RESEARCH ARTICLE

REVISED *In silico* study on RNA structures of intronic mutations of beta-globin gene [version 2; peer review: 1 approved, 1 not approved]

Nur Imaniati Sumantri¹, Kenny Lischer^{2,3}, Dian Rachma Wijayanti⁴, Tomy Abuzairi ^{1,3}

¹Biomedical Engineering, Department of Electrical Engineering, Faculty of Engineering, Universitas Indonesia, Depok, 16424, Indonesia

²Bioprocess Technology, Department of Chemical Engineering, Faculty of Engineering, Universitas Indonesia, Depok, 16424, Indonesia

³Research Center for Biomedical Engineering, Faculty of Engineering, Universitas Indonesia, Depok, 16424, Indonesia

⁴Department of Biotechnology, Faculty of Science and Technology, Binawan University, Jakarta, 13630, Indonesia

V2 First published: 27 Jan 2020, 9:49
<https://doi.org/10.12688/f1000research.21953.1>
 Second version: 20 Jul 2020, 9:49
<https://doi.org/10.12688/f1000research.21953.2>
 Latest published: 01 Sep 2020, 9:49
<https://doi.org/10.12688/f1000research.21953.3>

Abstract

Background: Mutation of the beta-globin gene (HBB) interferes with primary mRNA transcription, leading to beta-thalassemia disease. The IVS1nt1 and IVS1nt5 mutations were reported as two of the most prevalent intronic mutations associated with beta-thalassemia major. These mutations may affect the mRNA structure of the human beta-globin (*HBB*) gene. However, the mechanism by which variation in *HBB* alters the mRNA structure remains unclear. The objective of this study was to unveil the secondary and tertiary conformation difference of the mutants compared to the wildtype using *in silico* analysis.

Methods: The sequence of *HBB* was obtained from Ensemble database and mutated manually at nucleotides 143 (IVS1nt1G>T) and 147 (IVS1nt5G>C). The RNA secondary and tertiary structure were performed by ViennaRNA Web Services and RNA Composer, respectively.

Results and Discussion: The results revealed the unique folding characteristics of each mutations for the secondary and tertiary structures. Based on the structure, unwanted folding occurred in the IVS1nt1G>T and IVS1nt5G>C mRNA structures compared to the wild-type structure. This finding was supported by the results of centroid-based analysis and RNA structure analysis, indicating that the larger loops in IVS1nt1 and IVS1nt5 result in an unstable structure. Our study found that intronic mutations affect the mRNA structure of *HBB* by altering its folding mechanism.

Open Peer Review

Approval Status

	1	2
version 3 (revision) 01 Sep 2020	 view	
version 2 (revision) 20 Jul 2020	 view	 view
version 1 27 Jan 2020	 view	 view

1. Zhichao Miao , European Bioinformatics Institute (EMBL-EBI), Cambridge, UK
2. Kholis Abdurachim Audah , Swiss German University, Tangerang, Indonesia
 Indonesian Society for Bioinformatics and Biodiversity, Tangerang, Indonesia
 Indonesian Society for Biochemistry and Molecular Biology, Tangerang, Indonesia
 Swiss German University, Tangerang,

Keywords

Intervening sequence, IVS1nt5, IVS1nt1, beta globin gene, RNA structure, beta thalassemia major

Indonesia

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Tomy Abuzairi (tomy.abuzairi@ui.ac.id)

Author roles: Sumantri NI: Formal Analysis, Methodology, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; Lischer K: Conceptualization, Formal Analysis, Validation; Wijayanti DR: Conceptualization, Validation, Writing – Original Draft Preparation; Abuzairi T: Conceptualization, Supervision, Visualization, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This work has been supported by QQ Research Grant 2019 (No. NKB-0327/UN2.R3.1/HKP.05.00/2019) from Universitas Indonesia.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2020 Sumantri NI *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Sumantri NI, Lischer K, Wijayanti DR and Abuzairi T. *In silico study on RNA structures of intronic mutations of beta-globin gene [version 2; peer review: 1 approved, 1 not approved]* F1000Research 2020, 9:49 <https://doi.org/10.12688/f1000research.21953.2>

First published: 27 Jan 2020, 9:49 <https://doi.org/10.12688/f1000research.21953.1>

REVISED Amendments from Version 1

The manuscript was revised according to the reviewers' comments, as follows:

1. Methods: the tertiary structure were predicted using new tools.
2. Results: the full sequence of tertiary structure were displayed.
3. Discussion: clinical implication was discussed to support structural alteration effect.
4. Discussion: more discussion and additional reference were made on mutants structure and functional impact.
5. Data: Figure 5 was updated.

Any further responses from the reviewers can be found at the end of the article

Introduction

Thalassemia is a hereditary blood disorder that induces the production of an abnormal form or inadequate amount of hemoglobin in the body. Hemoglobinopathy has been identified in approximately 71% of all countries around the world, and more than 50% of cases of β -thalassemia occur in Southeast Asia¹. β -thalassemia is one of the most prevalent blood disorders worldwide. It can result in various traits and the coinheritance of thalassemia minor, intermedia, or major depends on many factors. People with β -thalassemia exhibit reduced hemoglobin production, and low levels of hemoglobin results in a lack of oxygen supply throughout the body².

β -thalassemia is caused by mutations in the human beta-globin (*HBB*) gene, which is responsible for producing β -globin, a subunit of hemoglobin³. Most mutations associated with β -thalassemia are caused by the substitution of one or a limited number of nucleotides in *HBB*. These mutations affect the functions of the gene including transcription, RNA processing or translation of β -globin mRNA. β zero (β^0)-thalassemia is caused by mutations in *HBB* that stop the production of beta-globin. Conversely, other mutations only reduce the amount of beta-globin protein produced, a condition termed β plus (β^+)-thalassemia⁴.

People with β^0 - and β^+ -thalassemia have been diagnosed with thalassemia major and thalassemia intermedia, respectively. Thalassemia major has a more severe phenotype than thalassemia intermedia. Most patients with thalassemia major die at a young age⁵. Unfortunately, as many as 23,239 babies are born with inherited β -thalassemia major every year¹.

Conformational changes in regulatory RNA induce specific human diseases. More than 95% of mutations result in only small local changes in the conformation of RNA. The same phenotype (disease) can be caused by different mutations with varying degrees of effects on the overall RNA structure⁶.

Numerous variants of non-deletional β -thalassemia are caused by mutations that interfere with processing of the primary mRNA transcript⁷. These mutations affect AG or GT dinucleotides

at the exon-intron splice junction. These mutations eventually ablate regular splicing and induce β^0 -thalassemia, resulting in a β -thalassemia major phenotype⁴. Mutations in intervening sequence 1 (IVS1) at nucleotides 143 (change from guanine to thymine, IVS1nt1G>T) and 147 (change from guanine to cytosine, IVS1nt5G>C) are common mutations in Southeast Asia that result in β -thalassemia minor and major, respectively⁸.

A study on mutations that interfere with mRNA transcription may explain the mechanism by which the mRNA structure is changed by mutations in *HBB*. Specific *in silico* RNA analysis and visualization tools to study these structural changes have been developed, such as ViennaRNA Web Services, RNA Composer, and UCSF Chimera^{9–12}. ViennaRNA Web Services and RNA Composer are web interface packages that offer structure modeling for RNA. This study evaluated the effects of the intronic *HBB* mutations IVS1nt1 and IVS1nt5 on the RNA structure using specific *in silico* tools.

Methods

Data access

The complete genome sequence of *HBB* was obtained from Ensembl for transcriptional regulation detail by selecting transcript HBB-201 ENST00000335295.4¹³. Concerning mutations in IVS1, introns 1–2 were mutated manually at nucleotides 143 (IVS1nt1G>T) and 147 (IVS1nt5G>C). The sequences of wild-type and mutant *HBB* were aligned using BioEdit version 7.2¹⁴. The complete wild-type and mutant sequences were transcribed into mRNA using Nucleic Acid Converter (<https://skaminsky115.github.io/nac/>).

Secondary structure predictions

We used the RNAfold Server from ViennaRNA Web Services to predict the RNA secondary structures⁹. By selecting this applet, thermodynamic parameters, including centroids, partition function, and minimum free energy (MFE), were automatically calculated by the software as considered for structure folding. To create the RNA best structure tree plot, the wild-type and mutants sequences and structural elucidations were forwarded to the Barriers server, which is part of the ViennaRNA Web Services package^{15,16}. The tools were utilized with their respective default parameters.

Free energy minimization and base pairing probabilities are the most used methods in predicting single-sequence structure using the following equations¹⁷. When a structure **i** is at equilibrium with the unpaired state, the equilibrium is described by the equilibrium constant **K_i** as presented in Equation 1:

$$K_i = \frac{[\text{Structure } i]}{[\text{Unpaired state}]} = e^{-\Delta G_i^0 / RT} \quad (1)$$

The Gibbs free energy change for structure **i**, namely ΔG_i , quantifies the favorability of a given secondary structure compared with the unpaired state. Similarly, the equilibrium between two structures **i** and **j** is described in Equation 2 as follows:

$$\frac{[\text{Structure } i]}{[\text{Structure } j]} = \frac{K_i}{K_j} = e^{-(\Delta G_i^0 - \Delta G_j^0) / RT} \quad (2)$$

Thus, the Gibbs free energy change quantifies the favorability of a structure at a given temperature. The structure with the lowest Gibbs free energy change will be the most prevalent in solution at equilibrium.

Supplementing a free energy minimization calculation with a partition function calculation can identify the pairs that are more likely to be correct. The partition function Q is defined as the sum of the equilibrium constants K_i for all possible structures.

$$Q = \sum_{\text{all structures}} K_i = \sum_{\text{all structures}} e^{-\Delta G_i^0 / RT} \quad (3)$$

Thus, the probability of a particular structure i being found in solution can be calculated according to Equation 4 as follows:

$$P_i = \frac{e^{-\Delta G_i^0 / RT}}{Q} \quad (4)$$

Furthermore, the probability of a base pair i - j can be calculated by summing the equilibrium constants for structures that contain that pair and dividing by the partition function.

$$P(i-j \text{ basepair}) = \sum_k \frac{e^{-\Delta G_k^0 / RT}}{Q} \quad (5)$$

In Equation 5, the sum of the index k is taken over all structures that have the base pair i - j .

3D tertiary structure visualization

The 3dRNA v2.0 from Biomolecular Physics and Modeling Group Xiao Lab (<http://biophy.hust.edu.cn/services.html>) was used to construct the tertiary structures. Whole genome sequence and 2D structure with dot-bracket form, obtained from RNAfold Server, were entered to generate 3D structures for wild-type and each mutant¹⁰. The prediction files were downloaded in the .pdb format. The interactive model was visualized using the University of California at San Francisco (UCSF) Chimera version 1.13.1¹⁸.

Results and discussion

Intronic mutations of *HBB* can lead to β -thalassemia major. These mutations have spread widely around the world¹⁹. In this study, we observed the effects of intronic mutations on *HBB* mRNA structure. *HBB* is encoded by 3931 nucleotides located on chromosome 11 with its various regulatory genes. This gene contains 1608 bp and consists of three exons and two introns. Introns 1–2 cover nucleotides 143–272. This region is known as IVS1 (Figure 1a). In these introns, we highlighted two mutations, namely IVS1nt5G>C and IVS1nt1G>T, located at nucleotides 143 and 147, respectively (Figure 1b).

Our observation revealed differences in the MFE and other thermodynamic characteristics between wild-type and mutant *HBB*. We found that different point mutations occurred in the same IVS region, reflecting unique features. MFE, centroid, and partition function properties are depicted in a mountain plot in Figure 2.

The MFE has important roles for secondary structure prediction. Lower values are associated with better structures. Conformation of MFE secondary structure by RNA folding is irreversible because the sequence space sets up the shape space²⁰. Centroid-based analysis is an alternative method for predicting the secondary structure. This calculation yields 30% less prediction errors than the MFE²¹. The stability of RNA molecules fluctuates in different low energy structure. Therefore, we need to analyze the equilibrium ensemble of RNA structures via the partition function²². The partition function measures the quality of secondary structure by determining the confidence in base pairs predicted by free energy minimization²³.

Based on the mountain plot result, the mutations of IVS1nt1G>T and IVS1nt5G>C changes the mRNA structure of *HBB*. However, IVS1nt5G>C leads to a much larger free energy change versus the wild-type. A previous study found that IVS1nt1G>T and IVS1nt5G>C take place in splice junction and

Figure 1. Regional and sequence detail of mutant human beta-globin (*HBB*) gene. (A) Details of *HBB* showing the target mutations. (B) Sequence alignment of wild-type and mutant *HBB* was performed using BioEdit version 7.2.

Figure 2. Mountain plots of the thermodynamic characteristics of the predicted RNA secondary structures of the wild-type and mutant HBB. (A) Centroid, (B) minimum free energy, and (C) partition function.

consensus splice site, respectively, that might result in splice junction disturbances²⁴. Therefore, we believe that this disturbance may affect the mRNA structure.

Although introns are known as non-coding regions, their alteration influences mRNA expression, which can lead to several diseases. Using the RNAfold applet from the ViennaRNA package, we analyzed the thermodynamic characteristics and structural changes of wild-type and mutant *HBB* (Figure 3 and Figure 4)^{7,9}.

The secondary structure of mRNA is beneficial for tracing the refolding pathways of an RNA molecule. The next question that arises is how the secondary structure of mRNA is affected by mutation of *HBB*. To obtain the mRNA structure, the

RNAfold system was used (see Methods). The prediction of the mRNA secondary structure is based on the centroid-based pair probability (Figure 3). It was found that the three structures have a distinct structural backbone. Based on the structure, unwanted folding occurred in the IVS1nt1G>T and IVS1nt5G>C mRNA structures compared with the wild-type structure. The principle of this method is to create the secondary structure with minimal base pair distance to all other secondary structures in the Boltzmann ensemble⁹. The blue box in Figure 4 highlighted the main alteration formed within the three sequences. However, the structure of IVS1nt5G>C seems to be the best predicted according to the red color in the base-pairing probability. To further clarify the effect of folding, barrier tree plot analysis was performed (Figure 4). The result revealed that changes in the population occur earlier in

Figure 3. Centroid-based pair probability of the human beta-globin gene. (A) Wild-type, (B) IVS1nt1G>T, and (C) IVS1nt5G>C. Structural changes of RNA are highlighted by blue boxes.

Figure 4. Barrier tree plot for best RNA structure for the human beta-globin gene. (A) Wild-type, (B) IVS1nt1G>T, and (C) IVS1nt5G>C.

IVS1nt5G>C carriers than in people carrying wild-type or IVS1nt1G>T *HBB*. The earlier changes in the population result in different populations toward the end of the barrier tree plot.

These intronic *HBB* mutations may lead to critical folding changes that could drastically affect the tertiary structure of the beta-globin chain as well as the quaternary structure. To further clarify the effects of the mutations, 3D models of wild-type and mutant and *HBB* were visualized using UCSF Chimera software (Figure 5). In general, the tertiary structure of each sequence was unique. The point mutation gave tremendous folding changes in the 3D model. Once the base shifted, base pairing mechanism was influenced massively. Previous research reported the predicted pathogenic effects of these intronic mutations using the HSF server. IVS1nt5G>C was predicted to be a silencer motif new site, intronic identifying elements and exonic splicing silencer (ESS) site were broken. ESS provides binding sites for proteins, promoting exon exclusion, and helps the spliceosome to decoy splice sites²⁵. In addition, Seo *et al.* (2013) stated that most of the mutations affecting splicing disrupted the highly conserved donor and acceptor sites (GT/AG) at the exon–intron junctions and polypyrimidine tract, and the branch-point sequence may be disrupted by mutations affecting the splicing sites²⁶.

RNA structure stability depends on metal or ligand binding, the alteration of which can affect the binding site activity and structure stability. The properties of RNA structures related to multiple ligand binding, ligand binding-induced conformational changes, and the ion-stabilized catalytic core have been explored. The study by Miao *et al.* stated that the positive charge of metal cations plays compensates for the negative charge of the RNA phosphate backbone²⁷. In addition to that function, metal ions also bind to extremely specific locations on RNA 3D structures²⁷. All of these findings reveal that IVS1nt5G>C significantly changes the mRNA structure. However, the protein structure of HBB still needs to be elucidated to

understand the effects of the mutations on protein features after translation.

The mRNA plays important roles in protein regulation. RNAs start to form after the transcription process is completed. Once done, the structure of RNA is critical for its activities. Both IVS1nt5G>C and IVS1nt1G>T give implications to the clinical condition due to protein instability. There are some disruptions in producing beta globin chains by these mutations, hence the beta thalassemia major (BTM) patients produce beta globin chain in a limited number or even not at all. This mechanism results in unbalanced alpha and beta globin chains in hemoglobin that expedite the hemolysis and ineffective erythropoiesis contributing to reduced production of mature red blood cells²⁸. Thus, the BTM patients requires lifelong blood transfusions and commonly suffer complications, such as endocrine complications and cardiac complications²⁹.

This study examined the effects of mutations in terms of the minimum energy and the impact on the secondary and tertiary structures of *HBB* mRNA. Changes in the RNA formation may lead to disturbances of the protein structure. Protein expression can be influenced by mRNA folding, other local and global sequence features, and modulation of mRNA stability by codon usage³⁰. Regarding the length of introns 1–2, the mutations occur in small RNA-affected regions where analysis is important for detection, annotation, quantification, and *de novo* discovery of putative small RNA molecules³¹. However, it will be interesting to clarify the effects of splicing, debranching, and trimming on the mRNA structure in future research.

Conclusion

Our research revealed the impact of intronic mutations on the secondary and tertiary structure of mRNA. Different point mutations occurring in the same IVS region were associated with unique characteristics. We have expanded our understanding of the thermodynamic aspects of these mutants. The limitations of this study included our inability to clarify the interaction of

Figure 5. Tertiary structures of wild-type (upper), IVS1nt1 G>T (middle), and IVS1nt5 G>C human beta-globin (lower). This prediction shows complex structural folding considering the base pairing probability.

secondary structures due to the lack of an RNA analysis platform function for intronic mutations. RNA molecular simulation and wet lab experiments are required to elucidate the detailed features of the intronic mutants and their interactions with ions or ligands to further assess their functional failure, especially in genetic disease.

Data availability

Underlying data

Figshare: DNA sequence of wild-type and intronic mutation human beta-globin gene, <https://doi.org/10.6084/m9.figshare.11565726.v2>³².

Figshare: RNA sequence of wild-type and intronic mutation human beta-globin gene, <https://doi.org/10.6084/m9.figshare.11566596.v1>³³.

Figshare: Mountain plot of wild-type and intronic mutation human beta-globin gene, <https://doi.org/10.6084/m9.figshare.11556675.v3>³⁴.

Figshare: Centroid base pair probability of wild-type and intronic mutation human beta-globin gene, <https://doi.org/10.6084/m9.figshare.11565711.v2>³⁵.

Data are available under the terms of the [Creative Commons Attribution 4.0 International license](#) (CC-BY 4.0).

Acknowledgements

The authors would like to thank Enago (www.enago.com) for the English language review.

References

- Modell B, Darlison M: **Global epidemiology of haemoglobin disorders and derived service indicators.** *Bull World Health Organ.* 2008; **86**(6): 480–487. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Musharraf SG, Iqbal A, Ansari SH, *et al.*: **β -Thalassemia Patients Revealed a Significant Change of Untargeted Metabolites in Comparison to Healthy Individuals.** *Sci Rep.* 2017; **7**: 42249. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ngo DA, Steinberg MH: **Genomic approaches to identifying targets for treating β hemoglobinopathies.** *BMC Med Genomics.* 2015; **8**: 44. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Thein SL: **The molecular basis of β -thalassemia.** *Cold Spring Harb Perspect Med.* 2013; **3**(5): a011700. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Fuchareon S, Winichagoon P: **Haemoglobinopathies in southeast Asia.** *Indian J Med Res.* 2011; **134**: 498–506. [PubMed Abstract](#) | [Free Full Text](#)
- Halvorsen M, Martin JS, Broadaway S, *et al.*: **Disease-associated mutations that alter the RNA structural ensemble.** *PLoS Genet.* 2010; **6**(8): e1001074. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Thein SL: **Molecular basis of β thalassemia and potential therapeutic targets.** *Blood Cells Mol Dis.* 2018; **70**: 54–65. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Li CK: **New trend in the epidemiology of thalassaemia.** *Best Pract Res Clin Obstet Gynaecol.* 2017; **39**: 16–26. [PubMed Abstract](#) | [Publisher Full Text](#)
- Gruber AR, Lorenz R, Bernhart SH, *et al.*: **The Vienna RNA websuite.** *Nucleic Acids Res.* 2008; **36**(Web Server issue): W70–W74. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Wang J, Wang J, Huang Y, *et al.*: **3dRNA v2.0: An Updated Web Server for RNA 3D Structure Prediction.** *Int J Mol Sci.* 2019; **20**(17): pii: E4116. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Lorenz R, Wolfinger MT, Tanzer A, *et al.*: **Predicting RNA secondary structures from sequence and probing data.** *Methods.* 2016; **103**: 86–98. [PubMed Abstract](#) | [Publisher Full Text](#)
- Fallmann J, Will S, Engelhardt J, *et al.*: **Recent advances in RNA folding.** *J Biotechnol.* 2017; **261**: 97–104. [PubMed Abstract](#) | [Publisher Full Text](#)
- Hunt SE, McLaren W, Gil L, *et al.*: **Ensembl variation resources.** *Database (Oxford).* 2018; **2018**: bay119. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Hall TA: **BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT.** *Nucleic Acids Symp Ser.* 1999; **41**: 95–98. [Reference Source](#)
- Flamm C, Hofacker IL, Stadler PF, *et al.*: **Barrier trees of degenerate landscapes.** *Z Phys Chem.* 2002; **216**(2): 155. [Publisher Full Text](#)
- Wolfinger MT, Svrcek-Seiler WA, Flamm C, *et al.*: **Efficient computation of RNA folding dynamics.** *J Phys A Math Gen.* 2004; **37**: 4731–4741. [Publisher Full Text](#)
- Seetin MG, Mathews DH: **RNA structure prediction: an overview of methods.** In: Keiler K. (eds) *Bacterial Regulatory RNA.* Methods Mol Biol. Humana Press, Totowa, NJ. 2012; **37**: 99–122. [PubMed Abstract](#) | [Publisher Full Text](#)
- Pettersen EF, Goddard TD, Huang CC, *et al.*: **UCSF Chimera—a visualization system for exploratory research and analysis.** *J Comput Chem.* 2004; **25**(13): 1605–1612. [PubMed Abstract](#) | [Publisher Full Text](#)
- De Sanctis V, Kattamis C, Canatan D, *et al.*: **β -Thalassemia Distribution in the Old World: an Ancient Disease Seen from a Historical Standpoint.** *Mediterr J Hematol Infect Dis.* 2017; **9**(1): e2017018. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Lorenz R, Bernhart SH, Höner Zu Siederdisen C, *et al.*: **ViennaRNA Package 2.0.** *Algorithms Mol Biol.* 2011; **6**: 26. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Schuster P: **Prediction of RNA secondary structures: from theory to models and real molecules.** *Rep Prog Phys.* 2006; **69**(5). [Publisher Full Text](#)
- Ding Y, Chan CY, Lawrence CE: **RNA secondary structure prediction by centroids in a Boltzmann weighted ensemble.** *RNA.* 2005; **11**(8): 1157–1166. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Hofacker IL, Lorenz R: **Predicting RNA structure: advances and limitations.** In: Waldsich C, editor. *RNA folding methods and protocols.* Methods Mol Biol. Humana Press, New York. 2014; **1086**: 1–19. [PubMed Abstract](#) | [Publisher Full Text](#)
- Thein SL, Wood WG: **The molecular basis of β thalassemia, $\delta\beta$ thalassemia, and hereditary persistence of fetal hemoglobin.** In *Disorders of hemoglobin: Genetics, pathophysiology, and clinical management.* Ed. Steinberg MH, Forget BG, Higgs DR, Nagel RL, 2009; 323–356. Cambridge University Press, Cambridge, UK. [Publisher Full Text](#)
- Mahdieh N, Rabbani B: **Beta thalassemia in 31,734 cases with HBB gene mutations: Pathogenic and structural analysis of the common mutations; Iran as the crossroads of the Middle East.** *Blood Rev.* 2016; **30**(6): 493–508. [PubMed Abstract](#) | [Publisher Full Text](#)
- Seo S, Takayama K, Uno K, *et al.*: **Functional analysis of deep intronic SNP rs13438494 in intron 24 of PCLO gene.** *PLoS One.* 2013; **8**(10): e76960. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Miao E, Westhof Z: **RNA Structure: Advances and Assessment of 3D Structure Prediction.** *Annu Rev Biophys.* 2017; **46**: 483–503. [PubMed Abstract](#) | [Publisher Full Text](#)
- Oikonomidou PR, Rivella S: **What can we learn from ineffective erythropoiesis in thalassemia?** *Blood Rev.* 2018; **32**(2): 130–143. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Borgna-Pignatti C, Gamberini MR: **Complications of thalassemia major and their treatment.** *Expert Rev Hematol.* 2011; **4**(3): 353–366. [PubMed Abstract](#) | [Publisher Full Text](#)
- Boël G, Letso R, Neely H, *et al.*: **Codon influence on protein expression in *E. coli* correlates with mRNA levels.** *Nature.* 2016; **529**(7586): 358–363. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

31. Pantano L, Pantano F, Marti E, *et al.*: **Visualization of the small RNA transcriptome using seqclusterViz [version 2; peer review: 2 approved]**. *F1000Res*. 2019; 8: pii: ISCB Comm J-232.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
32. Abuzairi T: **DNA sequence of wild-type and intronic mutation human beta-globin gene**. *figshare*. Dataset. 2020.
<http://www.doi.org/10.6084/m9.figshare.11565726.v2>
33. Abuzairi T: **RNA sequence of wild-type and intronic mutation human beta-globin gene**. *figshare*. Dataset. 2020.
<http://www.doi.org/10.6084/m9.figshare.11566596.v1>
34. Abuzairi T: **Mountain plot of wild-type and intronic mutation human beta-globin gene**. *figshare*. Figure. 2020.
<http://www.doi.org/10.6084/m9.figshare.11556675.v3>
35. Abuzairi T: **Centroid base pair probability of wild-type and intronic mutation human beta-globin gene**. *figshare*. Figure. 2020.
<http://www.doi.org/10.6084/m9.figshare.11565711.v2>

Open Peer Review

Current Peer Review Status:

Version 2

Reviewer Report 05 August 2020

<https://doi.org/10.5256/f1000research.26502.r67437>

© 2020 Audah K. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Kholis Abdurachim Audah

¹ Department of Biomedical Engineering, Swiss German University, Tangerang, Indonesia

² Indonesian Society for Bioinformatics and Biodiversity, Tangerang, Indonesia

³ Indonesian Society for Biochemistry and Molecular Biology, Tangerang, Indonesia

⁴ Academic Research and Community Services, Swiss German University, Tangerang, Indonesia

I accept the revised version and no further comment from my side so far.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: My area of research is in protein engineering, protein-protein interactions/ biomolecules interactions/ structure-function relationships which include but not limited to DNA manipulation, DNA mutagenesis, cloning and expression, protein purification (various chromatography methods), protein characterization employing different biophysical methods (CD, MALDI-MS, Fluorescence Spectroscopy), Enzyme kinetics. My research interest is in the area of infectious diseases and drug discovery.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 28 July 2020

<https://doi.org/10.5256/f1000research.26502.r67438>

© 2020 Miao Z. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Zhichao Miao

European Bioinformatics Institute (EMBL-EBI), Cambridge, UK

The authors have made some changes according to the other reviewer but have not addressed my comments yet.

I would like to highlight the comments below:

- It is better to support the computational method with experimental clues.
- The authors should show the paired positions in Figure 1.
- Figure 3, it is better to use multiple sequence alignment (MSA) to predict secondary structure and show the co-evolution.
- For 3D structure prediction, it is difficult for an automatic method (e.g., RNAComposer) to handle a structure of more than 150nt without a good secondary structure assignment. This paper tries to use RNAComposer to predict a structure of >350nt. This is beyond the current capability of automatic prediction. From Figure 5, we can find severe atomic clashes, which are not expected in 3D structures. To improve, I suggest to focus on the mutational region, use the MSA-based 2D prediction result to predict 3D structure, and use energy minimization (e.g., Rosetta) to optimize the 3D structure. However, the only information from 3D structure prediction is the feasibility of the 2D structure rather than the functional inferences expected from this paper.
- "RNA Composer" should be "RNAComposer"

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: RNA structure prediction, single cell omics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

Author Response 17 Aug 2020

Tomy Abuzairi, Universitas Indonesia, Depok, Indonesia

It is better to support the computational method with experimental clues.

Response: We thank the reviewer for their interest in the study and their comments. We agree with this comment, but unfortunately there has been some technical limitations, so that we just conduct in silico analysis for a moment.

The authors should show the paired positions in Figure 1.

Response: We have revised Figure 1 into multiple alignment displayed with complete sequences. Little additional method is also written in this version.

For 3D structure prediction, it is difficult for an automatic method (e.g., RNAComposer) to

handle a structure of more than 150nt without a good secondary structure assignment. This paper tries to use RNAComposer to predict a structure of >350nt. This is beyond the current capability of automatic prediction. From Figure 5, we can find severe atomic clashes, which are not expected in 3D structures. To improve, I suggest to focus on the mutational region, use the MSA-based 2D prediction result to predict 3D structure, and use energy minimization (e.g., Rosetta) to optimize the 3D structure. However, the only information from 3D structure prediction is the feasibility of the 2D structure rather than the functional inferences expected from this paper.

Response: Thank you for the suggestion. We already revised the method to predict whole sequences as shown in Figure 5 in the version 2 of this paper. We used the 3dRNA v2.0 from Biomolecular Physics and Modeling Group Xiao Lab (<http://biophy.hust.edu.cn/services.html>) which constructs 3D structure based on 2D structure obtained from RNAfold Server (used in secondary structure prediction method).

"RNA Composer" should be "RNAComposer"

Response: Well noted. We will put more concern to avoid typos.

Competing Interests: No competing interests were disclosed.

Author Response 28 Aug 2020

Tomy Abuzairi, Universitas Indonesia, Depok, Indonesia

It is better to support the computational method with experimental clues.

Response: We thank the reviewer for their interest in the study and their comments. We agree with this comment, but unfortunately there has been some technical limitations, so that we just conduct in silico analysis for a moment.

The authors should show the paired positions in Figure 1.

Response: We have revised Figure 1 into multiple alignment displayed with complete sequences. Little additional method is also written in this version.

For 3D structure prediction, it is difficult for an automatic method (e.g., RNAComposer) to handle a structure of more than 150nt without a good secondary structure assignment. This paper tries to use RNAComposer to predict a structure of >350nt. This is beyond the current capability of automatic prediction. From Figure 5, we can find severe atomic clashes, which are not expected in 3D structures. To improve, I suggest to focus on the mutational region, use the MSA-based 2D prediction result to predict 3D structure, and use energy minimization (e.g., Rosetta) to optimize the 3D structure. However, the only information from 3D structure prediction is the feasibility of the 2D structure rather than the functional inferences expected from this paper.

Response: Thank you for the suggestion. We already revised the method to predict whole sequences as shown in Figure 5 in the version 2 of this paper. We used the 3dRNA v2.0 from Biomolecular Physics and Modeling Group Xiao Lab (<http://biophy.hust.edu.cn/services.html>) which constructs 3D structure based on 2D structure obtained from RNAfold Server (used in secondary structure prediction

method).

"RNA Composer" should be "RNAComposer"

Response: Well noted. We will put more concern to avoid typos.

Competing Interests: No competing interests were disclosed

Version 1

Reviewer Report 11 March 2020

<https://doi.org/10.5256/f1000research.24205.r60477>

© 2020 Audah K. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Kholis Abdurachim Audah

¹ Department of Biomedical Engineering, Swiss German University, Tangerang, Indonesia

² Indonesian Society for Bioinformatics and Biodiversity, Tangerang, Indonesia

³ Indonesian Society for Biochemistry and Molecular Biology, Tangerang, Indonesia

⁴ Academic Research and Community Services, Swiss German University, Tangerang, Indonesia

- It would be very helpful if the authors included the full structure of the wild type and compare it with the two constructed mutants' structure.
- Figure 2 showed a subtle difference in MFE and other thermodynamics characteristics, is there any value that can be presented as a result of statistical analysis or based on best fitting model?
- Figure 3 seems to have significant importance in terms of the structure stability (clinical consequence, for example). Any supporting literature?
- However, the author did not elaborate more on this. For example, IVS1nt5G>T mRNA structures that showed a significant difference (blue-circled). This needs to be discussed more.
- Again, the value needs to be shown here: "In the secondary structure, the higher the number of the larger loops, the more unstable the structure".
- Figure 5 seems to differ significantly between WT and the mutants, but the author stated otherwise? Except this statement: "excluding a small change as presented in the lower left part". This needs to be clarified and explain/discuss any structure-function relationship alteration.

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Partly

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Not applicable

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: My area of research is in protein engineering, protein-protein interactions/ biomolecules interactions/ structure-function relationships which include but not limited to DNA manipulation, DNA mutagenesis, cloning and expression, protein purification (various chromatography methods), protein characterization employing different biophysical methods (CD, MALDI-MS, FLuorescence Spectroscopy), Enzyme kinetics. My research interest is in the area of infectious diseases and drug discovery.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 09 May 2020

Tomy Abuzairi, Universitas Indonesia, Depok, Indonesia

It would be very helpful if the authors included the full structure of the wild type and compare it with the two constructed mutants' structure.

Response: We thank the reviewer for their interest in the study and their comments. We agree with this comment. The secondary structure prediction (Figure 3) already displayed the full structure of wildtype and mutants. Meanwhile, we have revised Figure 5 as full sequence RNA tertiary structure that was predicted using 3dRNA from <http://biophy.hust.edu.cn/services.html>

Figure 2 showed a subtle difference in MFE and other thermodynamics characteristics, is there any value that can be presented as a result of statistical analysis or based on best

fitting model

Response: We show all the value of thermodynamic characteristics in the mountain plot to show the different on the every position paired. By using plot, we can easily compare the structures. For example, we can easily compare wild-type, IVS1nt1G>T, and IVS1nt5G>C for the MFE as shown in Figure 2(b) that shows IVS1nt5G>C has the lowest value compare with wild-type and IVS1nt1G>T. Therefore, the plot is sufficient to show subtle difference in thermodynamic characteristics.

Figure 3 seems to have significant importance in terms of the structure stability (clinical consequence, for example). Any supporting literature?

Response: We agree with this and have incorporated your suggestion throughout the manuscript.

“The principle of this methods is to create the secondary structure with minimal base pair distance to all other secondary structures in the Boltzmann ensemble⁹.”

“Both IVS1nt5G>C and IVS1nt1G>T give implications to the clinical condition due to protein instability. There are some disruptions in producing beta globin chains by these mutations, hence the beta thalassemia major (BTM) patients produce beta globin chain in a limited number or even not at all. This mechanism results in unbalanced alpha and beta globin chains in haemoglobin that expedite the haemolysis and ineffective erythropoiesis contributing to reduced production of mature red blood cells²⁸. Thus, the BTM patients requires lifelong blood transfusions and commonly suffer complications, such as endocrine complications and cardiac complications²⁹”

However, the author did not elaborate more on this. For example, IVS1nt5G>C mRNA structures that showed a significant difference (blue-circled). This needs to be discussed more.

Response: We have updated our discussion for general alteration mainly happens in IVS1nt5G>C.

“The blue box in Figure 4 highlighted the main alteration formed within the three sequences. However, the structure of IVS1nt5G>C seems to be the best predicted according to the red color in the base-pairing probability.”

Again, the value needs to be shown here: “In the secondary structure, the higher the number of the larger loops, the more unstable the structure”.

Response: The prediction tool does not provide loop counting of the secondary structure, hence we omit the sentence and create additional sentences to support our results.

“In general, the tertiary structure of each sequence was unique. The point mutation gave tremendous folding changes in the 3D model. Once the base shifted, base pairing mechanism was influenced massively.”

Figure 5 seems to differ significantly between WT and the mutants, but the author stated

otherwise? Except this statement: "excluding a small change as presented in the lower left part". This needs to be clarified and explain/discuss any structure-function relationship alteration.

Response: This figure has been revised totally into full sequence 3D structure, while the structure-function relation already discussed, e.g. "TVS1nt5G>C was predicted to be a silencer motif new site, intronic identifying elements and exonic splicing silencer (ESS) site were broken".

Competing Interests: The authors declare that we have no competed of interests

Reviewer Report 12 February 2020

<https://doi.org/10.5256/f1000research.24205.r59162>

© 2020 Miao Z. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Zhichao Miao

European Bioinformatics Institute (EMBL-EBI), Cambridge, UK

Nur and colleagues report a computational study of the beta-globin gene mRNA structure based on secondary and tertiary structure prediction. However, the prediction is not reliable enough to fully support the conclusion. Therefore, I suggest to keep the paper in record but not use it as a conclusion. Major and minor comments are listed here:

Major comments:

- The paper is fully based on computational prediction. However, both the prediction approaches are not yet accurate enough to draw a conclusion without experimental validation.
- For secondary structure prediction, MFE (RNAfold) only consider the cis-Watson-Crick base pair situation without pseudoknots. However, different sequences that can be aligned together are expected to have the same secondary structure. Using the sequence co-evolution information may further improve secondary structure prediction accuracy. That is why ViennaRNA has RNAalifold to predict 2D structure from sequence alignment. Considering the modeling only includes point mutations, one cannot rule out the possibility that both wild type and mutants use the same structure but the mutation point use a base pair conformation different from cis-Watson-Crick (e.g., WC-sugar). In this case, the main topology of the structure should stay the same. The authors show the mutation in Figure 1, but do not show the positions paired with the mutational position.
- According to the 2D structures predicted in Figure 3, the proposed structures are not likely to fold in 3D. To improve, I suggest: 1) Predict the existence of Pseudoknots; 2) Check if the

mutants can use the same 2D structure according to the base pair; 3) Use multiple sequence alignment (MSA) to predict secondary structure and show the co-evolution.

- For 3D structure prediction, it is difficult for an automatic method (e.g., RNAComposer) to handle a structure of more than 150nt without a good secondary structure assignment. This paper tries to use RNAComposer to predict a structure of >350nt. This is beyond the current capability of automatic prediction. From Figure 5, we can find severe atomic clashes, which are not expected in 3D structures. To improve, I suggest to focus on the mutational region, use the MSA-based 2D prediction result to predict 3D structure, and use energy minimization (e.g., Rosetta) to optimize the 3D structure. However, the only information from 3D structure prediction is the feasibility of the 2D structure rather than the functional inferences expected from this paper.

Minor comments:

1. The free energy calculation is not part of this paper. The authors should just cite the related publications rather introducing the ideas.
2. The 3D structure prediction method should be spelt as "RNAComposer".
3. Figure 1B should include the base pair information (2D structure) of the alignment.
4. The y-axis of Figure 2 needs more explanation.
5. Figure 4 needs more explanation. Its meaning is not clear.
6. It could be good to share the predicted structure information to increase the reproducibility.

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

No

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Not applicable

Are all the source data underlying the results available to ensure full reproducibility?

Partly

Are the conclusions drawn adequately supported by the results?

No

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: RNA structure prediction, single cell omics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research