 is a DNA graph, then D_n is a DNA graph.

Theorem 3.3. The adjoint of a directed path is a DNA graph.

Proof. **Claim.** P_n is a DNA graph, where P_n is a directed path with n vertices.

Let $P_n = (v_1, v_2, \ldots, v_n)$. We label vertex v_i as follows:

$$(1, 1, \ldots, 1, 2, 2, \ldots, 2),$$

where the number of 1 is $(3n + 1 - i)$ and the number of 2 is $(i - 1)$.

It is easy to see that all labels are different, and there exists an arc from v_i to v_{i+1}. We claim that there is no arc $(v_i, v_j) \in P_n$,where $|i-j| \geqslant 2$. In fact, without loss of generality, let $j \geqslant 2+i$. We have

$$(l_2(v_i), \ldots, l_{3n}(v_i)) = (1, 1, 1, \ldots, 1, 2, 2, \ldots, 2),$$

where the number of 1 is $(3n-i)$ and the number of 2 is $(i-1)$.

$$(l_1(v_j), \ldots, l_{3n-1}(v_j)) = (1, 1, \ldots, 1, 2, 2, \ldots, 2),$$

where the number of 1 is $(3n+1-j)$ and the number of 2 is $(j-2)$. Since $j \geqslant 2+i$, we have

$$3n-i > 3n+1-j.$$

By definition 2.2, there is no arc from v_i to v_j. By theorem 3.1 the theorem follows.

Theorem 3.4. The adjoint of a directed cycle is a DNA graph.

Proof. **Claim:** C_n is a DNA graph, where C_n is a directed cycle.

In this proof, if $i \equiv 0 \pmod 4$, we denote $i \equiv 4 \pmod 4$. Define $C_n = (v_1, v_2, \ldots, v_n)$.

Case 1. Suppose $4|n$, let $n = 4m$. We label the vertices of C_n as follows:

Step 1: Define

$$(l_1(v_1), l_2(v_1), \ldots, l_n(v_1)) = (1, 1, \ldots, 1, 2, 2, \ldots, 2, 3, 3, \ldots, 3, 4, 4, \ldots, 4),$$

where the number of k is m, $k = 1, 2, 3, 4$.

Step 2: Suppose that the label of v_i is $(l_1(v_i), l_2(v_i), \ldots, l_n(v_i))$, define the label of v_{i+1} as follows:

$$(l_1(v_{i+1}), l_2(v_{i+1}), \ldots, l_{n-1}(v_{i+1}), l_n(v_{i+1})) = (l_2(v_i), l_3(v_i), \ldots, l_n(v_i), l_1(v_i)).$$

By this definition we have $l_j(v_i) \in \{1, 2, 3, 4\}$, where $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, n$. Obviously, all labels are different, there exists an arc from v_i to v_{i+1}, where $i = 1, 2, \ldots, n-1$. Since

$$(l_1(v_n), l_2(v_n), \ldots, l_n(v_n)) = (4, 1, 1, \ldots, 1, 2, 2, \ldots, 2, 3, 3, \ldots, 3, 4, 4, \ldots, 4),$$
$$(l_1(v_1), l_2(v_1), \ldots, l_n(v_1)) = (1, 1, \ldots, 1, 2, 2, \ldots, 2, 3, 3, \ldots, 3, 4, 4, \ldots, 4),$$

we know that there is an arc from v_n to v_1.

Suppose there exists an arc from v_i to v_j. Let the label of v_i be

$$(l_1(v_i), l_2(v_i), \ldots, l_n(v_i)).$$

Thus, the label of v_j is

$$(l_2(v_i), l_3(v_i), \ldots, l_n(v_i), l_n(v_j)).$$

By the symmetry of 1, 2, 3 and 4, without loss of generality, let $l_1(v_i) = 1$. By step 1 and step 2 we know that in every label of vertex the number of k is m, thus, $l_n(v_j) = 1 = l_1(v_i)$. Since we have proved that all the labels are different, and the label of v_{i+1} is

$$(l_2(v_i), l_3(v_i), \ldots, l_n(v_i), l_1(v_i)),$$

we know that $j = i+1$. Thus, there exists an arc from v_i to v_{i+1} only. Therefore, C_n is a DNA graph, where $n = 4m$.

Case 2. Suppose $n \not\equiv 0 \pmod 4$. Define the label of v_i as follows:

$$l_j(v_i) = \begin{cases} i + j \pmod 4, & \text{if} \quad i + j \leqslant n, \\ i + j - n \pmod 4, & \text{if} \quad i + j > n, \end{cases}$$

where $i = 1, 2, \ldots, n$; $j = 1, 2, \ldots, n$.

It is easy to see that $l_j(v_i) \in \{1, 2, 3, 4\}$, where $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, n$.

Assume that there exist two vertices v_i and v_{i+p} such that

$$(l_1(v_i), l_2(v_i), \ldots, l_n(v_i)) = (l_1(v_{i+p}), l_2(v_{i+p}), \ldots, l_n(v_{i+p})),$$

where $1 \leqslant p \leqslant n - i$. Thus, $l_n(v_i) = l_n(v_{i+p}) \pmod 4$, that is, $i \equiv i + p \pmod 4$, we have $4|p$.

Similarly, $l_{n+1-p-i}(v_{i+p}) = l_{n+1-p-i}(v_i) \pmod 4$, that is, $1 \equiv n + 1 - p$ (mod 4). Because we have proved that $4|p$, we have $4|n$, which contradicts with $n \not\equiv 0$ (mod 4). Hence, all labels are different.

By the definition of $l_j(v_i)$, we have

$$(l_2(v_i), l_3(v_i), \ldots, l_n(v_i)) = (l_1(v_{i+1}), l_2(v_{i+1}), \ldots, l_{n-1}(v_{i+1})).$$

Hence, there exists an arc from v_i to v_{i+1}, where $1 \leqslant i \leqslant n - 1$.

Similarly, the label of v_n is as follows:

$$(l_1(v_n), l_2(v_n), \ldots, l_t(v_n), \ldots, l_n(v_n)) = (1, 2, \ldots, t \pmod 4, \ldots, n \pmod 4)),$$

the label of v_1 is as follows:

$$(l_1(v_1), l_2(v_1), \ldots, l_{t-1}(v_1), \ldots, l_{n-1}(v_1), l_n(v_1)) = (2, 3, \ldots, t \pmod 4, \ldots, n \pmod 4), 1).$$

By definition 2.2 there is an arc from v_n to v_1.

Suppose there is an arc from v_i to v_j, let the label of v_i be $(l_1(v_i), l_2(v_i), \ldots, l_n(v_i))$. By definition 2.2 the label of v_j is as follows: $(l_2(v_i), \ldots, l_n(v_i), l_n(v_j))$. By mathematical induction and the definition of $l_j(v_i)$ defined above, we know that the number of i's in every label of vertex is constant, where $i = 1, 2, 3, 4$. Thus, $l_n(v_j) = l_1(v_i)$. Therefore, the label of v_j is as follows:

$$(l_2(v_i), \ldots, l_n(v_i), l_1(v_i)),$$

which is the label of v_{i+1}. Because we have proved that all labels are different, we obtain $j = i + 1$. Hence, C_n is a DNA graph. By theorem 3.1 the theorem follows.

The following theorem is clear.

Theorem 3.5. Let D be a directed graph, $a = (u, v) \in A(D)$, D' be the adjoint of D, we have

$$d_D^+(v) = d_{D'}^+(a), d_D^-(u) = d_{D'}^-(a).$$

By Theorem 3.5 we have the following corollary.

Corollary 3.6. Let D' be the adjoint of D, we have

$$\Delta^+(D) \geqslant \Delta^+(D'), \Delta^-(D) \geqslant \Delta^-(D'),$$
$$\delta^+(D) \leqslant \delta^+(D'), \delta^-(D) \leqslant \delta^-(D'),$$

where $\Delta^+(D)$ stands for the maximum outdegree of D, $\Delta^-(D)$ stands for the maximum indegree of D, $\delta^+(D)$ stands for the minimum outdegree of D, $\delta^-(D)$ stands for the minimum indegree of D. Similar meanings for D'.

Theorem 3.7. The connected self-adjoints are A_n and $\overleftarrow{A}_n$, where A_n is defined in definition 2.7, $\overleftarrow{A}_n$ is defined in definition 2.8.

Proof. **Claim 1.** If D is self-adjoint, we have

$$|V(D)| = |A(D)|.$$

In fact, let D' be the adjoint of D, we have

$$|V(D')| = |A(D)|.$$

Because D' is isomorphic to D, we have

$$|V(D')| = |V(D)|.$$

Hence,

$$|V(D)| = |A(D)|.$$

The claim holds.

Since D is connected, we know that the underlying graph of D contains a spanning tree T. Because $|E(T)| = |V(D)| - 1$, we have

$$|E(T)| = |A(D)| - 1.$$

Thus, the underlying graph of D contains a unique cycle $v_1v_2, \ldots, v_p$ with subtrees $T(v_1), T(v_2), \ldots, T(v_p)$, where $T(v_j)$ is a subtree rooted at v_j and

$$V(T(v_j)) \cap \{v_1, v_2, \ldots, v_p\} = \{v_j\},$$

where $j = 1, 2, \ldots, p$.

Claim 2. Let $T'(v_j)$ be the adjoint of $T(v_j)$, then the underlying graph of $T'(v_j)$ contains no cycle.

Otherwise, suppose there exists a cycle $C \in T^{'}(v_j)$, for every vertex $a \in C$, by Definition 2.5 vertex a corresponds to an arc $(u, v) \in A(T(v_j))$, in this way along the cycle $C \in T^{'}(v_j)$ we find a cycle in the underlying graph of $T(v_j)$, which is a contradiction.

Claim 3. D contains a unique directed cycle $(v_1, v_2, \ldots, v_p)$.

By contradiction. Suppose the underlying graph of D contains a unique cycle $v_1v_2, \ldots, v_p$ which is not a directed cycle. Without loss of generality, let $a_1 = (v_1, v_2)$, $a_2 = (v_3, v_2)$. By definition 2.5 we have

$$(a_1, a_2) \notin A(D'), \;\; (a_2, a_1) \notin A(D').$$

Hence, there is no cycle in the underlying graph of $D^{'}$ formed by the arcs in the unique cycle $v_1v_2, \ldots, v_p$ of D.

Because D' is isomorphic to D and D contains a unique cycle, then D' must contain a unique cycle. By claim 2 and the above analysis, we know that the unique cycle in D' must be formed by some arcs from some $T(v_j)$s and some arcs of the unique cycle $v_1v_2, \ldots, v_p \in D$. Let C be the unique cycle $v_1v_2, \ldots, v_p$ in the underlying graph of D.

Case 1. Suppose D contains $a_i = (v_i, v_{i+1}) \in C$, $a_{i-1} = (v_{i-1}, v_i) \in C$, $b_1 = (x_1, v_i) \in T(v_i)$, $b_2 = (v_i, x_2) \in T(v_i)$. By definition 2.5 we have

$$(b_1, a_i) \in D', \; (b_1, b_2) \in D', \; (a_{i-1}, b_2) \in D', \; (a_{i-1}, a_i) \in D'.$$

Hence, the underlying graph of D' contains a unique cycle $b_1a_ia_{i-1}b_2$. Since D' is isomorphic to D, we know that 4-cycle $b_1a_ia_{i-1}b_2 \in D'$ must be isomorphic to the unique cycle $C \in D$. Because $C \in D$ contains a directed 3-path (v_{i-1}, v_i, v_{i+1}), 4-cycle $b_1a_ia_{i-1}b_2 \in D'$ contains no directed 3-path, which is a contradiction.

Case 2. Suppose there exist $a_i = (v_i, v_{i+1}) \in C$, $a_{i-1} = (v_i, v_{i-1}) \in C$, $b_1 = (x_1, v_i) \in T(v_i)$, $b_2 = (v_i, x_2) \in T(v_i)$, then the underlying graph of D' contains no cycle formed by a_i, a_{i-1}, b_1 and b_2.

If there were $a_i = (v_i, v_{i+1}) \in C$, $a_{i-1} = (v_i, v_{i-1}) \in C$, $b_1 = (x_1, v_i) \in T(v_i)$, $b_2 = (v_i, x_2) \in T(v_i)$, $b_3 = (x_3, v_i) \in T(v_i)$, by definition 2.5 there would be at least two cycles in the underlying graph of D': $b_3a_ib_1a_{i-1}$ and $b_3a_ib_1b_2$. Because the underlying graph of D contains exactly one cycle, we know that D' cannot be isomorphic to D, which is a contradiction.

If there were $a_i = (v_i, v_{i+1}) \in C$, $a_{i-1} = (v_i, v_{i-1}) \in C$, $b_1 = (x_1, v_i) \in T(v_i)$, $b_2 = (v_i, x_2) \in T(v_i)$, $b_4 = (v_i, x_4) \in T(v_i)$, we can discuss similarly.

Case 3. Suppose $m = 0$ or $k = 0$ for every $S_{v_i,m,k}$, where $S_{v_i,m,k}$ is defined in definition 2.7.

Because the cycle C of D is not a directed cycle, we know that the unique cycle in the underlying graph of D' must contain some vertices formed from arcs of $T(v_i) \in D$ and all arcs of $C \in D$. Thus, the cycle in the underlying graph of D' would be longer than the cycle in the underlying graph of D, which is a contradiction. Thus, D' can not be isomorphic to D, which is a contradiction. claim 3 holds.

By claim 3 we denote $a_1 = (v_1, v_2)$, $a_2 = (v_2, v_3), \dots, a_j = (v_j, v_{j+1}), \dots, a_p = (v_p, v_1)$. By the definition of subtree $T(v_j)$ in this proof, we know that $T(v_j)$ is constructed recursively as follows:

Step 1: $S_{v_j,m_j,i_j} \subseteq T(v_j)$, where S_{v_j,m_j,i_j} is defined in definition 2.7, $m_j \geqslant 0$, $i_j \geqslant 0$, v_j is the vertex of directed cycle $(v_1, v_2, \dots, v_p)$.

In fact, there must be $m_j = 0$ or $i_j = 0$. Otherwise, suppose $m_j \geqslant 1$ and $i_j \geqslant 1$. Without loss of generality, let $a_{j1j} = (x_{j1}, v_j) \in T(v_j)$ and $a_{jj1} = (v_j, y_{j1}) \in T(v_j)$. By definition 2.5 we know that there are two cycles in the underlying graph of D': $a_1a_2, \dots, a_p$ and $a_{j1j}a_ja_{j-1}a_{jj1}$. By claim 3 we know that D cannot be isomorphic to D', which is a contradiction.

Step 2: For every $u \in T(v_j)$, if $d^+(u) = 1$ and $d^-(u) = 0$ or $d^-(u) = 1$ and $d^+(u) = 0$, we paste $S_{u,m,i}$ at u of $T(v_j)$, where $m \geqslant 0$, $i \geqslant 0$, $V(S_{u,m,i}) \cap V(D) = \{u\}$.

Claim 4. In step 2, if $d^-(u) = 1$ with $d^+(u) = 0$, and $S_{u,m,i}$ is pasted to $T(v_j)$ at u, $i \neq 1$, we have $m = 0$.

Case 1. Suppose $i = 0$. If $m \geqslant 1$, then the underlying graph of D' is disconnected, which is a contradiction. Hence, $m = 0$.

Case 2. Suppose $i \geqslant 2$. If $m \geqslant 1$, there exist $b_1 = (x, u) \in S_{u,m,i}$, $b_2 = (u, y_1) \in S_{u,m,i}$, $b_3 = (u, y_2) \in S_{u,m,i}$. Since $d^-(u) = 1$, there exists $b = (v, u) \in T(v_j)$. Thus, in the underlying graph of D' there are two cycles: $bb_2b_1b_3$ and $a_1a_2, \dots, a_p$, which contradicts with claim 3. Hence, $m = 0$. Claim 4 holds.

Similarly, we have

Claim 5. In step 2, if $d^+(u) = 1$ with $d^-(u) = 0$, and $S_{u,m,i}$ is pasted to $T(v_j)$ at u, $m \neq 1$, we have $i = 0$.

Define

$$d_D(S_{u,m,i} \to u) = \sum_{v \in S_{u,m,i}} d_D(u, v) = m + i,$$

$$d_D(T(v_j) \to C) = \sum_{v \in T(v_j)} d_D(v, v_j),$$

$$d_D(T \to C) = \sum_{v_j \in C} d_D(T(v_j), C),$$

where C is the unique directed cycle $(v_1, v_2, \dots, v_p) \in D$, $C \cap T(v_j) = \{v_j\}$, $d_D(x, y)$ is the distance between vertices x and y in the underlying graph of D.

Similarly, we define $d_{D'}(S'_{a,m,i} \to a)$, $d_{D'}(T(a_j) \to C')$ and $d_{D'}(T' \to C')$, where C' is the unique cycle $(a_1, a_2, \dots, a_p)$ in D', $T' = D' - \{(a_1, a_2), (a_2, a_3), \dots, (a_p, a_1)\}$.

Claim 6. If D contains one of the following cases: (1). $d^-(u) = 1$ with $d^+(u) = 0$, and $S_{u,m,i} \subseteq T(v_j)$, where $i = 1$, $m \geqslant 1$. (2). $d^+(u) = 1$ with $d^-(u) = 0$, and $S_{u,m,i} \subseteq T(v_j)$, where $m = 1$, $i \geqslant 1$, we have

$$d_D(T \to C) < d_{D'}(T' \to C').$$

In fact, let $V(S_{u,m,i}) = \{u, x_1, x_2, \dots, x_m, y_1, y_2, \dots, y_i\}$, $A(S_{u,m,i}) = \{a_{x1}, a_{x2}, \dots, a_{xm}, a_{y1}, a_{y2}, \dots, a_{yi}\}$, where $a_{xk} = (x_k, u)$, $a_{yh} = (u, y_h)$, $1 \leqslant k \leqslant m$, $1 \leqslant h \leqslant i$.

Suppose $d^-(u) = 1$ with $d^+(u) = 0$, and $S_{u,m,i} \subseteq T(v_j)$. Because $d^-(u) = 1$, there exists an arc $a = (v, u) \in D$. Let $S_{u,m,i} \in D$ correspond to $S'_{a,m,i} \in D'$:

$$\begin{aligned} V(S'_{a,m,i}) &= \{a, a_{x1}, \dots, a_{xm}, a_{y1}, \dots, a_{yi}\}, \\ A(S'_{a,m,i}) &= \{(a, a_{y1}), \dots, (a, a_{yi}), (a_{x1}, a_{y1}) \\ &\qquad , \dots, (a_{x1}, a_{yi}), \dots, (a_{xm}, a_{y1}), \dots, (a_{xm}, a_{yi})\} \end{aligned}$$

When $i \neq 1$, by claim 4 we have $m = 0$. It is easy to see that

$$d_D(S_{u,m,i} \to u) = d_{D'}(S'_{u,m,i} \to a).$$

When $i = 1$, it is easy to see that

$$d_D(S_{u,m,i} \to u) < d_{D'}(S'_{a,m,i} \to a).$$

Similarly, suppose $d^+(u) = 1$ with $d^-(u) = 0$, and $S_{u,m,i} \subseteq T(v_j)$.

When $m \neq 1$, we have

$$d_D(S_{u,m,i} \to u) = d_{D'}(S'_{a,m,i} \to a).$$

When $m = 1$, we have

$$d_D(S_{u,m,i} \to u) < d_{D'}(S'_{a,m,i} \to a).$$

Because we have proved that $T(v_j)$ is constructed by $S_{u,m,i}$ recursively in steps 1 and 2, T is composed by $T(v_j)$ adhered to a cycle $C = (v_1, v_2, \dots, v_p)$, it follows that

$$d_D(T \to C) < d_{D'}(T' \to C').$$

Hence, claim 6 holds.

By claims 4–6 we know that $T(v_j)$ is consisted either by some $S_{u,m,0}$ completely or by some $S_{u,0,i}$ completely. Thus, if we know the structure of S_{v_j,m_j,i_j}, where v_j is a vertex of cycle $C = (v_1, v_2, \dots, v_p)$, the structure of $T(v_j)$ is known. Hence, for simplicity, in the following we assume that D is composed as follows:

(1) A directed cycle $(v_1, v_2, \dots, v_p)$.

(2) $T(v_j) = S_{v_j,m_j,i_j}$, where $j = 1, 2, \dots, p$.

Claim 7. For every S_{v_j,m_j,i_j}, we have $m_j = 0$ or $i_j = 0$, where $j = 1, 2, \dots, p$. In fact, we have proved this claim in step 1 following claim 3.

By claim 7 we know that D is composed as follows:

A directed cycle $C_p = (v_1, v_2, \dots, v_p)$ patched at vertex v_j with $T(v_j)$, $j = 1, 2, \dots, p$, where $T(v_j) = S_{v_j,m_j,0}$ or $T(v_j) = S_{v_j,0,i_j}$. Denote $a_1 = (v_1, v_2)$, $a_2 = (v_2, v_3), \dots, a_p = (v_p, v_1)$. Let $T(v_k) \in D$, define the distance between $T(v_k)$ and $T(v_j)$ as follows:

$$d(T(v_k), T(v_j); D) = \min\{|k-j|, p-|k-j|\}.$$

Then,

$$d(T(v_k), D) = \sum_{v_j \in C_p} d(T(v_k), T(v_j); D).$$

$$d(D) = \sum_{v_k \in C_p} d(T(v_k), D).$$

Similarly, we define $d(T(a_k), T(a_j); D')$, $d(T(a_k), D')$ and $d(D')$. Because $T(v_j) = S_{v_j,m_j,0} \in D$ is isomorphic to $T(a_i) \in D'$, for convenience, we use $T(v_j)$ to replace $T(a_i) \in D'$. Similarly, when $i_j \neq 0$ we use $T(v_j) = S_{v_j,0,i_j} \in D$ to replace $T(a_{i-1}) \in D'$; when $i_j = 0$ we use $T(v_j) = S_{v_j,0,i_j} \in D$ to replace $T(a_i) \in D'$.

By definition 2.5, the following claim is clear.

Claim 8.

(1). If $T(v_k) = S_{v_k,m_k,0}$, $T(v_j) = S_{v_j,0,i_j}$, $i_j \neq 0$, then,

$$d(T(v_k), T(v_j); D) = \begin{cases} d(T(v_k), T(v_j); D'), & \text{if } 2|k-j| = p+1, \\ d(T(v_k), T(v_j); D') + 1, & \text{otherwise.} \end{cases}$$

(2). If $T(v_k) = S_{v_k,m_k,0}$, $T(v_j) = S_{v_j,m_j,0}$, then

$$d(T(v_k), T(v_j); D) = d(T(v_k), T(v_j); D').$$

(3). If $T(v_k) = S_{v_k,0,i_k}$, $T(v_j) = S_{v_j,0,i_j}$, then

$$d(T(v_k), T(v_j); D) = d(T(v_k), T(v_j); D').$$

By claim 8 the following claim is obvious.

Claim 9. If there are $T(v_l) = S_{v_l,m_l,0} \in D$ and $T(v_j) = S_{v_j,0,i_j} \in D$, where $m_l \geqslant 1$, $i_j \geqslant 1$, we have

$$d(T(v_k), D) \geqslant d(T(v_k), D').$$

Especially, for $k = |l - j| + \min\{l, j\} - 1$, we have

$$d(T(v_k), D) \geqslant d(T(v_k), D') + 1.$$

Claim 10. Let D' be the adjoint of D. If D is self-adjoint, then D must be A_n or $\overleftarrow{A}_n$.

Otherwise, suppose $T(v_l) = S_{v_l,m_l,0} \in D$ and $T(v_j) = S_{v_j,0,i_j} \in D$, where $m_l \geqslant 1$, $i_j \geqslant 1$. By claim 9 we have

$$d(D) > d(D').$$

Thus, D can not be isomorphic to D', which is a contradiction. Claim 10 holds.

Claim 11. A_n is self-adjoint, where A_n is defined in definition 2.7.

We use mathematical induction to prove this claim.

(1). Suppose $n=0$, $A_0 = C_p$. Let $V(C_p) = \{v_1, v_2, \dots, v_p\}$, $A(C_p) = \{a_1, a_2, \dots, a_p\}$, where $a_i = (v_i, v_{i+1})$, $i = 1, 2, \dots, p$.

We define two mappings from A_0 to A'_0 as follows:

$$\begin{aligned} f_0 \colon V(A_0) &\to V(A'_0), \\ v_i &\to a_i. \\ g_0 \colon A(A_0) &\to A(A'_0), \\ a_i &\to (a_i, a_{i+1}), \end{aligned}$$

where $i = 1, 2, \dots p$. It is easy to see that f_0 and g_0 are two bijections. Hence, A_0 is self-adjoint.

Similarly, we can prove that A_1 is self-adjoint.

(2). Suppose A_t is self$-$adjoint, where $t \geqslant 1$. That is, there exist two bijections $f_t \colon V(A_t) \to V(A'_t)$ and $g_t \colon A(A_t) \to A(A'_t)$.

Now we consider A_{t+1}. We define two mappings f_{t+1} and g_{t+1} as follows:

At first, let $f_{t+1}|_{A_t} = f_t$, $g_{t+1}|_{A_t} = g_t$.

Second, by definition 2.7 we paste $S_{u,m,0}$ to A_t for every vertex $u \in V(A_t)$ with $d^+_{A_t}(u) = 1$ and $d^-_{A_t}(u) = 0$. Let $a = (u, v) \in A(A_t)$, $V(S_{u,m,0}) = \{u, x_1, x_2, \dots, x_m\}$, $A(S_{u,m,0}) = \{a_{x_1u}, a_{x_2u}, \dots, a_{x_mu}\}$, where $a_{x_iu} = (x_i, u)$, $i = 1, 2, \dots, m$.

We define $f_{t+1}|_{S_{u,m,0}} : V(S_{u,m,0}) \to V(S'_{u,m,0})$ and $g_{t+1}|_{S_{u,m,0}} : A(S_{u,m,0}) \to A(S'_{u,m,0})$ as follows:

$$\begin{aligned} f_{t+1}|_{S_{u,m,0}} &\colon x_i \to a_{x_iu}, u \to a, \\ g_{t+1}|_{S_{u,m,0}} &\colon a_{x_iu} \to (a_{x_iu}, a), \end{aligned}$$

where $i = 1, 2, \dots, m$.

It is easy to see that f_{t+1} and g_{t+1} are two bijections from A_{t+1} to A'_{t+1}. Hence, A_n is self-adjoint.

Similarly, we can prove that $\overleftarrow{A_n}$ is self-adjoint. The theorem follows.

References

[1] J.D. Watson and F.H.C. Crick, Nature 173 (1953) 737–738.
[2] Yu.p. Lysov et al., Dokl. Acad. Sci. USSR 303 (1988) 1508–1511.
[3] P.A. Pevzner, J. Biomol. Struct. Dyn. 7 (1989) 63–73.
[4] J. Blazewicz, A. Hertz, D. Kobler and D. de Werra, Discrete Appl. Math. 98 (1999) 1–19.
[5] M.S. Waterman, Applications of Combinatorics to Molecular Biology, in: *Handbook of Combinatorics*, ed. R. Graham et al. (Elsevier Science, B. V., 1995).
[6] C. Berge, *Graphes* (Paris, Dunod, 1970).
[7] J.A. Bondy and U.S.R. Murty, *Graph Theory with Applications* (Macmillan Press Ltd., London, 1976).