International Journal of Pure and Applied Mathematics
Volume 114 No. 3 2017, 483-501
ISSN: 1311-8080 (printed version); ISSN: 1314-3395 (on-line version)
url: http://www.ijpam.eu
doi: 10.12732/ijpam.v114i3.6

ijpam.eu

FUZZY CODE ON RNA SECONDARY STRUCTURE

Ajay Kumar Saw[1 §], Soumyadeep Nandi[2], Binod Chandra Tripathy[3]

[1]Mathematical Sciences Division
Institute of Advanced Study in Science and Technology
Guwahati, 781035, Assam, INDIA

[2]Life Science Division
Institute of Advanced Study in Science and Technology
Guwahati, 781035, Assam, INDIA

[3]Department of Mathematics
Tripura University
Agartala, 799022, Tripura, INDIA

Abstract: In this paper, we developed a fuzzy code technique for molecular phylogenetic analysis. This proposed theory has potential to encode or decode information related to the evolution of sequences traversing from one stage to another in phylogenetic trees. Using this novel methodology we have encoded the sequence of RNA molecule of each species in phylogenetic trees which folds into three-dimensional structure due to transcription, termed as secondary structure.After encoding RNA sequence into the fuzzy code, we wrote mathematical formulation of RNA secondary structure. In addition we establish relation between RNA sequence and their secondary structure. We constructed the fuzzy neural network, fraction of neural neighbour in sequence space for differentiating compatible sequences. We have used technique involution metric, symmetric group, symmetric difference, etc; to establish a difference in secondary structures.

AMS Subject Classification: 03E72, 20Bx, 68R10, 20A15, 30Lxx

Key Words: fuzzy theory, group theory, metric space, phylogenetic trees, RNA sequenceing, RNA secondary structure

Received: December 3, 2016
Revised: May 12, 2017
Published: May 23, 2017

[§]Correspondence author

1. Introduction

Medicine at the turn of the century is characterised by the deepest change. It has ever been subjected to its history, i.e. its transformation from a healing profession to a branch of biotechnology. Viewed from an evolutionary perspective, these changes appear as an aspect of a Darwin-Lamarckian auto evolution of life on earth. The nucleic acids DNA and RNA as the genetic material of living beings and viruses play the pivotal role in this arena. This can be understood from the following example. Suppose we have a sequence of RNA nucleotide, after chemical changes or mutation take place during evolution from one generation to another, does not give us enough information about which terms of nucleotide sequence get affected. It means there exist uncertain condition in DNA or RNA sequence when they move from one generation to another.

To the above problem, a reliable technique is required, which can handle uncertainty condition. In this study, we present a novel methodology based on fuzzy theory [7]. Fuzzy sets were introduced by Lotfi A. Zadeh in 1965 as an extension of the classical notion of set. Fuzzy sets are sets whose elements have degrees of membership in terms of numeric value belong to $[0, 1]$. A polynucleotide of DNA or RNA molecule is a linear polymer that consists of many smaller units called its building block or monomers. Nucleotide of DNA or RNA is transformed into fuzzy sets termed as fuzzy nucleotide or fuzzy code. A Linear polymer of fuzzy nucleotide is termed as a fuzzy polynucleotide. Previous studies have shown that fuzzy theory[22, 23] is used as a tool in bioinformatics. In this study, we used the fuzzy concept in phylogenetic analysis. We have encoded a sequence of RNA into fuzzy code in the phylogenetic tree and conversely decoded. We used the fuzzy code as an information preserving tools in phylogenetic analysis, it preserved number of times of mutation of a nucleotide at each position, at each stage.

In this study, we have constructed fuzzy graph [9], where fuzzy nucleotide acts as vertices and interaction between two fuzzy nucleotide [17, 18, 19] treated as an edge, which is hydrogen bond between them. We have treated RNA secondary structure [8] or RNA fuzzy structure as a fuzzy graph. Each RNA molecule folds into a three dimensional structure, which determines it's biochemical function. After constructing the edges and vertices in a biochemical structure, we studied further some possible aspect of fuzzy polynucleotides space and fuzzy structure space of fixed length 'n' to explore some biochemical property. Folding of RNA polynucleotide [10] or secondary structure is viewed as a map that assigns a uniquely defined base pairing pattern to every sequence. The mapping was non-invertible since many fuzzy polynucleotides folded into the same minimum free energy secondary structure, see for instance [11]. We used involution metrics given by Reidys and stadler [5], symmetric group [30], symmetric difference method for predicting the difference between two secondary structure.

2. Fuzzy nucleotide

DNA and RNA are a linear polymer of nucleotides, and we are referred as polynucleotide. In this section, we represent RNA nucleotide as a fuzzy nucleotide.

Let $X = \{A, U, G, C\}$ be a non-empty set of nucleotides.

Fuzzy nucleotide W in X is characterised by it's membership function, $\mu_W : X \to [0, 1]$ and $\mu_W(x)$ is interpreted as the degree of membership of element x in fuzzy set W, for each $x \in X$ under certain condition:

$$\sum_{x \in X} \mu_W(x) \leqslant 1. \tag{2.1}$$

It is clear that W is completely determined by the set of tuples,
$W = \{(x, \mu_W(x)) \textit{ such that } x \in X,\ \Sigma_{x\in X}\ \mu_W(x) \leqslant 1\}$.
So it can also be written as

$$W = \{(a_1, a_2, a_3, a_4) \textit{ such that } \Sigma_{j=1}^{4} a_j \leqslant 1, a_j \in [0, 1]\} \tag{2.2}$$

We can write each fuzzy nucleotide as a four dimensional tuples. Each coordinate represents the degree of membership of particular nucleotide which belongs to set X. Here we fix the position of each nucleotide in coordinate-wise respectively. In which first, second, third and fourth coordinate represent the degree of membership of nucleotide A, U, G and C respectively.

Particularly in equation (2.2), if we put $a_1 = 1, a_2 = a_3 = a_4 = 0$ then $W = A$,
$a_2 = 1, a_1 = a_3 = a_4 = 0$, then $W = U$,
$a_3 = 1, a_1 = a_2 = a_4 = 0$, then $W = G$,
$a_4 = 1, a_1 = a_2 = a_3 = 0$ then $W = C$.
So we can represent a nucleotide regarding fixed code.
$A = (1, 0, 0, 0)$,
$U = (0, 1, 0, 0)$,
$G = (0, 0, 1, 0)$,
$C = (0, 0, 0, 1)$.

Fuzzy theory is a mathematical concept which comes under the category of uncertainty theory. Due to this characteristic, this theory is useful for handling the degree of presence of nucleotides in fuzzy system. We used fuzzy nucleotide which has the potential to express uncertainty due to the mutation in DNA or RNA sequence traversing from one generation to another. Fuzzy nucleotide express the degree of membership of each nucleotide as shown in equation(2.2). The fixed code is a particular case of fuzzy code, which expresses certainty for each nucleotide. Using this ideology, we encoded and decoded the evolution of RNA sequence under molecular phylogenetic principle. In next section, we focus on the encoding evolution of sequences in fuzzy codes and vice versa.

3. Encoding and Decoding of fuzzy polynucleotides in Molecular phylogenetic

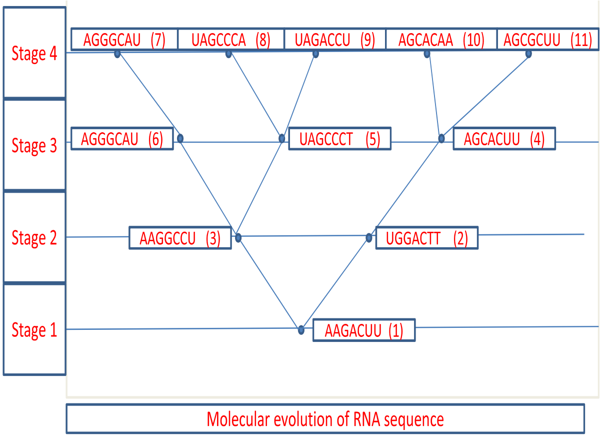

Figure 1: Molecular evolution of RNA sequence.

The molecular phylogenetic structure is generated from character datasets that provide evolutionary content and context. Evolution is modelled as a process that changes the state of a character, such as the type of nucleotides (AUGC) at a specific location in an RNA sequence. Each character is a function that maps a set of taxa to distinct states. It means, in the evolution of RNA or DNA sequence there are some stages from ancestral sequence to their descendant sequence.Molecular evolution of RNA sequences in the phylogenetic tree represents a change in nucleotide position in their sequences, while it moving from one species to another species with respect to time. A phylogenetic tree is a graphical representation of the evolution of species which contains different stages as shown in figure(3.1). Figure(3.1) illustrates how a molecular sequence might evolve over time as a result of multiple mutations that results in a small, but evolutionarily important changes in a nucleotide sequence. Such changes over time may eventually modulate the function of the protein within divergent species [28]. Molecular evolutionary of RNA sequence at each stage represents the uniqueness of species characteristic, which is our sequence of interest. In figure(3.1) , node 1 is ancestral sequence, node 2 and node 3 are descendants of node 1 , node 5 and node 6 are descendants of node 3 and so on. In this study, we encoded or decoded the uniqueness of sequencing regarding information preserving coding. **Note**, in this paper we have used RNA sequences as the character data. However, phylogenetic trees can be accurately estimated from many different types of molecular data.

3.1. Encoding RNA sequence into fuzzy code

Let

$$V_n^m = \{(w_1^m, w_2^m, w_3^m ... w_n^m) \textit{ such that } m \in \{1, 2, 3...p\} = M\} \tag{3.1}$$

is fuzzy polynucleotide of fixed length n at each stage m,

where $w_i^m = \{(a_1^m, a_2^m, a_3^m, a_4^m) \textit{ such that } \Sigma_{j=1}^4 a_j^m \leq 1,\ a_j^m \in [0, 1]\}$ is a fuzzy nucleotide of position i at stage m. ....(3.1.1)

and a_j^m = (repetition of jth coordinate at ith position from initial stage upto m stage)/m.

w_i^m can also be obtained by using previous information $w_i^{(m-1)}$ under following condition:

w_i^m= $[w_i^{(m-1)}$+ (fixed code of nucleotide at stage m)/$(m-1)] * ((m-1)/m)$ for $m > 1$. ....(3.1.2)

3.2. Decoding RNA fuzzy code into sequence

Condition:-

1. Following method has been applicable for those fuzzy codes which are not fixed codes.

2.Fuzzy code w_i^m except fixed code, suppose we want to know their nucleotide then it is mandatory to have knowledge about previous fuzzy code at position i, stage $(m-1)$ i.e w_i^{m-1}.

Method: case 1. If it is fixed code, then its nucleotide is trivial.
(1,0,0,0) = A,
(0,1,0,0)= U,
(0,0,1,0)=G,
(0,0,0,1)=C ,
(0,0,0,0) = unknown.

Case 2. If it is not fixed code, eg:(0.5,0.25,0,0.25), then how to know which nucleotide this code will represent . Let V_n^m, V_n^{m-1} be two sequence of fuzzy codes of length n, stage m and length n, stage $(m-1)$.Then
$V_n^{m-1} = w_1^{m-1}, w_2^{m-1}, w_3^{m-1}, ...w_n^{m-1}$ & $V_n^m = w_1^m, w_2^m, w_3^m, ...w_n^m$,
where $w_i^m = \{(a_1^m, a_2^m, a_3^m, a_4^m)\ such\ that\ \Sigma_{j=1}^4 a_j^m \leqslant 1, a_j^m \in [0,1)\}$
Consider a function

$$b_j^m = \begin{cases} 0; & \text{if } a_j^m - a_j^{m-1} \leq 0 \\ 1\ ; & \text{if } a_j^m - a_j^{m-1} > 0 \end{cases}$$

and

$$z_i^m = \{(b_1^m, b_2^m, b_3^m, b_4^m)\ such\ that \Sigma_{j=1}^4 b_j^m = 1, b_j^m \in \{0,1\}\}. \qquad (3.2)$$

Then, we assign $w_i^m = z_i^m$.

3.3. Validation of Encoding and Decoding in molecular phylogenetic trees

Suppose we have RNA sequence of length 4 moving from Stage 1 to 5 as shown in the table(1). Using equation(3.1), we encode RNA sequence into their fuzzy code, which shown in the table(2). Similarly using equation(3.2), we decode fuzzy code into RNA sequence.

stage 1	A	A	U	C
stage 2	A	A	G	A
stage 3	G	C	A	-
stage 4	G	U	A	U
stage 5	A	C	C	A

table (1)

stage 1	1000	1000	0100	0001
stage 2	1000	1000	0,0.5,0.5,0	0.5,0,0,0.5
stage 3	0.6,0,0.3,0	0.6,0,0,0.3	0.3,0.3,0.3,0	0.3,0,0,0.3
stage 4	0.5,0,0.5,0	0.5,0.25,0,0.25	0.5,0.25,0.25,0	0.25,0.25,0,0.25
stage 5	0.6,0,0.4,0	0.4,0.2,0,0.4	0.4,0.2,0.2,0.2	0.4,0.2,0,0.2

table (2)

In the table (1), stage 1 represents an ancestral sequence, stage 2 descendant of stage 1 and so on. With the help of encoding technique using equation(3.1), we can preserve mutation information as shown in the table(2). In the table (2),each position while moving at stages $m = 1\ to\ 5$ express changes in their nucleotides $\{A, U, G, C\}$. It means each code at position i and stage m tells about their nucleotide changes moving through evolution from one species to another species. By the help of table (1)and (2), we can write fuzzy code of figure(3.1).Fuzzy code preserves information of their nucleotides mutation at each stage, which is given in equation(3.1). Due to the information preserving property, it helps to determine mutation spots in the evolutionary structure of RNA sequences. This spot can play a better role in the sequence alignment to identify homologies. Suppose we are interested to know about the nucleotide type at any particular stage say m, example: at stage 4 in the table(2). Then by the help of equation(3.2) using the decoding technique, we can find nucleotide sequence. Similarly, we can find in the phylogenetic tree also, if it is written in fuzzy code form. Phylogenetic structure of RNA sequence is one kind of graphical approach. Therefore, it is very obvious, how to analysis the evolution of fuzzy polynucleotide of RNA sequence and their secondary structure with the help of graph theory. In next section, we constructed the fuzzy graph, where fuzzy nucleotide act as a vertex and interaction between two nucleotides acts as an edge.

4. Fuzzy Graph of RNA sequence

It is quite well known that graphs are simply models of relations. A graph is a convenient way of representing information involving the relationship between objects. The objects are represented by vertices and relations by edges. Using this standard logic, we constructed a fuzzy graph for each fuzzy polynucleotide, where fuzzy nucleotide acted as vertex and interaction between two fuzzy nucleotides acted as edges. In phylogenetic tree of RNA sequences of different species, we characterised sequence in terms of their length of sequence and stage. i.e, V_n^m is a fuzzy polynucleotide of length n at each stage m.

(4.1.) Let $V_n^m = \{w_1^m, w_2^m, w_3^m, ...w_n^m$ such that $m \in (1, 2, 3, 4...p) = M\}$ is fuzzy polynucleotide of length n at stage m, where each

$w_i^m = \{(a_1^m, a_2^m, a_3^m, a_4^m)\ such\ that\ \Sigma_{j=1}^4 a_j^m \leqslant 1, a_j^m \in [0, 1]\}$ is a fuzzy nucleotide position i at stage m which actually express variation of degree of membership of same set of nucleotides X and $\{a_j^m\}$ represent degree of membership of particular nucleotide at stage m for each position $j \in \{1, 2, 3, 4\}$. We assign unique degree of membership of fuzzy nucleotide at position $i \in (1, 2, 3...n)$ at stage m,

$$w_i^m(X) = \{a_j^m \ such\ that\ a_j^m - a_j^{m-1} > 0\ for\ all\ j \in \{1, 2, 3, 4\}\} \tag{4.1}$$

[**Note:-** $w_i^m(X) = \{\phi\}$, it means vacant position.]

(4.2.) Let $\mathbf{e}(w_i^m\ ,\ w_t^m)$ = interaction between two fuzzy nucleotide w_i^m and w_t^m, which is symmetric fuzzy relation on $V_n^m \times V_n^m$ = interaction between their decoded nucleotide =$\mathbf{e}(z_i^m, z_t^m)$ [from equation(3.2)] .

Mathematically written as,

$$e : z_i^m \times z_t^m \rightarrow [0, 1] \text{ for all } \ i \neq t \in \{1, 2, 3...n\}$$

However, complementary bases $A - U$,$G - C$ and $G - U$ form stable base pairs with each other using hydrogen bonds. Due to this base pairing property, the RNA sequence forms secondary structure. RNA secondary structure is one type of graphical representation. In this section, we defined all preliminary related to graph theory in fuzzy code form. In next section, we are constructing RNA secondary structure using a fuzzy system.

5. Fuzzy code on RNA secondary structure

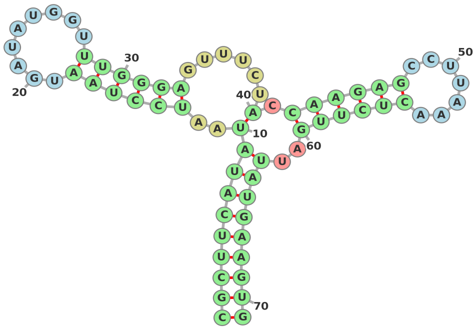

Figure 2: Secondary structure of RNA.

RNA form secondary structures because of WatsonCrick and GU wobble base pairs property that can be expressed by fuzzy codes. Figure(5.1) [29] is a pictorial representation of RNA secondary structure which has sequence length 70. Due to the base pairing property, it folds into itself and forms variety of secondary structure. In this section, we have constructed the mathematical formulation of secondary structure using the fuzzy system. A fuzzy polynucleotide is a sequence of fuzzy nucleotides or fuzzy codes. **A secondary structure [1]of RNA is a fuzzy structure** $H_n^m = (V_n^m, E)$, where V_n^m is a fuzzy polynucleotide of length n at stage m in which each fuzzy nucleotide acts as vertice associated with adjacency matrix $E = e(z_i^m, z_t^m)_{1 \leqslant i,t \leqslant n}$ fulfilling three condition:

(a) $e(z_i^m, z_{i+1}^m) \rightarrow (0, 1]$ for all $1 \leqslant i \leqslant n-1$.

(b) For each i, there is atmost one $t \neq i-1, i+1$ with

$$e(z_i^m, z_t^m) = \{1, if z_i^m \times z_t^m \in (AU, UA, UG, GU, GC, CG)\} \tag{5.1}$$

(c) If e(z_i^m , z_t^m) and e(z_k^m , z_l^m) be two interaction of fixed codes of nucleotides and $i < k < t$ then $i < l < t$.

Condition (c) gives guarantee that the fuzzy structure free from knot or pseudoknots. Vertices of the fuzzy polynucleotide are the individual fuzzy nucleotide in the order defined by the RNA sequence of fuzzy polynucleotide V_n^m, which is a string of length n at stage m over a nucleotide alphabet $X = \{A, U, G, C\}$. Edge $e(z_i^m, z_t^m)$ with $z_t^m \in \{z_{i-1}^m, z_{i+1}^m\}$ constitute the ribose-phosphate backbone of the nucleotides. An edge e(z_i^m , z_t^m) with $z_t^m \neq \{z_{i-1}^m, z_{i+1}^m\}$ is called base pair of the secondary structure. Each z_i^m which is connected only to it's neighbours in the backbone z_{i-1}^m and z_{i+1}^m is called unpaired nucleotides of fuzzy structure. The number of base pair and unpaired fuzzy nucleotides in fuzzy structure H_n^m are denoted by $n_p(H_n^m)$ and $n_u(H_n^m)$, respectively. Accordingly the chain length of the fuzzy structure, it can also be expressed in terms of base pair and unpaired fuzzy nucleotides,

$$n = n_u(H_n^m) + 2n_p(H_n^m) \tag{5.2}$$

Each RNA fuzzy polynucleotide folds due to base pairing property form RNA fuzzy structure (secondary structure). So there is a one-one correspondence relation between a fuzzy polynucleotide of length n at stage m and their fuzzy structure as defined above. But now there is a need which requires extension of this conventionally consider one-one correspondence relation between them, because folding of RNA fuzzy polynucleotide into fuzzy structure is viewed as a map that assign a unique defined base pairing pattern (AU, UA, GC, CG, UG, GU) to every fuzzy polynucleotide.It is worth to point out that the relation between the fuzzy polynucleotide and fuzzy structure is introduced only via the pairing rules.

The fuzzy structure is characterised by the only pairing scheme

$$\pi(H_n^m) = \{[w_i^m, w_t^m] \text{ such that } e(z_i^m, z_t^m) = 1 \text{ and } t \neq i-1, i+1\}, \quad (5.3)$$

where $[w_i^m, w_t^m]$ represent only i^{th} , t^{th} position of fuzzy nucleotide at stage m.

Two fuzzy structure(secondary structure) are considered as the same class, if and only if their base pairing position is same and independent of base nucleotide which belongs to any one of set (AU, UA, GC, CG, GU, UG).Similarly, a set of the fuzzy structure is considered into the same class if and only if their base pairing position is same and independent of base nucleotide which belongs to any one of set (AU, UA, GC, CG, GU, UG). We select one from this class which has minimal free energy level. Through the following condition, we convert the one-one correspondence map into many-one map between fuzzy polynucleotide space and fuzzy structure space. Therefore, a mapping is non-invertible since many sequences fold into the same minimal free energy fuzzy structure. Keeping this aspect in mind, in next section we established many-to-one relation between RNA fuzzy polynucleotide space and their fuzzy structure space instead of conventionally consider the one-one correspondence relation between them.

5.1. RNA fuzzy polynucleotide space and their structure space

RNA fuzzy polynucleotide space is the collection of all evolutionary stage of RNA fuzzy polynucleotide in phylogenetic tree with constant chain length n at each stage $m \in M$ over the nucleotides X is a generalised hyperspace denoted by Q_n^M of dimension $4n$, because each fuzzy nucleotide has dimension 4. Each fuzzy polynucleotide folds in itself due to the base pairing property form fuzzy structure. The collection of all fuzzy structure is called fuzzy structure space is denoted by F_n^M

The mapping from fuzzy polynucleotide space onto fuzzy structure space is many-to-one.We deal model of mapping with the pre-images of particular fuzzy structure in fuzzy polynucleotide space; these are fuzzy neural networks which a subspace of Q_n^M.

Since subspace of fuzzy polynucleotides fold into minimal free energy fuzzy structure. The fuzzy structure is characterised by same base pairing position (from equation 5.3). So, we took minimal free energy fuzzy structure say $H_n^{m_1}$, where $m_1 \in M$.

A subspace of $Q_n^M(X)$ is compatible with fuzzy structure $H_n^{m_1}$ if condition of equation (5.1) is fulfilled for all $\pi(H_n^{m_1})$ (equation (5.3)). In other words, the fuzzy nucleotides at i^{th} and t^{th} position of compatible fuzzy polynucleotide are capable to form base pair, when pair $[w_i^{m_1}, w_t^{m_1}] \in \pi[H_n^{m_1}]$. The set of all compatible fuzzy polynucleotides is denoted by $D(H_n^{m_1})$. The cardnality of this fuzzy polynucleotides which are compatible with fuzzy structure $H_n^{m_1}$ is given by,

$$|D(H_n^{m_1})| = |X|^{n_u(H_n^{m_1})}|\beta|^{n_p(H_n^{m_1})} \tag{5.4}$$

where

$$X = \{A, U, G, C\}, \quad \beta = \{z_i^{m_1} z_t^{m_1}$$

such that

$$e(z_i^{m_1}, z_t^{m_1}) = 1\} = \{AU, UA, UG, GU, GC, CG\}.$$

Consider a combinatory map

$$f : Q_n^M \to F_n^M, \tag{5.5}$$

where F_n^M denote the space of all RNA fuzzy structure which formed by the Q_n^M through base pairing property. Since mapping is many-one. Pre-image of $H_n^{m_1}$, $f^{-1}(H_n^{m_1})$ which consists of all fuzzy polynucleotides folding into the fuzzy structure $H_n^{m_1}$, is contained into compatible fuzzy polynucleotides.

Let $H_n^{m_1}$ be a fuzzy structure, then the subspace of Q_n^M compatible with $H_n^{m_1}$ denoted as $C(H_n^{m_1})$ is given by,

$$C(H_n^{m_1}) \cong Q^Y_{n_u(H_n^{m_1})}(X) \times Q^Y_{n_p(H_n^{m_1})}(\beta), \quad where\ Y \subseteq M. \tag{5.6}$$

Two fuzzy polynucleotides belong to $D(H_n^{m_1})$ are neighbour if they differ either :

- In a single position which is unpaired fuzzy nucleotide in fuzzy structure $H_n^{m_1}$, or
- In two positions $w_i^{m_1}$ and $w_t^{m_1}$which form a base pair $[w_i^{m_1}\ ,\ w_t^{m_1}] \in H_n^{m_1}$.

Two subspace $C(H_n^{m_1})$ and $C'(H_n^{m_2})$ of Q_n^M are isomorphic [2] if and only if their fuzzy structure has the same number of unpaired fuzzy nucleotides and base pair fuzzy nucleotides with respect to position.

The fuzzy neural network is modelled as a subspace of fuzzy polynucleotide space to fulfil a complex frame for the derivation of analytical result that can be used as a reference for a fuzzy neural network of RNA [3, 4]. The construction of fuzzy network model are the set of compatible fuzzy polynucleotides of a given fuzzy structure $H_n^{m_1}$. Fuzzy Polynucleotide is selected at randomly from a set of compatible fuzzy polynucleotides, but unpaired fuzzy nucleotides and base pair fuzzy nucleotides were distinguished. There were two elementary moves of compatible fuzzy polynucleotides, they were base exchange for unpaired fuzzy nucleotides and base exchanges for paired fuzzy nucleotides. Each fuzzy polynucleotide V_n^m has a certain number of fuzzy neural neighbours u_i^m in the unpaired base exchange neighbourhood and p_l^m in the paired base exchange neighbourhood.

The fraction of neural neighbour of length n moving from stage m to m' or m' to m within compatible polynucleotides of the fuzzy structure $H_n^{m_1}$ is given by,

$$(\lambda_u)^{m \leftrightarrow m'} = \frac{Total\ number\ of\ unpaired\ mutation\ moving\ from\ stage\ m\ to\ m'}{(X-1)n_u}$$

$$and \tag{5.7}$$

$$(\lambda_p)^{m \leftrightarrow m'} = \frac{Total\ number\ of\ paired\ mutation\ moving\ from\ stage\ m\ to\ m'}{(\beta-1)n_p},$$

respectively.

Averaging over all polynucleotides of the fuzzy neural network which is the set of compatible fuzzy polynucleotides of a given fuzzy structure $H_n^{m_1}$ is given by,

$$\lambda_u = \frac{1}{(Y-1)!}\Sigma_{m,m' \in Y}(\lambda_u)^{m \leftrightarrow m'} and \quad \lambda_p = \frac{1}{(Y-1)!}\Sigma_{m,m' \in Y}(\lambda_p)^{m \leftrightarrow m'}, \tag{5.8}$$

where Y = cardinality of fuzzy neural network which is subset of set M .

In this section, we established a many-one mapping between RNA fuzzy polynucleotide space and fuzzy structure space. The difference between fuzzy polynucleotide space and fuzzy structure space are due to base pairing property. The difference between two fuzzy polynucleotide are due to the difference in unpaired bases with respect to position and difference between two fuzzy structure are due to WatsonCrick and GU wobble pairs with respect to position. We analysed fuzzy neural network of compatible fuzzy polynucleotide sequences, which is a collection of the evolution of primary sequences of some stages $Y \subset M$ in fuzzy code form, which folds into minimal free energy fuzzy structure. We developed an equation to calculate the fraction of neural neighbour between two fuzzy polynucleotides at different stages having fixed length within the fuzzy neural network which mapped to minimal free energy fuzzy structure. In next section, analysed RNA fuzzy structure space, which forms through fuzzy polynucleotide space using WatsonCrick and GU wobble pairs.

5.2. Involution metric for the space of RNA fuzzy structure

For every $H_n^m = ([n], \pi(H_n^m)) \in F_n^M$ and we define a many-one mapping,

$$P : F_n^M \to S_n \tag{5.9}$$

satisfying the condition $P([w_i^m, w_t^m]) = (i, t)$

such that $P(H_n^m) = \prod[w_i^m, w_t^m] = \prod(i, t) \in S_n$ for every H_n^m, where $[w_i^m, w_t^m] \in \pi(H_n^m)$(from equation 5.3) and (i, t) denotes the transposition of symmetric group S_n [30].

$P(H_n^m)$ is the involution function for every $H_n^m \in F_n^M$, see [5].

Proposition 5.2.1. *The mapping $d_{inv} : F_n^M \times F_n^M \to \mathbf{N}$ (set of natural numbers) for every $(H_n^m, H_n^{m'}) \in (F_n^M)^2$ to the least number of transposition, $d_{inv}(H_n^m, H_n^{m'})$ which exactly represents the permutation $P(H_n^m)P(H_n^{m'})$ is a metric and (F_n^M, d_{inv}) are metric spaces see [6, 20, 21].*

Our aim is to compute explicitly this metric. For this we are going to introduce some notations:

Given two fuzzy structure of same length, $H_n^m = ([n], \pi(H_n^m))$ and $H_n^{m'} = ([n], \pi(H_n^{m'}))$. Their symmetric difference of fuzzy structure $H_n^m \vartriangle H_n^{m'} = ([n], \pi(H_n^m) \vartriangle \pi(H_n^{m'}))$, where $\pi(H_n^m) \vartriangle \pi(H_n^{m'}) = (\pi(H_n^m) \cup \pi(H_n^{m'})) - (\pi(H_n^m) \cap \pi(H_n^{m'}))$.

Note. $\pi(H_n^m) \vartriangle \pi(H_n^{m'})$ may not satisfy base pair property of fuzzy structure $H_n^m \vartriangle H_n^{m'}$.

The component of the fuzzy structure are those subsets

$$\{w_1^x, w_2^x, w_3^x, w_4^x....w_r^x\} \subset [n],\ r \leqslant n,$$

and $x \in \{m, m'\}$, such that

$$[w_1^x, w_2^x], [w_2^x, w_3^x]....[w_{r-1}^x, w_r^x] \in \pi(H_n^m) \vartriangle \pi(H_n^{m'})$$

and maximal with this property.

The unique bond condition implies, if $\{w_1^x, w_2^x, w_3^x, w_4^x....w_r^x\}$ is component then $\{[w_1^x, w_2^x], [w_3^x, w_4^x], [w_5^x, w_6^x], ...\} \in \pi(H_n^m)$, where $x = m$,

$$\{[w_2^x, w_3^x], [w_4^x, w_5^x], ...\} \in \pi(H_n^{m'}),$$

$x = m'$, vice versa.

Cardinality of components are it's length, which is the number of fuzzy nucleotides present in the component. Component is said to be closed if $r \geqslant 4$, even and $[w_1^x, w_r^x] \in \pi(H_n^m) \cup \pi(H_n^{m'})$. Components are open in rest of the other cases. Base pair of fuzzy nucleotides $[w_i^x, w_t^x] \in \pi(H_n^m) \vartriangle \pi(H_n^{m'})$ are involved in a component when it's vertex w_i^x, w_t^x belong to these components. Every base pair of fuzzy nucleotides in $\pi(H_n^m) \vartriangle \pi(H_n^{m'})$ are involved in one and only one component. Each base pair of fuzzy nucleotides sharing base pair in same component.

Proposition 5.2.2. *For every*

$$\{H_n^m = ([n], \pi(H_n^m)) \ , \ H_n^{m'} = ([n], \pi(H_n^{m'}))\} \in F_n^M,$$

$$d_{inv}(H_n^m, H_n^{m'}) = |\pi(H_n^m) \vartriangle \pi(H_n^{m'})| - 2\Omega(H_n^m, H_n^{m'}),$$

where $\Omega(H_n^m, H_n^{m'})$ *is the number of closed component in* $H_n^m \vartriangle H_n^{m'}$.

Proof. From Proposition 5.2.1: $d_{inv} = n(P(H_n^m)P(H_n^{m'}))$, where $n(P(H_n^m) P(H_n^{m'}))$ is the least number of transposition. But in this proposition we are going to compute d_{inv} explicitly.

Base pair of fuzzy nucleotide, suppose $[w_1^x, w_2^x] \in \pi(H_n^m) \cap \pi(H_n^{m'})$ then transposition appear in both $P(H_n^m)$ and $P(H_n^{m'})$. So that they cancel each other in product of transposition $P(H_n^m)P(H_n^{m'})$. Similarly, in symmetric difference $P(H_n^m) \vartriangle P(H_n^{m'}) = (P(H_n^m)) \cup P(H_n^{m'})) - (P(H_n^m) \cap P(H_n^{m'}))$, it reduces all common base pair of fuzzy nucleotides. On the other side, if two transpositions appear in the product $P(H_n^m)P(H_n^{m'})$ are distinct but not disjoint, then indexes involved in them belong to the same component of $H_n^m \vartriangle H_n^{m'}$. Since two different transpositions always commute, this permit us to reorganize the transposition in the product $P(H_n^m)P(H_n^{m'})$ by assembling them into subproducts corresponding to components of $H_n^m \vartriangle H_n^{m'}$. Particularly, for every component C of $H_n^m \vartriangle H_n^{m'}$.

For every $i = 1, 2$ we get:

$$P(C, H_n^m) = \prod_{[w_k^m, w_l^m] \in \pi(H_n^m), [w_k^m, w_l^m] \in C} (k, l) \quad (from\ equation\ (5.9)) \tag{5.10}$$

then,

$$P(H_n^m)P(H_n^{m'}) = \prod_{C \in \{Component of H_n^m \vartriangle H_n^{m'}\}} P(C, H_n^m)P(C, H_n^{m'}). \tag{5.11}$$

Since the components of $H_n^m \vartriangle H_n^{m'}$ are pairwise disjoint, so sum of the least number of transposition for each component of $H_n^m \vartriangle H_n^{m'}$ are equal to the least number of transposition of $P(H_n^m)P(H_n^{m'})$.

We should now compute for each component:

Case 1. If C is open component of length $r \geqslant 2$.

- If $r = 2$ then it is only one transposition either belong to $P(C, H_n^m)$ or $P(C, H_n^{m'})$ which trivially least number of transposition.
- If $r > 2$ and even then,

$P(C, H_n^m)P(C, H_n^{m'}) = (1,2)(3,4)(5,6)...(r-1,r) \quad (2,3)(4,5)(6,7)...(r-2, r-1)$

$$= \left(\begin{smallmatrix} 1 & 2 & 3 & 4 & \cdots & r-1 & r \\ 2 & 1 & 4 & 3 & \cdots & r & r-1 \end{smallmatrix}\right)\left(\begin{smallmatrix} 1 & 2 & 3 & 4 & 5 & \cdots & r-1 & r \\ 1 & 3 & 2 & 5 & 4 & \cdots & r-2 & r \end{smallmatrix}\right) = \left(\begin{smallmatrix} 1 & 2 & 3 & 4 & 5 & \cdots & r-1 & r \\ 2 & 4 & 1 & 6 & 3 & \cdots & r-3 & r-1 \end{smallmatrix}\right)$$

[due to permutation composition]

$= (1,2,4,6...r, r-1, r-3, r-5...7,5,3)$ [cyclic representation of S_n].

- If $r > 2$ and odd then,

$P(C, H_n^m)P(C, H_n^{m'}) = (1,2)(3,4)(5,6)...(r-2,r-1) \quad (2,3)(4,5)(6,7)...(r-1, r)$

$$= \left(\begin{smallmatrix} 1 & 2 & 3 & 4 & \cdots & r-2 & r-1 & r \\ 2 & 1 & 4 & 3 & \cdots & r-1 & r-2 & r \end{smallmatrix}\right)\left(\begin{smallmatrix} 1 & 2 & 3 & 4 & 5 & \cdots & r-1 & r \\ 1 & 3 & 2 & 5 & 4 & \cdots & r & r-1 \end{smallmatrix}\right) = \left(\begin{smallmatrix} 1 & 2 & 3 & 4 & 5 & \cdots & r-1 & r \\ 2 & 4 & 1 & 6 & 3 & \cdots & r & r-2 \end{smallmatrix}\right)$$

$$= (1,2,4,6...r-1, r, r-2, r-4...7,5,3)$$

.

A r-cyclic permutation(both even or odd) can be expressed as (r-1) least number of transposition,which is exactly the number of base pair of $\pi(H_n^m) \triangle \pi(H_n^{m'})$ involved in this component.

Case 2. If components are closed then $r \geqslant 4$, even and

$$[w_1^x, w_r^x] \in \pi(H_n^m) \cup \pi(H_n^{m'}),$$

say

$$C = \{w_1^x, w_2^x, w_3^x, w_4^x....w_r^x\}$$

with $\{[w_1^x, w_2^x], [w_3^x, w_4^x], [w_5^x, w_6^x], ...[w_{r-1}^x, w_r^x]\} \in \pi(H_n^m)$ where $x = m$ and $\{[w_2^x, w_3^x], [w_4^x, w_5^x], ...[w_r^x, w_1^x]\} \in \pi(H_n^{m'})$ where $x = m'$, vice versa.

Then,

$P(C, H_n^m)P(C, H_n^{m'}) = (1,2)(3,4)(5,6)...(r-1,r) \quad (2,3)(4,5)(6,7)...(r-2, r-1)(r,1)$

$$= \left(\begin{smallmatrix} 1 & 2 & 3 & 4 & \cdots & r-1 & r \\ 2 & 1 & 4 & 3 & \cdots & r & r-1 \end{smallmatrix}\right)\left(\begin{smallmatrix} 1 & 2 & 3 & 4 & 5 & \cdots & r-1 & r \\ r & 3 & 2 & 5 & 4 & \cdots & r-2 & 1 \end{smallmatrix}\right) = \left(\begin{smallmatrix} 1 & 2 & 3 & 4 & 5 & \cdots & r-1 & r \\ r-1 & 4 & 1 & 6 & 3 & \cdots & r-3 & 2 \end{smallmatrix}\right)$$

$$= (1, r-1, r-3, r-5, ...5,3)(2,4,6,...r).$$

This is product of two disjoint cycle each of length $r/2$.Each cyclic permutation represent $(r/2)-1$ number of transpositions. So total number of transposition in $P(C,H_n^m)P(C,H_n^{m'})$ is less than 2 to the involved in the number of base pair of
$\pi(H_n^m) \vartriangle \pi(H_n^{m'})$ in this component.

Hence

$$d_{inv}(H_n^m, H_n^{m'}) = |\{Base\ pair\ of\ \pi(H_n^m) \vartriangle \pi(H_n^{m'})$$

$|\{involved\ in\ open\ component\}| +$
$|\{Base\ pair\ of\ \pi(H_n^m) \vartriangle \pi(H_n^{m'})\ involved\ in\ closed\ component\}|$
$-2 * (number\ of\ closed\ component)$
$= |\pi(H_n^m) \vartriangle \pi(H_n^{m'})| - 2\Omega(H_n^m, H_n^{m'})$. Hence proved.

In this section, we analysed the difference between two RNA fuzzy structure in fuzzy structure space taking help involution metric space, symmetric group, symmetric difference, etc. Here, we established one-one correspondence relation between base pair in the fuzzy structure using symmetric difference method and transposition in symmetric group, because both represent abstractly similar things, One represents base pair of fuzzy nucleotides in their structure and another represent pair of fuzzy nucleotide position termed as transposition in symmetric group. Through this logic in proposition(5.2.2), we calculated difference between two fuzzy structure via two independent methods and establish a relation between them.

6. Discussion

In this study, we converted nucleotide into ordered fuzzy set and sequence of nucleotides or polynucleotide as fuzzy polynucleotide. In this study, the fuzzy code is the degree of membership of fuzzy set derived from equation(2.2). Evolution occurs through various genetic events, including transversion substitution, transition substitution, recombination, insertion, deletion etc. RNA sequences are frequently used in constructing molecular phylogenetic. Molecular evolutionary of RNA sequence at each stage represents the uniqueness of species characteristic, which was our sequence of interest.Through the following aspect, we wrote unique fuzzy code or encode for each RNA sequence as shown in figure(3.1) with the help of table(1) and table(2). Thus, we have represented each node of phylogenetic trees in terms of fuzzy code. Fuzzy code preserved information of their nucleotides mutation at each stage, which mathematically expressed in equation(3.1,3.1.1,3.1.2). Due to the information preserving property, it helps to determine mutation spots in the evolutionary structure of RNA sequences. This spot can play a better role in the sequence alignment to identify homologies. In reverse, with the help of the decoding technique, we determine the RNA sequence at any stage from their fuzzy code using equation (3.2).

Also, we have constructed a fuzzy graph for the secondary structure representation, where fuzzy nucleotide acted as vertices derived from equation (4.1) and interaction between two fuzzy nucleotides treated as edge , which is hydrogen bond between them. We treated the collection of all encoded fuzzy code of phylogenetic structure as a fuzzy polynucleotide space denoted by Q_n^M. Each fuzzy polynucleotide folds because of WatsonCrick and GU wobble base pairs property to form fuzzy structure(secondary structure). We took collection of all fuzzy structure termed as fuzzy structure space denoted by F_n^M. We developed fuzzy model of secondary structure derived from equation (5.1). Folding of RNA fuzzy polynucleotide into fuzzy structures is viewed as a map that assigns a uniquely defined base pairing pattern to every polynucleotide derived from equation (5.3). As it is conventionally understood for one sequence, we got one secondary structure due to base pair property. We considered many-one mapping instead of conventionally consider one-one mapping because many sequences fold into the same minimum free energy fuzzy structure.

We considered preimage of fuzzy structure as a fuzzy neural network, which is a subspace of fuzzy polynucleotide space. We developed a method for determining the cardinality of fuzzy neural network using equation (5.6), fraction of neural neighbour between two fuzzy polynucleotides using equation (5.7) and average neural neighbour of fuzzy neural network using equation (5.8), which is a subspace of fuzzy polynucleotide space. We analysed the difference between two RNA fuzzy structure in fuzzy structure space through involution metric space, symmetric group, symmetric difference etc; and established a relation between them, which is broadly explained in proposition(5.2.1) and (5.2.2).

References

[1] Michael Waterman, Secondary structures of single stranded nucleic acids, *Adv. Math.(Suppl. Studies)*, **1** (1978), 167-212.

[2] Waterman, M. S , Introduction to Computational Biology: Maps, Sequences, and Genomes, *Biometrics*,(1998). **doi:** 10.2307/2534039.

[3] Forst, C. V., C. Reidys and J.Webe.r , Evolutionary dynamics and optimization: Neutral networks as model-landscapes for RNA secondary-structure folding-landscapes, *ECAL*, (1995). **doi:** 10.1007/3-540-59496-5-294.

[4] Reidys, Christian M, Random Induced Subgraphs of Generalized n-Cubes, *Adv. Appl. Math* , **19** (1997), 360-377. **doi:** 10.1006/aama.1997.0553.

[5] Christian Reidys and Peter F. Stadler, Bio-molecular shapes and algebraic structures, *Computers and Chemistry*, **20** (1996), 85-94. **doi:** 10.1016/S0097-8485(96)80010-6.

[6] F. Rossell, Reidys' and Stadler's metricsfor {RNA} contact structures, *Mathematical and Computer Modelling*, **40** (2004), 771-776. **doi:** 10.1016/j.mcm.2004.10.008.

[7] L.A. Zadeh , Fuzzy sets, *Information and Control*, **8** (1965), 338-353. **doi:** 10.1016/S0019-9958(65)90241-X.

[8] Reidys, Christian and Forst, Christian V. and Schuster, Peter, Replication and mutation on neutral networks, *Bulletin of Mathematical Biology*, **63** (2001), 57-94. **doi:** 10.1006/bulm.2000.0206.

[9] Prabir Bhattacharya, Some remarks on fuzzy graphs, *Pattern Recognition Letters*, **6** (1987), 297-302. **doi:** 10.1016/0167-8655(87)90012-2.

[10] Aditi Gupta and Michael Gribskov, The Role of {RNA} Sequence and Structure in RNAProtein Interactions, *Journal of Molecular Biology* , **409** (2011), 574-587. **doi:** 10.1016/j.jmb.2011.04.007.

[11] Mathews, David H., Predicting a set of minimal free energy RNA secondary structures common to two sequences, *Bioinformatics*,**21** (2005), 2246-2253. **doi:** 10.1093/bioinformatics/bti349.

[12] Zaheri, Maryam and Dib, Linda and Salamin, Nicolas, A Generalized Mechanistic Codon Mode, *Molecular Biology and Evolution*, **31** (2014), 2528-2541. **doi:** 10.1093/molbev/msu196.

[13] Tan, Taison and Bogarad, Leonard D. and Deem, Michael W., Modulation of Base-Specific Mutation and Recombination Rates Enables Functional Adaptation Within the Context of the Genetic Code,*Journal of Molecular Evolution*,**59**(2004),385-399. **doi:** 10.1007/s00239-004-2633-8.

[14] Amy N. Langville and William J. Stewart, The Kronecker product and stochastic automata networks,*Journal of Computational and Applied Mathematics*,**167**(2004),429 - 447.**doi:** 10.1016/j.cam.2003.10.010.

[15] Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter, *Molecular Biology of the Cell*, Garland Science, New York(2002).

[16] Emile Zuckerkandl and Linus Pauling, Molecular disease, evolution, and genic heterogeneity, *Horizons in Biochemistry*, Academic Press, New York, (1962),189-225.

[17] Nasser, Sara and Breland, Adrienne and Harris, Frederick C. and Nicolescu, Monica and Vert, Gregory L.,Fuzzy Genome Sequence Assembly for Single and Environmental Genomes ,*Fuzzy Systems in Bioinformatics and Computational Biology*,Springer Berlin Heidelberg, (2009), 19-44. **doi:** 10.1007/978-3-540-89968-6-2.

[18] Kazem Sadegh-Zadeh,Fuzzy genomes, *Artificial Intelligence in Medicine*, **18**(2000),1-28. **doi:** 10.1016/S0933-3657(99)00032-9.

[19] Torres, Angela and Nieto, Juan J, The fuzzy polynucleotide space: basic properties, *Bioinformatics*, **19** (2003), 587-592. **doi:** 10.1093/bioinformatics/btg032.

[20] J.J. Nieto and A. Torres and M.M. Vzquez-Trasande, A metric space to study differences between polynucleotides, *Applied Mathematics Letters*,**16**(2003),1289 - 1294. **doi:** 10.1016/S0893-9659(03)90131-5.

[21] Nieto, Juan J. and Torres, A. and Georgiou, D. N. and Karakasidis, T. E., Fuzzy polynucleotide spaces and metrics,*Bulletin of Mathematical Biology*,**68**(2006),703–725. **doi:** 10.1007/s11538-005-9020-5.

[22] Dong Xu and R. Bondugula and M. Popescu and J. Keller, Bioinformatics and Fuzzy Logic ,*IEEE International Conference on Fuzzy Systems*,(2006),817-824. **doi:** 10.1109/FUZZY.2006.1681805.

[23] Friedrich Steimann, On the use and usefulness of fuzzy sets in medical {AI} ,*Artificial Intelligence in Medicine* ,**21**(2001),131 - 137. **doi:** 10.1016/S0933-3657(00)00077-4.

[24] Drake, John W. and Charlesworth, Brian and Charlesworth, Deborah and Crow, James F., Rates of Spontaneous Mutation, *Genetics*, **148**(1998),1667-1686.

[25] Schneider, Stefan and Excoffier, Laurent., Estimation of Past Demographic Parameters From the Distribution of Pairwise Differences When the Mutation Rates Vary Among Sites: Application to Human Mitochondrial DNA, *Genetics*,**152**(1999),1079-1089. **doi:** 10.1214/aoms/1177728268.

[26] Nachman, Michael W. and Crowell, Susan L.,Estimate of the Mutation Rate per Nucleotide in Humans,*Genetics*,**156**(2000),297-304.

[27] Roach, Jared C. and Glusman, Gustavo and Smit, Arian F. A. and Huff, Chad D. and Hubley, Robert and Shannon, Paul T. and Rowen, Lee and Pant, Krishna P. and Goodman, Nathan and Bamshad, Michael and Shendure, Jay and Drmanac, Radoje and Jorde, Lynn B. and Hood, Leroy and Galas, David J., Analysis of Genetic Inheritance in a Family Quartet by Whole-Genome Sequencing, *Science*,**328**(2010),636-639. **doi:** 10.1126/science.1186802.

[28] De Bruyn, Alexandre and Martin, Darren P. and Lefeuvre, Pierre, Phylogenetic Reconstruction Methods: An Overview, *Molecular Plant Taxonomy: Methods and Protocols*,Humana Press,(2014),257–277. **doi:** 10.1007/978-1-62703-767-9-13.

[29] Kerpedjiev, Peter and Hammer, Stefan and Hofacker, Ivo L.,Forna (force-directed RNA): Simple and effective online RNA secondary structure diagrams, *Bioinformatics*,(2015). **doi:** 10.1093/bioinformatics/btv372.

[30] Inui, Teturo and Tanabe, Yukito and Onodera, Yositaka, The Symmetric Group, *Group Theory and Its Applications in Physics*,Springer Berlin Heidelberg,(1990)333-359. **doi:** 10.1007/978-3-642-80021-4-15.