Authors' Response to Peer Reviews

Authors' Response to Peer Review of "Exploring the Accuracy of Ab Initio Prediction Methods for Viral Pseudoknotted RNA Structures (Preprint)"

Vasco Medeiros[1], BSc, MSc, AS; Jennifer Pearl[2], BSc, AS; Mia Carboni[3], BSc; Stamatia Zafeiri[1], BSc, MSc

[1]Stevenage Bioscience Catalyst, Stevenage, United Kingdom

[2]Stanford University, Santa Clara, CA, United States

[3]University of Turin, Torino, Italy

Corresponding Author:
Vasco Medeiros, BSc, MSc, AS
Stevenage Bioscience Catalyst
Gunnels Wood Rd Stevenage
Stevenage, SG1 2FX
United Kingdom
Phone: 44 07534150352
Email: vasco.miguel.medeiros@gmail.com

Live PREreview [1]

Summary

The study [2] examines the performance of 5 RNA-folding engines for predicting complex viral pseudoknotted RNA structures. This research fills a critical gap in the field by comparing the efficiency of minimal free energy (MFE) and maximum expected accuracy (MEA) using a curated dataset of 26 viral RNA sequences with known secondary structures. Contrary to prevailing assumptions favoring MEA models, their findings reveal that pKiss, an MFE-folding engine, outperforms Vsfold 5 in terms of the sensitivity, positive predictive value (PPV), and F_1-score, while laying emphasis on the importance of the PPV and sensitivity parameters in understanding and determining the superior accuracy of pKiss to predict correct base pairs and minimize incorrect predictions. The authors also point out that the engine still needed additional data to achieve high accuracy as well as a better understanding of thermodynamics at the intracellular level.

The statistical analyses used to evaluate the results were 2-way ANOVA and Tukey multiple comparisons test, which provided robust insights into the performance differences among the tested engines. The research integrates bioinformatics with statistics and advanced data science methodologies to promote our understanding of computational RNA biology. The study provides important insights into the relative advantages and disadvantages of both approaches in predicting pseudoknotted RNA structures by contrasting MFE models and MEA models. It also highlights avenues for future research to focus on the development of more sophisticated energy models and MFE engines, like pKiss, to enhance prediction capabilities, especially in the context of viral replication and gene regulation, which may lead to a better understanding of the functional roles of pseudoknotted RNA structures. Overall, this research contributes significantly to the field of computational and molecular biology.

Below, we list major and minor concerns that were discussed by participants of the live review, and where

possible, we provide suggestions on how to address those issues.

List of Major Concerns and Feedback

It would be helpful to provide more context on why percent error was chosen as the primary metric for evaluating different engines, considering alternatives like mean absolute error (MAE) and mean squared error (MSE) could enhance the analysis. For instance, MAE is robust against outliers, making it a valuable metric, especially when outlier removal is part of the process. Although MAE is less sensitive to extreme values, it can offer a useful qualitative check on the models. On the other hand, the mean MSE's sensitivity to outliers can be advantageous when the spread of the forecast is important. Including these metrics could provide a more comprehensive evaluation.

Response: It should be noted that the percent error (%) in this case is in fact analogous to the MAE (as described in equation 1) and is represented as such in Figure 4A. Hence, the two are used interchangeably, which is clarified on pages 8 and 10 and within the supplementary materials on page 23. The reviewer's suggestion to use MSE, in addition to MAE was heeded, and the following values for MSE were calculated and exhibited in the newly generated Figure 4B. The utility of MSE and MAE was expounded on in the Discussion section on page 15. It should be noted that a prerequisite for applying MSE is that all data be normally distributed, which was tested for and confirmed.

It should also be noted that MSE (or root mean square error) equals the residuals' SD, calculating the variance between the observed and expected values. This makes it unnecessary to calculate and display the SD of the MSE within Figure 4B.

The authors have conducted a comprehensive and insightful study, revealing important differences in prediction accuracy between Vsfold 5 and pKiss. One area that could further enhance the manuscript is the exploration of how auxiliary parameters (eg, Mg2+ binding, dangling end options, H-type penalties) are managed across the various RNA-folding engines utilized. For example, Vsfold 5, although being an MEA model, may encounter challenges if its handling of Mg2+ binding or dangling ends significantly diverges from what is optimal for the studied RNAs. The authors' observation in section 3.1 that "the low percent error exhibited by pKiss could be the result of the pseudoknot 'enforce' constraint, but it is more likely that this outcome was multivariable, equating to the Turner energy model used, and the sensitive auxiliary parameters enforced by the program" is particularly insightful. This highlights the complexity of RNA structure prediction algorithms. To build on these findings, a structured comparative analysis of parameter handling across different software tools could be highly beneficial. This analysis would not only clarify why certain engines performed better than others but also help in identifying best practices or potential biases in prediction methodologies. Such

an addition would significantly strengthen the study's conclusions and provide valuable guidance for future research in RNA structure prediction.

Response: This major criticism has been addressed. Though not all auxiliary parameters like Mg2+ binding and dangling end options were explored (given that the amount of data this would generate could warrant an entirely new paper/manuscript) several auxiliary parameters were modified and compared to the original percent error (%)/MAE of pKiss (Figure 4A).

These include overall pKiss function, altering the "enforce" setting to the "MFE" setting, which computes the single energetically best secondary structure (column 2, S2.1), altering the strategy from "pKiss C" (slow, low memory, but thorough) to "pKiss A" (fast but sloppy; column 3, S2.2), altering the exclusion of lonely base pairs to the inclusion of lonely base pairs (column 4, S2.8), altering the H-type penalty from 9 to 18 (column 5 S2.3), and altering the K-type penalty from 12 to 24 (column 5, S2.4). pKiss was chosen to explore these parameters, given that it is the most accurate out of all other folding software.

Testing these auxiliary parameters, displaying them in graphical format, and expounding upon them in the Discussion section of this report (page 15) helps eliminate potential biases and shows how sensitive these functions can be and why certain engines are better than others.

To build on these findings, a structured comparative analysis of parameter handling across different software tools could be highly beneficial. This analysis would not only clarify why certain engines performed better than others but also help in identifying best practices or potential biases in prediction methodologies. Such an addition would significantly strengthen the study's conclusions and provide valuable guidance for future research in RNA structure prediction.

In section 3.1 of the manuscript, no significant difference in percent error was identified. However, it does not specify the statistical test employed nor the method used for adjusting P values, which are essential details for validating the results. Additionally, the term "Vij" is introduced early in the manuscript but is not contextualized until page 13. Providing this context earlier would enhance the reader's understanding.

Response: Both before and after the PREreview criticisms were considered and implemented, the statistical tests employed in this paper were expounded on. Statistical analyses used throughout the paper (2-way ANOVA testing, outlier identification, normality, lognormality tests, etc) were summarized at the end of the Materials and Methods section (page 10). Moreover, any specific statistical test used for any dataset/figure was discussed in depth within the figure descriptions and, if relevant to the reader, discussed even more in depth within the Discussion section of the manuscript.

The term $V(i,j)$ (the real symmetric contact matrix) was contextualized earlier within the paper, as per the reviewers'

suggestion (pages 2 and 3), and the math behind its utility is explained in greater depth.

> *It would be beneficial if "false positive" and "false negative" were more clearly defined, particularly in the context of mRNA detection. To improve clarity, the authors might consider specifying that sensitivity is the appropriate measure for detecting mRNA among known positives, while specificity is the appropriate measure for detecting mRNA among known negatives, where the probability of false positives is 1 – specificity. Additionally, using the Youden index (J), which is defined as sensitivity + specificity – 1, could provide a helpful summary of detection accuracy. This index ranges from –1 (indicating 100% incorrect detection) to 1 (indicating 100% correct detection), offering a clear metric for assessing performance [3].*

Response: False positives (pairings that do not fall under Matthews' [4] parameters) and false negatives (base pairs missed by the prediction software) were more clearly defined on page 8 of the manuscript in the context of all metrics used and in the context of MAE/percent error (%). They were also discussed in greater depth on page 9 and on page 16, in the context of Youden's index. The specificity and sensitivity used to detect known negatives and known positives, respectively, are further discussed in greater depth within sections 2.4, 3.2, 3.3, and 4.2 of the original manuscript.

The reviewers' suggestion to use the Youden index (J; a more sensitive metric than F_1 scoring) was implemented. The equation for the Youden index (J) was provided on page 9, with figures for J values of the 26 experimentally derived pseudoknotted models from each of the 5 folding software displayed on page 13 (including raw values and normalized values), the results of which were then expounded on in the Discussion section.

While J is most often represented in a receiver operating characteristic curve, the authors are displaying the mean (SD) of J values across folding software for the sake of clarity and visual representation, given that 130 receiver operating characteristic curves would prove visually confusing. It should also be noted that data were normalized and put into graphical format (Figure 6B) again for visual clarification. This eliminates negative values for J (as negative values for J are not defined).

> *Providing the link to the dataset will allow better compliance with open science practices. Please add the link to the dataset as it appears to be missing from the reviewed version of the manuscript. When sharing the dataset, it would be important to also include the associated metadata and appropriate documentation that matches the methods described in the manuscript. For guidelines on how to share data so that it's as reusable as it can be, authors may refer to the Findability, Accessibility, Interoperability, and Reuse (FAIR) principles of data sharing [5].*

Response: The link to both the dataset, the RNA pseudoknot folding software web servers, and the manner in which each was implemented to said dataset have been provided, complying

with open science practices. They can be located within the supplemental materials, the DOI and URL of which can be found on page 17 of the original manuscript (DOI: 10.17605/OSF.IO/7QVKN [6]). They are located within the Open Science Framework an open-source cloud-based project management platform (for ease of access). The FAIR Principles of Data Sharing was also broached, referenced on page 4 of the manuscript.

> *Figure 5B displays the PPV as three distinct blocks rather than continuous values, with varying sensitivity within these blocks. This nonrandom binning of PPV suggests the need for further investigation to understand the underlying causes.*

Response: This abnormal trend is accounted for within the manuscript and expounded upon within the Discussion section (page 12, last paragraph). It is explained that, while most MFE folding engines tend to have lower PPVs than sensitivity values (due to thermodynamics imposed by MFE algorithms overshooting the number of canonical base pairings), pKiss and NUPACK do not follow this trend. This is because updated software such as this implements a more accurate assessment of the thermodynamic properties of the structure, removing unwanted pairs and improving overall performance [7].

> *In the Discussion section, the authors stated "We have provided evidence suggesting that MEA software is not always the optimal method of topological prediction when applied to short viral pseudoknotted RNA." This is a significant claim and would benefit greatly from specific references to support the evidence provided in the study. Citing the relevant figures and results that support this claim would significantly enhance comprehension and readability. For example, "As demonstrated in Figure 4, the MEA software Vsfold 5 exhibited higher percent errors in predicting knotted base pairs compared to MFE software like pKiss." Additionally, referencing previous studies that have reported similar findings or that discuss the limitations of MEA methods in RNA structure prediction in the Discussion section would strengthen the credibility of the authors' claims by showing that similar limitations have been observed by other researchers. This helps readers understand that the study is building upon existing knowledge. For instance, "Previous studies have also highlighted the limitations of MEA methods in RNA folding predictions, particularly for pseudoknotted structures (in-text citations)."*

Response: The authors recognize this as the most valuable criticism the reviewers have made and have addressed it accordingly. The length of the Discussion section was increased from 389 words to 1235 words, including a more in-depth analysis of the results and statistical significance of said results, as it applies to MEA software being suboptimal (at times) when compared to its MFE counterparts. Nine different references to in-text figures or supplementary materials were added; relevant literature was additionally cited, and previous studies were discussed and compared to the results of the paper.

However, it should also be noted that previous studies highlighting the limitations of MEA methods when applied to RNA folding have not been cited, given that this is a novel approach to exploring ab initio RNA prediction algorithms. Though RNA prediction algorithms have been explored in great depth in previous literature, none (to the author's knowledge) have compared MFE and MEA prediction software against one another, and none have explored this software when applied to short-stranded (20-150nt) viral pseudoknotted RNAs.

List of Minor Concerns and Feedback

Overall, the reviewers really appreciated how clearly the figures and results were presented. Below are some minor suggested improvements [7].

In the Abstract section: Please identify the abbreviation PPV as positive predictive value.

Response: The acronym for positive predictive value (PPV) was defined in the abstract as per the editor's suggestion (page 1).

Page 3, first paragraph after Figure 1: Definitions of pseudoknot should be referenced.

Response: On page 3, an official definition for a pseudoknot was referenced by Brierley et al [8]. Moreover, two more definitions were posited by the authors on the very same page.

Page 3, second paragraph after Figure 1: Please identify the NMR abbreviation as nuclear magnetic resonance.

Response: Nuclear magnetic resonance (NMR) was abbreviated, and the acronym was defined. Please note that, given the new additions to the manuscript, this clarification is now found on page 4 of the manuscript, rather than page 3.

Page 7: The manuscript acknowledges the skewness in the data and provides a rationale for its presence. It's noted that this skewness impacts the training and testing phases, often contributing to false positives and false negatives. It would be beneficial if the authors could elaborate on how they addressed data imbalance, particularly in relation to reducing false positives and false negatives. This additional detail would enhance the understanding of the methods used to manage data skewness and improve model performance.

Response: The reviewer is correct in stating that the manuscript acknowledges the skewness in data and that it provides a rationale for its presence on page 7:

"This skewness (regarding the class of RNA) is intentional, and true to nature, given that hairpin-type pseudoknots (H-type) are more common by far [8]."

However, the reviewer is incorrect in saying that this skewness impacts the training and testing phases (which were not discussed within the manuscript) and is incorrect in saying that it contributes to false positives and false negatives (which it does not). This would be the case for deep learning algorithms and folding software that is malleable, in the sense that it can be trained and altered when inserting varying inputs. However, the folding software provided by each of the 5 web servers do not harbor any deep learning algorithm/training/machine learning, unlike other software such as ATTfold (which is referenced in the paper, page 16).

Page 8, second paragraph: "Mathews et al. 2019" should be corrected to "Mathews, 2019" [4].

Response: "Mathews et al 2019" was indeed changed to "Mathews 2019" [4], as per the editors' suggestion, on page 8 of the manuscript.

*Page 8, equation 1: Add a "%" next to *100, giving the output of x%.*

The amendment was made, and equation 1 now has a % sign next to the 100 (100%).

Page 10, Figure 4: In the title, "accurcy" should be corrected to "accuracy."

Response: In Figure 4, the typo "accuracy" was indeed corrected. Please note that, given new additions to the manuscript, this amendment can now be found on page 11, while the amendment itself is located in Figure 4A.

Page 10, Figure 4: The bar of the SD of Vienna (knotted) is not presented.

Response: There being no SD bar present in the Vienna (knotted) control variable is actually correct. As stated on page 5 of the paper (as well as other instances within the manuscript), the Vienna RNAfold engine does not compute for pseudoknots, which is why it was implemented as a negative control. Its inability to compute pseudoknots axiomatically means that the percent error (%) for pseudoknot generation will always be 100%, leading to no SD whatsoever.

Page 10, Figure 4: The bars of the SD seem to be widely large, indicating significant variability in the results, so a test of the normality of data distribution should be performed before comparisons. This is also observed for the kinefold results in Figures 5 and 6.

Response: Tests for normality (gaussian) and lognormality were conducted on all relevant figures:

- Figure 4 (A-B): Kolmogorov-Smirnov tests
- Figure 5: Shapiro-Wilk test
- Figure 6 (A-B): Shapiro-Wilk test and Kolmogorov-Smirnov tests
- Figure 7: Kolmogorov-Smirnov test

Page 12, Figure 6B: The color bar on the heat maps is missing.

Response: Color bar on the heat map (now found on page 14 of the manuscript) was added.

References

1. Sadedri D, Nagesh V, Mahmoud RSG, Olatoye T, Arogundade FQ. Peer review of "Exploring the Accuracy of Ab Initio Prediction Methods for Viral Pseudoknotted RNA Structures (Preprint)". JMIRx Bio. 2024;2:e65154. [FREE Full text] [doi: 10.2196/65154]
2. Medeiros V, Pearl J, Carboni M, Zafeiri S. Exploring the accuracy of ab initio prediction methods for viral pseudoknotted RNA structures: retrospective cohort study. JMIRx Bio. 2024:e58899. [doi: 10.2196/58899]
3. Youden index. ScienceDirect. 2010. URL: https://www.sciencedirect.com/topics/medicine-and-dentistry/youden-index [accessed 2024-10-23]
4. Mathews DH. How to benchmark RNA secondary structure prediction accuracy. Methods. Jun 01, 2019;162-163:60-67. [FREE Full text] [doi: 10.1016/j.ymeth.2019.04.003] [Medline: 30951834]
5. FAIR principles. GO FAIR. URL: https://www.go-fair.org/fair-principles/ [accessed 2024-10-23]
6. Exploring the accuracy of ab initio prediction methods for viral pseudoknotted RNA structures. Open Science Framework. URL: https://osf.io/ujp5r [accessed 2024-10-23]
7. Do C, Woods D, Batzoglou S. CONTRAfold: RNA secondary structure prediction without physics-based models. Bioinformatics. Jul 15, 2006;22(14):e90-e98. [doi: 10.1093/bioinformatics/btl246] [Medline: 16873527]
8. Brierley I, Pennell S, Gilbert RJC. Viral RNA pseudoknots: versatile motifs in gene expression and replication. Nat Rev Microbiol. Aug 2007;5(8):598-610. [FREE Full text] [doi: 10.1038/nrmicro1704] [Medline: 17632571]

Abbreviations

FAIR: Findability, Accessibility, Interoperability, and Reuse
MAE: mean absolute error
MEA: maximum expected accuracy
MFE: minimal free energy
MSE: mean squared error
NMR: nuclear magnetic resonance
PPV: positive predictive value