J. Appl. Environ. Biol. Sci., 4(8S)106-113, 2014

ISSN: 2090-4274
Journal of Applied Environmental and Biological Sciences
www.textroad.com

Computational Prediction Of Micro-RNAs In Hepatitis B Virus Genome

Muhammad Bilal[11,#], Ayaz Ahmad[2,#,*], Khalid Khan[1], Nasir Ahmad[3], Ashfaq Ur Rehman[3], Muhammad Junaid[3], Amjad Iqbal[4], Waheed Ahmad[5], Abdul Wadood[3], Taj Ur Rehman[6]

[1] Department of Chemistry, Islamia College University Peshawar, Pakistan
[2] Department of Biotechnology, Abdul Wali Khan University Mardan, Pakistan
[3] Department of Biochemistry, Abdul Wali Khan University Mardan, Pakistan
[4] Department of Biochemistry, Abdul Wali Khan University Mardan, Pakistan
[5] Department of Computer Science, Abdul Wali Khan University Mardan, Pakistan
[6] Department of Chemistry, Abdul Wali Khan University Mardan, Pakistan

Received: September 1, 2014
Accepted: November 13, 2014

ABSTRACT

MicroRNAs (miRNAs) are 19—25 nucleotides long, single-stranded, endogenous non-coding RNA molecules that play crucial roles in the post-transcriptional regulation of gene expression by targeting messenger RNAs for cleavage or translational repression. Genomes of various organisms ranging from higher animals and plants to viruses transcribe miRNAs. Recent studies show that both virus and host encode miRNAs that can give benefit either to virus or to host, depending upon the specific interactions. Hepatitis B Virus (HBV) is involved in acute and chronic diseases of liver and can bring about hepatitis, liver cirrhosis and hepatocellular carcinoma (HCC). In this study, we analyzed HBV (genotype D) for miRNAs, computationally, since computational methods serve vigorous, better, and affordable tools for microRNAs identification. Initial searches through VMir software extracted 57 sequences with potential hairpin-like structures from HBV genome. MiPred program confirmed 10 candidates as real pre-miRNA like hairpin structures. Measurements of free energy and other parameters further excluded three candidates. Finally, MatureBayes web server v1.0 confirmed 12 mature miRNAs in 6 potential pre-miRNA candidates in HBV genome, including one which has been identified in earlier studies. These findings open new avenues for researchers to explore the role of these novel miRNAs in viral pathogenesis as well as in developing novel antiviral therapies.

KEYWORDS: MicroRNAs, miRNAs, Hepatitis, Hepatitis B Virus HBV.

INTRODUCTION

MiRNAs are endogenous non-coding small RNA molecules that regulate gene expression post-transcriptionally. Genomes of a variety of organisms ranging from higher animals and plants to some viruses have been reported to encode miRNAs. Though the complete picture of miRNAs' functions is yet to be explored, they have been reported to involve in regulating numerous cellular processes including differentiation, morphogenesis, organogenesis, and metabolism [1-5]. Initially, miRNAs are transcribed as long transcripts called primary miRNAs (pri-miRNAs) [6]. Drosha, a nuclear RNase III enzyme, processes the nuclear pri-miRNA into one to several, 60-70 nt long, miRNA precursors (pre-miRNAs) which acquire hairpin stem-loop structures while still inside the nucleus [7]. The resultant miRNA precursors are transported to the cellular cytoplasm for further processing by the exportin-5 [8, 9]. Subsequently, another enzyme, called Dicer cleaves the pre-miRNA into an imperfect dsRNA duplex [7, 10-12]. This duplex consists of two strands i.e. one is the mature miRNA strand, also called functional strand while the other is its paired strand. The functional strand of the duplex is then loaded into RNA-induced silencing complex (RISC), whereas the complementary strand is assumed to be degraded. The target mRNA is then either cleaved or repressed translationally by the RISC, depending upon the degree of complementarity between the target mRNA and the RISC bound miRNA [13, 14].

Recent studies on viruses suggest viral and host cell miRNAs as a new class of regulators of viral pathogenesis [15]. Virus-encoded miRNAs were reported to affect the expression of host genes and *vice versa*. For instance, the miRNAs embedded in the Herpes viruses genome control the expression of not only viral proteins but also the host cell proteins, thereby facilitating their pathogenic cycles [16].

Hepatitis B virus is categorized as hepatotropic non-cytopathic DNA virus which belongs to a small,

[1] [#] These authors contributed equally to this work

* **Corresponding Author:** Ayaz Ahmad, Department of Biotechnology, Abdul Wali Khan University Mardan, Pakistan. Phone # 0092-937-542046, E-mail: ahdayazb5@awkum.edu.pk

enveloped DNA viruses family called *Hepadnaviridae*. It is major cause of acute and chronic infections of liver [17]. Over 0.35 billion chronic carriers of HBV have been reported around the globe. Chronic HBV infection has been the key factor greatly linked with high threat of liver cirrhosis, which is associated with hepatocellular carcinoma (HCC) [18, 19]. A clear picture of hepatitis B Biology as well as mechanism of pathogenesis is indispensable for controlling hepatitis B worldwide. Being hepatotropic, HBV replicates favorably inside hepatocytes. Even though the viral entrance pathway into hepatocytes remains to be explored, still it is believed that the large envelope protein's N-terminus is involved in cell attachment and entry [20]. The fusion of HBV and host cell membranes is followed by the transport of viral capsid to the nuclear pore. Here the shifting of relaxed circular genome (RC-DNA) of HBV into hepatocyte nucleus occurs. Inside the nucleus, cellular enzymes convert the rcDNA into covalently closed circular DNA (cccDNA) [21, 22]. The transcription of all viral RNAs uses cccDNA as template. HBV genome (3.2 kb) which contains four overlapping open reading frames, including gp1, gp2, gp3, and gp4, can express seven different hepatitis-B proteins, namely, DNA polymerase, three S proteins, one X protein, and two C proteins.

During the last decade, experimental techniques such as cDNA cloning followed by confirmation through Northern blotting have been commonly employed to discover most of miRNAs in various organisms, including viruses [23-26]. However, these procedures are slow, laborious and expensive [27]. In contrast, computer based strategies for the prediction of novel miRNAs represent vigorous, convenient and affordable techniques. In this study we carried out *in silico* analysis of the HBV genome and identified 12 mature miRNAs including one which was reported earlier.

MATERIALS AND METHODS

Genome Sequence Retrieval

Fasta format of complete genome sequence of HBV, genotype D, (accession number AJ344117) was retrieved from genome data bank (NCBI) (http://www.ncbi.nlm.nih.gov/nuccore/AJ344117.1). HBV genome contains 3182 nucleotides. The overall computational prediction procedure is represented in the form of a flowchart in Figure 1.

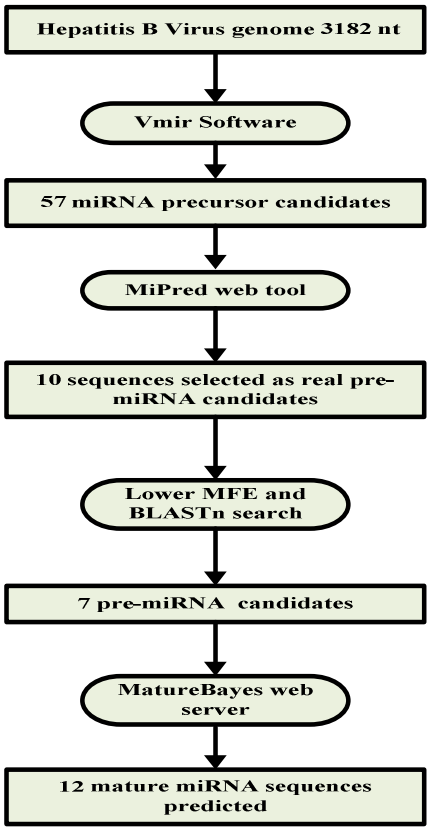

Figure 1: *Flowchart of the computational prediction process*

Pre-miRNA Extraction

The identification of pre-miRNA candidates was performed by specialized *ab initio* viral pre-miRNAs prediction software VMir on the basis of comparison to structural characteristics of already recognized pre-miRNA hairpins. In order to extract hairpin-structured miRNA precursors, the viral genome was scanned by VMir software (program version 2.3, scoring algorithm version 1.4) [28, 29]. Primarily, potential hairpin shaped sequences were mined as candidate miRNA precursors (pre-miRNAs).

Confirmation of Real Pre-miRNAs

In the next step, MiPred program with RF algorithm (http://www.bioinf.seu.edu.cn/miRNA/) was used to distinguish real miRNAs from pseudo ones [30].

Screening for Potential Unique Pre-miRNA Structures

The candidate pre-miRNAs sequences were analyzed for secondary structure prediction and minimum free energy (MFE) by RNAfold web server. Sequences with hairpin-like secondary structures, and having lower MFE (equal to or less than –25 kcal/mol) were selected as potential miRNA precursors. These sequences were further confirmed to be unique by conducting BLASTn searches.

Mature miRNA Prediction

Prediction of mature miRNAs in HBV, genotype D was performed on MatureBayes web server v1.0 (http://mirna.imbb.forth.gr/MatureBayes.html). This computer-based web tool uses a Naive Bayes classifier which is based on secondary structure and sequence features of the pre-miRANs for the prediction of mature miRNAs in any given pre-miRNA [31], MatureBayes compute the probabilistic start position of mature miRNA(s) by two alternatives.

RESULTS AND DISCUSSION

VMir is a low stringency software especially designed for the identification of viral miRNAs [28]. By using RNAfold algorithm, VMir executes structure prediction by minimal folding free energy and detects individual hairpins above a certain size limit (by default 45 nt) [29]. VMir assigns score to these hairpins which is based on statistical comparison to a reference set of recognized pre-miRNA hairpins [29]. In this study, fasta format of the Hepatitis B virus (genotype D) genome was uploaded into VMir Analyzer. The program was allowed to operate with its adjusted parameters for window size and step size. The viral genome was scanned in both the orientations for extraction of hairpin like sequences (pre-miRNAs). As a result, a total of 202 sequences as candidate miRNA precursors were initially detected by VMir Analyzer (Figure 2(a)). These 202 pre-miRNA candidates were passed through a filter. By adjusting the filter values for minimal scores and window counts, only 57 hairpins succeeded to pass through the window filter (Figure 2(b)). Figures 2(a) and 2(b) show the location and VMir scores for unfiltered and filtered hairpins, respectively.

2(a)

2(b)

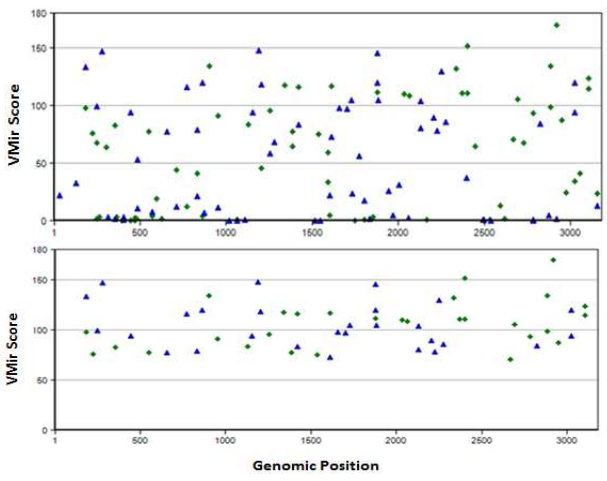

Figure2. ***(a)*** *VMir analysis of the HBV genome; showing that all hairpins are widely dispersed across the viral genome.* ***(b)*** *only those hairpins are shown which passed the filter and achieved a VMir score between 85 and 146 and located between nucleotides 200 and 3132. Hairpins are plotted according to genomic location and VMir score*

These 57 pre-miRNA candidates were then analyzed by MiPred (online web server) to differentiate the real pre-miRNAs from pseudo ones [30]. MiPred confirmed 11 sequences as real pre-miRNA candidates having fold-back hairpin shaped structures from a total of 57 filtered sequences. Measurement of lower minimum free energy (equal to or less than –25 kcal/mol) and BLASTn search further extracted 7 real pre-miRNA candidates. The secondary structure prediction of these sequences with potential hairpin-like structures was accomplished by RNAfold program (Figure 3). MatureBayes web server was used for the identification of mature miRNA sequences inside the pre-miRNA hairpin structures (Gkirtzou et al.). This computational tool predicted 12 mature miRNAs (Table1) in 6 potential pre-miRNA hairpin structure candidates (Figure 4).

Figure 3:

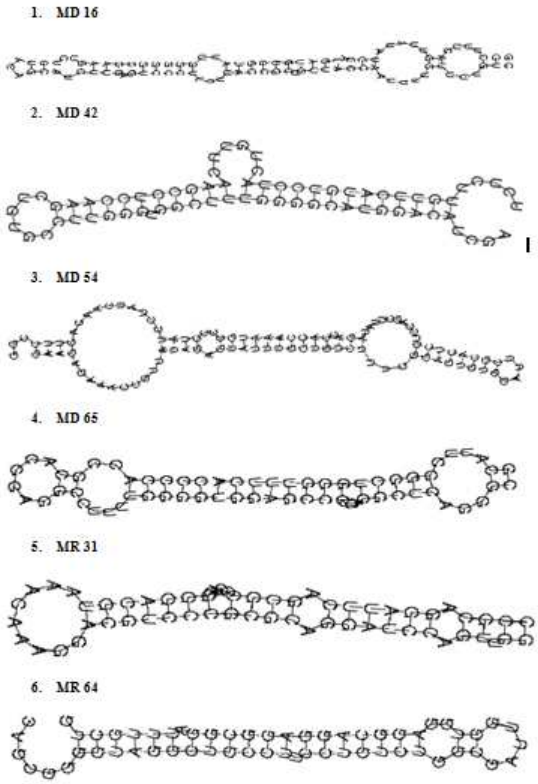

Figure 3: *Secondary structures of the pre-miRNA precursors using RNAfold program*

Although the exact functions of viral miRNAs (vmiRNAs) are yet to be unearthed, lines of evidences suggest that vmiRNAs have the potential to target both host and viral transcripts. Similar to other viral factors, vmiRNAs share their part in cellular reprogramming to (a) control the latent-lytic switch, (b) encouraging cell existence, growth and diversity for the purpose to support viral replication and (c) eluding immune recognition. Similarly, 1) vmiRNAs make a host cell environment conducive by down-regulating selected viral and cellular mRNAs in order to complete viral life cycle, 2) Viruses have the ability to either mimic host cell miRNAs or influence the expression of cellular miRNA for controlling existing regulatory pathways and 3) Viral replication can be directly influenced by cellular miRNAs, and mammalian RNA virus genomes can be directly targeted by several miRNAs. A virus requires retaining the infected cells of the host alive for a long time, which helps in completion of viral life cycle and thus making their existence possible. This session is significantly prolonged for viruses that initiate latent infections. Hence, viral miRNAs may promote virus replication by at least two ways i.e. extending cellular life time and escaping immune recognition.

Figure 4:

Figure 4: *Hairpin structures of HBV pre-miRNAs. The putative mature miRNAs sequences are shown in red*

Experimental techniques like cDNA cloning and Northern blotting have been widely used for the discovery of majority of miRNAs in various organisms [23-26]. Many miRNAs have also been identified computationally, as Computer based strategies represent robust, convenient and affordable techniques. These methods include programs which rely on features like hairpin stem loop structures of miRNAs identification and thermodynamics stability etc. [23-26].

Here we report the prediction of several miRNAs encoded by HBV genome by using computational tools which included softwares and publically available online web servers. VMir is a low stringency; updated *ab initio* computational algorithm especially designed for the detection of fold-back, hairpin shaped putative pre-miRNAs in viral genomes and has been successfully used for the prediction of miRNAs in the genomes of several viruses related to herpesvirus and polyomavirus families. VMir investigation of HBV genome indicates that candidates are overall extensively distributed across the viral genome (Figure 2(a)). Several user-adjustable quality filters have been included in VMir program that can be used to decrease the intricacy of the prediction and make the prediction easier [29]. After filtering, the 57 high scoring filtered hairpins (with scores between 85 and 146) appeared between nucleotides 200 and 3132 (Figure 2(b)).

Table 1: Mature miRNA sequences predicted by MatureBayes web tool

>Precursor hairpin: MD 16		
Sequence	**GGCUUUCAGUUAUAUGGAUGAUGUGGUAUUGGGGGCCAAGUCUGUACAGCAUCUUGAGUC CCUUUUUACCGCUGUUACCAAUUUUCUUUUGUC**	
Duplex	Position 20 (5' stem)	Sequence: AUGUGGUAUUGGGGGCCAAGUC
	Position 54 (3' stem)	Sequence: UGAGUCCCUUUUUACCGCUGUU
Mature 5'stem	Position 20	Sequence: AUGUGGUAUUGGGGGCCAAGUC
Mature 3'stem	Position 54	Sequence: UGAGUCCCUUUUUACCGCUGUU
>Precursor hairpin: MD 42		
Sequence	**UCUCUUGUUCAUGUCCUACUGUUCAAGCCUCCAAGCUGUGCCUUGGGUGGCUUUGGGGCA UGGACAUCGA**	
Duplex	Position 7 (5' stem)	Sequence: UUCAUGUCCUACUGUUCAAGCC
	Position 43 (3' stem)	Sequence: UGGGUGGCUUUGGGGCAUGGAC
Mature 5'stem	Position 11	Sequence: UGUCCUACUGUUCAAGCCUCCA
Mature 3'stem	Position 43	Sequence: UGGGUGGCUUUGGGGCAUGGAC
>Precursor hairpin: MD 54		
Sequence	**GAAGAGAAACCGUUAUAGAGUAUUUGGUGUCUUUCGGAGUGUGGAUUCGCACUCCUCCAG CUUAUAGACCACCAAAUGCCCCUAUCCUAUCAACACUUCCGG**	
Duplex	Position 15 (5' stem)	Sequence: UAGAGUAUUUGGUGUCUUUCGG
	Position 54 (3' stem)	Sequence: CUCCAGCUUAUAGACCACCAAA
Mature 5'stem	Position 15	Sequence: UAGAGUAUUUGGUGUCUUUCGG
Mature 3'stem	Position 64	Sequence: UAGACCACCAAAUGCCCCUAUC
>Precursor hairpin: MD 65		
Sequence	**GCAUUCGGGCUGGGUUUCACCCCACCGCACGGAGGCCUUUUGGGGUGGAGCCCUCAGGCU CAGGGC**	
Duplex	Position 10 (5' stem)	Sequence: : UGGGUUUCACCCCACCGCACGG
	Position 36 (3' stem)	Sequence: CUUUUGGGGUGGAGCCCUCAGG
Mature 5'stem	Position 10	Sequence: UGGGUUUCACCCCACCGCACGG
Mature 3'stem	Position 36	Sequence: CUUUUGGGGUGGAGCCCUCAGG
>Precursor hairpin: MR 31		
Sequence	**CCGCAGGAUUCAGCGCCGACGGGACGUAAACAAAGGACGUCCCGCGCAGGAUCCAGUUGG**	
Duplex	Position 2 (5' stem)	Sequence: GCAGGAUUCAGCGCCGACGGGA
	Position 37 (3' stem)	Sequence: CGUCCCGCGCAGGAUCCAGUUG
Mature 5'stem	Position 8	Sequence: UUCAGCGCCGACGGGACGUAAA
Mature 3'stem	Position 37	Sequence: CGUCCCGCGCAGGAUCCAGUUG
>Precursor hairpin: MR 64		
Sequence	**CAGCGGGGUAGGCUGCCUUCCUGUCUGGCGAUUGGUGGAGGCAGGAGGCGGAUUUGCUG**	
Duplex	Position 2 (5' stem)	Sequence: GCGGGGUAGGCUGCCUUCCUGU
	Position 36 (3' stem)	Sequence: GGAGGCAGGAGGCGGAUUUGCU
Mature 5'stem	Position 23	Sequence: UCUGGCGAUUGGUGGAGGCAGG
Mature 3'stem	Position 36	Sequence: GGAGGCAGGAGGCGGAUUUGCU

The characteristic feature of majority of pre-miRNAs to fold into typical stem-loop, fold-back structures and acquire hairpin shapes in many genomes makes it hard to differentiate real pre-miRNAs from pseudo ones [30]. In order to distinguish the real pre-miRNAs from pseudo ones, we used MiPred program, a hybrid tool with combined features like local contiguous structure-sequence composition, MFE and Monte Carlo randomization test (Xue et al., 2005). MiPred outperform some of the existing web tools (Triplet-SVM-classifier, miRabela and ProMiR II) in terms of specificity (98.21%) and sensitivity (95.09%) for prediction of real pre-miRNA like hairpins [30]. MiPred confirmed 11 sequences as real pre-miRNA like hairpins in HBV genome.

Further, the prediction of mature miRNAs in pre-miRNA candidates was accomplished by MatureBayes web tool that incorporate a specialized Naive Bayes classifier, which perform the task by taking in consideration the sequence and secondary structure statistics related to pre-miRNAs.. It is of great significance to predict the starting position of mature miRNA within pre-miRNA because the 2^{nd} to 8^{th} nt of the mature miRNA, also called seed region is important for discovering their targets in genomes [31]. MatureBayes can predict the start position of the mature miRNA and/or the miRNA- miRNA* duplex with high accuracy, significantly outperforming the two existing tools i.e. BayesMiRNAfind and ProMiR [31]. MatureBayes predicted 12 mature miRNAs in HBV genome (Table 1), including the one which have been already identified by Jin and co-workers [32]. The positions of these mature miRNAs in the respective 6 potential hairpins are shown in figure 4.

Medical strategies like antiviral nucleoside/nucleotide analogs and the use of interferon (IFN) was established for treating chronically infected patients. However, currently existing therapies for the termination of HBV infection are insufficient in the majority of patients. Improved knowledge regarding HBV–host interaction is mandatory for new antiviral therapeutic strategies. The association of viral-encoded miRNAs with the pathogenic characteristics of the virus highlights the biological importance of miRNAs in evolving therapeutic targets in a broad spectrum of diseases and is expected to develop into a novel armada of more powerful and mechanism-oriented therapeutics. *In silico* prediction of miRNAs is only the first step of miRNA study and should be followed by other investigations like target and function analysis for a comprehensive understanding of its biological roles.

ACKNOWLEDGEMENTS

The authors would like to thank Professor Zuhong Lu of Southeast University, Nanjing, PR China for his help in MiPred program. Research in Dr. Ayaz Ahmad's Lab is funded by Abdul Wali Khan University Mardan (AWKUM), Higher Education Commission (HEC) of Pakistan and Relief International (RI).

REFERENCES

1. Poy, M.N., M. Spranger, and M. Stoffel, *microRNAs and the regulation of glucose and lipid metabolism.* Diabetes, Obesity and Metabolism, 2007. **9**(s2): p. 67-73.
2. Zhao, Y. and D. Srivastava, *A developmental view of microRNA function.* Trends in biochemical sciences, 2007. **32**(4): p. 189-197.
3. Baltimore, D., et al., *MicroRNAs: new regulators of immune cell development and function.* Nature immunology, 2008. **9**(8): p. 839-845.
4. Gao, F.-B., *Posttranscriptional control of neuronal development by microRNA networks.* Trends in neurosciences, 2008. **31**(1): p. 20-26.
5. Griffiths-Jones, S., et al., *miRBase: tools for microRNA genomics.* Nucleic acids research, 2008. **36**(suppl 1): p. D154-D158.
6. Lee, Y., et al., *MicroRNA maturation: stepwise processing and subcellular localization.* The EMBO journal, 2002. **21**(17): p. 4663-4670.
7. Lee, Y., et al., *The nuclear RNase III Drosha initiates microRNA processing.* nature, 2003. **425**(6956): p. 415-419.
8. Yi, R., et al., *Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs.* Genes & development, 2003. **17**(24): p. 3011-3016.
9. Lund, E., et al., *Nuclear export of microRNA precursors.* Science, 2004. **303**(5654): p. 95-98.

10. HutvÃ¡gner, G.r., et al., *A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA.* Science, 2001. **293**(5531): p. 834-838.
11. Grishok, A., et al., *Genes and Mechanisms Related to RNA Interference Regulate Expression of the Small Temporal RNAs that Control< i> C. elegans</i> Developmental Timing.* Cell, 2001. **106**(1): p. 23-34.
12. Ketting, R.F., et al., *Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans.* Genes & development, 2001. **15**(20): p. 2654-2659.
13. HutvÃ¡gner, G.r. and P.D. Zamore, *A microRNA in a multiple-turnover RNAi enzyme complex.* Science, 2002. **297**(5589): p. 2056-2060.
14. Okamura, K., et al., *Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways.* Genes & development, 2004. **18**(14): p. 1655-1666.
15. Grundhoff, A. and C.S. Sullivan, *Virus-encoded microRNAs.* Virology. **411**(2): p. 325-343.
16. Lin, Z. and E.K. Flemington, *miRNAs in the pathogenesis of oncogenic human viruses.* Cancer letters. **305**(2): p. 186-199.
17. Seeger, C. and W.S. Mason, *Hepatitis B virus biology.* Microbiology and Molecular Biology Reviews, 2000. **64**(1): p. 51-68.
18. Lai, C.-L. and M.-F. Yuen, *Chronic hepatitis Bâ€"new goals, new treatment.* New England Journal of Medicine, 2008. **359**(23): p. 2488-2491.
19. Yang, H.-I., et al., *Associations between hepatitis B virus genotype and mutants and the risk of hepatocellular carcinoma.* Journal of the National Cancer Institute, 2008. **100**(16): p. 1134-1143.
20. LepÃ¨re-Douard, C., et al., *The first transmembrane domain of the hepatitis B virus large envelope protein is crucial for infectivity.* Journal of virology, 2009. **83**(22): p. 11819-11829.
21. Weiser, B., et al., *Closed circular viral DNA and asymmetrical heterogeneous forms in livers from animals infected with ground squirrel hepatitis virus.* Journal of virology, 1983. **48**(1): p. 1-9.
22. Wei, Y., et al., *Molecular biology of the hepatitis B virus and role of the X gene.* Pathologie Biologie. **58**(4): p. 267-272.
23. Pfeffer, S.b., et al., *Identification of virus-encoded microRNAs.* Science, 2004. **304**(5671): p. 734-736.
24. Cai, X., et al., *Kaposi's sarcoma-associated herpesvirus expresses an array of viral microRNAs in latently infected cells.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(15): p. 5570-5575.
25. Samols, M.A., et al., *Cloning and identification of a microRNA cluster within the latency-associated region of Kaposi's sarcoma-associated herpesvirus.* Journal of virology, 2005. **79**(14): p. 9301-9305.
26. Pfeffer, S.b., et al., *Identification of microRNAs of the herpesvirus family.* Nature methods, 2005. **2**(4): p. 269-276.
27. Cui, C., et al., *Prediction and identification of herpes simplex virus 1-encoded microRNAs.* Journal of virology, 2006. **80**(11): p. 5499-5508.
28. Sullivan, C.S. and A. Grundhoff, *Identification of viral microRNAs.* Methods in enzymology, 2007. **427**: p. 1-23.
29. Grundhoff, A., C.S. Sullivan, and D. Ganem, *A combined computational and microarray-based approach identifies novel microRNAs encoded by human gamma-herpesviruses.* Rna, 2006. **12**(5): p. 733-750.
30. Xue, C., et al., *Classification of real and pseudo microRNA precursors using local structure-sequence features and support vector machine.* BMC bioinformatics, 2005. **6**(1): p. 310.
31. Gkirtzou, K., et al., *MatureBayes: a probabilistic algorithm for identifying the mature miRNA within novel precursors.* PloS one. **5**(8): p. e11843.
32. Jin, W.-B., et al., *HBV-encoded microRNA candidate and its target.* Computational biology and chemistry, 2007. **31**(2): p. 124-126.