# Analysis of key genes in *Mycobacterium ulcerans* reveals conserved RNA structural motifs and regions with apparent pressure to remain unstructured

Warren B. Rouse[1], Jessica Gart[2], Lauren Peysakhova[2]
and Walter N. Moss[1]*

[1]Roy J. Carver Department of Biochemistry, Biophysics, and Molecular Biology, Iowa State
University, Ames, IA, United States, [2]Science and Engineering Research Program (SERP), Staten
Island Technical High School, Staten Island, NY, United States

Buruli Ulcer is a neglected tropical disease that results in disfiguring and
dangerous lesions in affected persons across a wide geographic area,
including much of West Africa. The causative agent of Buruli Ulcer is
*Mycobacterium ulcerans*, a relative of the bacterium that causes tuberculosis
and leprosy. Few therapeutic options exist for the treatment of this disease
beyond antibiotics in the early stages, which are frequently ineffective, and
surgical removal in the later stage. In this study we analyze six genes in
*Mycobacterium ulcerans* that have high potential of therapeutic targeting.
We focus our analysis on a combined in silico and comparative sequence
study of potential RNA secondary structure across these genes. The result of
this work was the comprehensive local RNA structural landscape across each
of these significant genes. This revealed multiple sites of ordered and evolved
RNA structure interspersed between sequences that either have no bias for
structure or, indeed, appear to be ordered to be unstructured and (potentially)
accessible. In addition to providing data that could be of interest to basic
biology, our results provide guides for efforts aimed at targeting this pathogen
at the RNA level. We explore this latter possibility through the *in silico* analysis of
antisense oligonucleotides that could potentially be used to target
pathogen RNA.

KEYWORDS

Buruli ulcer, neglected tropical disease, mycobacterium, RNA, ASO, secondary
structure, therapeutics

## Introduction

Buruli Ulcer (BU) is a skin-related neglected tropical disease (NTD) caused by the family of bacteria responsible for tuberculosis and leprosy (1). The bacterial pathogen, *Mycobacterium ulcerans* (*M. ulcerans*), is the causative agent of this disease, which mainly affects the skin and, on rare occasions, the bone. BU has been reported in 33 countries, 14 of which regularly report to the World Health Organization (2). *M. ulcerans* thrives in temperatures between 29–33°C and requires a low oxygen concentration of 2.5% (3, 4), thus, the bacteria is able to easily spread in areas near the equator such as West Africa and other tropical areas in the Western Pacific (5, 6).

Although this pathogen spreads quite readily, there is limited knowledge on its mode of transmission (5). However, it is known that *M. ulcerans* relies on the virulence factor mycolactone for pathogenesis (7, 8). Mycolactone, an exotoxin secreted by *M. ulcerans*, has three major biological functions: cytotoxicity, resulting in large-scale skin ulceration; immunosuppression, resulting in limited local inflammatory response, and analgesic effects, resulting in limited or no pain associated with damage (9, 10). *M. ulcerans'* ability to block and degrade the nerves of infected tissue, and therefore inhibit the feeling of pain, makes it particularly dangerous (11). Due to the long incubation period and lack of symptoms, the infected often unknowingly carry the disease, prohibiting early diagnosis that is vital to effective treatment (7).

Current treatment options include surgical removal of the skin, oral antibiotics, and dressing of the skin (12). Surgery and skin grafting are only effective for early cases of BU where small areas of the skin are affected and can be safely removed (13). However, at later stages of the disease, large areas of the skin become damaged and surgery is not a safe solution (13). Three antibiotic options have been utilized to treat BU including rifampin, clarithromycin, and streptomycin (13), however, oral antibiotics have not proven to be ideal treatments as patients who take these medications are frequently hospitalized due to adverse effects (7).

Although BU is devastating to people in many developing countries, very few studies have been conducted to find novel treatments. Since it only affects a small portion of the world (6, 13), therapeutics would have a relatively small return on investment—thus leaving BU a neglected disease. With this in mind, RNA-targeting therapeutics are emerging as exciting new treatments, which can take the form of small-molecules (14–16) or antisense oligonucleotides [ASOs (17, 18)]. The ASO approach could be particularly well suited as a strategy for treating BU and other neglected diseases, due to their rapid and relatively low cost to develop (19, 20). Additionally, as RNA-based therapies grow in popularity, the projected future cost associated with their deployment is expected to decline (21).

ASOs are relatively simple yet promising therapeutics that can target RNA *via* sequence complementarity. Once bound to the target sequence, they can prevent protein production by blocking the ribosome, masking important functional elements, or causing degradation of the bound transcript (22). Significant to this current study, ASOs have previously been used as antimicrobial agents (18, 22), which offers hope that this method could be applied to BU. Targeting of RNA, using both small-molecules and ASOs, is greatly facilitated by analysis of target secondary and tertiary structure (16). Analysis of RNA structure has proven that it is an important component of gene regulation in human and pathogen genomes (23–26); however, to date, no analyses of RNA structure have been performed in *M. ulcerans*. A better understanding of RNA structure in this pathogen would thus provide information that could prove useful in treating BU, as RNA structures have been shown to be important to proper regulation of translation in other bacterial strains (27).

In this study, we apply an in silico sequence analysis pipeline for prediction of RNA secondary structure, ScanFold, that is based on the identification of small local motifs with unusually high propensity for ordered thermodynamic stability (28). This is accomplished by measuring the stability of evolved/ordered sequences vs in silico randomized sequences with the same nucleotide composition. This approach was optimized on viral human pathogens: HIV, Zika, human Herpesviridae and, most recently, SARS-CoV-2 (23). Significantly, this latter analysis of SARS-CoV-2 led to the discovery (or highlighting) of structural motifs that are being targeted with both small molecules (29–31) and ASOs (32, 33). In this current analysis of *M. ulcerans*, we enhance our approach to make predictions that are of particular relevance to identification of efficient ASOs.

## Methods

The *M. ulcerans* agy99 genome (NC_008611.1) and associated gff3 data as well as the pMUM001 virulence plasmid (NC_005916.1) and associated gff3 data were downloaded from NCBI (Supplementary File S1). Based on previous work by Butt et al. (34), six essential genes with no host (human) homology were chosen for analyses. The chosen target genes are *Mul_RS01615*, *Mul_RS01365*, *Mul_RS04730*, *Mul_RS09540*, *Mul_RS04200* and *Mul_RS00210*, or Targets 1-6 respectively. All genes except *Mul_RS00210* were found in the NC_008611.1 genome and *Mul_RS00210* was found in the NC_005916.1 virulence plasmid.

These genomes and associated gff3 files were loaded into IGV-ScanFold (version 0.2.0) (https://github.com/ResearchIT/ IGV-ScanFold/releases/) to view the sequence of *M. ulcerans* as a whole. After searching for the genes, they were initially scanned in their entirety by highlighting the gene and using the

"ScanFold>Run ScanFold on selected region" option. For this analysis the default parameters were changed to a 120 nt window size, a 1 nt step size, 100 randomizations per window, mononucleotide shuffling, 37°C temperature, competition of 1 (to demand that only one unique base pair per nucleotide is possible), either forward or reverse strand depending on the gene polarity, and the RNAfold folding algorithm was selected. Metrics obtained from ScanFold include the MFE (a measure of thermodynamic stability), ΔG z-score (a measure of ordered stability that can indicate potential function), and ensemble diversity or ED (a measure of predicted structure's conformational volatility). The MFE is estimated by the change in predicted Gibb's folding free energy (the ΔG°) going from a fully denatured (random coil) RNA to an ordered 2D structure, where more negative values indicate increasing favorability and greater stability. The ΔG z-scores identify structures that have propensity for ordered stability, where negative values indicate the number of standard deviations more stable the native sequence is compared to any randomized version of it (35, 36). The ED uses the RNA partition function to compare the distance between the optimal thermodynamic structure and suboptimal conformations (37). This distance is measured as the number of base pairs different between Boltzmann weighted conformations and is averaged across the ensemble. Lower ED values indicate a single dominant conformation, while higher EDs suggest multiple conformations or a lack of defined structure (38). Additionally, arc diagrams depicting weighted z-score structures are produced where blue, green, and yellow arcs indicate z-scores ≤ -2, ≤ -1, and < 0, respectively. For more information on these metrics, see the ScanFold methods paper (39). After initial analysis, several genes were found to have regions of low z-score near the 3′ and 5′ ends. All genes were re-scanned using the same parameters with an additional 50-300 nucleotides upstream and downstream of the start and stop codon, respectively. This allowed us to note any patterns in stability immediately outside the targets. As an additional layer of structure prediction validation, these genes, including the upstream and downstream sequences, were also scanned using dinucleotide shuffling. The results of mono- and dinucleotide shuffling were compared using the "ct_compare.py" script (https://github.com/moss-lab/SARS-CoV-2) and were found to have only minor differences. The average difference in percent of paired nucleotides was 2.29% and 0.66% for -1 and -2 z-scores pairs respectively, and the average percent similarity or consistency in pairing partners was 86.63% and 97.93% for -1 and -2 z-scores pairs respectively. All structures analyzed in this work were predicted using both shuffling methods, further validating that they may be potentially functional.

To further assess the potential functionality of structures found in these genes, covariation analysis was completed using the cm-builder analysis pipeline (40). All 21 structures with a z-score ≤ -2 were analyzed for covariation using the cm-builder

perl script (40). This script builds off the RNAFramework toolkit (41, 42) and utilizes Infernal (here using release 1.1.2 (43);) to build and search for covariation models from each predicted ScanFold secondary structure. Briefly, cm-builder uses Infernal to generate a covariance model (CM) using the sequence and structure of the selected motif. A database of sequences is built for Infernal using a BLAST search (44). For *M. ulcerans*, the nt collection database was searched for each of the selected motifs found in the target genes using the following parameters: BLASTn (Somewhat similar), Organism (bacteria, taxid:2), and a max sequence number of 5000. Infernal then takes the BLAST data and aligns the sequences removing any redundant and severely truncated alignments. The resulting set of homologs from the alignment of BLAST hits is then aligned to the original CM, which is used to build a new CM. The whole process is repeated three times. Finally, the resulting alignment is refactored to remove gap-only positions and include only bases spanning the first to the last base-paired residue. The final alignment file is analyzed using R-scape v2 (41) where APC corrected *G*-test statistics (45) are used to identify motifs showing significantly covarying base-pairs using the default E value of 0.05. Statistically significant covariation is indicative of evidence for a conserved RNA base pair, but the lack of covariance does not indicate that the structure has no potential for function. All covariation results were analyzed to evaluate their significance or power, and all statistically significant base pairs were annotated on the 2D models made using VARNA (46) (Supplementary File S2).

After running ScanFold to define regions of unusual thermodynamic stability, and cm-builder to identify statistically significant covarying base pairs, the sequences and predicted secondary structures of each gene were loaded into OligoWalk (47) and ran using default parameters (Mode: Break Local Structure, Oligo Chemistry: DNA, Oligo Length: 18nt, and Oligo Concentration: 1uM). The overall ΔG, duplex ΔG, intra-oligo ΔG, and inter-oligo ΔG values for 18 nucleotide complementary oligonucleotides tiled across each gene sequence were obtained for further analysis. Additionally, using the sequence fasta file, final partners file, and log files from IGV-ScanFold an in-house python script, "ScanFold_Oligo_Metrics.py", was used to partition the data into 18 nt fragments and determine the average ED, MFE, and z-score for each. (https://github.com/moss-lab/ScanFold_Oligo_Metrics/blob/main/ScanFold_Oligo_Metrics.py). This was accomplished using a tiled window approach, which allowed for the complementary analysis of ScanFold and OligoWalk metrics for the same fragments. For more information on the data generated by this script, see the ReadMe file on GitHub. Using Oligowalk and ScanFold data, bar graphs were made for every gene (Supplementary File S3). All data used to create these graphs and analyze the results, can be found in (Supplementary File S4). Following the work of Matveeva et. al., oligonucleotides that had intra-oligo values greater than -1.1 kcal/mol, inter-oligo

values greater than -8 kcal/mol, duplex ΔG values of less than -15 kcal/mol, and negative overall ΔG values were considered as optimal ASO targets to begin further analysis. This data was then compared to the ScanFold fragment data and trends were noted.

# Results

We focused on six genes found to be essential to *M. ulcerans* (34, 48) with no known homology to human hosts: *Mul_RS01615*, *Mul_RS01365*, *Mul_RS04730*, *Mul_RS09540*, *Mul_RS04200* and *Mul_RS00210*. These were selected for their essentiality to *M. ulcerans*, their homology to genes targeted in other pathogens, and their lack of host homology (34, 48). These were subjected to the full ScanFold pipeline as integrated in IGV-ScanFold. In the first step, ScanFold-Scan was used to define the local RNA folding landscape of each gene (Supplementary File S5). Here, a scanning analysis window of 120 nucleotides was moved one base at a time across each gene sequence while predicting several folding metrics for each window: the minimum free energy of folding (MFE, ΔG°), its associated base pairing model (secondary structure), ensemble diversity (ED), and a thermodynamic z-score that compares the MFE of the natively ordered RNA to the MFE of randomized versions of the same sequence to identify propensity for ordered stability as indicated by negative values. Details on all metrics can be found in the Methods Section. In the second stage, ScanFold-Fold, folding metrics are partitioned to each nucleotide and base pair, and consensus base pairs across all windows—weighted by their propensity for ordered stability (thermodynamic z-score) —are identified and output as distinct motifs (Supplementary File S6).

A summary of results across each analyzed gene is seen in Table 1. Here, the average ΔG values across all analyzed genes ranged from -48.40 kcal/mol to -33.23 kcal/mol. This difference in predicted RNA folding stability is directly correlated with GC %, as expected, which ranged from 68.14% to 61.32%. The average z-score and ensemble diversity (ED), however, do not follow trends for ΔG or GC%. The z-score quantifies how much greater-than-random the folding stability of an RNA sequence is, which is primarily dependent on the sequence order and not

its nucleotide composition. Likewise, the ED value indicates the diversity of potential structural conformations in an RNA's folding ensemble, which also appears to be an evolved property of ordered/functional RNA sequences (38). The average z-score ranged from -0.82 to −0.03, while the average ED ranged from 20.64 to 26.85. Here we see that the gene with the lowest (most favorable) ΔG and highest GC% (*Mul_RS01615*) has the highest (least favorable) average z-score and ED. The gene with the lowest average z-score (*Mul_RS04730*) also showed the highest percent of windows with z-score < -1 (39.55%). This did not hold true for z-scores < -2 (two standard deviations more stable than random), but the gene with the second lowest average z-score (*Mul_RS01365*) showed the highest percent of windows with z-scores < -2 (14.76%). The gene with the smallest fraction of its nucleotides spanned by low z-score windows was *Mul_RS01615* (19.0% and 4.4% for the -1 and -2 z-score cutoffs, respectively).

In the prediction of the ΔG for each analysis window, a model secondary structure is also generated. Across each gene, this resulted in many predicted base pairs where a specific nucleotide can be paired differently across multiple overlapping windows. As a result, many potential base pairing partners may be predicted per nucleotide. ScanFold-Fold predicts a single structural context (paired or unpaired) for each nucleotide based on its contributions to low z-score windows (indicating ordered stability and potential functionality). The genes that had the greatest percentages of low z-score base pairs were *Mul_RS04730* and *Mul_RS01365*, where *Mul_RS04730* had 25.90% of its base pairs predicted with average z-score < -1 and *Mul_RS01365* had 10.12% of its base pairs predicted with average z-score < -2 (Table 1). Additionally, ScanFold-Fold extracts discrete structural motifs (i.e., single hairpins or multi-branched stem loops) containing at least one base pair with an average z-score < -2. This results in a list of motifs for each gene, where the longest gene *Mul_RS00210* had the greatest number of motifs (8 motifs) and the much shorter gene *Mul_RS01365* had the second greatest number of motifs (7 motifs).

In summary, all predictions (mono- and dinucleotide shuffling) indicate a particular importance for functional RNA

**TABLE 1** Summary of ScanFold data across each target gene.

| Gene | GC% | ΔG | ED | Average ZS | Windows | ZS <-1 | ZS <-2 | Total BP | BP ZS <-1 | BP ZS <-2 | Motifs |
|------|-----|-----|-----|-----------|---------|--------|--------|----------|-----------|-----------|--------|
| *Mul_RS01365* | 63.92% | -44.69 | 25.86 | -0.45 | 1064 | 24.44% | **14.76%** | 7763 | 17.92% | **10.12%** | 7 |
| *Mul_RS04730* | 63.90% | -44.1 | 23.39 | **-0.82** | 1062 | **39.55%** | 10.36% | 7277 | **25.90%** | 7.09% | 2 |
| *Mul_RS04200* | 64.50% | -45.14 | 23.87 | -0.32 | 3590 | 22.17% | 3.34% | 25523 | 11.87% | 0.37% | 1 |
| *Mul_RS09540* | 61.20% | -40.61 | **20.64** | -0.3 | 248 | 32.26% | 8.06% | 2043 | 11.75% | 0.64% | 1 |
| *Mul_RS00210* | 61.32% | -33.23 | 23.32 | -0.44 | **7757** | 27.12% | 7.48% | **45399** | 17.76% | 2.49% | **8** |
| *Mul_RS01615* | **68.14%** | **-48.4** | 26.85 | -0.03 | 1843 | 19.00% | 4.40% | 9686 | 13.13% | 1.71% | 2 |

The ScanFold data as found across all windows spanning each gene. From left to right the data corresponds to percent GC content, average MFE, average ED, average z-score, number of windows generated, percent of windows with z-score < -1, percent of windows with z-score < -2, total number of possible base pairs, percent of base pairs with z-score < -1, percent of base pairs with z-score < -2, and number of motifs with a z-score < -2.

secondary structures encoded within the genes for virulence factor production and cell wall biosynthesis (Supplementary Files S5). The ScanFold-Fold results also give us a means of generating motifs of interest for further analysis (Supplementary File S6).

## Analysis of RNA folding in *Mul_RS01365*

*Mul_RS01365* encodes a protein that is homologous to the *desA2* (stearoyl-ACP desaturase) protein of *M. tuberculosis*. This protein is thought to function by catalyzing the initial conversion of saturated fatty acids to unsaturated fatty acids in lipid metabolism.

A summary of ScanFold data is shown on Figure 1, where blue, green, and yellow arcs indicate base pairing with average z-scores ≤ -2, ≤ -1, and < 0, respectively. The overall thermodynamic stability remains flat across *Mul_RS01365* until scans reach the 3′ end of the gene, where stability increases (more negative ΔG in windows overlapping these nucleotides). Despite fairly monotonous trends in ΔG, evidence for regions of ordered RNA stability (negative z-scores) are clustered into two distinct regions at the 5′ and 3′ ends of the coding region, respectively. Interestingly, the core coding region is spanned by an area with positive z-scores, which indicates that the ordered/evolved sequence is less stable than predicted based on nucleotide content: i.e., the evolved sequence may be ordered to be unstructured or accessible. This region of unusual instability correlates with spikes in the ensemble diversity, which indicates a volatile ensemble of potential RNA secondary structures or a lack of stable structure. Conversely, the previously mentioned low z-score clusters overlap regions of low ensemble diversity, indicating one (or a few similar) conformation(s).
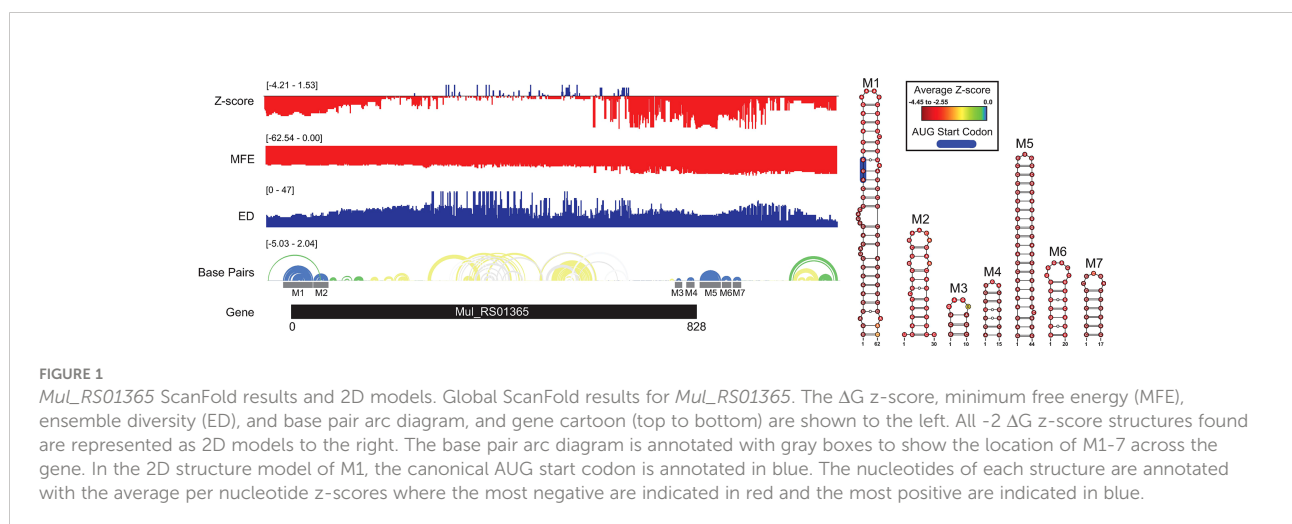
When the low z-score windows were evaluated by ScanFold-Fold, seven distinct motifs (M1-7) were identified with exceptionally low (< -2) z-score-weighted base pairs (Figure 1). These motifs comprise two upstream hairpins (M1 and M2) and five downstream hairpins that span the translational start and stop codons, respectively. M1 is notable for containing the start codon embedded within a stable helix and for being the longest thermodynamically stable hairpin predicted for this gene. When evaluated for their conservation across mycobacterial species, none of the proposed base pairs were found to have statistically significant covariation (Supplementary File S2). The proposed structures were, however, found to be present/conserved across a wide array of species.

In M1, for example, the hairpin structure is preserved across pathogenic species of mycobacteria (Figure 2). Conservation is highest in the terminal hairpin loop region that contains the start codon, and trails off toward the basal stem, where deletions and inconsistent mutations (that ablate canonical base pairing potential) would be predicted to weaken the basal stem. The core hairpin is best preserved (100% preservation of base pairing) in the medically significant strains, *M. gordonae*, *kansasii*, *tuberculosis*, and *bovis* while inconsistent mutations begin accumulating in *M. fortuitum*, where a U>C mutation disrupts a single AU pair (Figure 2). M. avium had the most inconsistent mutations in the core hairpin, that would be predicted to weaken or break four base pairs (out of 15). Conversely, while no compensatory mutations are observed in the core hairpin, six species had at least one consistent change, where single point mutations preserve base pairing; *M. Leprae* had two consistent mutations, however, these were offset by several inconsistent changes.

## Analysis of RNA folding in *Mul_RS04730*

*Mul_RS04730* encodes a protein homologous to the *rpoA* (alpha chain of RNA polymerase) of *M. tuberculosis*. This



**FIGURE 1**
*Mul_RS01365* ScanFold results and 2D models. Global ScanFold results for *Mul_RS01365*. The ΔG z-score, minimum free energy (MFE), ensemble diversity (ED), and base pair arc diagram, and gene cartoon (top to bottom) are shown to the left. All -2 ΔG z-score structures found are represented as 2D models to the right. The base pair arc diagram is annotated with gray boxes to show the location of M1-7 across the gene. In the 2D structure model of M1, the canonical AUG start codon is annotated in blue. The nucleotides of each structure are annotated with the average per nucleotide z-scores where the most negative are indicated in red and the most positive are indicated in blue.
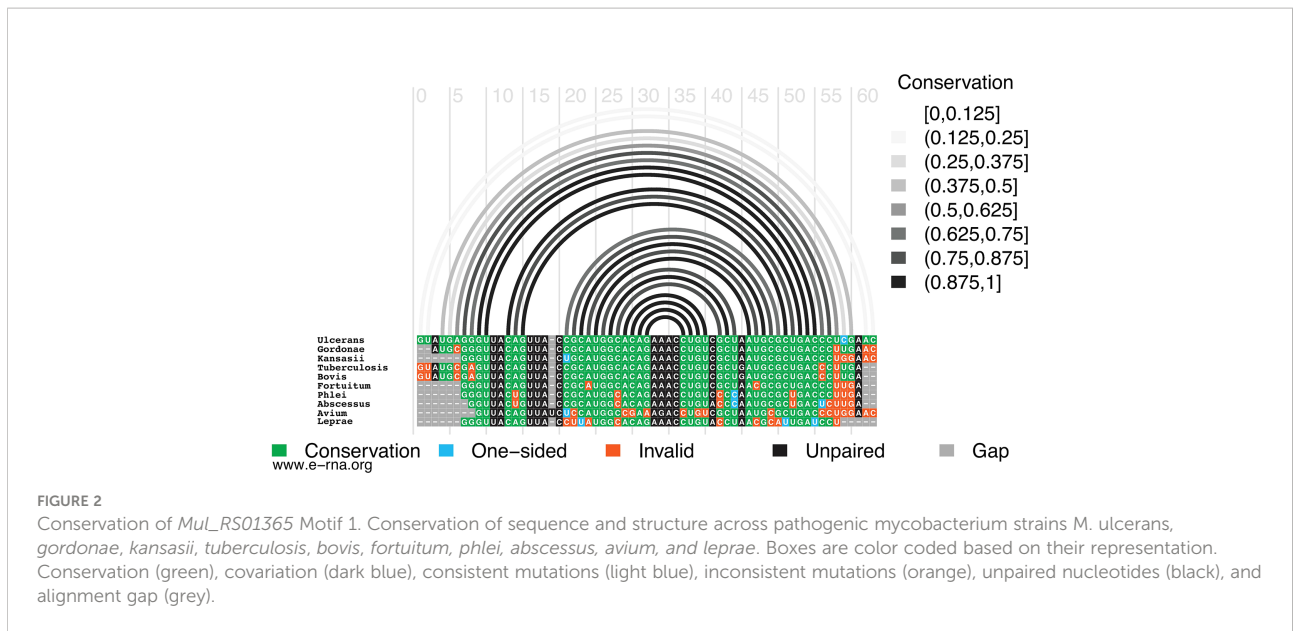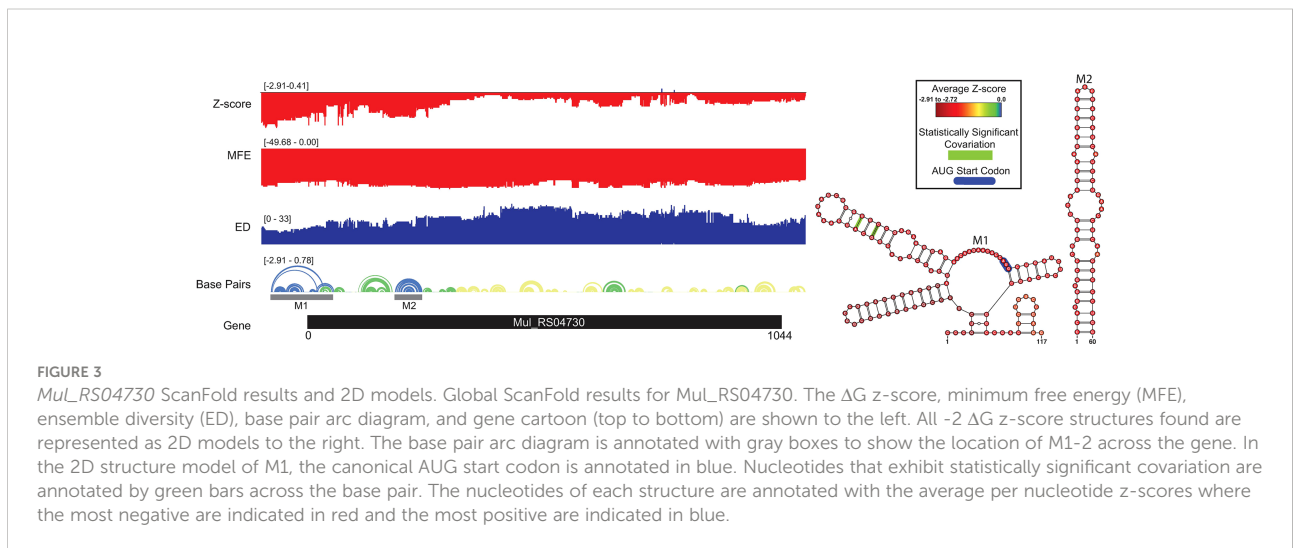
**FIGURE 2**
Conservation of *Mul_RS01365* Motif 1. Conservation of sequence and structure across pathogenic mycobacterium strains M. ulcerans, *gordonae, kansasii, tuberculosis, bovis, fortuitum, phlei, abscessus, avium, and leprae*. Boxes are color coded based on their representation. Conservation (green), covariation (dark blue), consistent mutations (light blue), inconsistent mutations (orange), unpaired nucleotides (black), and alignment gap (grey).

protein functions as a component of the DNA-dependent RNA polymerase responsible for bacterial genome replication.

A summary of ScanFold data is shown in Figure 3 following the same color scheme mentioned previously. Here, the overall thermodynamic stability (ΔG) remains flat across the entirety of the transcript. Despite the flatline trend in ΔG, evidence of ordered RNA stability (negative z-scores) is clustered into one distinct region around the 5′ end of the coding region. The coding region is predominantly spanned by structures with z-scores of 0 to -1 but a few structures with z-scores ≤ -2 and ≤ -1 are interspersed. This indicates that the ordered/evolved sequence is overall only slightly more stable than predicted based on nucleotide content: i.e., the evolved sequence may be ordered to be structured. The regions of only slightly increased stability correlate with spikes in the ensemble diversity when

compared to those of the much more stable region near the 5′ end of the coding region. This is indicative of a volatile ensemble of potential RNA secondary structures or a lack of stable structure across the core coding sequence; whereas the more 5′ region shows a less volatile ensemble or more stable structures that may be potentially functional.

When the low z-score windows were evaluated by ScanFold-Fold, two distinct motifs (M1-2) were identified with exceptionally low (< -2) z-score-weighted base pairs (Figure 3). These motifs comprise one upstream multi-branch helix (M1) and one downstream hairpin (M2) that span the start site of translation and part of the coding sequence, respectively. M1 is notable for containing the start codon in a single stranded region between two hairpins of the multi-branch helix. When evaluated for their conservation across mycobacterial species, two of the proposed



**FIGURE 3**
*Mul_RS04730* ScanFold results and 2D models. Global ScanFold results for Mul_RS04730. The ΔG z-score, minimum free energy (MFE), ensemble diversity (ED), base pair arc diagram, and gene cartoon (top to bottom) are shown to the left. All -2 ΔG z-score structures found are represented as 2D models to the right. The base pair arc diagram is annotated with gray boxes to show the location of M1-2 across the gene. In the 2D structure model of M1, the canonical AUG start codon is annotated in blue. Nucleotides that exhibit statistically significant covariation are annotated by green bars across the base pair. The nucleotides of each structure are annotated with the average per nucleotide z-scores where the most negative are indicated in red and the most positive are indicated in blue.

base pairs were found to have statistically significant covariation with power ≥ 0.25 (Supplementary File S2 and Figure 3). In addition, part of the M1 structure and all of the M2 structure were found to be present/conserved across a wide array of species.

In M1, for example, the hairpin structure is preserved across pathogenic species of mycobacteria (Figure 4). Conservation is highest in the central hairpin that shows evidence of significant covariation and in the downstream hairpin containing the start codon. However, toward the 5′ end of the structure, conservation drops off, as homologous sequences were not found here. The small hairpin containing the last nucleotide of the start codon is best conserved (100% preservation of base pairing) in the medically significant strain, *M. gordonae*, *kansasii*, *fortuitum*, *phlei*, *abscessus*, *and avium*; while inconsistent mutations begin accumulating in M. *tuberculosis*, *bovis, and leprae* where a U>A mutation disrupts a single AU pair (Figure 4). *M. gordonae* had the most inconsistent nucleotides (with the proposed 2D structure) in the 5′ hairpin (10 mutations), while *M. tuberculosis* and *bovis* had the most inconsistent nucleotides in the core hairpin (5 mutations). Interestingly, two compensatory mutations are observed in the core hairpin, which match those found in the cm-builder analysis. Seven of the ten species had at least one consistent change, where single point mutations preserve base pairing. *M. kansasii* had two consistent mutations, however, these were offset by several inconsistent changes. Notably, *M. leprae*, another necrotizing bacterium, appeared to have the highest level of conservation across the entire structured region when compared to *M. ulcerans*. Overall, the core hairpin and next downstream hairpin containing the start codon appear to be quite conserved, and when analyzed against all other data seems to indicate this region could be potentially functional.
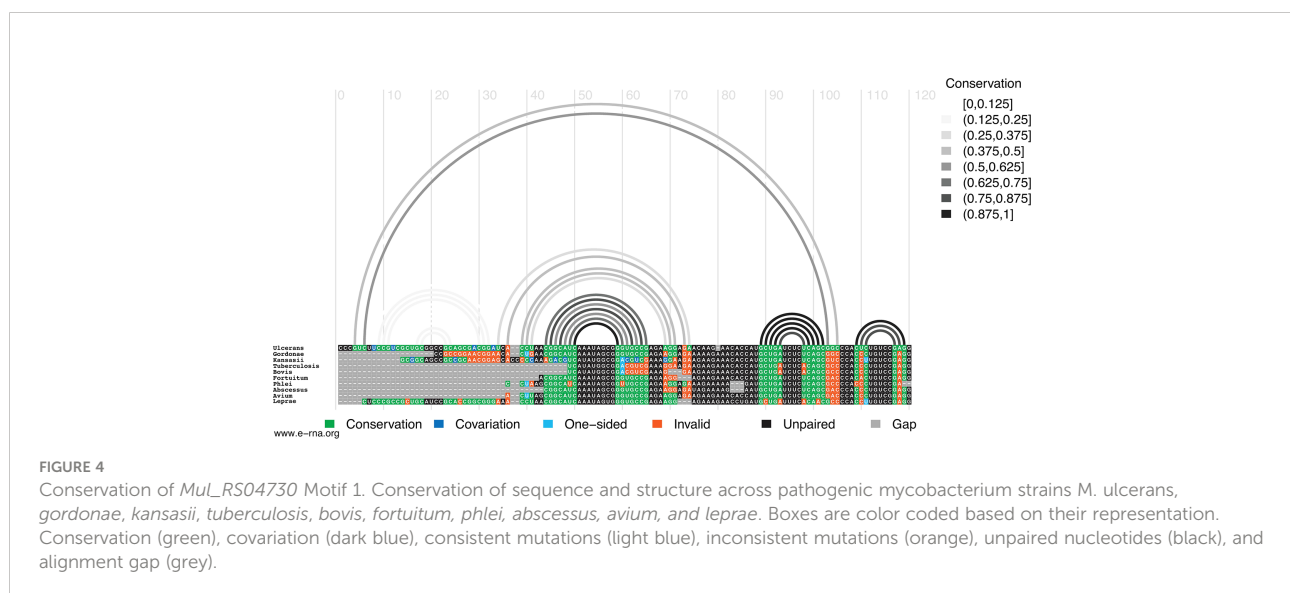
## Analysis of RNA folding in *Mul_RS04200*

*Mul_RS04200* encodes a protein homologous to the *rpoB* (beta chain of RNA polymerase) of *M. tuberculosis*. This protein functions as a component of the DNA-dependent RNA polymerase responsible for bacterial genome replication.

Interesting trends are seen in the ScanFold-Scan folding metrics partitioned per nucleotide of the gene (Supplementary Figure S1). The overall thermodynamic stability (ΔG°) remains flat across the entirety of the *Mul_RS04200* gene. Despite the flatline trend in ΔG°, there is evidence for regions of ordered RNA stability (negative z-scores). Z-scores remained relatively negative across the entire gene, but regions of lower z-scores were noted. Only one region in the middle of the gene showed a significant decrease in z-score below -2, while the majority of the gene's 3′ end displayed lower than average z-scores (indicating increased stability towards the end of the gene). In contrast to the 3′ end, a region of positive z-scores (potentially ordered to be unstructured) is observed approximately 750 nucleotides from the 5′ end.

The ScanFold-Fold motif found in the lowest z-score region (M1; Supplementary Figure S1) had base pairs with significantly low z-scores (< -2) which increased upstream and downstream of the hairpin. Notably, as base pairing extended out from the basal stem, the z-scores steadily increased until the final two bulges and terminal loop became only slightly negative.

## Analysis of RNA folding in *Mul_RS09540*

*Mul_RS09540* encodes a protein homologous to the *rpoZ* (omega component of RNA polymerase) of *M. tuberculosis*. This



FIGURE 4
Conservation of *Mul_RS04730* Motif 1. Conservation of sequence and structure across pathogenic mycobacterium strains M. ulcerans, *gordonae, kansasii, tuberculosis, bovis, fortuitum, phlei, abscessus, avium, and leprae*. Boxes are color coded based on their representation. Conservation (green), covariation (dark blue), consistent mutations (light blue), inconsistent mutations (orange), unpaired nucleotides (black), and alignment gap (grey).

protein functions as a component of the DNA-dependent RNA polymerase responsible for bacterial genome replication.

When analyzing the ScanFold-Scan folding metrics partitioned per nucleotide of the gene, interesting trends are seen (Supplementary Figure S2). Compared to the others, *Mul_RS09540* is the smallest gene that was analyzed. The overall thermodynamic stability (ΔG) remains uniform across the entirety of the gene, and it was found to have the fewest base pairs with z-scores < -2. Near the 3′ end, there are two distinct clusters of z-score values that increase into the positive range, thus indicating a region that may be evolved to have reduced structural propensity. This same unstable region correlates with high ensemble diversity, further indicating the lack of ordered and stable structure at the 3′ end. Conversely, in the middle of the gene, low z-score regions overlapped with regions of low ensemble diversity indicating that stronger RNA secondary structures are present near the core of the coding sequence. Interestingly, the most stable area of the structure seems to occur towards the middle of the gene rather than near the 3′ or 5′ ends.

The ScanFold-Fold motif found for this region (M1; Supplementary Figure S2) is predicted to form a 106 nucleotide hairpin that displays interesting trends in stability. In this structure, the most significantly low z-score base pairs were found in the first 5 base pairs of the basal stem. As base pairing extended out from the basal stem, the z-scores steadily increased until the final two bulges and terminal loop became only slightly negative.

## Analysis of RNA folding in *Mul_RS00210*

*Mul_RS00210* encodes a protein homologous to the *Pks7* (putative polyketide synthase) of *M. tuberculosis*. This protein likely functions in lipid metabolism where it is believed to be involved in an intermediate step for the synthesis of a polyketide molecule, mycolactone. This polyketide has been shown to act as the virulence factor essential for infection and the painless nature of the ulcers (48). Rather than falling within the bacterial genome, the *Mul_RS00210* gene occurs within the 174 kb virulence plasmid, pMUM001, that encodes a cluster of giant polyketide synthases (48). Until recently, this was an uncharacterized example of plasmid-mediated virulence in a Mycobacterium, and it is believed that the pathogenicity of *M. ulcerans* is due the acquisition of pMUM001 by horizontal gene transfer (48).

A summary of ScanFold data is shown on (Figure 5). The overall thermodynamic stability (ΔG) of *Mul_RS00210* is more variable than all other genes analyzed in this study. Evidence for regions of ordered RNA stability (negative z-scores) are clustered into many distinct regions throughout the coding region. Here, there are eight structured regions with z-scores < -2 and five small regions of positive z-scores interspersed (the first region is in the intergenic region). This indicates that the

ordered/evolved sequences making up these eight structures are much more stable than predicted based on nucleotide content: i.e., the evolved sequence may be ordered to have structure. These regions of high ordered stability (low z-score) correlate with dips in the ensemble diversity when compared to those found in regions of much lower ordered stability indicating that they are more stable and have a less volatile ensemble of potential RNA secondary structures that can form. The few regions of positive z-score may be indicative of regions that are evolved to be unstructured and therefore accessible to regulatory molecules.

When the low z-score windows were evaluated by ScanFold-Fold, eight distinct motifs (M1-8) were identified with exceptionally low (< -2) z-score-weighted base pairs (Figure 5). These motifs comprise seven hairpins (M1-7) and one multi-branch helix (M8). Here, all hairpins were found to have base pairs with significantly lower than average z-scores (< -2), whereas the multi-branch helix was found to only have significantly low z-scores in the basal stem. When evaluated for conservation across mycobacterial species, two of the proposed base pairs found in structures M1 and M2 were found to have statistically significant covariation with power of 0 to 0.1 and ≥ 0.25 respectively (Figure 5 and Supplementary File S2). In addition, all structures were found to be present/conserved across a wide array of species.

One unique feature of all -2 z-score structures found in this gene was their presence throughout the other two genes encoding the polyketide synthase subunits responsible for production of mycolactone (Figure 5). Predicted motifs are thus replicated multiple times throughout the plasmid—up to 15 times (Figure 5). Although this observed multiplication of structural elements is unique to our analysis, this finding is not entirely surprising as Stinear et al. found that these three genes have stretches of up to 27kb with near identical nucleotide sequence (99.7%). Additionally, of the 105-kb mycolactone locus, there is only ~9.5 kb of unique, nonrepetitive sequence (48).

## Analysis of RNA folding in *Mul_RS01615*

*Mul_RS01615* encodes a protein that is homologous to the *accD3* (putative acetyl CoA carboxylase carboxyl transferase-beta subunit) of *M. tuberculosis*. This protein is a component of the acetyl coenzyme A carboxylase complex and plays a functional role in lipid metabolism.

The overall thermodynamic stability (ΔG) stability varied across the gene, with a general trend toward lower stability (more positive values) toward the 3′ end (Figure 6). A distinct cluster of low z-score nucleotides occurs in the core coding region of the gene, which overlaps a cluster of distinctly low ensemble diversity values. The ScanFold-Fold motif built for this region, M1, contains a multibranch loop structure formed from
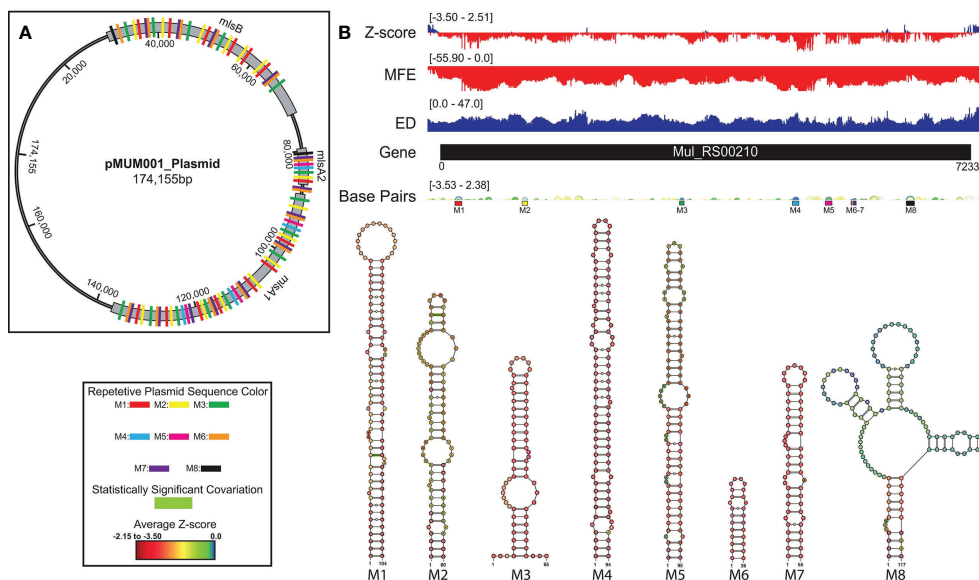
FIGURE 5

*Mul_RS00210* ScanFold results, 2D models, and structure occurrence throughout the pMUM001 plasmid. Global ScanFold results for *Mul_RS00210* including the location of repetitive sequence and structure elements across the pMUM001 virulence plasmid. **(A)** A schematic diagram of the 174kb pMUM001 virulence plasmid with the locations of the three largest genes that encode the subunits responsible for production of mycolactone. The colored boxes annotated across these three genes represent the locations of the sequences that form structures M1 (red), M2 (yellow), M3 (green), M4 (blue), M5 (pink), M6 (orange), M7 (purple), and M8 (black) in the Mul_RS00210 gene. **(B)** Global ScanFold results for Mul_RS00210 including (top to bottom) the ΔG z-score, minimum free energy (MFE), ensemble diversity (ED), base pair arc diagram, gene cartoon, and 2D models of -2 ΔG z-score structures. The base pair arc diagram is annotated with colored boxes (correlated to the colors in panel **(A)** to show the location of M1-8 across the gene. In the 2D structure models, nucleotides that exhibit statistically significant covariation are annotated by green bars across the base pair. The nucleotides of each structure are annotated with the average per nucleotide z-scores where the most negative are indicated in red and the most positive are indicated in blue.

two hairpin loops with low, but not significantly negative, z-score nucleotides and base pairs. The multibranch loop structure sits atop a long stem formed by significantly low (< -1) z-score base pairs and nucleotides, where the basal stem (composed of six base pairs) had the lowest (< -2) z-scores. Although the ensemble diversity was low across M1, the region spanning the

two hairpins of the multibranch loop were higher (Figure 6), suggesting potential conformational dynamics for the two hairpins. Notably, the region immediately downstream of M1 was spanned by positive z-scores, indicating a potential for ordered *instability*. A second structural motif, M2, was predicted immediately downstream of the annotated open
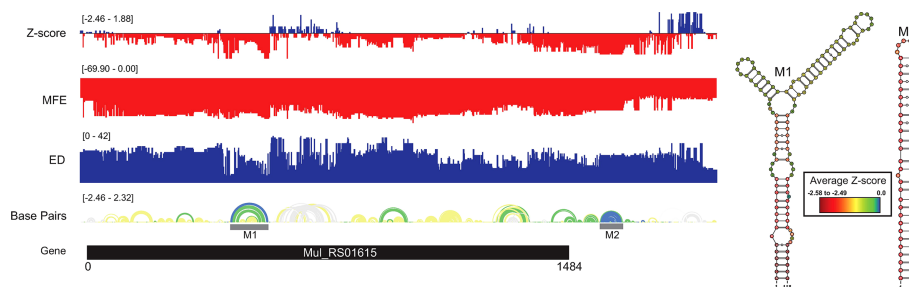


FIGURE 6

*Mul_RS01615* ScanFold results and 2D models. Global ScanFold results for *Mul_RS01615*. The ΔG z-score, minimum free energy (MFE), ensemble diversity (ED), base pair arc diagram, and gene cartoon (top to bottom) are shown to the left. The base pair arc diagram is annotated with gray boxes to show the location of M1-7 across the gene. All -2 ΔG z-score structures found are represented as 2D models to the right. The nucleotides of each structure are annotated with the average per nucleotide z-scores where the most negative are indicated in red and the most positive are indicated in blue.
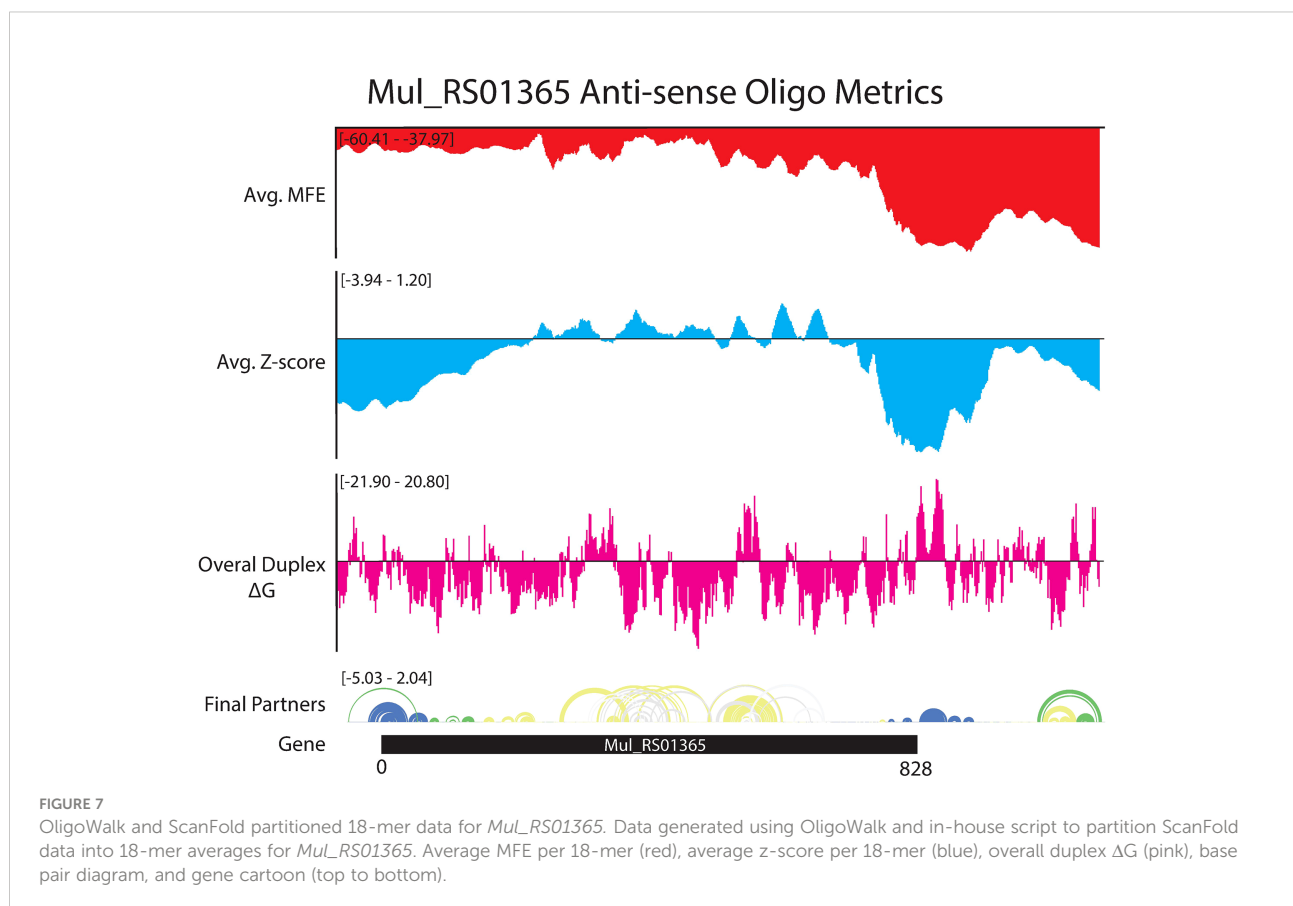
reading frame (ORF) for *Mul_RS01615* (Figure 6). This motif falls in a somewhat diffuse region of low z-scores that overlaps a cluster of moderate ensemble diversities; thus, while the base pairs and nucleotides comprising M2 have significantly low z-scores (<-2), the predicted conformational ensemble does not appear to be particularly tight (i.e., potential for dynamics). It is worth noting that the transcript annotations for *M. ulcerans* are not sufficient to determine if this motif falls within the 3′ UTR *of Mul_RS01615*. While both motifs in *Mul_RS01615* were found to have potentially homologous sequences and structures in other mycobacteria (Supplementary File S2), significant covariation was not observed.

## Analysis of antisense oligonucleotide accessibility

To explore the potential value of ScanFold data in identifying binding sites for short antisense oligonucleotides (ASO), we partitioned the ScanFold-Scan and -Fold results by averaging predicted metrics across short (18 nt) windows that approximate the size of potential ASO binding sites. This was further enhanced by predicting duplex binding affinities *via* the program OligoWalk (47) while considering the effects of

ScanFold predicted local *intra*molecular structure on *inter*molecular duplex formation (all data available in Supplementary File S4). To facilitate analysis, these results were also plotted vs. ScanFold-Fold predicted base pairs (Supplementary File S3). Numerous short stretches of sequence across each gene of interest, including potentially accessible regions with positive z-scores, overlap strong predicted duplex binding sites. While the results *in toto* are potentially valuable for aiding in the identification and design of ASOs vs. *M. ulcerans*, we focus our attention on the two genes that have predicted ordered structures that encompass the start sites of translation, which are particularly attractive sites for antimicrobial ASO therapeutics.

The ASO accessibility results for the *desA2* homolog, *Mul_RS01365*, are summarized in Figure 7. The trend toward enhanced thermodynamic stability of local RNA secondary structure in this gene is starkly illustrated, where significant dips in the partitioned averaged MFE and z-score overlap the translational stop site and ScanFold-Fold modeled low z-score structures. The inaccessibility of this region to ASOs is corroborated by OligoWalk predictions that show the overall duplex ΔGs are highly unfavorable: indeed, this region is predicted to have the most positive values across the gene. Conversely, the 5′ end, spanning the start codon, is predicted



**FIGURE 7**
OligoWalk and ScanFold partitioned 18-mer data for *Mul_RS01365*. Data generated using OligoWalk and in-house script to partition ScanFold data into 18-mer averages for *Mul_RS01365*. Average MFE per 18-mer (red), average z-score per 18-mer (blue), overall duplex ΔG (pink), base pair diagram, and gene cartoon (top to bottom).
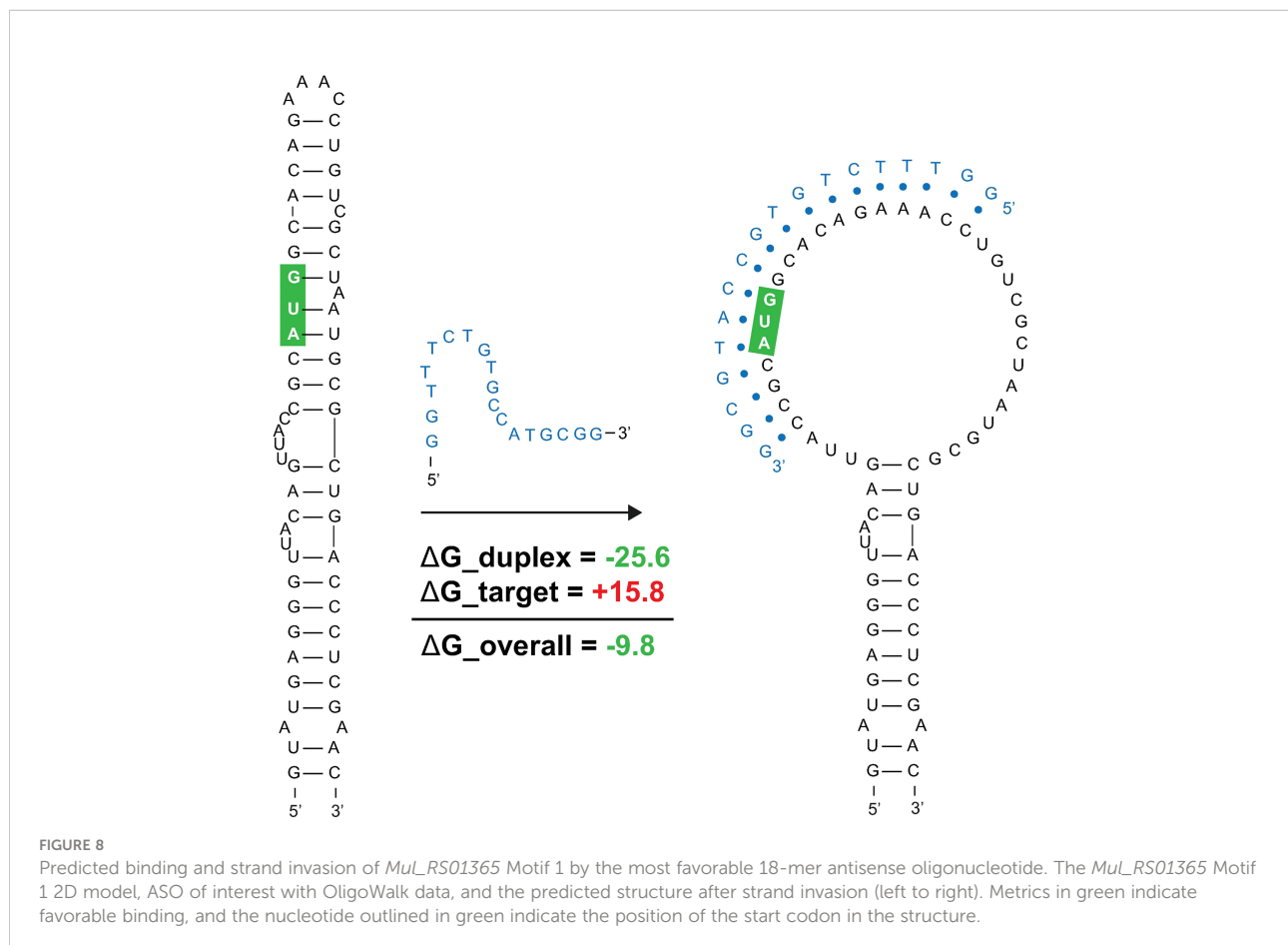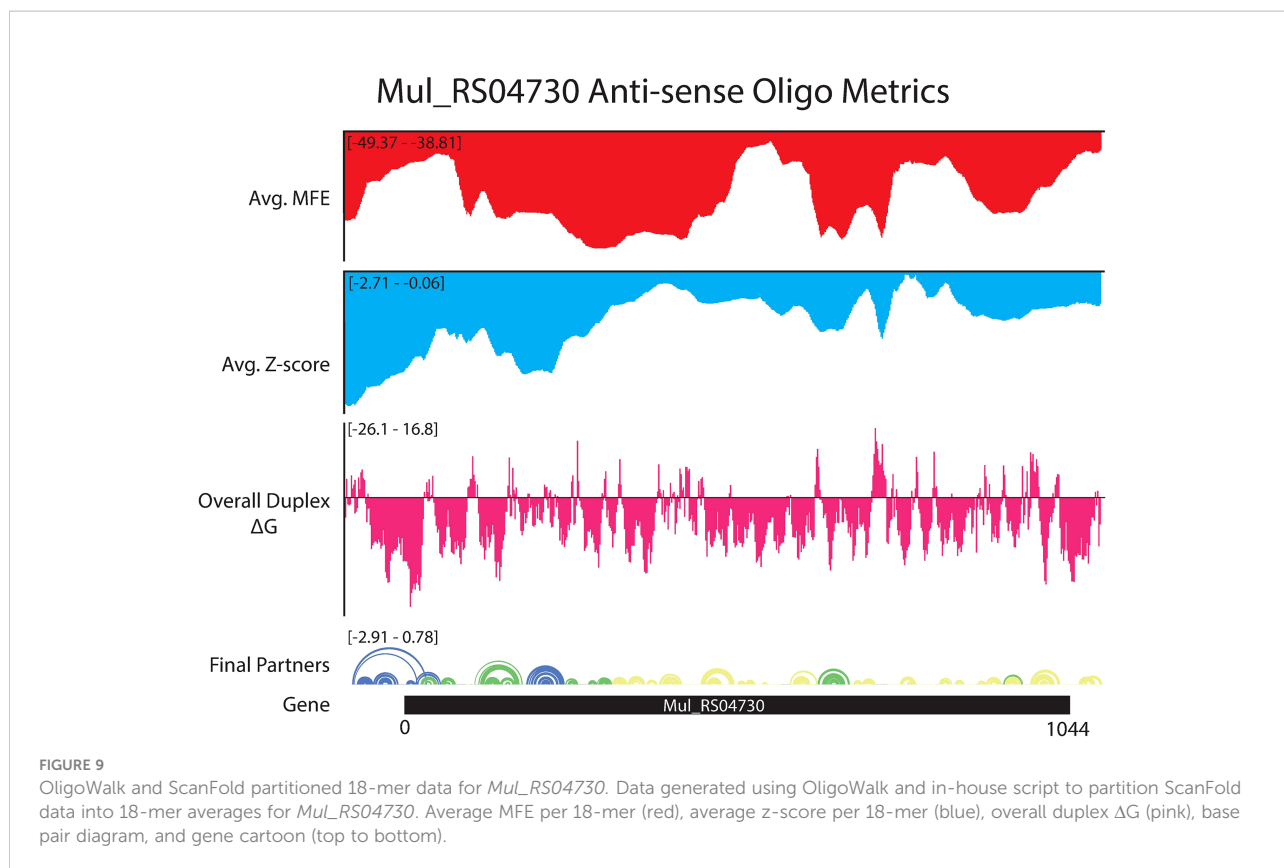
to have some of the least stable (least negative predicted average MFEs) local structure across the gene. Perhaps counterintuitively, there are dips in the average z-score and ensemble diversity that indicate ordered structure and indeed, ScanFold-Fold models show significantly ordered base pairing in this region (Figures 1, 7). Put another way, while the thermodynamic stability of this region is predicted to be low, it is still higher than expected given the nucleotide content, thus the ordered stability indicated by negative z-scores.

Favorable overall ASO duplex ΔGs span the start codon, despite it being contained in ordered structure (e.g., in M1, Figure 1), indicating a potential for "strand invasion", where an ASO can efficiently bind by breaking/replacing intramolecular helices with intermolecular base pairs. To explore this further, and to illustrate how our data could facilitate ASO design, the most favorable (lowest ΔG, taking into account disruption of target structure) ASO predicted to occlude the start site was identified (Figure 8). Here, an 18-mer ASO is predicted to bind to the mRNA with an overall ΔG of -9.8 kcal/mol, which considers the significant energy barrier (+15.8 kcal/mol) needed to disrupt 12 base pairs in the M1 hairpin structure. This disruption is predicted to totally ablate the terminal hairpin structure (Figure 8). Notably, other favorable ASO binding sites

flank this optimal one (Figure 8 and Supplementary File S4) and thus, the ASO sequence can be extended in either direction to enhance predicted stability: e.g., to add additional stabilizing base pairs to the looped-out nucleotides predicted to be released upon disruption of the terminal hairpin loop.

ASO accessibility results for the *rpoA* homolog *Mul_RS04730* are shown in Figure 9. The partitioned MFE varies across the mRNA, however, it is predicted to be less stable in the 18-mers that span the start codon. The z-score is lowest toward the 5′ end of the RNA and steadily increases toward the 3′ end, indicating 18-mers are less likely to be engaged in ordered RNA structures as one moves along the sequence. Importantly, the overall predicted ASO duplex stabilities were most favorable in the region spanning the start codon; indeed, the most stable predicted duplex across the entire mRNA spans the start site (Figure 9). The overall predicted ASO duplex ΔG is -15.0 kcal/mol, which considers the relatively low barrier (+7.4 kcal/mol) needed to invade the flanking hairpin structures in the target RNA (2 base pairs in each structure) (Figure 10). Similar to *Mul_RS01365*, multiple other ASOs are predicted to bind near this optimal site, allowing for more stable sequences to be designed: e.g., by extending the optimal predicted ASO to further invade the small downstream hairpin structure.



**FIGURE 8**
Predicted binding and strand invasion of *Mul_RS01365* Motif 1 by the most favorable 18-mer antisense oligonucleotide. The *Mul_RS01365* Motif 1 2D model, ASO of interest with OligoWalk data, and the predicted structure after strand invasion (left to right). Metrics in green indicate favorable binding, and the nucleotide outlined in green indicate the position of the start codon in the structure.

**FIGURE 9**

OligoWalk and ScanFold partitioned 18-mer data for *Mul_RS04730*. Data generated using OligoWalk and in-house script to partition ScanFold data into 18-mer averages for *Mul_RS04730*. Average MFE per 18-mer (red), average z-score per 18-mer (blue), overall duplex ΔG (pink), base pair diagram, and gene cartoon (top to bottom).
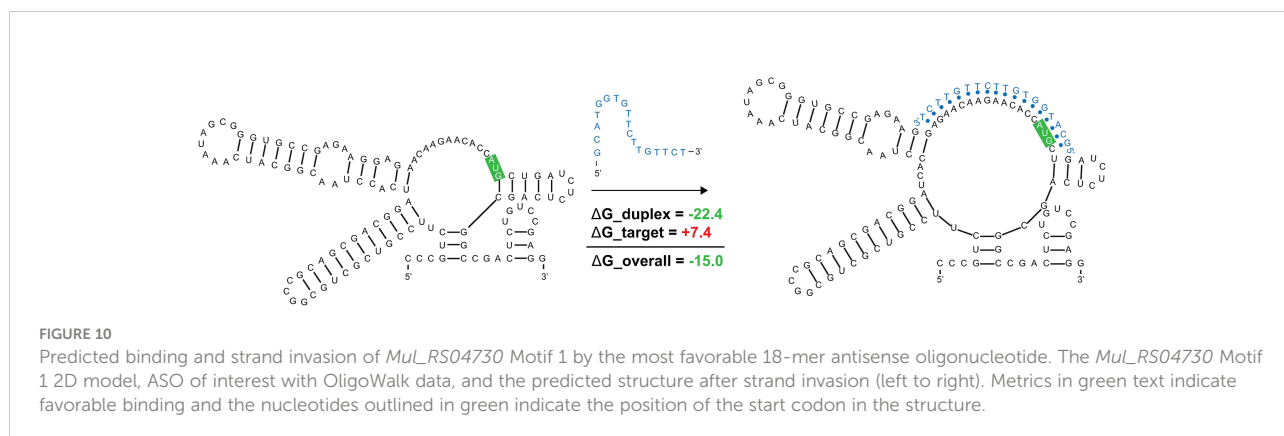
Beyond these interesting examples of ordered/structured RNAs that could be targeted with ASOs there are regions with apparent ordering for a lack of structure (positive z-scores) that could be especially accessible to ASOs (Supplementary File S3). For example, the 18-mer sequences that span the core coding region of *Mul_RS01365* are all predicted to have positive z-scores and relatively unfavorable MFEs (Figure 7). This central portion of the coding region also possesses the most favorable predicted ASO binding sites for the gene, suggesting that this region is an especially attractive site for ASO binding.

## Discussion

This study reports the first comprehensive analysis of RNA secondary structure in *M. ulcerans*. Here, we focused on a set of essential genes whose proteins have no homology to the human host. This is based on analyses of human and *M. ulcerans* pathways in the KEGG database and BLASTP homology searches of 20 different bacteria (34). This large set of 236 genes was also prioritized based on their druggablity and how well suited they are as potential therapeutic targets (34). Here, we



**FIGURE 10**

Predicted binding and strand invasion of *Mul_RS04730* Motif 1 by the most favorable 18-mer antisense oligonucleotide. The *Mul_RS04730* Motif 1 2D model, ASO of interest with OligoWalk data, and the predicted structure after strand invasion (left to right). Metrics in green text indicate favorable binding and the nucleotides outlined in green indicate the position of the start codon in the structure.

focused on six of these genes specifically based on their potential significance for drug targeting. Five of these genes are annotated as being essential to survival and replication (RNA synthesis and lipid metabolism) of the bacteria (34), while one is annotated to be required for pathogenicity and the painless nature of the ulcers (polyketide or mycolactone synthesis) (9, 48). We have specifically adapted our structural analysis pipelines to obtain data on these genes that may provide valuable results for advancing *M. ulcerans* mRNA therapeutics. Maps of the local RNA folding landscapes provide insights into the stability and potential for ordered/functional folding across each gene of interest. For example, looking at the ScanFold data for *Mul_RS01365* (both partitioned per-nucleotide and by 18-mers; Figures 1 and 7, respectively), we see thermodynamic stability (favorable/low MFE predictions) and ordered stability (low z-scores) at the 3′ end spanning the stop codon. This suggests potential roles for RNA structure and its thermodynamic stability in the termination of translation. Conversely, the coding region of this gene appears to be unstructured (as evidenced by mediocre MFEs and positive z-scores) perhaps to facilitate interactions with regulatory molecules or to promote rapid translation of this gene. Indeed, this latter idea is gaining traction as a mode for affecting protein folding (49): i.e., mRNA structural stability regulating the rate of translation to facilitate folding of proteins being translated.

While none of the analyzed genes had global biases for ordered structure (average z-score < -1; Table 1) clusters of ordered stability were present in each, where at least one defined structural motif could be predicted with exceptional (z-score < -2) base pairs, yielding 21 motifs in total across the six target genes when using mono- or dinucleotide shuffling. These sequences have been (presumably) ordered to fold over evolution and this proposition is supported by the overall conservation of secondary structure that was observed for each motif, as well as the identification of 4 statistically significant covarying base pairs. The functional roles of conserved, ordered RNA secondary structures in *M. ulcerans* and species with identified homologs can be diverse. For example, as noted above, structures in coding regions could affect translation rates. Structures may also be playing roles in modulating accessibility to regulatory molecules present in the bacterial cell or in mRNA turnover (e.g., modulating sensitivity to endogenous RNases). Significantly, two motifs span the translational start site: in *Mul_RS01365* and *Mul_RS04730* the start codons are modeled to lie within a helix and loop, respectively.

The instances where start sites are constrained within ordered structure present an interesting case for potential ASO therapeutics. Promising efforts at designing antimicrobial ASOs have focused on occluding the translational start sites of bacterial genes (50). For *Mul_RS01365* and *Mul_RS04730*, we predict oligos that can invade target intramolecular structure and, for *Mul_RS01365*, have the potential to greatly disrupt the ordered folding that may itself be functionally significant. That is to say,

an ASO targeting the start site of *Mul_RS01365* may have two potential modes of action: obscuring the start site to impede translation and disrupting a secondary structure that may itself play roles in regulating translation. Beyond these two structural motifs, we present 17 others which may also be amenable to ASO-mediated disruption or that may be targeted with RNA-targeting small-molecules—a promising area of research for novel therapeutics. Notably, the 8 identified motifs found in *Mul_RS00210* are replicated multiple times in the virulence plasmid, which increases the concentration of potential ligand binding sites. Disrupting the genes responsible for mycolactone production could attenuate the pathogenicity of *M. ulcerans*. Our data also suggest regions of unusual thermodynamic *instability* across all analyzed genes, which may be ordered for accessibility to trans-regulatory molecules. These sites may be particularly attractive for ASOs that can mask such sites from interacting partners and/or stimulate mRNA degradation.

It is worth noting that there are several potential limitations to our approaches. These limitations arise in part from the sparse data for *M. ulcerans* due to the "orphan" status of its associated disease. This paucity of data makes analyses of this pathogen challenging, including the analysis of functional RNA structure. For example, the identification of RNA secondary structure models is limited by the lack of covariation across the identified structures due to the limited diversity of sequences available. While structures are conserved, the little variability in sequence cannot identify significant, structure-preserving, covariation. There are few available large datasets; this includes genome sequencing of different mycobacterium strains (such as other *M. ulcerans* strains), biochemical probing data, and "omics" data such as RNA binding protein sites. Lastly, we are limited by a lack of experimental data to validate the effect of our predicted ASOs on survival and replication of *M. ulcerans*. Even with these potential limitations, the results presented here provides researchers with valuable data that can guide their studies to identify targets and design experiments. This highlights the need for additional work on this pathogen to collect more genome-wide datasets and develop better systems for *in vitro* studies (e.g. of putative ASO activity).

## Conclusion

We present the first in-depth analysis of RNA structure in six key genes of *M. ulcerans*, the microbe responsible for the neglected tropical disease, Buruli Ulcer. Our results are made public to advance a basic understanding of the RNA biology of this pathogen—by providing conserved structural motifs with high likelihood of function (which may, themselves, serve as potential therapeutic targets). As well, we hope to advance novel therapeutics against *M. ulcerans* by providing data to facilitate antisense oligonucleotide design.

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

## Author contributions

WBR was responsible for data collection and analysis as well as writing and editing the manuscript. JG and LP were responsible for picking the disease/organism of interest, data collection and analysis, as well as assisted in writing and editing the manuscript. WNM was responsible for coming up with the project and manuscript preparation. All authors contributed to the article and approved the submitted version.

## Funding

## Acknowledgments

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fitd.2022.1009362/full#supplementary-material

**SUPPLEMENTARY FIGURE 1**
All Mul_RS04200 ScanFold results. Figure showing ScanFold results for Mul_RS04200 including z-score, MFE, ED, base pair diagram, gene cartoon, and 2D model of the structure with a z-score <-2.

**SUPPLEMENTARY FIGURE 2**
All Mul_RS09540 ScanFold results. Figure showing ScanFold results for Mul_RS09540 including z-score, MFE, ED, base pair diagram, gene cartoon, and 2D model of the structure with a z-score <-2.

**SUPPLEMENTARY FILE S1**
M. ulcerans genomic data used in IGV-ScanFold. This file contains the M. ulcerans bacterial genome fasta, virulence plasmid fasta, and their associated gff3 genome annotations.

**SUPPLEMENTARY FILE S2**
All cm-builder covariation data. This file contains all the data required to run cm-builder, all the output files generated by INFERNAL and R-Scape, and results of power analysis.

**SUPPLEMENTARY FILE S3**
OligoWalk and 18-mer ScanFold bar charts. This file contains the OligoWalk and 18-mer partitioned ScanFold data as bar charts overlaid against the gene cartoon for all six genes studied.

**SUPPLEMENTARY FILE S4**
OligoWalk and 18-mer ScanFold raw data. This file contains the raw output data from OligoWalk and in-house script for portioning 18-mer ScanFold data for all six genes of interest.

**SUPPLEMENTARY FILE S5**
All ScanFold-Scan data. This file contains two folders with results from mono- and dinucleotide shuffling for each gene studied. These folders contain the raw ScanFold-Scan output data such as per nucleotide MFE, ED, z-score, input, and output fasta files, and out file.

**SUPPLEMENTARY FILE S6**
All ScanFold-Fold data. This file contains two folders with results from mono- and dinucleotide shuffling for each gene studied. These folders contain the raw ScanFold-Fold output data such as the log file, base pair track, final partners data, all dot bracket files, all CT files, extracted structures gff3 file, and the global VARNA 2D model.

## References

1. Walsh DS, Portaels F, Meyers WM. Buruli ulcer: Advances in understanding mycobacterium ulcerans infection. *Dermatol Clin* (2011) 29(1):1–8. doi: 10.1016/j.det.2010.09.006

2. Omansen TF, Erbowor-Becksen A, Yotsu R, van der Werf TS, Tiendrebeogo A, Grout L, et al. Global epidemiology of buruli ulcer, 2010-2017, and analysis of 2014 WHO programmatic targets. *Emerg Infect Dis* (2019) 25(12):2183–90. doi: 10.3201/eid2512.190427

3. Kent ML, Watral V, Wu M, Bermudez LE. *In vivo* and *in vitro* growth of mycobacterium marinum at homoeothermic temperatures. *FEMS Microbiol Lett* (2006) 257(1):69–75. doi: 10.1111/j.1574-6968.2006.00173.x

4. Clark HF, Shepard CC. Effect of environmental temperatures on infection with mycobacterium marinum (Balnei) of mice and a number of poikilothermic species. *J Bacteriol* (1963) 86:1057–69. doi: 10.1128/jb.86.5.1057-1069.1963

5. Merritt RW, Walker ED, Small PL, Wallace JR, Johnson PD, Benbow ME, et al. Ecology and transmission of buruli ulcer disease: A systematic review. *PloS Negl Trop Dis* (2010) 4(12):e911. doi: 10.1371/journal.pntd.0000911

6. Simpson H, Deribe K, Tabah EN, Peters A, Maman I, Frimpong M, et al. Mapping the global distribution of buruli ulcer: A systematic review with evidence consensus. *Lancet Glob Health* (2019) 7(7):e912–e22. doi: 10.1016/S2214-109X(19)30171-8

7. Yotsu RR, Suzuki K, Simmonds RE, Bedimo R, Ablordey A, Yeboah-Manu D, et al. Buruli ulcer: A review of the current knowledge. *Curr Trop Med Rep* (2018) 5 (4):247–56. doi: 10.1007/s40475-018-0166-2

8. George KM, Chatterjee D, Gunawardana G, Welty D, Hayman J, Lee R, et al. Mycolactone: A polyketide toxin from mycobacterium ulcerans required for virulence. *Science* (1999) 283(5403):854–7. doi: 10.1126/science.283.5403.854

9. Demangel C, Stinear TP, Cole ST. Buruli ulcer: reductive evolution enhances pathogenicity of mycobacterium ulcerans. *Nat Rev Microbiol* (2009) 7(1):50–60. doi: 10.1038/nrmicro2077

10. Hall B, Simmonds R. Pleiotropic molecular effects of the mycobacterium ulcerans virulence factor mycolactone underlying the cell death and immunosuppression seen in buruli ulcer. *Biochem Soc Trans* (2014) 42(1):177–83. doi: 10.1042/BST20130133

11. En J, Goto M, Nakanaga K, Higashi M, Ishii N, Saito H, et al. Mycolactone is responsible for the painlessness of mycobacterium ulcerans infection (buruli ulcer) in a murine study. *Infect Immun* (2008) 76(5):2002–7. doi: 10.1128/IAI.01588-07

12. Gordon CL, Buntine JA, Hayman JA, Lavender CJ, Fyfe JA, Hosking P, et al. All-oral antibiotic treatment for buruli ulcer: A report of four patients. *PloS Negl Trop Dis* (2010) 4(11):e770. doi: 10.1371/journal.pntd.0000770

13. Converse PJ, Nuermberger EL, Almeida DV, Grosset JH. Treating mycobacterium ulcerans disease (Buruli ulcer): from surgery to antibiotics, is the pill mightier than the knife? *Future Microbiol* (2011) 6(10):1185–98. doi: 10.2217/fmb.11.101

14. Falese JP, Donlic A, Hargrove AE. Targeting RNA with small molecules: from fundamental principles towards the clinic. *Chem Soc Rev* (2021) 50(4):2224–43. doi: 10.1039/D0CS01261K

15. Warner KD, Hajdin CE, Weeks KM. Principles for targeting RNA with drug-like small molecules. *Nat Rev Drug Discov* (2018) 17(8):547–58. doi: 10.1038/nrd.2018.93

16. Ursu A, Childs-Disney JL, Andrews RJ, O'Leary CA, Meyer SM, Angelbello AJ, et al. Design of small molecules targeting RNA structure from sequence. *Chem Soc Rev* (2020) 49(20):7252–70. doi: 10.1039/D0CS00455C

17. Rinaldi A. RNA To the rescue: RNA is one of the most promising targets for drug development given its wide variety of uses. *EMBO Rep* (2020) 21(7):e51013. doi: 10.15252/embr.202051013

18. Sully EK, Geller BL. Antisense antimicrobial therapeutics. *Curr Opin Microbiol* (2016) 33:47–55. doi: 10.1016/j.mib.2016.05.017

19. Bennett CF. Efficiency of antisense oligonucleotide drug discovery. *Antisense Nucleic Acid Drug Dev* (2002) 12(3):215–24. doi: 10.1089/108729002760220806

20. The cost of getting personal. *Nat Med* (2019) 25(12):1797. doi: 10.1038/s41591-019-0700-3

21. Roberts TC, Langer R, Wood MJA. Advances in oligonucleotide drug delivery. *Nat Rev Drug Discovery* (2020) 19(10):673–94. doi: 10.1038/s41573-020-0075-7

22. Batista-Duharte A, Sendra L, Herrero MJ, Tellez-Martinez D, Carlos IZ, Alino SF. Progress in the use of antisense oligonucleotides for vaccine improvement. *Biomolecules* (2020) 10(2):e00754–20. doi: 10.3390/biom10020316

23. Andrews RJ, O'Leary CA, Tompkins VS, Peterson JM, Haniff HS, Williams C, et al. A map of the SARS-CoV-2 RNA structurome. *NAR Genom Bioinform* (2021) 3(2):lqab043. doi: 10.1093/nargab/lqab043

24. Andrews RJ, O'Leary CA, Moss WN. A survey of RNA secondary structural propensity encoded within human herpesvirus genomes: Global comparisons and local motifs. *PeerJ* (2020) 8:e9882. doi: 10.7717/peerj.9882

25. Zhang P, Park HJ, Zhang J, Junn E, Andrews RJ, Velagapudi SP, et al. Translation of the intrinsically disordered protein alpha-synuclein is inhibited by a small molecule targeting its structured mRNA. *Proc Natl Acad Sci USA* (2020) 117 (3):1457–67. doi: 10.1073/pnas.1905057117

26. O'Leary CA, Andrews RJ, Tompkins VS, Chen JL, Childs-Disney JL, Disney MD, et al. RNA Structural analysis of the MYC mRNA reveals conserved motifs that affect gene expression. *PloS One* (2019) 14(6):e0213758. doi: 10.1371/journal.pone.0213758

27. Meyer MM. The role of mRNA structure in bacterial translational regulation. *Wiley Interdiscip Rev RNA* (2017) 8(1):153–60. doi: 10.1002/wrna.1370

28. Andrews RJ, Roche J, Moss WN. ScanFold: An approach for genome-wide discovery of local RNA structural elements-applications to zika virus and HIV. *PeerJ* (2018) 6:e6136. doi: 10.7717/peerj.6136

29. Haniff HS, Tong Y, Liu X, Chen JL, Suresh BM, Andrews RJ, et al. Targeting the SARS-CoV-2 RNA genome with small molecule binders and ribonuclease targeting chimera (RIBOTAC) degraders. *ACS Cent Sci* (2020) 6(10):1713–21. doi: 10.1021/acscentsci.0c00984

30. Wacker A, Weigand JE, Akabayov SR, Altincekic N, Bains JK, Banijamali E, et al. Secondary structure determination of conserved SARS-CoV-2 RNA elements by NMR spectroscopy. *Nucleic Acids Res* (2020) 48(22):12415–35. doi: 10.1093/nar/gkaa1013

31. Rangan R, Watkins AM, Chacon J, Kretsch R, Kladwang W, Zheludev IN, et al. *De novo* 3D models of SARS-CoV-2 RNA elements from consensus experimental secondary structures. *Nucleic Acids Res* (2021) 49(6):3092–108. doi: 10.1093/nar/gkab119

32. Lulla V, Wandel MP, Bandyra KJ, Ulferts R, Wu M, Dendooven T, et al. Targeting the conserved stem loop 2 motif in the SARS-CoV-2 genome. *J Virol* (2021) 95(14):e0066321. doi: 10.1128/JVI.00663-21

33. Chan AP, Choi Y, Schork NJ. Conserved genomic terminals of SARS-CoV-2 as coevolving functional elements and potential therapeutic targets. *mSphere* (2020) 5(6):e00754–20. doi: 10.1128/mSphere.00754-20

34. Butt AM, Nasrullah I, Tahir S, Tong Y. Comparative genomics analysis of mycobacterium ulcerans for the identification of putative essential genes and therapeutic candidates. *PloS One* (2012) 7(8):e43080. doi: 10.1371/journal.pone.0043080

35. Clote P, Ferre F, Kranakis E, Krizanc D. Structural RNA has lower folding energy than random RNA of the same dinucleotide frequency. *RNA* (2005) 11 (5):578–91. doi: 10.1261/rna.7220505

36. Freyhult E, Gardner PP, Moulton V. A comparison of RNA folding measures. *BMC Bioinf* (2005) 6:241. doi: 10.1186/1471-2105-6-241

37. Mathews DH. Using an RNA secondary structure partition function to determine confidence in base pairs predicted by free energy minimization. *RNA* (2004) 10(8):1178–90. doi: 10.1261/rna.7650904

38. Moss WN. The ensemble diversity of non-coding RNA structure is lower than random sequence. *Noncoding RNA Res* (2018) 3(3):100–7. doi: 10.1016/j.ncrna.2018.04.005

39. Andrews RJ, Baber L, Moss WN. Mapping the RNA structural landscape of viral genomes. *Methods* (2020) 183:57–67. doi: 10.1016/j.ymeth.2019.11.001

40. Manfredonia I, Nithin C, Ponce-Salvatierra A, Ghosh P, Wirecki TK, Marinus T, et al. Genome-wide mapping of SARS-CoV-2 RNA structures identifies therapeutically-relevant elements. *Nucleic Acids Res* (2020) 48 (22):12436–52. doi: 10.1093/nar/gkaa1053

41. Rivas E, Clements J, Eddy SR. A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs. *Nat Methods* (2017) 14(1):45–8. doi: 10.1038/nmeth.4066

42. Rivas E. Evolutionary conservation of RNA sequence and structure. *Wiley Interdiscip Rev RNA.* (2021) 12(5):e1649. doi: 10.1002/wrna.1649

43. Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* (2013) 29(22):2933–5. doi: 10.1093/bioinformatics/btt509

44. McGinnis S, Madden TL. BLAST: At the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res* (2004) 32:W20–5. doi: 10.1093/nar/gkh435

45. Woolf B. The log likelihood ratio test (the G-test); Methods and tables for tests of heterogeneity in contingency tables. *Ann Hum Genet* (1957) 21(4):397–409. doi: 10.1111/j.1469-1809.1972.tb00293.x

46. Darty K, Denise A, Ponty Y. VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics* (2009) 25(15):1974–5. doi: 10.1093/bioinformatics/btp250

47. Lu ZJ, Mathews DH. OligoWalk: An online siRNA design tool utilizing hybridization thermodynamics. *Nucleic Acids Res* (2008) 36:W104–8. doi: 10.1093/nar/gkn250

48. Stinear TP, Mve-Obiang A, Small PL, Frigui W, Pryor MJ, Brosch R, et al. Giant plasmid-encoded polyketide synthases produce the macrolide toxin of mycobacterium ulcerans. *Proc Natl Acad Sci USA* (2004) 101(5):1345–9. doi: 10.1073/pnas.0305877101

49. Faure G, Ogurtsov AY, Shabalina SA, Koonin EV. Role of mRNA structure in the control of protein folding. *Nucleic Acids Res* (2016) 44(22):10898–911. doi: 10.1093/nar/gkw671

50. Hegarty JP, Stewart DBSr. Advances in therapeutic bacterial antisense biotechnology. *Appl Microbiol Biotechnol* (2018) 102(3):1055–65. doi: 10.1007/s00253-017-8671-0