

## Article

# A Mathematical Analysis of HDV Genotypes: From Molecules to Cells

Rami Zakh <sup>1,†</sup>, Alexander Churkin <sup>2,\*</sup>, Franziska Totzeck <sup>3,†</sup>, Marina Parr <sup>3</sup>, Tamir Tuller <sup>4</sup>, Ohad Etzion <sup>5</sup>, Harel Dahari <sup>6</sup> , Michael Roggendorf <sup>7</sup>, Dmitriy Frishman <sup>3,\*</sup> and Danny Barash <sup>1,\*</sup> 

- <sup>1</sup> Department of Computer Science, Ben-Gurion University, Beer-Sheva 8410501, Israel; zakhr@post.bgu.ac.il  
<sup>2</sup> Department of Software Engineering, Sami Shamoon College of Engineering, Beer-Sheva 8410501, Israel  
<sup>3</sup> Department of Bioinformatics, Wissenschaftszentrum Weihenstephan, Technische Universität München, Maximus-von-Imhof-Forum 3, 85354 Freising, Germany; f.totzeck@tum.de (F.T.); mar.ark.parr@gmail.com (M.P.)  
<sup>4</sup> Department of Biomedical Engineering, Tel-Aviv University, Tel-Aviv 6997801, Israel; tamirtul@post.tau.ac.il  
<sup>5</sup> Soroka University Medical Center, Ben-Gurion University, Beer-Sheva 8410501, Israel; ohadet34@yahoo.com  
<sup>6</sup> Stritch School of Medicine, Loyola University Chicago, Maywood, IL 60153, USA; hdahari@luc.edu  
<sup>7</sup> Institute of Virology, Technische Universität München, 81675 Munich, Germany; michael.roggendorf@tum.de  
\* Correspondence: alexach3@sce.ac.il (A.C.); d.frishman@wzw.tum.de (D.F.); dbarash@cs.bgu.ac.il (D.B.)  
† Equal contribution.

**Abstract:** Hepatitis D virus (HDV) is classified according to eight genotypes. The various genotypes are included in the HDVdb database, where each HDV sequence is specified by its genotype. In this contribution, a mathematical analysis is performed on RNA sequences in HDVdb. The RNA folding predicted structures of the Genbank HDV genome sequences in HDVdb are classified according to their coarse-grain tree-graph representation. The analysis allows discarding in a simple and efficient way the vast majority of the sequences that exhibit a rod-like structure, which is important for the virus replication, to attempt to discover other biological functions by structure consideration. After the filtering, there remain only a small number of sequences that can be checked for their additional stem-loops besides the main one that is known to be responsible for virus replication. It is found that a few sequences contain an additional stem-loop that is responsible for RNA editing or other possible functions. These few sequences are grouped into two main classes, one that is well-known experimentally belonging to genotype 3 for patients from South America associated with RNA editing, and the other that is not known at present belonging to genotype 7 for patients from Cameroon. The possibility that another function besides virus replication reminiscent of the editing mechanism in HDV genotype 3 exists in HDV genotype 7 has not been explored before and is predicted by eigenvalue analysis. Finally, when comparing native and shuffled sequences, it is shown that HDV sequences belonging to all genotypes are accentuated in their mutational robustness and thermodynamic stability as compared to other viruses that were subjected to such an analysis.

**Keywords:** RNA graph representation; Laplacian eigenvalues; viral kinetic models; hepatitis delta virus genotypes; folding energy; mutational robustness



**Citation:** Zakh, R.; Churkin, A.; Totzeck, F.; Parr, M.; Tuller, T.; Etzion, O.; Dahari, H.; Roggendorf, M.; Frishman, D.; Barash, D. A Mathematical Analysis of HDV Genotypes: From Molecules to Cells. *Mathematics* **2021**, *9*, 2063. <https://doi.org/10.3390/math9172063>

Academic Editor: Eva H. Dulf

Received: 7 July 2021

Accepted: 24 August 2021

Published: 26 August 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

RNA secondary structures assume functional roles during the virus life cycle and are therefore a topic of considerable interest [1,2]. Over the years, specific functions of RNA structure motifs have been examined in many viruses (for example, in the hepatitis C virus (HCV) [3–5]) by combining computational and experimental approaches. In recent years, the advent of high-throughput structure-probing methods has opened up exciting opportunities in elucidating the viral RNA structure repertoire at a large scale [6]. A large percentage of viral RNA motifs tend to possess linear secondary structures similar to the ones depicted in [7], which are often stem-loop structures designated by SL for their identification. Stem-loop structures can be found in a multitude of viruses, e.g., [7], and in

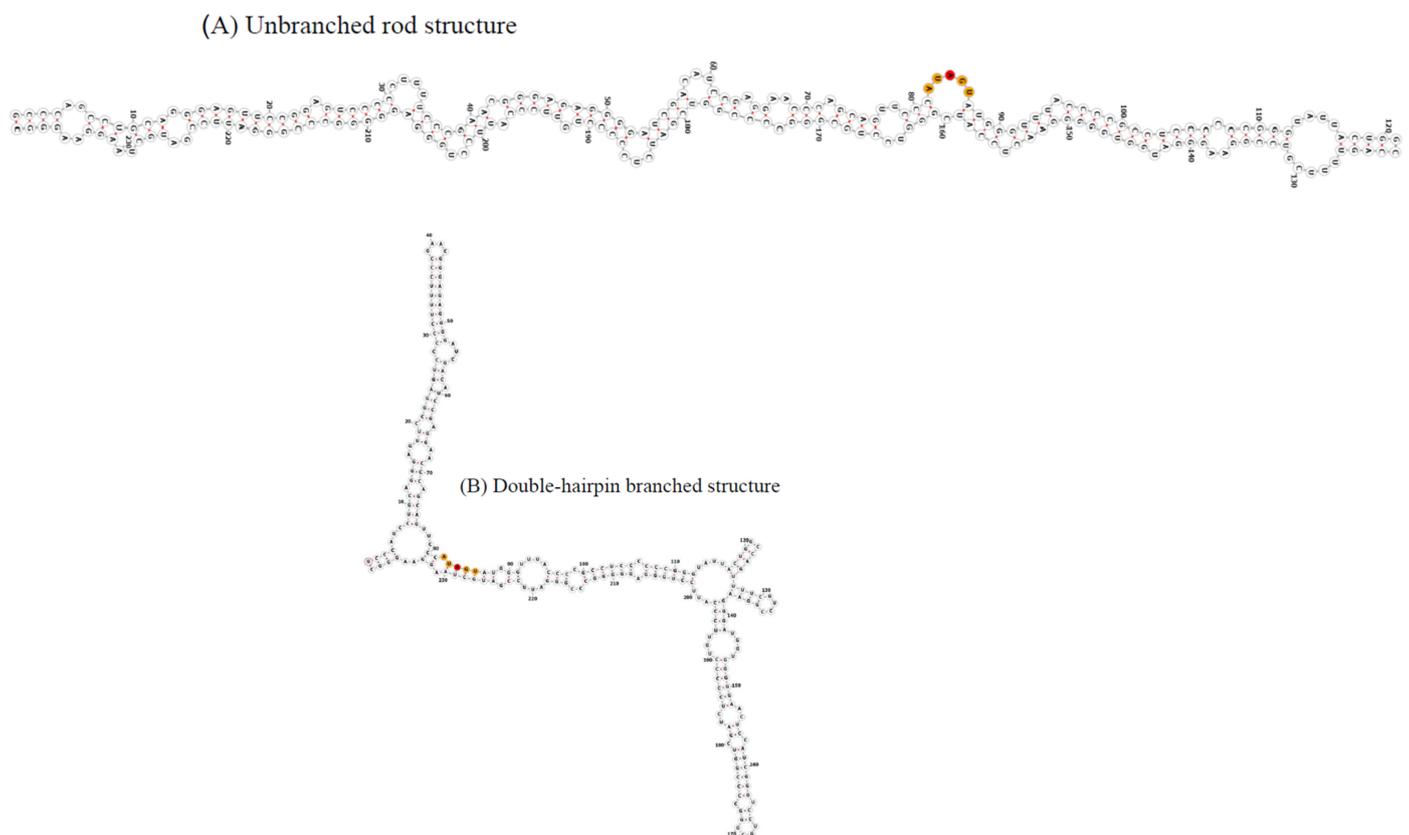
general, they are important from the evolutionary perspective [8]. In viruses, stem-loop structural motifs have been investigated in our previous article [9]. This contribution is a continuation of [9] that specifically considers hepatitis delta virus (HDV), which has the smallest human viral genome and uses the envelope of hepatitis B virus (HBV) to generate infectious particles. Two main forms of HDV infection have been described: (1) co-infection together with HBV, with a high rate of viral clearance in adults clinically similar to HBV mono-infection [10], or (2) super-infection in the presence of a pre-existing HBV infection. The latter results in a persistent chronic HDV infection in 70–90% of the cases and early on was shown to be associated with a high risk to develop cirrhosis and primary liver cancer (HCC) [11]. Since the discovery of this virus, eight genotypes with distinct geographical and ethnic regions and at least nine subtypes have been defined [12]. The lowest inter-genotype divergence of  $\geq 10\%$  was reported between HDV-5 and HDV-2, whereas the highest inter-genotype difference was estimated for HDV-3 [12], which is predominant in South America. For certain genotypes, differences in clinical outcomes have been observed [13]. Genotype 3 is associated with fulminant hepatitis epidemics with high lethality rates [14].

An important issue in modeling and analyzing RNAs is the representation of their secondary structure, desirably in a simplified and yet useful manner. Several approaches have been devised, among which some major pioneering ones are the full graph representation in which each nucleotide is a node [15], a coarse-grain tree-graph representation in which each motif is a node [16], and a full tree forming a homeomorphically irreducible tree [17]. All of the aforementioned representations have been implemented in the Vienna RNA package [18]. The full graph representation is equivalent to the dot-bracket representation in the Vienna RNA package [18–20] and the ct file in mfold/UNAFold [21–23].

In the context of RNA secondary structure analysis, coarse-grain tree-graphs have found a variety of uses [16,24–28]. It is also possible to generalize the coarse-grain representations to abstract shapes [29]. In [16,24], a coarse-grain representation of the RNA secondary structure was proposed, which was later named Shapiro's representation in the Vienna RNA package. In [25], topological indices were first suggested to be used for analyzing coarse-grain tree-graphs. In [26,27], it was found that the second smallest eigenvalue of the Laplacian matrix is able to provide a similarity measure for differentiating between various RNA tree-graph topologies. The smallest eigenvalue of the Laplacian matrix is identically zero and then the second smallest eigenvalue, which is called algebraic connectivity, provides a measure of how much the tree-graph is linear (a path) or compact (a star) as illustrated in [27]. This concept can be applied when filtering candidates in the process of RNA deleterious mutation prediction, which was used in the relevant prediction software RNAmute and its extension [28,30,31]. Aside from design applications that have to do with conformational switching and multistable RNAs [32], it will also be used here to detect large conformational switches that were reported in [33] by applying a similar idea. In passing, it is worthwhile noting that conformational switching is present in both HCV [34] and HIV [35], as examples, in structural elements of different length scales and with different mechanisms than the relevant one in HDV. Reverting to [26,27], seminal theorems by Fiedler and Merris [36,37] were shown to be applicable for the examination of how the coarse-grain tree-graph representing the RNA secondary structure is shaped. Following Fiedler's work on the algebraic connectivity in general graphs, Merris has shown how tree-graphs can be ordered by their algebraic connectivity [38], alongside a wider perspective of the Laplacian spectrum of a graph [39,40].

Following [9], in which the dataset was taken from [41], herein, the initial plan was to apply a similar methodology to perform structural classification in HDVdb [42]. Unlike in [9], the sequences in HDVdb are no longer small RNAs, and there are only 512 sequences. Nevertheless, when RNA folding prediction methods are applied on these sequences, stem-loop unbranched rod structures emerge that are mostly completely linear and can be represented by a tree-graph that is a path on  $n$  vertices. This prompted us to specifically examine the well-known conformation switch in HDV that was mentioned above [33] and

check whether our methodology that was outlined in [9] can detect more HDV genotypes other than the peculiar genotype 3 that is associated with such a conformational switch. Because all genotypes are represented in HDVdb [42], the topic of genotypes was highlighted in our analysis. Our goal was to filter out most sequences in HDVdb that only exhibit an unbranched rod structure in their folding prediction by energy minimization and only collect the sequences that exhibit a double-hairpin branched structure, classifying them according to their genotypes. We performed the filtering using an approach that will be described in detail, also addressing the advantages of our approach, in the Results and Discussion sections. In genotype 3, the unbranched rod structure displayed in Figure 1 (containing the essentials of Figure 2A of [43]) is responsible for virus replication, and the double-hairpin branched structure displayed in Figure 1 (containing the essentials of Figure 2B of [43]), is responsible for RNA editing. With the issue of genotypes as concentration, aside from examining RNA structures at the level of molecules, we also address a yet unexplored topic of how genotypes may affect HDV viral kinetics at the level of cells and provide a future perspective. Finally, as in [9], we also examine the mutational robustness and thermodynamic stability of HDVdb sequences/structures. We demonstrate for completeness that these sequences are mutationally robust and thermodynamically stable, as expected, in comparison to their corresponding shuffled sequences and their predicted structures.



**Figure 1.** (A) To the left, the unbranched rod structure of HDV, isolate Peru-1 (L22063), obtained by mfold prediction with the optimal solution taken, containing the most essential information of Figure 2A of [43]. (B) To the right, the double-hairpin branched structure of HDV, isolate Peru-1 (L22063), obtained by mfold prediction with the second suboptimal solution taken, containing the most essential information of Figure 2B of [43]. The five nucleotides “AUAGU” comprising the editing site are colored.

## 2. Materials and Methods

The mathematical analysis consists of two main components that have been put forth in [9]. The first component relies on filtering rod-like structures that are dominant in HDV because they correspond to virus replication. These rod-like structures can be distinguished by their unique coarse-grain tree-graph representation, which is a path on  $n$  vertices. Section 2.1 describes the filtering method that utilizes a formula for the second-lowest eigenvalue of the Laplacian matrix corresponding to a coarse-grain tree-graph representation that is a path on  $n$  vertices. Section 2.2 describes the second component that involves an analysis of mutational robustness and thermodynamic stability.

### 2.1. Defining the Laplacian Matrix of a Tree-Graph and Calculating Its Second Lowest Eigenvalue for a Path

The coarse-grain tree-graph representation of an RNA secondary structure, also known as the Shapiro representation [16], enables an initial analysis of the collection of RNA structures based on their constituent motifs and their compactness. The tree-graph can then be represented by its corresponding Laplacian matrix.

Let  $T = (V, E)$  be a tree with vertex set  $V = \{v_1, v_2, \dots, v_n\}$  and edge set  $E$ . Let us denote by  $d(v)$  the degree of  $v$  where  $v \in V$  is a vertex of  $T$ . The Laplacian matrix of  $T$  is  $L(T) = (l_{ij})$ , where:

$$l_{ij} = \begin{cases} d(v_i), & \text{if } i = j, \\ -1, & \text{if } v_i, v_j \in E \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

The eigenvalues of the Laplacian matrix are independent of the choice of labeling for the nodes in the tree-graph  $T$ , which only amounts to interchanges of the rows and columns. For example, with the orderly labeling of the linear tree-graph containing four nodes as in Figure 2, the Laplacian matrix  $L(T)$  becomes:

$$L_4 = \begin{pmatrix} 1 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 1 \end{pmatrix} \quad (2)$$

The smallest eigenvalue of the Laplacian matrix is identically zero. The second smallest eigenvalue, which is called the algebraic connectivity [36], provides a measure of how much the tree-graph is linear (a path) or compact (a star). In the following, we will derive a formula for the second smallest eigenvalue of a path in an alternative way to the one used in [36], directly utilizing the tridiagonal structure of the Laplacian matrix.

The Laplacian matrix is a slight deviation from a tridiagonal Toeplitz matrix, which stems from the fact that tree-graph extreme vertices are bounded to a single neighbor, rather than two. For example, the tridiagonal Toeplitz matrix corresponding to  $L_4$  above is:

$$L'_4 = \begin{pmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{pmatrix} \quad (3)$$

As noted in [9], calculating the eigenvalues of the Laplacian matrix could be approached in principle using a methodology similar to that of calculating the eigenvalues of a tridiagonal Toeplitz matrix. Calculating the eigenvalues of such a matrix (such as  $L'_4$ ) can be simplified by using a theorem whereby if  $L' = h(M)$  and the eigenvalues of  $M$  are

calculated and denoted as  $\lambda_1, \lambda_2, \dots, \lambda_n$ , then the eigenvalues of  $L'$  are, respectively,  $h(\lambda_1), h(\lambda_2), \dots, h(\lambda_n)$ . In a generalized form for the matrix  $L'_n$ , we obtain:

$$L'_n = \begin{pmatrix} 2 & 0 & \cdot & 0 & 0 \\ 0 & 2 & \cdot & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & 2 & 0 \\ 0 & 0 & \cdot & 0 & 2 \end{pmatrix} - \begin{pmatrix} 0 & 1 & \cdot & 0 & 0 \\ 1 & 0 & \cdot & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & 0 & 1 \\ 0 & 0 & \cdot & 1 & 0 \end{pmatrix} = 2 \bullet I_n - 1 \bullet M_n \tag{4}$$

where  $M_n$  is an  $n \times n$  tridiagonal matrix of the following form:

$$M_n = \begin{pmatrix} 0 & 1 & \cdot & 0 & 0 \\ 1 & 0 & \cdot & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & 0 & 1 \\ 0 & 0 & \cdot & 1 & 0 \end{pmatrix} \tag{5}$$

Because the eigenvalues of  $I_n$  are trivial, finding the eigenvalues of a tridiagonal Toeplitz matrix amounts to finding those of  $M_n$ . This can be completed in at least two ways. The first, which is longer, is to use elementary properties of determinants to directly derive the characteristic polynomial and then find its roots. As worked out in [44], denoting the characteristic polynomial of the matrix  $M_n$  by  $\varphi_n(x)$  and using the transformation  $x = 2 \bullet \cos \Theta$ , the characteristic polynomial becomes, after a considerable derivation:

$$\varphi_n(2 \bullet \cos \Theta) = \frac{\sin(n+1)\Theta}{\sin \Theta} \tag{6}$$

which is solved at  $\Theta = \frac{k\pi}{n+1}$  ( $k = 1, \dots, n$ ), yielding the eigenvalues:

$$\lambda_k(M_n) = 2 \bullet \cos \frac{k\pi}{n+1} \quad (k = 1, \dots, n) \tag{7}$$

The second way, which is shorter, is to note that the expansion of the characteristic polynomial satisfies the three-point recurrence relationship that matches the recursive formula for the Chebyshev polynomials of the first kind; see also [45]:

$$T_n(x) = x \bullet T_{n-1}(x) - T_{n-2}(x) \tag{8}$$

where  $T_n(x)$  is the Chebyshev polynomial of the first kind of order  $n$ , e.g.,  $T_1(x) = x$  and  $T_2(x) = x^2 - 1$ . For  $x = \cos \Theta$ , the following relationship holds [46]:

$$T_n(\cos \Theta) = \cos(n\Theta) \tag{9}$$

Thus, we obtain that the zeros of  $T_n(x)$  are the roots given by Equation (7). Therefore, using Equation (4), the eigenvalues of  $L'_n$  are given by:

$$\lambda_k(L'_n) = 2 \bullet \lambda_k(I_n) - 1 \bullet \lambda_k(M_n) = 2 - 2 \bullet \cos \frac{k\pi}{n+1} \quad (k = 1, \dots, n) \tag{10}$$

Looking back at the Laplacian matrix for a path on  $n$  vertices in our application:

$$L_n = \begin{pmatrix} (2-1) & -1 & \cdot & 0 & 0 \\ -1 & 2 & \cdot & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & 2 & -1 \\ 0 & 0 & \cdot & -1 & (2-1) \end{pmatrix} \tag{11}$$

which can also be written as:

$$L_n = \begin{pmatrix} 2 & 0 & \cdot & 0 & 0 \\ 0 & 2 & \cdot & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & 2 & 0 \\ 0 & 0 & \cdot & 0 & 2 \end{pmatrix} - \begin{pmatrix} 1 & 1 & \cdot & 0 & 0 \\ 1 & 0 & \cdot & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & 0 & 1 \\ 0 & 0 & \cdot & 1 & 1 \end{pmatrix} = 2 \bullet I_n - 1 \bullet P_n \tag{12}$$

where  $P_n$  is similar to  $M_n$  of Equation (5) except for two additional ones at the beginning and end of the main diagonal:

$$P_n = \begin{pmatrix} 1 & 1 & \cdot & 0 & 0 \\ 1 & 0 & \cdot & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & 0 & 1 \\ 0 & 0 & \cdot & 1 & 1 \end{pmatrix} \tag{13}$$

$P_n$  is a slight variation on  $M_n$ , and using a similar procedure, it is found (e.g., see [44]) that its eigenvalues are given as the roots of:

$$\varphi_n(2 \bullet \cos \Theta) = \frac{2 \bullet (\sin n\Theta) \bullet (\cos \Theta - 1)}{\sin \Theta} \tag{14}$$

which is solved at  $\Theta = \frac{k\pi}{n}$  ( $k = 0, 1, \dots, n - 1$ ), hence:

$$\lambda_k(P_n) = 2 \bullet \cos \frac{k\pi}{n} \quad (k = 0, 1, \dots, n - 1) \tag{15}$$

From Equations (12) and (15), it follows that the eigenvalues of  $L_n$  are:

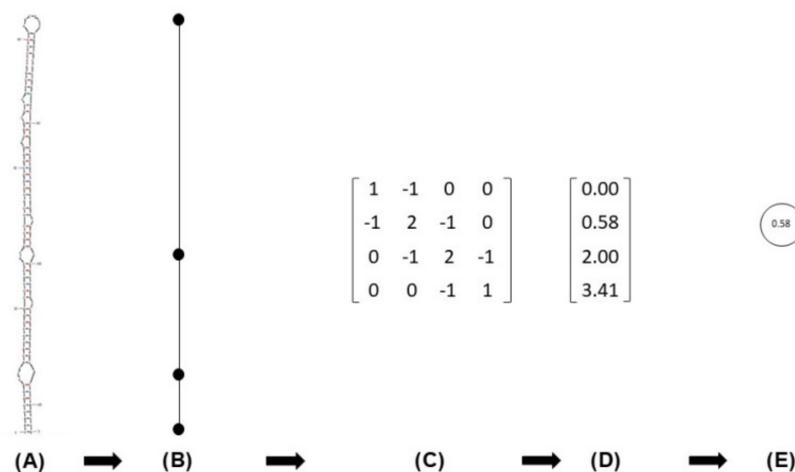
$$\lambda_k(L_n) = 2 \bullet \lambda_k(I_n) - 1 \bullet \lambda_k(P_n) = 2 - 2 \bullet \cos \frac{k\pi}{n} \quad (k = 0, 1, \dots, n - 1) \tag{16}$$

which leads to the trivial eigenvalue of zero for  $k = 0$ , and the smallest second eigenvalue for  $k = 1$  that is given by:

$$a(T) = 2(1 - \cos(\pi/n)) \tag{17}$$

The second smallest eigenvalue of the Laplacian matrix is called the algebraic connectivity [36] of  $T$  and labeled as  $a(T)$ . Some of the properties of  $a(T)$  that concern the application presented here are mentioned in the Appendix of [9], in addition to the illustrative calculation of  $a(T)$  for the RNA secondary structure example shown in Figure 2.

Note that by convention of the choice of tree-graph representation, those loops with single isolated nucleotides are not accounted for as nodes, but the 5'-3' ends are counted as a node. In the case of a star of four vertices, for example,  $a(T) = 1.0$ , which is the upper bound for the algebraic connectivity. A star applies for a tree-graph possessing three vertices or more ( $n \geq 3$ ), and the algebraic connectivity of a star is always unity [40]. The algebraic connectivity  $a(T)$  is characterized by some special properties described in the Appendix of [9] that are advantageous for the RNA secondary structure application that is presented here.



**Figure 2.** RNA secondary structure and coarser levels of representation. (A) Secondary structure of an HDV virusoid sequence taken from [47]. (B) Tree-graph representation of the secondary structure. (C) Laplacian matrix corresponding to the tree-graph representation. (D) Spectrum of the Laplacian matrix. (E) Second-smallest eigenvalue of the Laplacian matrix.

## 2.2. Mutational Robustness and Thermodynamic Stability

For quantitatively measuring mutational robustness, the neutrality  $\eta$  is calculated. Given an RNA sequence of length  $N$ , the neutrality is calculated by:

$$\eta = (N - d)/N \quad (18)$$

where  $d$  is the base-pair distance between secondary structure of the original sequence and secondary structure of the mutant, averaged over all  $3N$  one-mutant neighbors. The base-pair distance available in the Vienna RNA package is used to calculate the distance between two RNA secondary structures. The RNA secondary structures in this study were predicted by energy minimization approach [18,21] using RNAfold available in the Vienna RNA package [18–20], noting that similar predictions can be completed with mfold/UNAFold [21–23].

## 3. Results

### 3.1. Eigenvalue Analysis

As was conceptualized in [9], we start with an eigenvalue analysis of the HDV genome sequences that are available in [42]. Despite their length of several hundred nucleotides, the HDV genome sequences are known to be well predicted by energy minimization methods, tending to have linear rod-like structures. It is worthwhile noting that the genomes are circular, and because of the antigenome concept, a reverse complement should be taken before inserting the sequences as input to energy minimization software that relies on energy parameters [48] such as mfold (as used in the past in [43]) or RNAfold, choosing the circular folding option. MPGAfold [33] can also be used, but it is more involved to apply. Having experimented with some sequences that are relevant to the HDV editing that takes place in genotype 3 from patients in South America, the branched double-hairpin structures responsible for editing are found in mfold's suboptimal solutions as depicted in Figure 1 (containing the essentials of Figure 2B of [43]). There is a slight difference between the mfold and RNAfold solutions that worked to our advantage. While the branched double-hairpin structures are well captured in the manner that suboptimal solutions were devised in mfold [49], contrary to the detection of conformational switches by mutations application in [50], it would be more difficult to view them in the way that suboptimal solutions were devised in RNAsubopt [51]. Nevertheless, the optimal solutions of RNAfold with the structure ensemble are slightly more sensitive to deviations from the unbranched structure. Thus, while it is possible to extend our simulations to consider

the first few suboptimal solutions of mfold, we found that it is enough, for a convincing sample of sequences that we experimented with, to only consider the optimal solution of RNAfold as a way to detect branched double-hairpin structures successfully and then double-check the relevant sequences by examining their suboptimal solutions with mfold (in some cases, the branched double-hairpin structure will already appear in the optimal solution of mfold). The following table reports an eigenvalue analysis of all HDV genome sequences in [42] by calculating their second-smallest eigenvalue of the Laplacian matrix corresponding to the tree-graph representation of their folding prediction as illustrated in Figure 2. As explained above, for the folding prediction, we used RNAfold with only the optimal solution considered. The code for calculating the eigenvalues was written in Java and is available at <https://github.com/ChurkinAlex/RNAStructureEigenvalueCalculator> (accessed on 3 July 2021).

This will further be analyzed in Section 4 with respect to the formula  $a(T) = 2(1 - \cos(\pi/n))$  in Equation (17) for a linear tree-graph representation of the RNA secondary structure to show the tendency of HDV viruses to possess unbranched (rod-like) stem-loop structures represented by a path on  $n$  vertices. From the biological standpoint, it shows that the dominant structure is the unbranched structure responsible for virus replication. It should be noted that the number of branched structured sequences could have been calculated by alternative approaches such as going through each one of the vertices in the tree-graph and checking whether there is an external loop besides the ones on the two opposite ends. This is more cumbersome as compared to the mathematical approach as described above with using Equation (17), simply plugging the number of vertices  $n$  in the equation and checking whether the resultant algebraic connectivity is equal to the second smallest eigenvalue of the Laplacian matrices corresponding to the tree-graph representation obtained by our Java code. The running time to calculate the eigenvalues of the Laplacian matrices is notably very small because the tree-graph vertices represent RNA secondary structure motifs (e.g., bulges and loops), and therefore the size of all Laplacian matrices is smaller than  $90 \times 90$  according to the eigenvalue interval at the top row of Table 1. Thus, the mathematical approach, in this case, is both more elegant and efficient as compared to an alternative computational approach that is based on base-pairing or loop-type identification.

**Table 1.** Eigenvalue analysis: distribution of the HDV sequences in the various genotypes according to the second-smallest Laplacian eigenvalue.

Genotype	Eigenvalue (Interval)	Number of Sequences	Number of Branched Structured Sequences
All	0.0013–0.0031	512	187
Genotype 1 (“italiense”)	0.0015–0.0031	321	100
Genotype 2 (“japanense”)	0.0014–0.0021	24	5
Genotype 3 (“peruense”)	0.0013–0.0016	11	4
Genotype 4 (“taiwanense”)	0.0015–0.0020	37	23
Genotype 5 (“togense”)	0.0013–0.0017	23	7
Genotype 6 (“careense”)	0.0014–0.0020	15	9
Genotype 7 (“cameroonense”)	0.0013–0.0021	45	33
Genotype 8 (“senegalense”)	0.0014–0.0016	6	0
Undefined	0.0013–0.0020	30	6

In examining the second column of Table 1 that lists the eigenvalue interval values themselves, one cannot assert that one particular HDV genotype is more remote or peculiar than the other HDV genotypes. However, the second eigenvalue of the Laplacian matrix is just a single similarity measure, and tree edit distances are more informative for this purpose. Thus, we computed pairwise tree edit distances using RNAdistance in the Vienna RNA package to perform an analysis considering a few representatives of each HDV genotype in order to assess the similarity between HDV genotypes. The results are

displayed in Table 2. While the results indicate that there is no particular genotype more distant from other genotypes, which is in line with the eigenvalue analysis, the tree edit distance is a more comprehensive similarity measure for such purposes as compared to the second eigenvalue of the Laplacian matrix, which is not a distance.

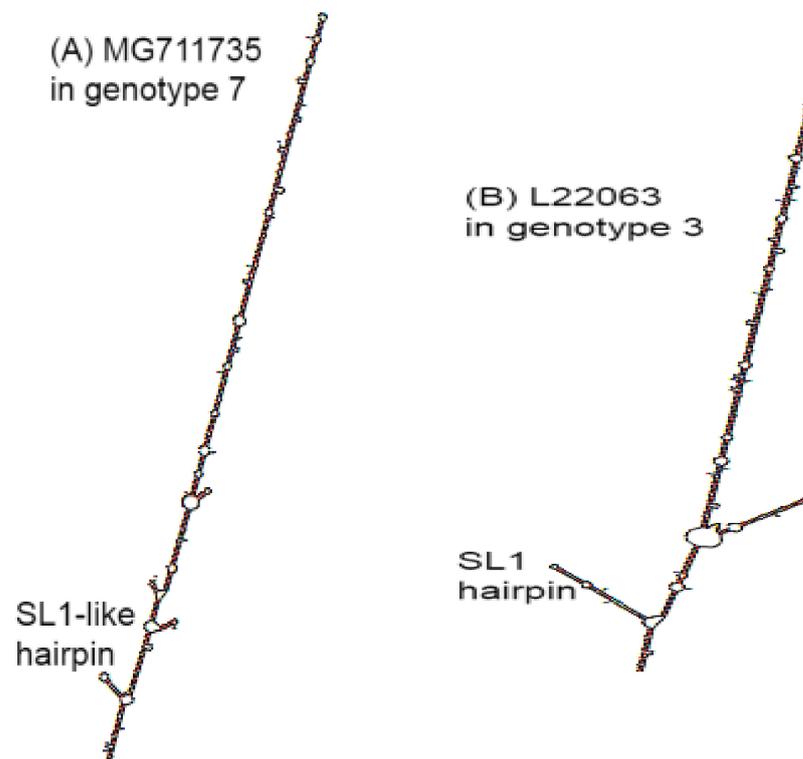
**Table 2.** Tree edit distance analysis: tree edit distances between the eight HDV genotypes.

	Genotype 1	Genotype 2	Genotype 3	Genotype 4	Genotype 5	Genotype 6	Genotype 7	Genotype 8
Genotype 1	0	457	463	446	423	416	432	423
Genotype 2	457	0	500	455	450	437	461	430
Genotype 3	463	500	0	469	470	475	471	454
Genotype 4	446	455	469	0	419	440	462	431
Genotype 5	423	450	470	419	0	399	433	378
Genotype 6	416	437	475	440	399	0	442	409
Genotype 7	432	461	471	462	433	442	0	395
Genotype 8	423	430	454	431	378	409	395	0

Although the second eigenvalue of the Laplacian matrix is not a distance, it is still a useful similarity measure. We observe that for the HDV viral genomes circular folding predictions that are represented by the Shapiro coarse-grain tree-graph representation, the values in Table 1 are relatively very small compared to the values calculated for other RNA secondary structures found in nature (e.g., [26,27]), and the properties of the algebraic connectivity listed in the Appendix of [9] can be used in a beneficial manner.

### 3.2. A Directed Structure-Based Search

Having observed in Table 1 that most of the HDV sequences possess unbranched rod-like structures (64% in total) that can be filtered out, we remain with a smaller pool of branched structured sequences. Within them, there is even a smaller pool of branched multi-hairpin structured sequences and a much smaller pool of branched multi-hairpin structure sequences in which the SL1 upper-left hairpin of Figure 1B (depicting the essentials drawn in Figure 2B of [43]) contains the sub-sequence “GAAC” in the external loop of the SL1 hairpin. We chose to concentrate on this sub-sequence because it was shown to be important for RNA editing (e.g., [43]), and it is known that the GAAC tetraloop is a frequent sequence in the GANC family that has function capabilities. We find that only a few structures contain this hairpin composition. The structures of genotype 3 from the Amazon rainforest in South America (Peru, Bolivia, Brazil, Ecuador, Venezuela) contain it, validating our method, as well as two outliers that can be easily discarded as such. However, we noticed another group of structures that contain it belonging to genotype 7 from Cameroon, which is further discussed in Section 4. Figure 3 illustrates how the conformational switching from an unbranched structure to a branched structure in genotype 7 displayed in Figure 3A resembles and differs from the known conformational switching that takes place in genotype 3 displayed in Figure 3B. The SL1-like hairpin in Figure 3 contains the sub-sequence “GAAC” in its external loop, and it is also noted that it appears in the optimal solution of mfold, while the SL1 hairpin in Figure 3 that is exactly composed of the sub-sequence “GAAC” in its external loop appears in the second suboptimal solution of mfold, raising the possibility that genotype 7 is more susceptible to conformational switching than genotype 3.



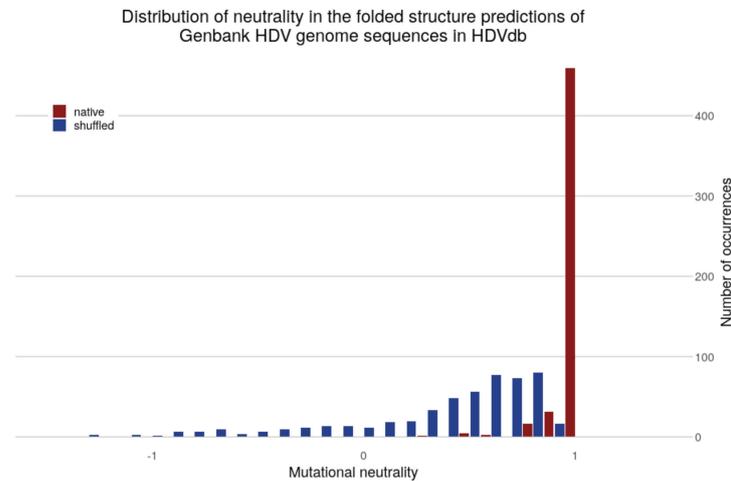
**Figure 3.** (A) Secondary structure prediction by mfold/UNAFold of MG711735 from genotype 7 (upper part) with an SL1-like hairpin that contains the sub-sequence “GAAC” as in SL1. Conformational switching from an unbranched structure to a branched structure appears in the optimal solution. (B) Secondary structure prediction by mfold/UNAFold of L22063 from genotype 3 (upper part) with the SL1 hairpin that is known to be responsible for RNA editing. Conformational switching from an unbranched structure to a branched structure appears in the second suboptimal solution.

It should also be noted that concerning the sub-sequence required for RNA editing, we have tried searching for the five nucleotides “AUAGU” (representing the AMBER/W site for genotype 3 in South America) that are mentioned in the caption of Figure 1 and are colored in the figure itself over the entire Genbank HDV genome sequences file of HDVdb. We found that they are most prevalent in genotype 7 in Cameroon with 26 out of 69 hits (38%), with genotype 3 in South America coming second with 10 out of 69 hits (14%), and with the rest of the genotypes in other geographical locations possessing much lower hits. However, because of the antigenome concept, finding the reverse complement of these five nucleotides near SL1 would have shown a precise similarity to the editing mechanism in genotype 3. In our case, alternative scenarios or mechanisms that are non-explored yet and utilize conformational switching are also possible.

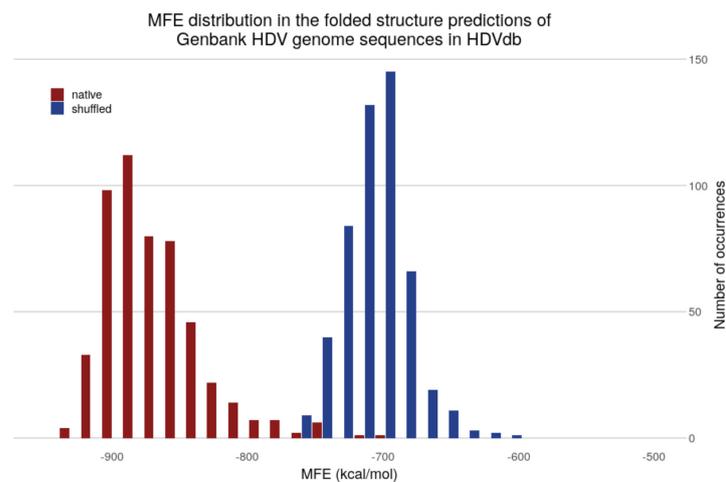
### 3.3. Mutational Robustness and Thermodynamic Stability

Finally, to conclude the analysis of HDV sequences, it is worthwhile verifying that the HDV native sequences are more mutationally robust and more thermodynamically stable than their corresponding shuffled HDV sequences as expected. Equation (18) is used to calculate the neutrality, and in all of the calculations, the RNAfold of the Vienna RNA Package 2.0 [20] is used for RNA structure prediction by energy minimization. Figure 4 depicts the neutrality distribution of native and shuffled HDV sequences for all structures in the dataset, having used Python’s multiprocessing module (“process” class) to speed up the calculation. Shuffling is performed by using shuffleseq from EMBOSS with dinucleotide shuffling. Figure 5 depicts the mean free energy (MFE) distribution for all structures in the dataset. The reported *p*-values in the captions are calculated from a T-test to check that the average neutrality values of the structures of native sequences are significantly larger than

the average neutrality values of the structures of shuffled sequences and the average MFE values of the structures of native sequences are significantly smaller than the average MFE values of the structures of shuffled sequences, respectively.



**Figure 4.** The neutrality distribution of all structures from our dataset comparing native and shuffled sequences. Shuffled sequences with dinucleotide shuffling were obtained using shuffleseq from EMBOSS. Reported  $p$ -value (see text) is less than 0.0001.



**Figure 5.** The MFE distribution of all structures in our dataset comparing native and shuffled sequences. Shuffled sequences with dinucleotide shuffling were obtained using shuffleseq from EMBOSS. Reported  $p$ -value (see text) is less than 0.0001.

#### 4. Discussion and Conclusions

The mathematical methods that are used herein to analyze HDV genotypes are conceptually within a similar framework as the ones we have used in [9], taking advantage of the properties of the Laplacian matrix and its second smallest eigenvalue to indicate how the coarse-grain representation of the RNA secondary structure is assumed in numerous scenarios but modified to our HDV application that aims to detect conformational switching. This also applies to the computational methods of using RNA folding prediction by energy minimization, by which we also calculate neutrality and MFE distributions. The results obtained in the previous section can now be analyzed, and as a consequence, a hypothesis can be formulated that suggests that the specific mechanism by which RNA editing occurs in HDV genotype 3 (as illustrated in detail, for example, in [52]) may not be limited to genotype 3 from the Amazon rainforest that originated this research [53]. By our

analysis with HDVdb [42], we may have possibly found traces for it as well in genotype 7 from Cameroon.

From Table 1, one can notice that most of the HDV genome sequences in [42] (64%) are predicted to fold to an unbranched (rod-like) stem-loop structure represented by a path on  $n$  vertices because this stem-loop structure is responsible for the virus replication. If we are seeking to detect the branched double-hairpin structure that is responsible for the virus RNA editing, we can preliminary filter out a large number of sequences for which their second smallest eigenvalue of the Laplacian matrix corresponds to Equation (17) for the algebraic connectivity of a path on  $n$  vertices. In general, from the perspective of the ordering of trees by their algebraic connectivity [38], all of the predicted folded structures tend to have a very low algebraic connectivity number close to zero because of the tendency to have a linear stem-loop structure (for HDV genome sequences, even longer in linearity than the examples in the literature that were cited in [9]). By examining the different genotypes, one can observe in Table 1 that genotype 7 is contrasting the behavior of the total number of sequences in the first row of Table 1, and about 73% of the genotype 7 sequences are predicted to fold to a branched structure, which may signal that this genotype may exhibit an interesting story in its ability to perform a conformational switch that relates to function. As a possibility, after further filtering according to the directed structure-based search as described in sub-Section 3.2, it is found that alongside the known sequences of genotype 3 from South America, a hairpin that appears similar to SL1 as depicted in Figure 3B can also be detected in genotype 7 from Cameroon, as depicted in Figure 3A. Three of these sequences from genotype 7, besides MG711735, are MG711804, MG711754 and LT604971 (if these sequences are inserted into mfold or RNAfold, the reverse complement should be taken beforehand, and circular folding should be selected).

For completion of our analysis method that was outlined in [9], by the comparison between native and shuffled sequences in [35], it is found in Figure 4 that native sequences are more mutationally robust than shuffled sequences. Additionally, Figure 5 shows that native sequences are more thermodynamically stable than shuffled sequences. This can serve as a verification that, indeed, the RNA sequences in [42] significantly exhibit characteristics of natural RNAs. Furthermore, it can be viewed from Figures 4 and 5 that HDV sequences are accentuated in their mutational robustness and thermodynamic stability as compared to HCV [54] and HIV [55] and, arguably, to other viruses, because of their very high amount of base pairing in their unbranched structure. It is also interesting to note in passing that from the evolutionary perspective, the viral RNA structures explored in this study and previously in [9] have maintained a high degree of conservation, whereas the high mutation rate of RNA viruses and their substantial capacity for adaptation could have made the sequences unidentifiable and likened them to “evolutionary losers” in the archaeology of coding RNA [56]. A mathematical analysis could also benefit the concepts that have been described in [56] and relate to ancient-like RNA elements having a specific structure in genomic viral RNAs.

Future work relating to mathematical modeling of HDV genotypes is to address HDV viral kinetics across the different genotypes by a simple differential equation model [57,58], which can be solved by standard mathematical software such as Matlab without the need for a sophisticated numerical solution as in, for example, [59,60], where the model is nonlinear. Some parameter values can be fixed based on [61], while others are estimated for each patient according to each participant’s viral kinetics. As in [62], for HCV genotypes 1 and 2, we hypothesize that HDV genotypes [63,64] can also potentially affect HDV viral kinetics.

The association between HDV genotypes and RNA folding has not yet been elucidated. Further research in this field may decipher the mechanisms responsible for some of the clinical phenomena observed in patients infected with HDV. These include the wide variability between patients in HDV replication rate, HDV vs. HBV dominance, the severity of clinical presentation following acute HDV infection, and the rate of disease progression

to cirrhosis and hepatocellular carcinoma. Some of these variations are linked with specific HDV genotypes, while others occur in patients infected with the same genotype. It is possible that sequence variations within (sub-genotypic) or across genotypes may yield different RNA folding patterns, which in turn could modify certain viral functional capacities such as editing and packaging efficiency as well as RNA–protein interactions. These conformational–functional alterations may bear an impact on viral infectivity, HBV–HDV interactions, and immune system activation, eventually translating into distinct clinical patterns of disease presentation and progression. The outcome of RNA–protein interaction is likely to provide some key insights into these open questions. More generally, the link between the cellular level that is addressed by viral kinetics modeling and the molecular level of examining primary and secondary structures of the RNA molecule might be more thoroughly understood by the prediction of RNA–protein interactions, a computational sub-field that has been developed in recent years and is actively being pursued these days, e.g., [65–67].

Future studies exploring structural–functional associations across HDV genotypes should center on the following perspective points: (1) the higher viral packaging efficiencies of genotype 1 vs. genotype 2 isolates; (2) the higher editing efficiencies of genotype 1 compared to those of genotype 2; (3) genotype 2 HDV infection is less frequently associated with fulminant hepatitis at the acute stage and follows a more benign long-term course (lower rate of progression to cirrhosis or hepatocellular carcinoma (HCC)) at the chronic stage as compared to genotype 1; (4) HDV genotype-dependent interactions with HBV (e.g., [68]).

To summarize, the biological and pathogenetic significance of HDV genotypes has not yet been elucidated in detail and is worthwhile exploring. Mathematical and computational methods can assist in these explorations and discover new findings. In this contribution, it is exemplified how a mathematical analysis of HDVdb [42] in which the aim was to explore how RNA tree-graphs are ordered by their algebraic connectivity assisted by the unique properties of the Laplacian matrices of tree-graphs and their second-lowest eigenvalue in Merris [38,69] can possibly yield new biological findings of clinical relevance. As per our analysis, we raise the possibility that HDV sequences of genotype 7 from Cameroon perform conformational switching that may affect their function, akin to a similar mechanism of RNA editing initially described in genotype 3 in South America. It is also possible that our analysis hints at another interesting biological function by way of conformational switching, other than RNA editing, that may take place in genotype 7. An analysis with a similar methodology can also be applied to other viruses and other biological agents of interest. From the pathogenetic point of view, a different clinical outcome of HDV infection has been observed for genotype 3 in comparison to the other genotypes. Due to the small numbers of genotype 7 cases with clinical information, such studies on disease outcome would be of interest.

Since the discovery of HDV, it was thought that this type of virus is only present in humans in an invariable association with HBV. However, recently, HDV-like sequences have been identified by metagenomic analyses in snakes (*Boa constrictor*), ducks (*Anas* species), rodents, fish, amphibians, and invertebrates (termites) without evidence of any HBV-like agent supporting infection. Most of these viruses have similar genomic features, including size and circular and unbranched rod-like structures. The snake-derived HDV protein bears 50% with L-HDAg. The duck-associated HDV protein shares only a homology of 32%. Detailed analysis of sequences and secondary structures as shown in this study for HDV genotypes will provide more insight on the relation of these agents to “human” HDV with respect to biology and possible pathogenesis.

Another open question is the interaction of certain HDV genotypes to corresponding genotypes of HBV. From the different genotypes of HBV, genotype F is the most prevalent in South America and is frequently associated with HDV super-infection. However, the more recent finding of the co-infection of HDV-3 with different genotypes of HBV suggests that the association between HDV-3 and HBV-F is not necessarily causally related to a more

severe clinical course of infection [68]. This indicates that the properties of the genome, including secondary structures, correlate with clinical outcomes. This may also be true for HDV-7 by the results of our mathematical analysis.

**Author Contributions:** Conceptualization, D.B.; methodology, R.Z., A.C., F.T., M.P., T.T., O.E., H.D., D.F. and D.B.; software, A.C., F.T. and R.Z.; investigation, A.C., F.T., R.Z., M.P., T.T., O.E., H.D., D.F. and D.B.; data curation, A.C. and F.T.; writing—original draft preparation, R.Z., A.C., F.T., M.P., T.T., O.E., H.D., D.F. and D.B.; writing—review and editing, R.Z., A.C., F.T., M.P., T.T., O.E., H.D., M.R., D.F. and D.B.; supervision, A.C., D.F. and D.B.; funding acquisition, T.T., H.D., D.F. and D.B. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the DFG, grant FR 1411/17-1, and the NIH grant R01AI146917.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** We acknowledge the help of Michael Kiening at an early stage of this work.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## References

- Hofacker, I.L.; Stadler, P.F.; Stocsits, R.R. Conserved RNA secondary structure in viral genomes: A survey. *Bioinformatics* **2004**, *20*, 1495–1499. [[CrossRef](#)]
- Marz, M.; Beerenwinkel, N.; Droste, C.; Fricke, M.; Frishman, D.; Hofacker, I.L.; Hoffmann, D.; Middendorf, M.; Rattei, T.; Stadler, P.F.; et al. Challenges in RNA virus bioinformatics. *Bioinformatics* **2014**, *30*, 1793–1799. [[CrossRef](#)]
- You, S.; Stump, D.D.; Branch, A.D.; Rice, C.M. A cis-acting replication element in the sequence encoding the NS5B RNA-dependent polymerase is required for hepatitis C virus RNA replication. *J. Virol.* **2004**, *78*, 1352–1356. [[CrossRef](#)]
- Tuplin, A.; Evans, D.J.; Simmonds, P. Detailed mapping of RNA secondary structures in core and NS5B-encoding region sequences of hepatitis C virus by RNase cleavage and novel bioinformatic prediction methods. *J. Gen. Virol.* **2004**, *85*, 3037–3047. [[CrossRef](#)]
- Vassilaki, N.; Friebe, P.; Meuleman, P.; Kallis, S.; Kaul, A.; Paranhos-Baccalà, G.; Leroux-Roels, G.; Mavromara, P.; Bartenschlager, R. Role of the hepatitis C virus core +1 open reading frame and core cis-acting RNA elements in viral RNA translation and replication. *J. Virol.* **2008**, *82*, 11503–11515. [[CrossRef](#)]
- Watters, K.E.; Choudhary, K.; Aviran, S.; Lucks, J.B.; Perry, K.L.; Thompson, J.R. Probing RNA structures in a positive sense RNA virus reveals selection pressures for structural elements. *Nucl. Acids Res.* **2018**, *46*, 2573–2584. [[CrossRef](#)] [[PubMed](#)]
- Ochsenreiter, R.; Hofacker, I.L.; Wolfinger, M.T. Functional RNA structures in the 3'UTR of tick-borne, insect-specific and no-known-vector Flaviviruses. *Viruses* **2019**, *11*, 298. [[CrossRef](#)] [[PubMed](#)]
- Trifonov, E.N.; Gabdank, I.; Barash, D.; Sobolevsky, Y. Primordia vita. Deconvolution from modern sequences. *Orig. Life. Evol. Biosph.* **2006**, *36*, 559–595. [[CrossRef](#)] [[PubMed](#)]
- Churkin, A.; Totzeck, F.; Zakh, R.; Parr, M.; Tuller, T.; Frishman, D.; Barash, D. A mathematical analysis of RNA structural motifs in viruses. *Mathematics* **2021**, *9*, 585. [[CrossRef](#)]
- Noureddin, M.; Gish, R. Hepatitis delta: Epidemiology, diagnosis and management 36 years after discovery. *Curr. Gastroenterol. Rep.* **2014**, *16*, 365. [[CrossRef](#)]
- Smedile, A.; Farci, P.; Verme, G.; Caredda, F.; Cargnel, A.; Caporaso, N.; Dentico, P.; Trepo, C.; Opolon, P.; Gimson, A.; et al. Influence of delta infection on severity of hepatitis B. *Lancet* **1982**, *2*, 945–947. [[CrossRef](#)]
- KarimZadeh, H.; Usman, Z.; Frishman, D.; Roggendorf, M. Genetic diversity of hepatitis D virus genotype-1 in Europe allows classification into subtypes. *J. Viral. Hepat.* **2019**, *26*, 900–910. [[CrossRef](#)] [[PubMed](#)]
- Roulot, D.; Brichler, S.; Layese, R.; BenAbdesselam, Z.; Zoulim, F.; Thibault, V.; Scholtes, C.; Roche, B.; Castelnau, C.; Poynard, T.; et al. Origin, HDV genotype and persistent viremia determine outcome and treatment response in patients with chronic hepatitis delta. *J. Hepatol.* **2020**, *73*, 1046–1062. [[CrossRef](#)]
- Gomes-Gouvê, M.S.; Soares, M.C.P.; Bensabath, G.; de Carvalho-Mello, I.M.V.G.; Brito, E.M.F.; Souza, O.S.C.; Queiroz, A.T.L.; Carrilho, F.J.; Pinho, J.R.R. Hepatitis B virus and hepatitis delta virus genotypes in outbreaks of fulminant hepatitis (Labrea black fever) in the western Brazilian Amazon region. *Gen. Virol.* **2009**, *90*, 2638–2643. [[CrossRef](#)]
- Waterman, M.S. Secondary structure of single stranded nucleic acids. *Adv. Math. Suppl. Stud.* **1978**, *1*, 167–212.
- Shapiro, B.A. An algorithm for comparing multiple RNA secondary structures. *Comput. Appl. Biosci.* **1988**, *4*, 387–393. [[CrossRef](#)]
- Fontana, W.; Konings, D.A.M.; Stadler, P.F.; Schuster, P. Statistics of RNA secondary structures. *Biopolymers* **1993**, *33*, 1389–1404. [[CrossRef](#)]

18. Hofacker, I.L.; Fontana, W.; Stadler, P.F.; Bonhoeffer, S.; Tacker, M.; Schuster, P. Fast folding and comparison of RNA secondary structures. *Monatsh. Chem.* **1994**, *124*, 167–188. [[CrossRef](#)]
19. Hofacker, I.L. Vienna RNA secondary structure server. *Nucl. Acids Res.* **2003**, *31*, 3429–3431. [[CrossRef](#)]
20. Lorentz, R.; Bernhart, S.H.; Höner zu Siederdisen, C.; Tafer, H.; Flamm, C.; Stadler, P.F.; Hofacker, I.L. ViennaRNA Package 2.0. *Algorithms Mol. Biol.* **2011**, *6*, 26. [[CrossRef](#)]
21. Zuker, M. Computer prediction of RNA secondary structure. *Meth. Enzym.* **1989**, *180*, 262–288.
22. Zuker, M. Mfold webserver for nucleic acid folding and hybridization prediction. *Nucl. Acids Res.* **2003**, *31*, 3406–3415. [[CrossRef](#)]
23. Markham, N.R.; Zuker, M. UNAFold: Software for nucleic acid folding and hybridization. *Methods Mol. Biol.* **2008**, *453*, 3–31.
24. Le, S.Y.; Nussinov, R.; Maizel, J.V. Tree graphs of RNA secondary structures and their comparison. *Comput. Appl. Biosci.* **1989**, *22*, 461–473. [[CrossRef](#)]
25. Benedetti, G.; Morosetti, S. A graph-topological approach to recognition of pattern and similarity in RNA secondary structures. *Biophys. Chem.* **1996**, *59*, 179–184. [[CrossRef](#)]
26. Barash, D. Deleterious mutation prediction in the secondary structure of RNAs. *Nucl. Acids Res.* **2003**, *31*, 6578–6584. [[CrossRef](#)] [[PubMed](#)]
27. Barash, D. Second eigenvalue of the Laplacian matrix for predicting RNA conformational switch by mutation. *Bioinformatics* **2004**, *20*, 1861–1869. [[CrossRef](#)] [[PubMed](#)]
28. Churkin, A.; Barash, D. RNAmute: RNA secondary structure mutation analysis tool. *BMC Bioinf.* **2006**, *7*, 221. [[CrossRef](#)] [[PubMed](#)]
29. Giegerich, R.; Voss, B.; Rehmsmeier, M. Abstract shapes of RNA. *Nucl. Acids Res.* **2004**, *32*, 4843–4851. [[CrossRef](#)]
30. Churkin, A.; Gabdank, I.; Barash, D. The RNAmute webserver for the mutational analysis of RNA secondary structures. *Nucl. Acids Res.* **2011**, *39*, W92–W99. [[CrossRef](#)] [[PubMed](#)]
31. Barash, D.; Churkin, A. Mutational analysis in RNAs: Comparing programs for RNA deleterious mutation prediction. *Brief. Bioinf.* **2011**, *12*, 104–114. [[CrossRef](#)] [[PubMed](#)]
32. Flamm, C.; Hofacker, I.L.; Maurer-Stroh, S.; Stadler, P.F.; Zehl, M. Design of multistable RNA molecules. *RNA* **2001**, *7*, 254–265. [[CrossRef](#)]
33. Linnstaedt, S.D.; Kaspar, W.K.; Shapiro, B.A.; Casey, J.L. The role of a metastable RNA secondary structure in hepatitis delta virus genotype III RNA editing. *RNA* **2006**, *12*, 1521–1533. [[CrossRef](#)] [[PubMed](#)]
34. Shetty, S.; Stefanovic, S.; Mihailescu, M.R. Hepatitis C virus RNA: Molecular switches mediated by long-range RNA-RNA interactions? *Nucl. Acids Res.* **2013**, *41*, 2526–2540. [[CrossRef](#)] [[PubMed](#)]
35. Abbink, T.E.M.; Ooms, M.; Joost Haasnoot, P.C.; Berkhout, B. The HIV-1 leader RNA conformational switch regulates RNA dimerization but does not regulate mRNA translation. *Biochemistry* **2005**, *44*, 9058–9066. [[CrossRef](#)] [[PubMed](#)]
36. Fiedler, M. Algebraic connectivity of graphs. *Czech. Math. J.* **1973**, *23*, 298–305. [[CrossRef](#)]
37. Grone, R.; Merris, R. Algebraic connectivity of trees. *Czech. Math. J.* **1987**, *37*, 660–670. [[CrossRef](#)]
38. Merris, R. Ordering trees by algebraic connectivity. *Graphs. Combin.* **1990**, *6*, 229–237.
39. Merris, R. Characteristic vertices of trees. *Lin. Multilin. Alg.* **1987**, *22*, 115–131. [[CrossRef](#)]
40. Grone, R.; Merris, R.; Sunder, V.S. The Laplacian spectrum of a graph. *SIAM J. Matrix. Anal. Appl.* **1990**, *11*, 218–238. [[CrossRef](#)]
41. Kiening, M.; Ochsenreiter, R.; Hellinger, H.J.; Rattei, T.; Hofacker, I.L.; Frishman, D. Conserved secondary structures in viral mRNAs. *Viruses* **2019**, *11*, 401. [[CrossRef](#)]
42. Usman, Z.; Velkov, S.; Protzer, U.; Roggendorf, M.; Frishman, D.; Karimzadeh, H. HDVdb: A comprehensive hepatitis D virus database. *Viruses* **2020**, *12*, 538. [[CrossRef](#)] [[PubMed](#)]
43. Casey, J.L. RNA editing in hepatitis delta virus genotype III requires a branched double-hairpin RNA structure. *J. Virol.* **2002**, *76*, 7385–7397. [[CrossRef](#)] [[PubMed](#)]
44. Elliott, J.F. The Characteristic Roots of Certain Real Symmetric Matrices. Master’s Thesis, University of Tennessee, Knoxville, TN, USA, 1953.
45. Gover, M.J.C. The eigenproblem of a tridiagonal 2-Toeplitz matrix. *Lin. Alg. Its Appl.* **1994**, *197*, 63–78. [[CrossRef](#)]
46. Abramowitz, M.; Stegun, I.A. *Handbook of Mathematical Functions*; Dover: New York, NY, USA, 1965.
47. Ivry, T.; Michal, S.; Avihoo, A.; Sapiro, G.; Barash, D. An image processing approach to computing distances between RNA secondary structures dot plots. *Alg. Mol. Biol.* **2009**, *4*, 4. [[CrossRef](#)]
48. Mathews, D.H.; Disney, M.D.; Childs, J.L.; Schroeder, S.J.; Zuker, M.; Turner, D.H. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 7287–7292. [[CrossRef](#)]
49. Zuker, M. On finding all suboptimal solutions. *Science* **1989**, *244*, 48–52. [[CrossRef](#)]
50. Churkin, A.; Barash, D. An efficient method for the prediction of deleterious multiple-point mutations in the secondary structure of RNAs using suboptimal folding solutions. *BMC Bioinf.* **2008**, *9*, 222. [[CrossRef](#)]
51. Wuchty, S.; Fontana, W.; Hofacker, I.L.; Schuster, P. Complete suboptimal folding of RNA and the stability of secondary structures. *Biopolymers* **1999**, *49*, 145–165. [[CrossRef](#)]
52. Linnstaedt, S.D. RNA Editing in Hepatitis Delta Virus. Ph.D. Thesis, Georgetown University, Washington, DC, USA, 2008.
53. Casey, J.L.; Brown, T.L.; Colan, E.J.; Wignall, S.; Gerin, J.L. A genotype of hepatitis D virus that occurs in northern South America. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 9016–9020. [[CrossRef](#)]

54. Churkin, A.; Cohen, M.; Shemer-Avni, Y.; Barash, D. Bioinformatic analysis of the neutrality of RNA secondary structure elements across genotypes reveals evidence for direct evolution of genetic robustness in HCV. *J. Bioinform. Comput. Biol.* **2010**, *8*, 1013–1026. [[CrossRef](#)] [[PubMed](#)]
55. Goz, E.; Tuller, T. Evidence of a direct evolution selection for strong folding and mutational robustness within HIV coding regions. *J. Comput. Biol.* **2016**, *23*, 641–650. [[CrossRef](#)] [[PubMed](#)]
56. Ariza-Mateos, A.; Briones, C.; Perales, C.; Domingo, E.; Gómez, J. The archaeology of coding RNA. *Ann. N. Y. Acad. Sci.* **2019**, *1447*, 119–134. [[CrossRef](#)]
57. Shekhtman, L.; Cottler, S.J.; Hershkovich, L.; Uprichard, S.L.; Bazinet, M.; Pantea, V.; Cebotarescu, V.; Cohuhari, L.; Jimbei, P.; Krawcztk, A.; et al. Modelling hepatitis D virus RNA and HBsAg dynamics during nucleic acid polymer monotherapy suggest rapid turnover of HBsAg. *Sci. Rep.* **2020**, *10*, 7837. [[CrossRef](#)] [[PubMed](#)]
58. Guedj, J.; Rotman, Y.; Cotler, S.J.; Koh, C.; Schmid, P.; Albrecht, J.; Haynes-Williams, V.; Liang, J.; Hoofnagle, J.H.; Heller, T.; et al. Understanding early serum hepatitis D virus and hepatitis B surface antigen kinetics during pegylated interferon-alpha therapy via mathematical modeling. *Hepatology* **2014**, *60*, 1902–1910. [[CrossRef](#)]
59. Barash, D. Nonlinear diffusion on an extended neighborhood. *Appl. Numer. Math.* **2005**, *52*, 1–11. [[CrossRef](#)]
60. Reinharz, V.; Dahari, L.; Barash, D. Numerical schemes for solving and optimizing multiscale models with age of hepatitis C virus dynamics. *Math. Biosci.* **2018**, *300*, 1–13. [[CrossRef](#)]
61. Koh, C.; Canini, L.; Dahari, H.; Zhao, X.; Uprichard, S.J.; Haynes-Williams, V.; Winters, M.A.; Subramanya, G.; Cooper, S.L.; Pinto, P.; et al. Oral prenylation inhibition with lonafarnib in chronic hepatitis D infection: A proof-of-concept randomized, double-blind, placebo-controlled phase 2A trial. *Lancet Infect. Dis.* **2015**, *15*, 1167–1174. [[CrossRef](#)]
62. Neumann, A.U.; Lam, N.P.; Dahari, H.; Davidian, M.; Wiley, T.E.; Mika, B.P.; Perelson, A.S.; Layden, T.J. Differences in viral dynamics between genotypes 1 and 2 of hepatitis C virus. *J. Infect. Dis.* **2000**, *182*, 28–35. [[CrossRef](#)]
63. Gilman, C.; Heller, T.; Koh, C. Chronic hepatitis delta: A state-of-the art review and new therapies. *World J. Gastroenterol.* **2019**, *25*, 4580–4597. [[CrossRef](#)]
64. Olivero, A.; Smedile, A. Hepatitis delta virus diagnosis. *Semin. Liver. Dis.* **2012**, *32*, 220–227. [[CrossRef](#)] [[PubMed](#)]
65. Forties, R.A.; Bundschuh, R. Modeling the interplay of single-stranded binding proteins and nucleic acid secondary structure. *Bioinformatics* **2010**, *26*, 61–67. [[CrossRef](#)] [[PubMed](#)]
66. Strack, R. Predicting RNA–protein binding affinity. *Nat. Methods* **2019**, *16*, 460.
67. Kappel, K.; Jarmoskaite, I.; Vaidyanathan, P.P.; Greenleaf, W.J.; Herschlag, D.; Das, R. Blind test of RNA-protein binding affinity prediction. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 8336–8341. [[CrossRef](#)]
68. Kay, A.; Melo da Silva, E.; Pedreira, H.; Negreiros, S.; Lobato, C.; Braga, W.; Muwonge, R.; Dény, P.; Reis, M.; Zoulim, F.; et al. HBV/HDV co-infection in the Western Brazilian Amazonia: An intriguing mutation among HDV genotype 3 carriers. *J. Viral. Hepat.* **2014**, *21*, 921–924. [[CrossRef](#)] [[PubMed](#)]
69. Merris, R. Laplacian matrices of graphs: A survey. *Lin. Alg. Its Appl.* **1994**, *197*, 143–176. [[CrossRef](#)]