Journal of Information Assurance and Security.
ISSN 1554-1010 Volume 9 (2014) pp. 255-267

Computational Study of secondary structure of Premature miRNA with the help of Rough sets

Joyshree Nath

[1] Indira Gandhi National Open University, St. Xavier's College(Autonomous), Kolkata
Joshreenath@gmail.com

Abstract: **In the past years quite a lot of research has been done on pre-mature and mature microRNAs (miRNAs). Scientists show huge interest in this domain, as this domain is still not explored fully. The miRNAs (both plant and animal) are believed to have huge genetic information. Quite an amount of work has been done in the following genres namely: discovering proper miRNA structures and functions, mapping their relation with genetic disorders, discovering correct miRNA targets and lots more. So in experimental or computational genomics generation of a perfectly correct miRNA structure is an important task. This paper is a systematic study to get a proper secondary structure of the pre-mature miRNA and get the mature miRNA sequence corresponding to it with the help of rough sets and finite state automation. A higher accuracy on obtaining secondary structures of pre-mature miRNAs using computational methods will pave a new way towards certifying novel miRNAs. This might reduce the cost of determining secondary structures through free energy minimizing methods and other experimental methods. In the present paper the authors have introduced rough sets and finite state automation to determine and represent the secondary structure of any given pre-mature miRNA sequence.**

Keywords: miRNA; genetic disorder;, pre-mature miRNA;, , structural behavior;, secondary structure;, rough set approach;, finite state automation in computational genomics;

I. Introduction

A voluminous amount of work is done in the domain of mature miRNAs and their target mRNAs. Lesser has been done in the field of premature miRNAs. In our previous work [1] and [2] we have dealt with proper computational quantification of pre-mature miRNAs for a few species, like, gorilla, human, and orangutan(pongo pygmaeus). This work concentrates on predicting the secondary structures of the miRNAs thus extracted in the premature miRNAs. Then the mature miRNAs are predicted from those structures. Whether those predicted structures match with those in the database of miRBase is also checked in this paper. In the next section algorithms used in the previous paper [2] are discussed. In section 2, the background study on miRNAs are given. In section 3, related works with respect to computational prediction of premature miRNA secondary structure is discussed. In section 4 certain basic concepts of rough sets are given. In section 5, the algorithm used in previous paper is shown. In section 6, the algorithm for the current paper is discussed. In section 7, the results obtained from using section 5 methods are discussed. In section 7 the conclusion and future scope of the method is discussed.

II. Background Study

A. Introduction to genes and non functional genes:

The sequence of DNA is normally known as a gene sequence. DNA or deoxy-ribonucleic acid is a long chain of nucleotides arranged in a double helical structure. The entire DNA sequence is not composed of genes. The functional part of the DNA ultimately leads to protein synthesis, a phenomena known as **central dogma**. There is also a non functional part of the DNA. A huge amount of this non functional gene sequence was earlier considered as **junk** by scientists. Junk DNA corresponded to the non-coding part of the gene that do not encode for protein sequences. Previously this part was not considered to be useless. But for around 10-12 years this area is exploited and a new kind of gene sequence called "microRNA(miRNA)" are found to be prevalent in this domain[3].

B. Biogenesis of miRNA

MiRNAs are a class of conserved short non-coding RNAs. Each miRNA is first transcribed as a longer RNA sequence called pri-miRNA. This resultant pri-miRNA is processed into hairpin like RNA molecule, called pre-miRNA. A ribonuclease called Dicer cleaves it to form a very short double-stranded RNA. This shorter pre-miRNA attaches to the protein complex RISC (RNAi-induced silencing complex. This complex degrades one of the strands leaving the other RNA strand, the mature strand. The mature strand binds to its target mRNA or messenger RNA sequence. Now this mRNA or messenger RNA carries the direction of protein synthesis by ribosomes. So by binding to the corresponding mRNA, a miRNA regulates the behavior of the post-transcriptional genes. Thus miRNAs could result in repressing the end protein products. These proteins constitute the major portion human body and supply energy for the body. So miRNAs have the power to effect the proper functioning of human body. In order to understand functional roles of the miRNAs, accurate prediction of miRNA-target pairs are necessary [4,5,6]. Proper functioning of miRNAs can lead to biological processes like lung, cardiac development and improper functioning and

expression level of miRNAs are found and predicted to be involved in diseases like, fragile X syndrome, dementia, Alzheimer's disease, etc [7]. Microarray profiling method measure the expression level of different miRNAs in human and other animals. Such experimental expression values can provide knowledge in the behavioral pattern of any particular disease in human body and other animals [8,9,10,11].

C. *MiRNA and gene expression*

MiRNA are small non coding RNAs which are a part of the (previously considered) unused DNA. It leads to the regulation of gene in the post-transcriptional stage. This means that, the genes which would be otherwise expressed would be under expressed or not expressed due to the effects of miRNAs. Along with gene expression, protein expression is also controlled by miRNAs, through their target mRNAs. Still a lot of experiments are done to know the exact procedure of this regulation [5],[6].

B. MiRNA are small non coding RNAs which are a part of the "unused DNA". It leads to the regulation of gene in the post-transcriptional stage. This means that, the genes which would otherwise be expressed, would be under or not expressed at all due to the effects of the miRNAs. Along with gene expression, protein expression is also controlled by miRNAs, through their target mRNAs. Still now a lot of experiments are going on to know the exact procedure of this regulation [5],[6].

D. MiRNA and diseases

It has been found that miR-133a1 is involved in the correct functioning of heart. So it can be pursued that abnormal expressions of this miRNA may lead to cardiac diseases. The expression level of miR-21 increases in cardiac hypertrophy. So somehow if the expression is deregulated experimentally, such cardiac state could possibly be arrested. The miRNAs like miR-17, miR-18, miR-20, miR-21, and quite others are found involved in various types of cancers in human body. Either they are up-regulated or down-regulated in higher stages of cancer. Experiments are continued to see how such factors can be utilized for treatments in cancer. Experiments suggest that over or under expression of miRNAs might be involved in diseases like, Schizophrenia, Alzheimer's disease, Parkinson's disease [6,7].

C. Chemicals and its effects on miRNAs

Over exposure to chemicals like cadmium, aluminum, arsenic, zinc, etc., which are harmful to body, lead to the modification of miRNA expressions. The miRNAs target mRNAs. The mRNAs again subsequently modulate protein formation. So it can be said that miRNAs indirectly modulate gene and protein expressions. Either gene and in turn protein formation gets too much inhibited or over-expressed by the effect of these chemicals. Thus heart attacks, neural disorders, limb malfunction, hormonal imbalance, cancer and even death can occur due to the harmful effect of these chemicals on the human miRNAs. Although how exactly the chemicals help in changing the expression of miRNAs is yet to be discovered [6,12,13].

D. MiRNAs and proper target prediction

Quite a lot of research is continued to find the proper mRNA targets and its implications (if any) on the diseases. It is found that the 2-8 nucleotides in the 5' end of the miRNA (called the seed region) targets it's corresponding complementary areas in the 3'UTR portion of the mRNA gene. The non seed regions, 5'UTR regions of mRNAs are also explored for target binding sites. But it is still maintained that the most functional target binding is along the seed site of miRNAs. Lots of target prediction algorithms are there to computationally and experimentally assess correct miRNA targets. For example, miRanda and TargetScan uses base pairing alignment of miRNAs and mRNAs as a category, RNAHybrid mainly considers the proper mRNA secondary structure suitable for it's respective miRNA, probabilistic computational model is used in NBMiRTar [5,6,9,14,15,16].

III. Rough Sets

A. Introduction to Rough Sets

Today we work with numerous data and thus numerous databases. This is also the case in the domain of bio informatics. We need to analyze such huge amount of raw data to extract pure information pattern and relationship between data from it. Rough set theory was developed by Zdzis law Pawlak in the early 1980's. The theory of rough sets is a mathematical tool for extracting knowledge from uncertain and incomplete and inconsistent data based information.Through this method basically raw data tables obtained from natural phenomenon, from various resources are analyzed and classified. The main goal of the rough set analysis is to synthesize approximation of concepts from the acquired data. In the data table, that is analyzed, each row represents an event or object. Every column represents an attribute or a property with respect to that object. This is a tool which is evolving nowadays to mine the huge amount of data that is available in every genre all over the world. After mining such huge amount of data we are liable to get the desired information based upon any valuable theory about the data can be revealed. Analysis of medical data is often concerned with treatment of incomplete knowledge, with management of inconsistent pieces of information. So it is quite likely that rough sets could bring up new data patterns and provide new medical knowledge. [17]

B. Some Important Definitions of Rough Sets

- Rough Sets help in finding minimal sets of data (data reduction) and in turn helps in easy and direct interpretation of huge amount of data.
- If there is an information system S then in terms of rough sets we can define S= (U, A) where U is a finite set of objects, U = $\{x_1, x_2, x_3, \ldots, x_n\}$ and A is a finite set of attributes (features), the attributes in A.
- Each f_q: U → V_q is an information function which assigns particular values from domains of attributes V_q to objects such that $fq(x_i) \in V_q$ for all xi ∈ U and q ∈ A.

- Objects characterized by the same information are considered indiscernible or similar. They form what is called equivalence classes. [18].
- Approximation of an information system S means, using only the information contained in the set of attributes A1, where A1 ⊆A . by constructing the lower and upper approximations of X, lower approximations ={x | $[x]_{A1} \subseteq X$} , where $X \subseteq U$ and upper approximations= {x | $[x]_{A1} \cap X \neq \varphi$}.
- The set difference of lower and upper approximation set gives something called the boundary region of X, and thus consists of those objects that we cannot decisively classify into X.
- A set is said to be rough if the boundary region is non-empty.[19]
- Some attributes of an information system may be redundant with respect to a specific classification [a] generated by attributes a⊆ A.
- With the help of dependency properties of attributes, one can find a reduced set of the attributes by removing redundant attributes, without loss of the classification power.
- A reduct is a set of necessary minimum data to maintain the original properties of the information table.
- Reducts keep only those attributes that preserve the indiscernibility relation and, consequently, set approximation.
- The core contains all the attributes that cannot be removed from the set attribute set A without changing the original classification [a] [20].

IV. Algorithm used previously to predict secondary structure of pre-mature miRNAEquations

A. Extracting datasets:

The miRNA sequences for Homo sapiens (hsa), Gorilla gorilla (ggo) and Pongo pygmaeus (ppy) from miRBase version 19 were extracted and saved in separate text files.

B. Predicting possible secondary structures from linear structures:

Every miRNA sequence is taken and with the following method the probable secondary structures are calculated.

- The initial point or nucleotide of the primary sequence, to be considered for the secondary structure calculation, is taken as s and the finishing point or nucleotide is taken f.

- The s and f point can be placed anywhere in the sequence, the only thing is that the position of f should be at least 4 nucleotides further s(f-s>3). This is because a minimum of 3 nucleotides will be required to form a hairpin loop of pre-mature miRNA. Two types of secondary structure formations are considered. One type is where s is incremented and f is kept fixed and another one is where f is incremented and s is fixed.

- Some amount of energy called the penalty energy for loops like hairpin and internal is added with the final negative energy score. This energy is considered for nucleotide number starting from minimum of 3 and 4 for hairpin and internal loop. For stacking energies six valid base pairs-AU,UA,CG,GC,GU,UG are considered in a stack, and their available experimentally validated energy values are used to calculate the final energy score.

- The interaction of nucleotides for which there are no values, a zero is considered and nothing is thus added for that interaction in the final energy value.

- The energy values for such structures are extracted from the tables obtained from [21, 22, 23, 24, 25]. All these energy values are combined to generate a rough set data table in the next section. The secondary structure with the minimum energy value is considered is the best probable secondary structure.

C. Extracting mature miRNA sequences

Mature miRNA sequences were extracted from the probable pre-mature miRNA secondary structures in this section. Very simple computational method was used and it was intended that machine learning would be clubbed with the current procedure and better dicer binding cleavage sites from the pre-mature miRNAs, so that mature miRNAs could be detected in better and efficient way. In our previous paper the concepts of [26] was considered and a window of 20-26 sequence length was placed over the miRNA strings. In this way many miRNA sequences were extracted starting from 3' end to and before the beginning of the hairpin loop of the premature miRNA. Also same window was maintained from the 5' end right after the end of the hairpin loop. Each time with the change in the window length a different miRNA sequence was extracted from the same region of premature miRNA. Then such sequences were blasted in miRBase database to check for sequence similarity.

D. Assessment of characteristics of mature miRNA sequences

The computationally obtained mature miRNAs were passed through the same statistical measures and ranges of the values were obtained. The measures used were calculating the frequency distribution of the nucleotides and measuring the GC content of the nucleotides. These resultant values were tallied with the range of actual values obtained from the actual mature miRNAs obtained from miRBase database.

E. Using a Finite State Machine to represent the selected secondary structure of the pre-mature miRNAs

In the previous paper a finite state machine was used to represent the generated secondary structure of the premature miRNAs through finite state automata. Only dinucleotides, like AG, AU, CU, etc. were taken as valid states to represent the states of the corresponding state transition table. Thus the final state for that corresponding sequence is obtained. This FSM that is designed will be incorporated in generating the

secondary structure of the premature miRNAs in the future works.

V. Results obtained in the previous work

In this sections the author has given results obtained from different datasets.

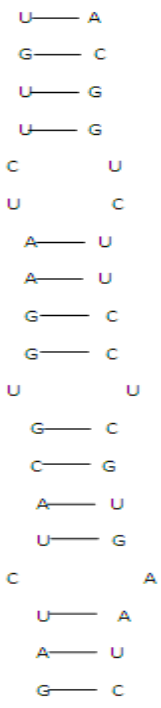
A. Extracting datasets

The pre-mature miRNA sequences were extracted for the three species, Homo sapiens (hsa), Gorilla gorilla (ggo) and Pongo pygmaeus (ppy) from miRBase version 19 and saved in separate text files.
For example, Gorilla gorilla (ggo) the content of 5.txt came as the sequence for, ggo-let-7f MI0020620
AAAACAUUGCUCUAUCAGAGUGAGGUAGUAGAUU GUAUAGUUGUGGGGUAGUGAUUUUACCCUGUUCA GGAGAUAACUAUACAAUCUAUUGCCUUCCCUGAGG AGUAGAC

B. Predicting all possible secondary structures and selecting the one which is potentially the most stable

Let us take an example of ggo sequence and see the possible secondary structures and the mature miRNA sequences subsequently.
The ggo sequence considered is of, ggo-mir-18a MI0002966

UGUUCUAAGGUGCAUCUAGUGCAGAUAGUGAAGU AGAUUAGCAUCUACUGCCCUAAGUGCUCCUUCUGG CA
Taking s at the first 5’ nucleotide, that is U and taking f at the last 3’nucleotide, that is, A, the secondary structure comes like this.



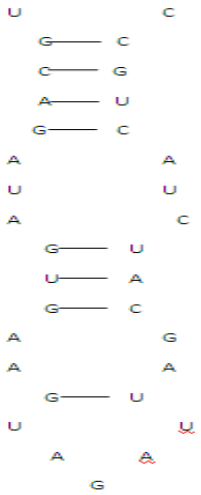


Figure-1:Secondary structure formed of premature miRNA sequence ggo-mir-18a.

The bonding in Figure-1 is showed with a ‘-‘ sign between valid base pairs only. This structure has the most valid stack base pairs. The total energy for this structure comes as ~ -24.56, after adding the energy penalties for hairpin and internal loops. The negative value shows that it is quite stable and it almost matches with the secondary structure from miRBase database.

C. Extraction of mature miRNA sequences and Blasted sequences

Now using [26], the possible mature miRNA sequences were extracted from this above secondary structure of length 20-26 from both 5’end and 3’end. Few of the predicted sequences were UCUUCCUCGUGAAUCCGUCA, CGUGAAUCCCGUCAUCUACGAU,CUUCCUCGUGAA UCCGUCAUC, from 3’end. Among these the later one when blasted in miRBase, for sequence similarity, got a sequence alignment with homo sapiens, which could be worth exploring by biologists.

Accession	ID	Query start	Query end	Subject start	Subject end	Strand	Score	Evalue	Alignment
MIMAT0021117	hsa-miR-5187-5p	3	21	4	22	-	59	8.6	Align

Alignment of Query to mature miRNAs

Query: 3-21 hsa-miR-5187-5p: 4-22 score: 59 evalue: 8.6

```
UserSeq        21 uccucgugaauccgucauc 3
                  ||| | | |||||| |||||
hsa-miR-5187-5p 4 uccacuucaaucccucauc 22
```

Search parameters

Search algorithm:	BLASTN
Sequence database:	mature
Evalue cutoff:	10
Max alignments:	100
Word size:	4
Match score:	+5
Mismatch penalty:	-4
Species filter:	ath, cel, dme, hsa, mmu

Figure-2: Blast results for the mature miRNA sequence CUUCCUCGUGAAUCCGUCAUC obtained in miRBase.

Also the sequences UAAGGUGCAUCUAGUGCAGAUA, (exact match with the miRBase database of mature miRNA), and GCAUCUAGUGCAGAUAGUGAG,

CUAAGGUGCAUCUAGUGCAGAUAG from 5'end were obtained.

a. Estimation of few characteristics of such mature miRNA sequences : Estimation of GC Content

The formula for GC content calculation is [(G+C)/(A+C+G+U)]*100 . For the first predicted sequence=>UCUUCCUCGUGAAUCCGUCAUC the GC content comes as 50%. For the second predicted sequence=> UAAGGUGCAUCUAGUGCAGAUA(miRBase validated sequence) the GC content comes as 41%, which is close to the previous sequence.

b. Estimation of few characteristics of such mature miRNA sequences : Frequency analysis of the nucleotides

First predicted sequence=UCUUCCUCGUGAAUCCGUCAUC
Length of Sequence=22

Character	Frequency	Percentage
A	1	4.55
AA	1	4.55
C	4	18.18
CC	2	9.09
G	3	13.64
U	6	27.27
UU	1	4.55

Second predicted sequence= UAAGGUGCAUCUAGUGCAGAUA(miRBase validated sequence)
Length of Sequence=22

Character	Frequency	Percentage
A	5	22.73
AA	1	4.55
C	3	13.64
G	4	18.18
GG	1	4.55
U	6	27.27

D. Finite State Automation of different structure in the following nucleotides

UGUUCUAAGGUGCAUCUAGUGCAGAUAGUGAAGU AGAUUAGCAUCUACUGCCCUAAGUGCUCCUUCUGG CA

We form the table-VI(a)

Invalid	Valid Pair		Invalid	Valid transition	Transition Energy
	U	A		UA→GC	-2.4
	G	C		GC→UG	-1.5
	U	G		UG→UG	+0.3
	U	G			
C			U		
U			C		
	A	U		AU→AU	-1.1
	A	U		AU→GC	-2.2
	G	C		GC→GC	-3.4
	G	C			
U			U		
	G	C		GC→CG	-3.3
	C	G		CG→AU	-3.08
	A	U		AU→UG	-0.6
	U	G			
C			A		
				UA→AU	-1.3
				AU→UA	-2.2
U			C		
	G	C		GC→CG	-3.3
	C	G		CG→AU	-3.08
	A	U		AU→GC	-2.2
	G	C			

We represent the Finite State Automation as follows:
Q=set of non-empty states={UA,AU,UG,CG,GC,GU,CU}
Σ=Set of non-empty energy={-1.3,-2.4,-2.2,-3.3,-3.08,-1.10,-0.6,-1.50,+0.3,0.0}
Initial state={UA}
Final state={GC}

Table-VI(b) : Transition table of Finite state automation

Present State	Next State Energy									
	-1.3	-2.4	-2.2	-3.3	-3.08	-1.10	-0.6	-1.5	+0.3	0.0
→UA	AU	GC								
GU	UA									
AU			GC			AU	UG			
CU										AU
UG									UG	CU
CG					AU					
GC			GC	CG					UG	

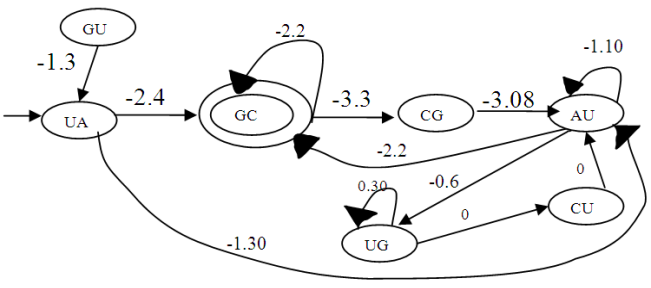


Figure-3:Finite State Automation of Secondary Structure of a pre-mature miRNA of ggo

VI. Algorithm Used in this Work

A. Preparing datasets

At first the energy value tables were extracted from [21,22,23,24,25,27] for nucleotide interactions in hairpin loops, internal loops, bulge loops, single base interactions and most importantly stacking base interactions. Below there are 4 tables each depicting a type energy value interaction between the nucleotides. In the next subsection we show the combined energy tables forming the initial rough set data table to work upon. The following energy values are calculated taking two things into consideration-1) the valid base pairs the Watson-Crick base pairs and G:U wobble pair and 2)these energy values are calculated experimentally at 37ºC temperature

TABLE –I : FREE ENERGIES FOR STACKING BASE INTERACTIONS FOR VALID BASE PAIRS [21, 22, 23, 24, 25,27]

	AU	**UA**	**CG**	**GC**	**GU**	**UG**
AU	-0.9	-1.1	-2.2	-2.1	-0.6	-1.4
UA	-1.3	-0.9	-2.4	-2.1	-1.0	-1.3
CG	-2.1	-2.1	-3.3	-2.4	-1.4	-2.1
GC	-2.4	-2.2	-3.4	-3.3	-1.5	-2.5
GU	-1.3	-1.4	-2.5	-2.1	-0.5	1.3
UG	-1.0	-0.6	-1.5	-1.4	-0.3	-0.5

TABLE –II : FREE ENERGIES FOR MISMATCHED BASE PAIRS FORMING A HAIRPIN LOOP WITH A VALID CLOSING BASE PAIR [21, 22, 23, 24, 25,27]

	AA	AC	AG	AU	CA	CC	CG	CU	GA	GC	GG	GU	UA	UC	UG	UU
AU	-0.3	-0.5	-0.3	-0.3	-0.1	-0.2	-1.5	-0.2	-1.1	-1.2	-0.2	-0.2	-0.3	-0.3	-0.5	-1.1
UA	-1.5	-1.5	-1.4	-1.8	-1.0	-0.9	-2.9	-0.8	-2.2	-2.0	-1.6	-1.1	-1.7	-1.4	-1.8	-2.0
CG	-1.5	-1.5	-1.4	-1.8	-1.0	-0.9	-2.9	-0.8	-2.2	-2.0	-1.6	-1.1	-1.7	-1.4	-1.8	-2.0
GC	-1.1	-1.5	-1.3	-2.1	-1.1	-0.7	-2.4	-0.5	-2.4	-2.9	-1.4	-1.2	-1.9	-1.0	-2.2	-1.5
GU	0.2	-0.3	-0.3	-0.3	-0.1	-0.2	-1.5	-0.2	-0.9	-1.1	-0.3	0.0	-0.3	-0.3	-0.4	-1.1
UG	-0.5	-0.3	-0.6	-0.5	-0.2	-0.1	-1.7	0.0	-0.8	-1.2	-0.3	-0.7	-0.6	-0.1	-0.6	-0.8

TABLE –III : FREE ENERGIES FOR MISMATCHED BASE PAIRS FORMING A INTERNAL LOOP WITH A VALID CLOSING BASE PAIR [21, 22, 23, 24, 25,27]

	AA	AC	AG	AU	CA	CC	CG	CU	GA	GC	GG	GU	UA	UC	UG	UU
AU	0.7	0.7	-0.4	0.7	0.7	0.7	0.7	0.7	-0.4	0.7	0.7	0.7	0.7	0.7	0.7	0.0
UA	0.7	0.7	-0.4	0.7	0.7	0.7	0.7	0.7	-0.4	0.7	0.7	0.7	0.7	0.7	0.7	0.0
CG	-0.0	-0.0	-1.1	-0.0	-0.0	-0.0	-0.0	-0.0	-1.1	-0.0	-0.0	-0.0	-0.0	-0.0	-0.0	-0.7
GC	-0.0	-0.0	-1.1	-0.0	-0.0	-0.0	-0.0	-0.0	-1.1	-0.0	-0.0	-0.0	-0.0	-0.0	-0.0	-0.7
GU	0.7	0.7	-0.4	0.7	0.7	0.7	0.7	0.7	-0.4	0.7	0.7	0.7	0.7	0.7	0.7	0.0
UG	0.7	0.7	-0.4	0.7	0.7	0.7	0.7	0.7	-0.4	0.7	0.7	0.7	0.7	0.7	0.7	0.0

TABLE –IV (A): FREE ENERGIES FOR MISMATCHED BASE PAIRS FORMING SINGLE BASE INTERACTIONS WITH VALID BASE PAIRS AT 5’ END [21, 22, 23, 24, 25,27]

	A	C	G	U
AU	-0.8	-0.5	-0.8	-0.6
UA	-0.7	-0.1	-0.7	-0.1
CG	-1.7	-0.8	-1.7	-1.2
GC	-1.1	-0.4	-1.3	-0.6
GU	-0.8	-0.5	-0.8	-0.6
UG	-0.7	-0.1	-0.7	-0.1

TABLE –IV (A): FREE ENERGIES FOR MISMATCHED BASE PAIRS FORMING SINGLE BASE INTERACTIONS WITH VALID BASE PAIRS AT 3’ END [21, 22, 23, 24, 25,27]

	A	C	G	U
AU	-0.3	-0.1	-0.2	-0.2
UA	-0.3	-0.3	-0.4	-0.2
CG	-0.2	-0.3	-0.0	-0.0
GC	-0.5	-0.3	-0.2	-0.1
GU	-0.3	-0.1	-0.2	-0.2
UG	-0.3	-0.3	-0.4	-0.2

B. *Conversion of the dataset into Rough Set data table*

On one hand a table(Table 6) is prepared with O column gives the objects as defined below. The column A1 gives the attribute value for each object with respect to adenine nucleotide in a certain condition (as defined below) upto object $x82_{G1}$. The column A2 gives the attribute value for each object with respect to cytosine nucleotide in a certain condition (as defined below) upto object $x82_{G1}$. The column A3 gives the attribute value for each object with respect to guanine nucleotide in a certain condition (as defined below) upto object $x82_{G1}$. The column A4 gives the attribute value for each object with respect to uracil nucleotide in a certain condition (as defined below) upto object $x82_{G1}$. If in any column no experimental energy value is obtained or yet

validated then that column value is kept '-'. So wherever there is a '-' it signifies no energy value applicable for that cell.

The following are the energy values or objects for stacking interactions between the watson crick base pairs and G:U wobble pair.

$x1_A$→ energy value or object for having adenine stacked on AU with its only other valid nucleotide companion uracil.
$x2_C$→ energy value or object for having cytosine stacked on AU with its only other valid nucleotide companion guanine.
$x3_G$→ energy value or object for having guanine stacked on AU with its two other valid nucleotide companions guanine and uracil.
$x4_U$→ energy value or object for having uracil stacked on AU with its two other valid nucleotide companions guanine, cytosine.

$x5_A$→ energy value or object for having adenine stacked on UA with its only other valid nucleotide companion uracil.
$x6_C$→ energy value or object for having cytosine stacked on UA with its only other valid nucleotide companion guanine.
$x7_G$→ energy value or object for having guanine stacked on UA with its two other valid nucleotide companions guanine and uracil.
$x8_U$→ energy value or object for having uracil stacked on CG with its two other valid nucleotide companions guanine, cytosine.
$x9_A$→ energy value or object for having adenine stacked on CG with its only other valid nucleotide companion uracil.
$x10_C$→ energy value or object for having cytosine stacked on CG with its only other valid nucleotide companion guanine.
$x11_G$→ energy value or object for having guanine stacked on CG with its two other valid nucleotide companions guanine and uracil.
$x12_U$→ energy value or object for having uracil stacked on CG with its two other valid nucleotide companions guanine, cytosine.

$x13_A$→ energy value or object for having adenine stacked on GC with its only other valid nucleotide companion uracil.
$x14_C$→ energy value or object for having cytosine stacked on GC with its only other valid nucleotide companion guanine.
$x15_G$→ energy value or object for having guanine stacked on GC with its two other valid nucleotide companions guanine and uracil.
$x16_U$→ energy value or object for having uracil stacked on GC with its two other valid nucleotide companions guanine, cytosine.

$x17_A$→ energy value or object for having adenine stacked on GU with its only other valid nucleotide companion uracil.
$x18_C$→ energy value or object for having cytosine stacked on GU with its only other valid nucleotide companion guanine.
$x19_G$→ energy value or object for having guanine stacked on GU with its two other valid nucleotide companions guanine and uracil.
$x20_U$→ energy value or object for having uracil stacked on GU with its two other valid nucleotide companions guanine, cytosine.

$x21_A$→ energy value or object for having adenine stacked on UG with its only other valid nucleotide companion uracil.
$x22_C$→ energy value or object for having cytosine stacked on UG with its only other valid nucleotide companion guanine.
$x23_G$→ energy value or object for having guanine stacked on UG with its two other valid nucleotide companions guanine and uracil.
$x24_U$→ energy value or object for having uracil stacked on UG with its two other valid nucleotide companions guanine, cytosine.

The following are the energy values or objects for having a hairpin loop with different valid (watson-crick and G:U wobble base pair) closing base pairs and having any base pair other than Watson-Crick and G:U wobble base pair as base pairs.

$x25_A$→energy value or object for having a hairpin loop with adjoining adenine nucleotide and AU as the closing base pair.
$x26_C$→ energy value or object for having a hairpin loop with adjoining cytosine nucleotide and AU as the closing base pair.
$x27_G$→ energy value or object for having a hairpin loop with adjoining guanine nucleotide and AU as the closing base pair.
$x28_U$→ energy value or object for having a hairpin loop with adjoining uracil nucleotide and AU as the closing base pair.
$x29_A$→ energy value or object for having a hairpin loop with adjoining adenine nucleotide and UA as the closing base pair.
$x30_C$→ energy value or object for having a hairpin loop with adjoining cytosine nucleotide and UA as the closing base pair.
$x31_G$→ energy value or object for having a hairpin loop with adjoining guanine nucleotide and UA as the closing base pair.
$x32_U$→ energy value or object for having a hairpin loop with adjoining uracil nucleotide and UA as the closing base pair.

$x33_A$→ energy value or object for having a hairpin loop with adjoining adenine nucleotide and CG as the closing base pair.
$x34_C$→ energy value or object for having a hairpin loop with adjoining cytosine nucleotide and CG as the closing base pair.
$x35_G$→ energy value or object for having a hairpin loop with adjoining guanine nucleotide and CG as the closing base pair.
$x36_U$→ energy value or object for having a hairpin loop with adjoining uracil nucleotide and CG as the closing base pair.

$x37_A$→ energy value or object for having a hairpin loop with adjoining adenine nucleotide and GC as the closing base pair.
$x38_C$→ energy value or object for having a hairpin loop with adjoining cytosine nucleotide and GC as the closing base pair.
$x39_G$→ energy value or object for having a hairpin loop with adjoining guanine nucleotide and GC as the closing base pair.
$x40_U$→ energy value or object for having a hairpin loop with adjoining uracil nucleotide and GC as the closing base pair.

$x41_A$→ energy value or object for having a hairpin loop with adjoining adenine nucleotide and GU as the closing base pair.
$x42_C$→ energy value or object for having a hairpin loop with adjoining cytosine nucleotide and GU as the closing base pair.
$x43_G$→ energy value or object for having a hairpin loop with adjoining guanine nucleotide and GU as the closing base pair.
$x44_U$→ energy value or object for having a hairpin loop with adjoining uracil nucleotide and GU as the closing base pair.
$x45_A$→ energy value or object for having a hairpin loop with adjoining adenine nucleotide and UG as the closing base pair.

$x46_C$→ energy value or object for having a hairpin loop with adjoining cytosine nucleotide and UG as the closing base pair.
$x47_G$→ energy value or object for having a hairpin loop with adjoining guanine nucleotide and UG as the closing base pair.
$x48_U$→ energy value or object for having a hairpin loop with adjoining uracil nucleotide and UG as the closing base pair.

The following are the energy values or objects for having an internal loop with different valid (watson-crick and G:U wobble base pair) closing base pairs and having any base pair other than watson-crick and G:U wobble base pair as base pairs.

$x49_A$→ energy value or object for having a internal loop with adjoining adenine nucleotide and AU as the closing base pair.
$x50_C$→ energy value or object for having a internal loop with adjoining cytosine nucleotide and AU as the closing base pair.
$x51_G$→ energy value or object for having a internal loop with adjoining guanine nucleotide and AU as the closing base pair.
$x52_U$→ energy value or object for having a internal loop with adjoining uracil nucleotide and AU as the closing base pair.
$x53_A$→ energy value or object for having a internal loop with adjoining adenine nucleotide and UA as the closing base pair.
$x54_C$→ energy value or object for having a internal loop with adjoining cytosine nucleotide and UA as the closing base pair.
$x55_G$→ energy value or object for having a internal loop with adjoining guanine nucleotide and UA as the closing base pair.
$x56_U$→ energy value or object for having a internal loop with adjoining uracil nucleotide and UA as the closing base pair.

$x57_A$→ energy value or object for having a internal loop with adjoining adenine nucleotide and CG as the closing base pair.
$x58_C$→ energy value or object for having a internal loop with adjoining cytosine nucleotide and CG as the closing base pair.
$x59_G$→ energy value or object for having a internal loop with adjoining guanine nucleotide and CG as the closing base pair.
$x60_U$→ energy value or object for having a internal loop with adjoining uracil nucleotide and CG as the closing base pair.
$x61_A$→ energy value or object for having a internal loop with adjoining adenine nucleotide and GC as the closing base pair.
$x62_C$→ energy value or object for having a internal loop with adjoining cytosine nucleotide and GC as the closing base pair.
$x63_G$→ energy value or object for having a internal loop with adjoining guanine nucleotide and GC as the closing base pair.
$x64_U$→ energy value or object for having a internal loop with adjoining uracil nucleotide and GC as the closing base pair.

$x65_A$→ energy value or object for having a internal loop with adjoining adenine nucleotide and GU as the closing base pair.
$x66_C$→ energy value or object for having a internal loop with adjoining cytosine nucleotide and GU as the closing base pair.
$x67_G$→ energy value or object for having a internal loop with adjoining guanine nucleotide and GU as the closing base pair.
$x68_U$→ energy value or object for having a internal loop with adjoining uracil nucleotide and GU as the closing base pair.
$x69_A$→ energy value or object for having a internal loop with adjoining adenine nucleotide and GU as the closing base pair.
$x70_C$→ energy value or object for having a internal loop with adjoining cytosine nucleotide and UG as the closing base pair.
$x71_G$→ energy value or object for having a internal loop with adjoining guanine nucleotide and UG as the closing base pair.
$x72_U$→ energy value or object for having a internal loop with adjoining uracil nucleotide and UG as the closing base pair.

The following are the energy values or objects for having a single base with different valid (watson-crick and G:U wobble base pair) closing base pairs and having any single nucleotide (A/C/G/U) adjoined to it.

$x73_A$→ energy value or object for having AU as the closing base pair, with a single nucleotide attached to A.
$x74_C$→ energy value or object for having CG as the closing base pair, with a single nucleotide attached to C.
$x75_G$→ energy value or object for having GC as the closing base pair, with a single nucleotide attached to G.
$x76_U$→ energy value or object for having UA or UG (as both lead to same 4 values) as the closing base pair, with a single nucleotide attached to U.
$x77_{G1}$→ energy value or object for having GU as the closing base pair, with a single nucleotide attached to G.

$x78_A$→ energy value or object for having UA as the closing base pair, with a single nucleotide attached to A.
$x79_C$→ energy value or object for having GC as the closing base pair, with a single nucleotide attached to C.
$x80_G$→ energy value or object for having CG as the closing base pair, with a single nucleotide attached to G.
$x81_U$→ energy value or object for having AU or GU (as both lead to same 4 values)as the closing base pair, with a single nucleotide attached to U.
$x82_{G1}$→ energy value or object for having UG as the closing base pair, with a single nucleotide attached to G.

TABLE –V: COMBINED ENERGY VALUE TABLE

O1	A1	A2	A3	A4
$x1_A$	-	-	-	-0.9
$x2_C$	-	-	-2.2	-
$x3_G$	-	-2.1	-	-0.6
$x4_U$	-1.1	-	-1.4	-
$x5_A$	-	-	-	-1.3
$x6_C$	-	-	-2.4	-
$x7_G$	-	-2.1	-	-1
$x8_U$	-0.9	-	-1.3	-
$x9_A$	-	-	-	-2.1
$x10_C$	-	-	-3.3	-
$x11_G$	-	-2.4	-	-1.4
$x12_U$	-2.1	-	-2.1	-
$x13_A$	-	-	-	-2.4
$x14_C$	-	-	-3.4	-
$x15_G$	-	-3.3	-	-1.5
$x16_U$	-2.2	-	-2.5	-
$x17_A$	-	-	-	-1.3
$x18_C$	-	-	-2.5	-
$x19_G$	-	-2.1	-	-0.5
$x20_U$	-1.4	-	1.3	-

$x21_A$	-	-	-	-1
$x22_C$	-	-	-1.5	-
$x23_G$	-	-1.4	-	-0.3
$x24_U$	-0.6	-	-0.5	-
$x25_A$	-0.3	-0.5	-0.3	-0.3
$x26_C$	-0.1	-0.2	-1.5	-0.2
$x27_G$	-1.1	-1.2	-0.2	-0.2
$x28_U$	-0.3	-0.3	-0.6	-1.1
$x29_A$	-1.5	-1.5	-1.4	-1.8
$x30_C$	-1	-0.9	-2.9	-0.8
$x31_G$	-2.2	-2	-1.6	-1.1
$x32_U$	-1.7	-1.4	-1.8	-2
$x33_A$	-1.5	-1.5	-1.4	-1.8
$x34_C$	-1	-0.9	-2.9	-0.8
$x35_G$	-2.2	-2	-1.6	-1.1
$x36_U$	-1.7	-1.4	-1.8	-2
$x37_A$	-1.1	-1.5	-1.3	-2.1
$x38_C$	-1.1	-0.7	-2.4	-0.5
$x39_G$	-2.4	-2.9	-1.4	-1.2
$x40_U$	-1.9	-1	-2.2	-1.5
$x41_A$	0.2	-0.3	-0.3	-0.3
$x42_C$	-0.1	-0.2	-1.5	-0.2
$x43_G$	-0.9	-1.1	-0.3	0
$x44_U$	-0.3	-0.3	-0.4	-1.1
$x45_A$	-0.5	-0.3	-0.6	-0.5
$x46_C$	-0.2	-0.1	-1.7	0
$x47_G$	-0.8	-1.2	-0.3	-0.7
$x48_U$	-0.6	-0.1	-0.6	-0.8
$x49_A$	0.7	0.7	-0.4	0.7
$x50_C$	0.7	0.7	0.7	0.7
$x51_G$	-0.4	0.7	0.7	0.7
$x52_U$	0.7	0.7	0.7	0
$x53_A$	0.7	0.7	-0.4	0.7
$x54_C$	0.7	0.7	0.7	0.7
$x55_G$	-0.4	0.7	0.7	0.7
$x56_U$	0.7	0.7	0.7	0
$x57_A$	0	0	-1.1	0
$x58_C$	0	0	0	0
$x59_G$	-1.1	0	0	0
$x60_U$	0	0	0	-0.7
$x61_A$	0	0	-1.1	0
$x62_C$	0	0	0	0
$x63_G$	-1.1	0	0	0
$x64_U$	0	0	0	-0.7
$x65_A$	0.7	0.7	-0.4	0.7
$x66_C$	0.7	0.7	0.7	0.7
$x67_G$	-0.4	0.7	0.7	0.7
$x68_U$	0.7	0.7	0.7	0
$x69_A$	0.7	0.7	-0.4	0.7
$x70_C$	0.7	0.7	0.7	0.7
$x71_G$	-0.4	0.7	0.7	0.7
$x72_U$	0.7	0.7	0.7	0
$x73_A$	-0.8	-0.5	-0.8	-0.6
$x74_C$	-1.7	-0.8	-1.7	-1.2
$x75_G$	-1.1	-0.4	-1.3	-0.6
$x76_U$	-0.7	-0.1	-0.7	-0.1
$x77_{G1}$	-0.8	-0.5	-0.8	-0.6
$x78_A$	-0.3	-0.3	-0.4	-0.2
$x79_C$	-0.5	-0.3	-0.2	-0.1
$x80_G$	-0.2	-0.3	0	0
$x81_U$	-0.3	-0.1	-0.2	-0.2
$x82_{G1}$	-0.3	-0.3	-0.4	-0.2

The rough set data table used below is prepared using the subset of the same objects used before. This time only a selective set of objects are used. The energy interactions that are not considered for the rough set calculation are only the valid stacking interactions of the watson crick base pairs. They are excluded as for any stable pre-mature miRNA secondary structure abundance of Watson crick pairs are needed. This is an established fact. Now we want to concentrate on how the other energy interactions contribute in establishing a stable secondary structure of pre-mature miRNA. This subset of objects and attributes selected can be considered as the features based on which a particular secondary structure will be selected. The lesser is the nucleotide bonding energy, the more strength does the structure have.

For Table V, mentioned above, equivalence classes were formed. These are the classes containing objects having same cell for each attribute used. The results after calculating the equivalence classes the Table V are shown in the results section.

C. Calculation of lower and upper approximations and accuracy of the dataset

Let us consider a X such that X=O1={$x1_A$, $x2_C$……, $x82_{G1}$} in Table V. The indiscernibility relations, that is, the non-empty subsets of the conditional attributes are derived[27]. This comes as follows:
The subsets from conditional attributes come as {A1}, {A2}, {A3},{A4},{A1,A2},{A1,A3},{A1,A4},{A2, A3},},{A2, A4},{A3,A4},},{A1,A2,A3},{A2,A3,A4}, {A1,A2,A4}, {A1,A3,A4},{A1,A2,A3,A4}. The calculated indiscernible sets are shown in the result section.
Calculation of lower approximation ($\underline{BX}$) is by comparing the set of all elementary classes formed above for a certain set of decision parameter. Here the decision parameter or the decision attribute will be "Yes". The decision attribute corresponds to probable stable secondary structure. Now lower approximation set consists of all elements that surely belong to the set X. Again to calculate upper approximation (BX) the same comparison is done but this time the final upper

approximation set consists of those elements that has at least one element common from X and the range of elementary sets. So the upper approximation of the set constitutes of all elements that possibly belong to the set. The accuracy ($\mu_B(X)$) of the final set is calculated by $\mu_B(X)=card(\underline{BX}) / card(BX)$[28]. If $\mu_B(X) < 1$, then the set X is roughly definable in O1 [where $O=\{x1_A,x2_C\ldots\ldots x82_{G1}\}$] or else X is a crisp set in O1. The results are discussed in detail in results section. [29].

D. Calculation of Reduct and Core

The calculation of reduct and core is done from discernibility matrix, D [28], or, by checking the dependency of objects on attributes. Here the later option is used, that is, It is checked if removed any one or more attributes from the rough set table V, one can get the same number of equivalence or elementary classes as before. This subset of attributes generating the same elementary classes as before are the reducts and the intersection of all reducts gives the core.

In the future extension of this paper it is intended that a rough set based classifier will be designed to classify the different properties of the pre-mature miRNAs and thus generate and predict any property on pre-mature miRNAs.

VII. Results and Discussions:

In this sections the authors have given results obtained their prepare dataset.

A. Conversion of the dataset into Rough Set data table

For all the attribute combinations as mentioned in section VI(C) indiscernibility relations. These are the classes containing objects having same cell value for each conditional attribute used. Here only the table of indiscernibility set {A1,A2,A3,A4} is given as that is only needed for getting a deeper view into the pre-mature miRNA secondary structure. All the elementary classes are mentioned in this table.

TABLE –VI: ROUGH SET DATA TABLE WITH ALL ELEMENTARY CLASSES AND INDISCERNIBILITIY RELATIONS

U	A1	A2	A3	A4
$\{x5_A, x17_A\}$	-	-	-	-1.3
$\{x26_C, x42_C\}$	-0.1	-0.2	-1.5	-0.2
$\{x29_A,x33_A\}$	-1.5	-1.5	-1.4	-1.8
$\{x30_C, x34_C\}$	-1	-0.9	-2.9	-0.8
$\{x31_G,x35_G\}$	-2.2	-2	-1.6	-1.1
$\{x32_U,x36_U\}$	-1.7	-1.4	-1.8	-2
$\{x49_A, x53_A, x65_A, x69_A\}$	0.7	0.7	-0.4	0.7
$\{x50_C, x54_C, x66_C, x70_C\}$	0.7	0.7	0.7	0.7
$\{x51_G, x55_G, x67_G, x71_G\}$	-0.4	0.7	0.7	0.7
$\{x52_U, x56_U, x68_U, x72_U\}$	0.7	0.7	0.7	0
$\{x57_A,x61_A\}$	0	0	-1.1	0
$\{x58_C, x62_C\}$	0	0	0	0
$\{x59_G,x63_G\}$	-1.1	0	0	0
$\{x60_U,x64_U\}$	0	0	0	-0.7
$\{x73_A,x77_{G1}\}$	-0.8	-0.5	-0.8	-0.6
$\{x76_U,x78_{U1}\}$	-0.7	-0.1	-0.7	-0.1
$\{x78_A,x82_{G1}\}$	-0.3	-0.3	-0.4	-0.2
$\{x81_U\}$	-0.3	-0.1	-0.2	-0.2
$\{x1_A\}$	-	-	-	-0.9
$\{x2_C\}$	-	-	-2.2	-
$\{x3_G\}$	-	-2.1	-	-0.6
$\{x4_U\}$	-1.1	-	-1.4	-
$\{x6_C\}$	-	-	-2.4	-
$\{x7_G\}$	-	-2.1	-	-1
$\{x8_U\}$	-0.9	-	-1.3	-
$\{x9_A\}$	-	-	-	-2.1
$\{x10_C\}$	-	-	-3.3	-
$\{x11_G\}$	-	-2.4	-	-1.4
$\{x12_U\}$	-2.1	-	-2.1	-
$\{x13_A\}$	-	-	-	-2.4
$\{x14_C\}$	-	-	-3.4	-
$\{x15_G\}$	-	-3.3	-	-1.5
$\{x16_U\}$	-2.2	-	-2.5	-
$\{x18_C\}$	-	-	-2.5	-
$\{x19_G\}$	-	-2.1	-	-0.5
$\{x20_U\}$	-1.4	-	1.3	-
$\{x21_A\}$	-	-	-	-1
$\{x22_C\}$	-	-	-1.5	-
$\{x23_G\}$	-	-1.4	-	-0.3
$\{x24_U\}$	-0.6	-	-0.5	-
$\{x25_A\}$	-0.3	-0.5	-0.3	-0.3
$\{x28_U\}$	-0.3	-0.3	-0.6	-1.1
$\{x37_A\}$	-1.1	-1.5	-1.3	-2.1

{x38$_C$}	-1.1	-0.7	-2.4	-0.5
{x39 $_G$}	-2.4	-2.9	-1.4	-1.2
{x40 $_U$}	-1.9	-1	-2.2	-1.5
{x41 $_A$}	0.2	-0.3	-0.3	-0.3
{x42$_C$}	-0.1	-0.2	-1.5	-0.2
{x43 $_G$}	-0.9	-1.1	-0.3	0
{x44 $_U$}	-0.3	-0.3	-0.4	-1.1
{x45 $_A$}	-0.5	-0.3	-0.6	-0.5
{x46$_C$}	-0.2	-0.1	-1.7	0
{x47 $_G$}	-0.8	-1.2	-0.3	-0.7
{x48$_U$}	-0.6	-0.1	-0.6	-0.8
{x74$_C$}	-1.7	-0.8	-1.7	-1.2
{x75 $_G$}	-1.1	-0.4	-1.3	-0.6
x79$_C$	-0.5	-0.3	-0.2	-0.1
x80$_G$	-0.2	-0.3	0	0

For Table 7 there 4 equivalence classes based on the 4 decision attributes.

The equivalence classes or some of the indiscernible relations are

IND{A1}={{x31$_G$, x35$_G$, x16$_U$},{x32$_U$, x36$_U$, x74$_C$},{x29$_A$, x33$_A$}, {x37$_A$, x27$_G$, x38$_C$, x75$_G$, x59$_G$, x63$_G$, x4$_U$},{x30$_C$, x34$_C$}, {x43$_G$, x8$_U$},{x47$_G$, x73$_A$, x77$_{G1}$},{x48$_U$, x24$_U$},{x45$_A$, x79$_C$},{ x51$_G$, x55$_G$, x67$_G$, x71$_G$},{x25$_A$, x28$_U$, x44$_U$, x78$_A$, x82$_{G1}$, x81$_U$},{x80$_G$, x46$_C$},{x26$_C$, x42$_C$},{x57$_A$, x61$_A$, x60$_U$, x64$_U$, x58$_C$, x62$_C$},{x49$_A$, x53$_A$, x65$_A$, x69$_A$, x52$_U$, x56$_U$, x68$_U$, x72$_U$, x50$_C$, x54$_C$, x66$_C$, x70$_C$},{x15$_G$, x11$_G$, x7$_G$, x3$_G$, x19$_G$, x23$_G$, x14$_C$, x10$_C$, x18$_C$, x6$_C$, x2$_C$, x22$_C$, x13$_A$, x9$_A$, x5$_A$, x17$_A$, x21$_A$, x1$_A$}},

IND{A2}={{x31$_G$, x35$_G$}, {x16$_U$, x12$_U$}, {x32$_U$, x36$_U$, x23$_G$}, {x20$_U$, x4$_U$, x8$_U$, x24$_U$, x14$_C$, x10$_C$, x18$_C$, x6$_C$, x2$_C$, x22$_C$, x13$_A$, x9$_A$, x5$_A$, x17$_A$, x21$_A$, x1$_A$}, {x29$_A$, x33$_A$, x37$_A$ },{x59$_G$, x63$_G$}, {x30$_C$, x34$_C$}, {x27$_G$, x47$_G$}, {x73$_A$, x77$_{G1,}$ x25$_A$}, {x48$_U$, x76$_U$, x81$_U$, x80$_G$, x46$_C$, x41$_A$}, {x26$_C$, x42$_C$}, {x57$_A$, x61$_A$, x60$_U$, x64$_U$, x58$_C$, x62$_C$},{x55$_G$, x67$_G$, x71$_G$}, {x28$_U$, x44$_U$, x78$_A$, x82$_{G1}$, x80$_G$, x45$_A$, x79$_C$}, {x49$_A$, x53$_A$, x65$_A$, x69$_A$, x52$_U$, x56$_U$, x68$_U$, x72$_U$, x50$_C$, x54$_C$, x66$_C$, x70$_C$},{x7$_G$, x3$_G$, x19$_G$}, {x14$_C$, x10$_C$, x18$_C$, x6$_C$, x2$_C$, x22$_C$, x13$_A$, x9$_A$, x5$_A$, x17$_A$, x21$_A$, x1$_A$}},

IND{A3}={{x30$_C$, x34$_C$},{x16$_U$, x18$_C$},{x38$_C$, x6$_C$},{ x40$_U$, x2$_C$},{x32$_U$, x36$_U$},{x74$_C$, x46$_C$},{x31$_G$,x35$_G$},{ x22$_C$, x26$_C$, x42$_C$},{x39$_G$, x29$_A$, x33$_A$, x4$_U$},{x37$_A$, x75$_G$, x8$_U$}, {x57$_A$, x61$_A$},{x73$_A$, x77$_{G1}$},{x48$_U$, x45$_A$, x28$_U$},{x44$_U$, x78$_A$, x82$_{G1}$, x49$_A$, x53$_A$, x65$_A$, x69$_A$},{x43$_G$, x47$_G$,, x25$_A$, x41$_A$},{x27$_G$, x79$_C$, x81$_U$},{x59$_G$, x63$_G$, x80$_G$, x60$_U$, x64$_U$, x58$_C$, x62$_C$},{x51$_G$, x55$_G$, x67$_G$, x71$_G$, x52$_U$, x56$_U$, x68$_U$, x72$_U$, x50$_C$, x54$_C$, x66$_C$, x70$_C$},{x15$_G$, x11$_G$, x7$_G$, x3$_G$, x19$_G$, x23$_G$, x14$_C$, x10$_C$, x18$_C$, x6$_C$, x13$_A$, x9$_A$, x5$_A$, x17$_A$, x21$_A$, x1$_A$}},

IND{A4}={{x37$_A$,x9$_A$},{x32$_U$,x36$_U$},{x29$_A$, x33$_A$},{ x40$_U$, x15$_G$},{x5$_A$, x17$_A$},{ x39$_G$, x74$_C$}, {x31$_G$,x35$_G$,x28$_U$, x44$_U$}, {x7$_G$, x21$_A$},{x30$_C$, x34$_C$, x48$_U$},{x47$_G$, x60$_U$, x64$_U$},{ x75$_G$,x73$_A$,x77$_{G1}$,x3$_G$},{x38$_C$,x45$_A$,x19$_G$},{,x25$_A$,x41$_A$, x23$_G$ },{ x27$_G$, x78$_A$, x82$_{G1}$, x81$_U$, x26$_C$, x42$_C$},{ x76$_U$,x79$_C$},{ x59$_G$, x63$_G$, x43$_G$, x46$_C$,x80$_G$, x57$_A$, x61$_A$,x58$_C$, x62$_C$,x52$_U$, x56$_U$, x68$_U$, x72$_U$},{x51$_G$, x55$_G$, x67$_G$, x71$_G$, x49$_A$, x53$_A$, x65$_A$, x69$_A$,x50$_C$, x54$_C$,x66$_C$,x70$_C$},{x12$_U$, x16$_U$, x20$_U$,x4$_U$, x8$_U$,x24$_U$, x14$_C$, x10$_C$, x18$_C$, x6$_C$, x2$_C$, x22$_C$}},

IND{A1,A2}={{ x31$_G$,x35$_G$},{x32$_U$,x36$_U$},{x29$_A$, x33$_A$}, {x59$_G$, x63$_G$},{x30$_C$, x34$_C$},{ x73$_A$,x77$_{G1}$},{x45$_A$, x79$_C$},{ x51$_G$, x55$_G$, x67$_G$, x71$_G$},{ x28$_U$, x44$_U$, x78$_A$, x82$_{G1}$},{x26$_C$, x42$_C$},{ x57$_A$, x61$_A$, x60$_U$, x64$_U$, x58$_C$, x62$_C$},{ x49$_A$, x53$_A$, x65$_A$, x69$_A$, x52$_U$, x56$_U$, x68$_U$, x72$_U$, x50$_C$, x54$_C$,x66$_C$,x70$_C$},{x7$_G$, x3$_G$, x19$_G$},{ x14$_C$, x10$_C$, x18$_C$, x6$_C$, x2$_C$, x22$_C$, x13$_A$, x9$_A$, x5$_A$, x17$_A$, x21$_A$, x1$_A$}},

IND{A1,A3}={{ x31$_G$,x35$_G$},{x32$_U$,x36$_U$},{x29$_A$, x33$_A$}, {x59$_G$, x63$_G$},{x30$_C$, x34$_C$},{ x73$_A$,x77$_{G1}$},{x45$_A$, x79$_C$},{ x51$_G$, x55$_G$, x67$_G$, x71$_G$},{ x28$_U$, x44$_U$, x78$_A$, x82$_{G1}$},{x26$_C$, x42$_C$},{ x57$_A$, x61$_A$, x60$_U$, x64$_U$, x58$_C$, x62$_C$},{ x49$_A$, x53$_A$, x65$_A$, x69$_A$, x52$_U$, x56$_U$, x68$_U$, x72$_U$, x50$_C$, x54$_C$,x66$_C$,x70$_C$},{x7$_G$, x3$_G$, x19$_G$},{ x14$_C$, x10$_C$, x18$_C$, x6$_C$, x2$_C$, x22$_C$, x13$_A$, x9$_A$, x5$_A$, x17$_A$, x21$_A$, x1$_A$}},

IND{A1,A4}={{ x22$_C$,x77$_{G1}$},{x81$_U$,x80$_G$},{x26$_C$, x10$_C$}, {x11$_G$, x13$_A$},{x6$_C$, x14$_C$},{x16$_U$,x39$_G$},{ x15$_G$, x41$_A$, x75$_G$, x20$_U$},{x42$_C$,x53$_A$},{x12$_U$,x59$_G$,x34$_C$},{x66$_C$,x37$_A$},{x21$_A$,x65$_A$},{x7$_G$, x9$_A$},{x23$_G$, x3$_G$, x68$_U$, x72$_U$},{x58$_C$, x52$_U$, x61$_A$, x62$_C$,x27$_G$, x29$_A$, x40$_U$,x31$_G$},{x19$_G$, x56$_U$, x63$_G$, x43$_G$},{x32$_U$,x46$_C$},{ x79$_C$, x17$_A$},{ x78$_A$, x4$_U$, x55$_G$, x8$_U$, x82$_{G1}$, x44$_U$}},

IND{A2,A3}={{x35$_G$,x32$_U$,x4$_U$},{x71$_G$,x79$_C$},{x24$_U$, x55$_G$},{x51$_G$,x25$_A$},},{x45$_A$, x67$_G$},{x5$_A$, x2$_C$},{x48$_U$, x42$_C$ },{x6$_C$,x26$_C$,x22$_C$},{x77$_{G1}$,x80$_G$},{x60$_U$,x1$_A$},{x49$_A$, x65$_A$},{x50$_C$, x53$_A$, x54$_C$, x66$_C$, x69$_A$, x70$_C$},{ x58$_C$, x52$_U$, x15$_G$, x19$_G$ },{ x61$_A$, x62$_C$, x41$_A$, x56$_U$, x68$_U$, x72$_U$, x11$_G$, x7$_G$, x3$_G$, x23$_G$, x9$_A$,x13$_A$},{ x75$_G$, x63$_G$},{x39$_G$, x12$_U$, x74$_C$, x37$_A$,x59$_G$, x34$_C$ }},

IND{A2,A4}={{x81$_U$,x80$_G$},{x26$_C$, x10$_C$},{ x67$_G$, x60$_U$}, {x35$_G$,x6$_C$},{x71$_G$,x69$_A$},{x30$_C$,x70$_C$},{x74$_C$,x15$_G$}, {x20$_U$,x42$_C$},{x66$_C$,x34$_C$},{x37$_A$,x21$_A$,x65$_A$,x7$_G$,x9$_A$,x68$_U$},{ x72$_U$,x3$_G$,x23$_G$,x49$_A$},{x19$_G$, x56$_U$, x63$_G$, x43$_G$, x58$_C$, x52$_U$, x61$_A$, x62$_C$,x27$_G$,x29$_A$,x40$_U$,x31$_G$},{x64$_U$, x28$_U$},{ x79$_C$, x17$_A$,x38$_C$, x33$_A$, x45$_A$, x5$_A$, x78$_A$, x4$_U$,x44$_U$, x55$_G$,x8$_U$, x82$_{G1}$, x44$_U$}},

IND{A3,A4}={{x30$_C$,x34$_C$},{x16$_U$,x18$_C$},{x32$_U$,x36$_U$}, {x31$_G$,x35$_G$},{x26$_C$, x42$_C$}, {x29$_A$, x33$_A$},{x57$_A$,x61$_A$},{ x73$_A$,x77$_{G1}$},{x78$_A$, x82$_{G1}$}, {x49$_A$,x53$_A$,x65$_A$,x69$_A$}, {x25$_A$, x41$_A$},{x27$_G$, x81$_U$}, {x64$_U$, x60$_U$}, {x58$_C$, x59$_G$, x63$_G$, x62$_C$,x80$_G$},{x52$_U$,x56$_U$,x68$_U$,x72$_U$},{x50$_C$,x51$_G$,x54$_C$,x55$_G$, x66$_C$, x67$_G$, x70$_C$, x71$_G$},{ x5$_A$, x17$_A$},{ x7$_G$, x21$_A$}},

IND{A1,A2,A4}={{x31$_G$,x35$_G$},{x32$_U$,x36$_U$},{x29$_A$,x33$_A$}, {x59$_G$, x63$_G$},{x30$_C$, x34$_C$},{x73$_A$,x77$_{G1}$},{x45$_A$, x79$_C$},{ x51$_G$, x55$_G$, x67$_G$, x71$_G$},{x28$_U$, x44$_U$}, {x78$_A$, x82$_{G1}$},{x26$_C$,x42$_C$},{x60$_U$,x64$_U$},{x57$_A$,x58$_C$,x61$_A$, x62$_C$}, {x49$_A$, x53$_A$, x65$_A$, x69$_A$, x52$_U$, x56$_U$, x68$_U$, x72$_U$, x50$_C$,

$x54_C$,$x66_C$,$x70_C$},{$x5_A$, $x17_A$},{$x14_C$, $x10_C$, $x18_C$, $x6_C$, $x2_C$, $x22_C$ }},

IND{A1,A2,A3}={{$x77_{G1}$,$x45_A$},{$x76_U$,$x24_U$},{$x61_A$,$x28_U$}, {$x56_U$, $x51_G$},{$x73_A$, $x48_U$ },{$x15_G$, $x7_G$},{$x60_U$, $x62_C$, $x67_G$, $x13_A$},{ $x47_G$,$x21_A$, $x23_G$},{$x75_G$, $x65_A$},{$x80_G$, $x72_U$},{$x52_U$, $x68_U$, $x50_C$, $x54_C$},{$x79_C$, $x64_U$, $x59_G$, $x63_G$, $x66_C$, $x70_C$, $x20_U$,$x9_A$},{ $x30_C$, $x38_C$, $x42_C$}, {$x14_C$, $x16_U$, $x36_U$, $x40_U$, $x35_G$, $x29_A$}},

IND{A1,A2,A3,A4}={{$x31_G$,$x35_G$},{$x32_U$,$x36_U$},{$x29_A$,$x33_A$}, {$x59_G$, $x63_G$}, {$x30_C$, $x34_C$}, {$x73_A$,$x77_{G1}$}, {$x51_G$, $x55_G$, $x67_G$, $x71_G$}, {$x78_A$, $x82_{G1}$},{$x26_C$,$x42_C$}, {$x60_U$,$x64_U$},{$x58_C$,$x62_C$}, {$x49_A$, $x53_A$, $x65_A$, $x69_A$},{$x52_U$, $x56_U$,$x68_U$,$x72_U$}, {$x5_A$, $x17_A$}}.

B. Calculation of lower and upper approximations and accuracy of the dataset

Now the set X here is denoted by the initial set O1 that we have taken to represent all kinds of objects for the secondary structure of a miRNA sequence. For Table 9 lower approximation (BX)= {$x17_A$, $x18_C$, $x19_G$, $x20_U$, $x21_A$, $x22_C$, $x23_G$, $x24_U$, $x31_G$, $x32_U$, $x35_G$, $x36_U$, $x39_G$, $x40_U$, $x43_G$, $x44_U$, $x47_G$, $x48_U$, $x51_G$, $x52_U$, $x55_G$, $x56_U$, $x59_G$, $x60_U$, $x63_G$, $x64_U$, $x67_G$, $x68_U$, $x71_G$, $x72_U$}.
Now all the energy parameters mentioned in table V enhance in stable secondary structure of a pre-m,ature miRNA in a smaller or bigger way. So here some extra parameters are considered which hinder the total energy value of a stable secondary structure to be negative. These energy values range from $x83_E$ to $x112_E$ depicting the penalizing energy values to be added over the total energy of the pre-mature miRNA secondary structure. These energy values lead to a decision parameter "No". So the upper approximation (BX) for each of the 4 sets={ $x1_A$, $x2_C$ ….. $x112_E$ }
Thus the accuracy ($\mu_B(X)$) = card(BX) / card(BX)=82/112=0.73Since $\mu_B(X) < 1$, then the set X is roughly definable in O1 [where O1={$x1_A$,$x2_C$…….$x114_E$}] .

C. Calculation of Reduct and Core

Based on the equivalence classes of Table VI, the reduct of comes as {A2,A3,A4} and the core is {A2}. This time the decision attribute is not taken into consideration.

D. Generation of decision rules and prediction of correct secondary structures

The generation of decision rules is yet to be done for this classification algorithm. This calculation will be incorporated in the future rough set based classifier. These rules will help in feature selection of new and novel pre-mature or mature miRNAs and suitable classification of them accordingly.

VIII. Conclusion and Future Scope:

In the current paper, the authors have gone one step further from the previous work [30] and tried to assess the secondary structure of the pre-mature miRNAs using rough sets. With the equivalence classes repetitive energy value calculation becomes less. In future decision rules will be generated and a proper rough set based classifier will be developed. The features thus classified to predict proper decision rules from those classification will be incorporated in the classifier.

References

[1] Nath, Joyshree, and Asoke Nath. "A new algorithm for quantitative deciphering of pre-mature MiRNAs using some statistical parameters." Information and Communication Technologies (WICT), 2012 World Congress on. IEEE, 2012.
[2] Quantitative Deciphering of Pre-Mature and Pattern Recognition of Mature Mirnas using Some Statistical Parameters Joyshree Nath, Asoke Nath, International Journal of Computer Information Systems and Industrial Management Applications,pp. 184-195,vol-6, 2014.
[3] Joyshree Nath, Kronecker Decision Diagram in Multi-valued Quantum Computing and its application in Bio informatics, In Proceedings of the National Conference on "Quantum correlations: Foundation and Applications" on 4th and 5th March,2014, held jointly by Department of Physics, Vidyasagar College for Women and Phsyics and Applied Mathematics Unit, ISI on the topic of quantum computing and bioinformatics held in ISI,Kolkata.
[4] Mitra, Ramkrishna, and Sanghamitra Bandyopadhyay. "MultiMiTar: A novel multi objective optimization based miRNA-target prediction method." *PloS one*6.9 (2011): e24583.

[5] Bandyopadhyay, Sanghamitra, and Ramkrishna Mitra. "TargetMiner: microRNA target prediction with systematic identification of tissue-specific negative examples." *Bioinformatics* 25.20 (2009): 2625-2631.

[6] A Comprehensive Study on Animal miRNAs : A Computational approach to explore its implications
in Biological and Chemical environment , Joyshree Nath, International Journal of Advanced Computer Research, Vol-3, No.1, issue-8, pp:153-158(2013).
[7] Bhattacharyya, Malay, and Sanghamitra Bandyopadhyay. "Studying the differential co-expression of microRNAs reveals significant role of white matter in early Alzheimer's progression." *Molecular BioSystems* 9.3 (2013): 457-466.

[8] S. Bandyopadhyay and M. Bhattacharyya, A Novel Method of Studying the Disease Regulatory Activities of MicroRNAs, Current Bioinformatics, 4(3):234-241, 2009.
[9] M. Bhattacharyya and S. Bandyopadhyay, Computational Discovery of Different Categories of Human Oncogenic MicroRNAs, In Proceedings of the 1st IFIP International Conference on Bioinformatics (BIOINFO), Surat, India, March 25-28, no. 94, 2010.

[10] S. Bandyopadhyay and M. Bhattacharyya, PuTmiR: A database for extracting neighboring transcription factors of human microRNAs, BMC Bioinformatics, 11:190, 2010.
[11] Bhattacharyya, Malay, and Sanghamitra Bandyopadhyay. "Integration of co-expression networks for gene clustering." *Advances in Pattern Recognition, 2009. ICAPR'09. Seventh International Conference on*. IEEE, 2009.
[12] Advances in microRNAs: implications for gene therapists., Marquez RT, McCaffrey AP., Hum Gene Ther. 2008;19:27–38.

[13] Wu, Xudong, and Yijiang Song. "Preferential regulation of miRNA targets by environmental chemicals in the human genome." BMC genomics 12.1 (2011): 244.

[14] A Comprehensive Study of Target Prediction Algorithms for Animal MicroRNAs(miRNAs), International Journal of Computer Applications(0975-8887, USA), Vol-40, No15(Feb),2012.

[15] Nath, Joyshree, and Asoke Nath. "A new algorithm for quantitative deciphering of pre-mature MiRNAs using some statistical parameters." *Information and Communication Technologies (WICT), 2012 World Congress on*. IEEE, 2012.

[16] Nath, Joyshree, and Asoke Nath. "A Modified Algorithm for Quantifying of Pre-mature MiRNAs using Some Fractal Parameters." *International Journal of Advanced Computer Research* 2.4 (2012).

[17] Hassanien, Aboul Ella, and Jafar MH Ali. "Rough set approach for generation of classification rules of breast cancer data." *Informatica* 15.1 (2004): 23-38.

[18] Pawlak, Zdzisław. "Rough sets." *International Journal of Computer & Information Sciences* 11.5 (1982): 341-356.

[19] Suraj, Zbigniew. "An Introduction to Rough Set Theory and Its Applications." *ICENCO, Cairo, Egypt* (2004).

[20] Swiniarski, Roman W., and Andrzej Skowron. "Rough set methods in feature selection and recognition." *Pattern recognition letters* 24.6 (2003): 833-849.

[21] Mathews, David H., et al. "Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure." Journal of molecular biology 288.5 (1999): 911-940.

[22] Zuker, Michael, David H. Mathews, and Douglas H. Turner. "Algorithms and thermodynamics for RNA secondary structure prediction: a practical guide."RNA biochemistry and biotechnology. Springer Netherlands, 1999. 11-43.

[23] Dima, Ruxandra I., Changbong Hyeon, and D. Thirumalai. "Extracting stacking interaction parameters for RNA from the data set of native structures." Journal of molecular biology 347.1 (2005): 53-69.

[24] web.njit.edu/~wangj/rna/slides/Adrian2008.pdf .

[25] Theory and Practice of Computational RNA Biology Kay Nieselt, Eric Westhof and Robert Giegerich (eds.)

[26] Park, Jong-Eun, et al. "Dicer recognizes the 5 [prime] end of RNA for efficient and accurate processing." Nature 475.7355 (2011): 201-205.

[27] Komorowski, Jan, et al. "Rough sets: A tutorial." *Rough fuzzy hybridization: A new trend in decision-making* (1999): 3-98.

[28] Walczak, B., and D. L. Massart. "Rough sets theory." *Chemometrics and intelligent laboratory systems* 47.1 (1999): 1-16.

[29] Pawlak, Zdzisław. *Rough sets: Theoretical aspects of reasoning about data*. Vol. 9. Springer, 1991.

[30] Computational Study of secondary structure of Premature miRNA and inclusion of Finite State Automation, joyshree Nath, Asoke Nath, published in IEEE conference proceedings : WICT-20 13 held at Hanoi in Dec 14-18(20 13), page 292-299(20 13).

[31] Applications of Boolean Functions for mapping the mutation of mRNA and Protein structure : A new Computational approach, Joyshree Nath, Proceedings of The 2013 International Conference on BioInformatics & Computational Biology,WorldComp 2013 held at Las Vegas, USA ,Jul 22-25(2013), Page 238-245.

Author Biographies

The author is M.Tech in Information and Technology. The author is currently doing research work in computational genomics. She has around 5 years of teaching experience.