A DNA sequence can be identified with a word over the alphabet $W = \{A, C, G, T\}$. He and Wang (2004) presented an algebraic method for analyzing DNA sequences. They proved that under a natural action of the symmetric group $S_4$ on DNA matrices, $\alpha \Pi = \Pi \alpha$, where $\alpha \in S_4$ and $\Pi$ is a DNA matrix constructed from a DNA sequence. In this paper, the general form of models, which are satisfied the mentioned equality, is obtained. It is shown that the Z-curve, H-curve and Randić’s model are special cases of this general form.

Key words: DNA sequence, DNA sequence matrix, Z-curve, H-curve.

INTRODUCTION

The extremely long DNA molecule is actually made of a long string of chemical building blocks called nucleotides. There are four different nucleotides, which are labeled adenine (A), thymine (T), guanine (G), and cytosine (C). A DNA sequence is a succession of letters A, C, G and T representing these four nucleotide bases of a DNA strand.

DNA computing is a form of computing which uses DNA, biochemistry and molecular biology, instead of the traditional silicon-based computer technologies. Generally, in DNA computing, the DNA sequences used for the computation should be critically designed in order to reduce error that could occur during computation. In this area of research, it is important to find a geometric method to represent a given DNA sequence. There are too many methods for mapping the symbolic characters A, T, C, G into numerical values. Some researchers used one-dimensional and others 2D, 3D or higher dimensional models to investigate DNA sequence. The data of some models are real, complex or even quaternion numbers [1-5]. Notice that by a model we consider a system of vectors with a unique 3D perception of our given DNA sequence.
For a positive integer \( n \), \( \Xi_n \) denotes the set of all DNA sequences of length \( n \) and \( \Xi = \bigcup_{n \in \mathbb{N}} \Xi_n \). Suppose \( \Sigma = \{A, C, G, T\} \), \( w = x_1x_2...x_n \) is a DNA sequence of length \( n \) and \( L_i(w) = |\{j \mid 1 \leq j \leq i \text{ and } x_j = L\}| \), where \( L \in \Sigma \). Therefore, for a given DNA sequence \( w \) of \( n \) bases, \( A_k, T_k, C_k \) and \( G_k \), \( 1 \leq k \leq n \), denote the number of bases \( A, T, C \) and \( G \), respectively.

The first visualization method for DNA sequences was introduced by Hamori and Ruskin named \( H\text{-Curve} \) [6]. In this method, the following system of vectors is considered:

\[
\begin{align*}
  x_i &= 2(T_i + A_i) - i \\
  y_i &= 2(A_i + G_i) - i \\
  z_i &= 2(T_i + G_i) - i
\end{align*}
\]  

In 1994, Zhang and Zhang proposed the \( Z\text{-curve} \) models for visualizing DNA sequences [7, 8]. Then the coordinates of \( Z\text{-curve} \) are determined by the following equalities:

\[
\begin{align*}
  x_i &= 2(A_i + G_i) - i \\
  y_i &= 2(A_i + C_i) - i \\
  z_i &= 2(A_i + T_i) - i
\end{align*}
\]  

Since 2000, Randić’s research group had proposed several visualization schemes for DNA sequences [9-16]. In one of his method, the four vertices associated with a regular tetrahedron are assigned to four nucleotides. The mapping between four nucleotides and corresponding 3D coordinates is shown below:

\[
\begin{align*}
  x_i &= 2(T_i + A_i) - i \\
  y_i &= 2(T_i + G_i) - i \\
  z_i &= 2(T_i + C_i) - i
\end{align*}
\]  

Throughout this paper our other notation is standard. They are appearing as in the same way as in the following [5, 17].

**RESULTS AND DISCUSSION**

He and Wang [17] considered the coordinates of a graphical representation of DNA sequence introduced by Zhang [7]. They presented an action of the symmetric group \( S_4 \) on \( Z\text{-curves} \) and the DNA matrices. The aim of this section is to prove all of these models is satisfied the Proposition 3.1. [17]. On the other hand, the general form of models satisfied the mentioned result is given. To explain, we consider a DNA sequence \( w \) and define its DNA sequence matrix \( \Pi(w) \) as follows:
Let $M_n$ denote the set of all DNA sequence matrices with $n$ columns and $\Lambda = \bigcup_{n \geq 1} M_n$. Clearly, $\Pi$ defined a one-to-one correspondence between $\Xi$ and $\Lambda$. For every $v \in \Xi_n$ and $w \in \Xi_m$, one can see that $vw \in \Xi_{m+n}$, where $vw$ is a DNA sequence constructed by $v$ and $w$ by juxtaposition. For the sake of completeness, we state below the Proposition 3.1 of the paper [17] which is crucial throughout the paper.

**Theorem 1.** Suppose $\sigma \in S_4$ and $P, Q \in \Lambda$. Then

(i) $\Pi \sigma = \sigma \Pi$,

(ii) $\sigma (P * Q) = \sigma (P) * \sigma (Q)$.

Suppose $w$ is a DNA sequence of length $n$ containing all nucleotides A, C, G and T, and $\delta_1(i), \delta_2(i), \delta_3(i), \delta_4(i) \in \{A_i, C_i, G_i, T_i\}$, where $1 \leq i \leq n$. Here $A_i = A_i(w)$, $C_i = C_i(w)$, $G_i = G_i(w)$ and $T_i = T_i(w)$. Since $A_i + C_i + G_i + T_i = i$, we can consider a general form of a coordinate system for $w$ as follows:

$$
\begin{align*}
&x_i = \alpha_1 \delta_1 + \alpha_2 \delta_2 + \alpha_3 \delta_3 + \alpha i \\
y_i = \beta_1 \delta_1 + \beta_2 \delta_2 + \beta_3 \delta_3 + \beta i \\
z_i = \gamma_1 \delta_1 + \gamma_2 \delta_2 + \gamma_3 \delta_3 + \gamma i
\end{align*}
$$

Suppose system (4) is satisfied the Theorem 1. Then it is easy to see that, Theorem 1(ii) is an immediate consequence of Theorem 1(i) and so we must check the equation $\Pi \sigma = \sigma \Pi$, for each $\sigma \in S_4$. He and Wang [17] presented an action of the symmetric group $S_4$ on DNA sequences. We now present new generators for this group compatible with system (4). Clearly, $S_4 = \langle (\delta_1 \delta_4), (\delta_2 \delta_4), (\delta_3 \delta_4) \rangle$. Suppose $O(3)$ denotes the set of all $3 \times 3$ orthogonal matrices on real numbers. Define:

$$
\Phi(\delta_2 \delta_4) = \begin{pmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \Phi(\delta_3 \delta_4) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -1 & 0 \end{pmatrix}, \Phi(\delta_1 \delta_4) = \begin{pmatrix} 0 & -1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}
$$

Extend these values to an injective homomorphism $\Phi : S_4 \longrightarrow O(3)$. Define an action of the symmetric group $S_4$ on the set of all DNA sequence matrices by $A^\alpha := A \Phi(\alpha)$, where $\alpha \in S_4$ and $A \in \Lambda$. Suppose the Theorem 1(i) is true for permutations $\chi_1$ and $\chi_2$. Then
\[
\Pi(\chi_1 \chi_2) = \Phi(\chi_1 \chi_2) \Pi(w) \\
= (\Phi(\chi_1) \Phi(\chi_2)) \Pi(w) \\
= \Phi(\chi_1) \Phi(\chi_2) \Pi(w) \\
= \Phi(\chi_2) \Pi(\chi_1) \\
= (\Pi(\chi_1)) \chi_2.
\]

This implies that to prove \( \Pi \sigma = \sigma \Pi \), for each \( \sigma \in S_4 \), it is enough to show that for each generator \( \chi_1 \) and \( \chi_2 \), \( \Pi(\chi_1 \chi_2) = (\Pi(\chi_1)) \chi_2 \). On the other hand,

\[
\Pi(v) = \begin{pmatrix}
\chi_1 \\
\chi_2 \\
\gamma_1
\end{pmatrix} = \begin{pmatrix}
\alpha_1 \beta_1 + \alpha_2 \beta_2 + \alpha_3 \beta_3 + \alpha \beta \\
\beta_1 \beta_1 + \beta_2 \beta_2 + \beta_3 \beta_3 + \beta \gamma \\
\gamma_1 \gamma_1 + \gamma_2 \gamma_2 + \gamma_3 \gamma_3 + \gamma \gamma
\end{pmatrix}
\]

Thus,

\[
\Phi(\delta_2 \delta_4) \Pi(v) = \begin{pmatrix}
-\gamma_1 \\
-\gamma_2 \\
\gamma_1 \gamma_1 + \gamma_2 \gamma_2 + \gamma_3 \gamma_3 + \gamma \gamma
\end{pmatrix} = \begin{pmatrix}
-\beta_1 \beta_1 - \beta_2 \beta_2 - \beta_3 \beta_3 - \beta \gamma \\
-\alpha_1 \beta_1 - \alpha_2 \beta_2 - \alpha_3 \beta_3 - \alpha \gamma \\
\gamma_1 \gamma_1 + \gamma_2 \gamma_2 + \gamma_3 \gamma_3 + \gamma \gamma
\end{pmatrix}
\]

So, by equation \( \Pi \sigma = \sigma \Pi \), we have:

\[
\begin{pmatrix}
-\beta_1 \beta_1 - \beta_2 \beta_2 - \beta_3 \beta_3 - \beta \gamma \\
-\alpha_1 \beta_1 - \alpha_2 \beta_2 - \alpha_3 \beta_3 - \alpha \gamma \\
\gamma_1 \gamma_1 + \gamma_2 \gamma_2 + \gamma_3 \gamma_3 + \gamma \gamma
\end{pmatrix} = \begin{pmatrix}
\alpha_1 \beta_1 + \alpha_2 \beta_2 + \alpha_3 \beta_3 + \alpha \beta \\
\beta_1 \beta_1 + \beta_2 \beta_2 + \beta_3 \beta_3 + \beta \gamma \\
\gamma_1 \gamma_1 + \gamma_2 \gamma_2 + \gamma_3 \gamma_3 + \gamma \gamma
\end{pmatrix}
\]

This shows that,

\[
\alpha_1 \beta_1 + (\alpha_2 + \beta_2) \beta_2 + (\alpha_3 + \beta_3) \beta_3 + \alpha_1 \beta_4 + (\alpha + \beta) \gamma = 0 \\
\alpha_1 \beta_1 + (\alpha_2 + \beta_2) \beta_2 + (\alpha_3 + \beta_3) \beta_3 + \alpha_1 \beta_4 + (\alpha + \beta) \gamma = 0
\]

and so

\[
\gamma_1 = 0, \beta_1 = \alpha_1 = \beta_2 + \alpha_2 = \beta_3 + \alpha_3 = - (\alpha + \beta).
\]

On the other hand,

\[
\Phi(\delta_2 \delta_4) \Pi(v) = \begin{pmatrix}
\gamma_1 \\
\gamma_2 \\
\gamma_1 \gamma_1 + \gamma_2 \gamma_2 + \gamma_3 \gamma_3 + \gamma \gamma
\end{pmatrix} = \begin{pmatrix}
-\gamma_1 \beta_1 - \gamma_2 \beta_2 - \gamma_3 \beta_3 - \gamma \gamma \\
\beta_1 \beta_1 + \beta_2 \beta_2 + \beta_3 \beta_3 + \beta \gamma \\
-\alpha_1 \beta_1 - \alpha_2 \beta_2 - \alpha_3 \beta_3 - \alpha \gamma
\end{pmatrix}
\]

This implies that,
\[
\begin{align*}
(\alpha_1 + \gamma_1)\beta_1 + \gamma_2\beta_2 + (\alpha_2 + \gamma_2)\beta_3 + \alpha_2 \beta_4 + (\alpha + \gamma) t &= u \\
(\alpha_1 + \gamma_1)\beta_1 + \alpha_2 \beta_3 + (\alpha_2 + \gamma_2)\beta_3 + \gamma_3 \beta_4 + (\alpha + \gamma) t &= 0, \\
-\beta_2 \beta_2 + \beta_2 \beta_4 &= 0
\end{align*}
\]

and hence,
\[
\beta_2 = 0, \quad \alpha_2 = \gamma_2 = \alpha_1 + \gamma_1 = \alpha_3 + \gamma_3 = - (\alpha + \gamma).
\]

Finally,
\[
\Phi(\beta_3, \beta_4) \Pi(v) = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} \alpha_1 \beta_1 + \alpha_2 \beta_3 + \alpha_3 \beta_4 + \alpha t \\
-\gamma_1 \beta_1 - \gamma_2 \beta_2 - \gamma_3 \beta_3 - \gamma t \end{pmatrix} = \begin{pmatrix} \beta_1 \beta_1 + \beta_2 \beta_4 + \beta_3 \beta_5 + \beta t \\
\gamma_1 \beta_1 + \gamma_2 \beta_2 + \gamma_3 \beta_3 + \gamma t \end{pmatrix}
\]

Therefore,
\[
(\beta_1 + \gamma_1)\beta_1 + (\beta_2 + \gamma_2)\beta_2 + (\beta_3 + \gamma_3)\beta_3 + (\beta + \gamma) t = 0 \\
(\beta_1 + \gamma_1)\beta_1 + (\beta_2 + \gamma_2)\beta_2 + (\beta_3 + \gamma_3)\beta_3 + (\beta + \gamma) t = 0, \\
-\alpha_3 \beta_3 + \alpha_2 \beta_4 = 0
\]

and so
\[
\alpha_3 = 0, \quad \beta_3 = \gamma_3 = \beta_2 + \gamma_2 = \beta_1 + \gamma_1 = -(\beta + \gamma).
\]

By substituting values given in Equations (5-7) in Equation (4), one can prove that \(\alpha = \beta = \gamma = A, \alpha_1 = \beta_1 = \gamma_1 = \alpha_2 = \beta_2 = \gamma_2 = B\) and \(B = -2A\). Therefore, the general form of the System (4) which is satisfied the Theorem 1(i) is as follows:

\[
\begin{align*}
\frac{dx}{dt} &= -2A(\theta_1 + \theta_2) + At \\
\frac{dy}{dt} &= -2A(\theta_1 + \theta_2) + At \\
\frac{dz}{dt} &= -2A(\theta_2 + \theta_3) + At
\end{align*}
\]

CONCLUSIONS

In this paper a general form of Z-curve, H-curve and Randić’s model is presented. It is proved that in a group theoretical language they are equivalent models for 3D visualization representation of sequences. Our main idea for this work is taken from the paper by He and Wang [17] and the book of Hsu [5].
REFERENCES