



THE COMPUTATIONAL DESIGN OF THE SPIKE GLYCOPROTEIN GENE siRNA OF SARS-COV-2

<u>ARLI ADITYA PARIKESIT¹</u>, Arif N.M Ansori², Viol D Kharisma³

¹Bioinformatics Department INDONESIA INTERNATIONAL INSTITUTE FOR LIFE SCIENCES ²Faculty of Veterinary Medicine, Universitas Airlangga, Surabaya, Indonesia

³Computational Virology and Complexity Sciences Research Unit, Division of Molecular Biology and Genetics, Generasi Biologi Indonesia Foundation,



Structural Bioinformatics in Drug Discovery

[Arli Aditya Parikesit] [arli.parikesit@i3l.ac.id]

BACKGROUND

Drug Discovery today are facing a serious challenge because of the increased cost and enormous amount of time taken to discover a new drug, and also because of rigorous competition amongst different pharmaceutical companies.

Drug Discovery & Development



Technology is impacting this process



CADD methods (COMPUTER AIDED DRUG DESIGN)





- To examine, evaluate and compare complex molecular structure
- To modify structure and assess geometric and energetic consequences of such modifications
- To perform conformational analysis
- To build macromolecules

CADD Methods ...

- To dock small molecules into macromolecules
- To observe the dynamics between the ligands and the macromolecules
- To map pharmacophore group or ligands
- To analyze relationship between chemical structure and biological activity
- To predict activity of compounds/analogues before the synthesis



MOTIVASI

- Berdasarkan Teori Dogma Sentral, Protein adalah regulator terpenting.
- Berarti, hanya sekitar 5% dari Genome Manusia yang penting, sisanya adalah 'sampah'
- Skema ini tidak dapat atasi tantangan biomedis termutakhir

John Mattick's Scheme



'Dunia RNA'

- Aliran Informasi Mattick menunjukkan berbagai tipe RNA berperan dalam regulasi gen
- Mereka adalah: snoRNA, longNCRNA, miRNA, dll
- 'last universal common ancestor' (LUCA) adalah virus-like dengan inti RNA
- Non Coding RNA (ncRNA) berperan dalam regulasi gen walau tidak ditranslasi ke protein

Contoh Struktur 2D RNA



Figure 10.1 The RNA secondary structure of signal recognition particle (SRP) RNA from the dog, Canis familiaris.

INFORMATION PLOT RNA





Figure 10.6 A mutual information plot of a tRNA alignment (top) shows four strong diagonals of covarying positions, corresponding to the four stems of the tRNA cloverleaf structure (bottom; the secondary structure of yeast phenylalanine tRNA is shown). Dashed lines indicate some of the additional tertiary contacts observed in the yeast tRNA-Phe crystal structure. Some of these tertiary contacts produce correlated pairs which can be seen weakly in the mutual information plot.

PERHITUNGAN ENERGI BEBAS RNA



overall $\Delta G = -4.6$ kcal/mol

Figure 10.10 An example ΔG calculation for an RNA stem loop (the wild type R17 coat protein binding site).

miRNA Biogenesis



Komputasi RNA

- 'Big Data' ncRNA seyogyanya diolah oleh praktisi bioinformatika.
- Tools Bioinformatika dapat digunakan untuk itu
- RNA Vienna Package adalah paket komputasi ncRNA
- Mengapa harus RNA Vienna? Karena sudah banyak dipublish pada jurnal internasional bereputasi, digunakan sebagai benchmark protokol komputasi



DYNAMIC PROGRAMMING (2)

a Recursive definition of the best score for a sub-sequence *i*, *j* looks at four possibilities:



b Dynamic programming algorithm for all sub-sequences *i*,*j*, from smallest to largest:



DYNAMIC PROGRAMMING (3)

3

Computational Chemistry with RNA Secondary Structures



Figure 1. Secondary structure of phenylalanine-tRNA from yeast as conventional drawing and in circular representation. The chords in the circular representation must not cross in secondary structure graphs.



Figure 2. Secondary structure elements that form the basis of the energy model for nucleic acids.

<u>https://ul.qucosa.de/api/qucosa%3A32602/attachment/ATT-0/</u>

NUSSINOV ALGORITHM (1)



Figure 10.1. The traceback path produced by the Nussinov folding algorithm for the sequence *GGGAAAUCC*. The scores on the optimal path are indicated circles. The starting point of the path is located at the top right-hand corner. The secondary structure (1) associated with the optimal path is shown on the right.

NUSSINOV ALGORITHM (2)



Figure 10.2. Eight alternative secondary structures (with three base pairs) of sequence GGGAAAUCC.

Representasi Struktur 2D RNA



Figure 1.5. Representations of RNA secondary structures. A) Circle plot. B) Conventional secondary structure plot. C) Mountain plot. D) Dot plot. E) Dot/bracket string notation. All plots represent the same structure, its the purine riboswitch (Rfam RF00167). Adopted from [Hofacker & Stadler 2007].

HOMOLOGY MODELING



factor/figures?lo=1&utm_source=google&utm_medium=organic

Vienna RNA Package

- Developed by Ivo Hofacker et al from University of Vienna, Austria, and Peter
 Stadler et al from University of Leipzig, Germany.
- Easy to use, and has web interface
- Offline tools is available in Linux platform -> It's cheap!

Vienna RNA Package Interface

	ViennaRNA W Institute for Theo	pretical Chemistry		•	
Structure prediction Folding Kinetics	Sequence Design	ncRNA Prediction	Genome Wide Screening	Other	
You are here: / RNA			Fo	nt size: 🙏 🔔 🔺	
The ViennaRNA Web Services					
This server provides programs, web services, ar offerings from our group see the main TBI web Web Servers					
Thermodynamic Structure Prediction		ncRNA Prediction			
 RNAfold server predicts minimum free energy structures a probabilities from single RNA or DNA sequer RNAalifold server 	ices.	structures in multip • RNAz server	etecting evolutionarily consen e sequence alignments.		
 predicts consensus secondary structures fro several related RNA or DNA sequences. You r alignment. RNAeval server provides a detailed thermodynamic descrip sequence/structure pair. RNAcofold server allows you to predict the secondary structu RNAup server 	need to upload an	evolutionarily conse sequence alignment Bcheck predicts rnpB gen RNAstrand server	es. Ict the reading direction of eve	s in multiple	
 several related RNA or DNA sequences. You ralignment. RNAeval server provides a detailed thermodynamic descriptsequence/structure pair. RNAcofold server allows you to predict the secondary structure 	need to upload an otion of a re of a dimer.	evolutionarily conse sequence alignment Bcheck predicts rnpB gen RNAstrand server allows you to pred	ved RNA secondary structure s. es. ct the reading direction of eve	s in multiple	

Figure 6: Vienna RNA Package Interface (<u>http://rna.tbi.univie.ac.at/</u>)



Ifold RNA is an online software that uses discrete molecular dynamics method to model the RNA 3D structure

Fig. 1. The structure of M-box riboswitch predicted by iFoldRNA v2 (sand color) is superimposed on the top of the crystal structure (PDB ID: 3pdr) (blue). RMSD between the predicted and the crystal structures is 7.7 Å. *P*-value, showing statistical significance of the prediction (Hajdin *et al.*, 2010), is less than 10^{-6} . RMSD was calculated using phosphate atoms only. INF = 0.725 (Parisien *et al.*, 2009). Experimental HRP data and base-pairing information were used (Ding *et al.*, 2012) (Color version of this figure is available at *Bioinformatics* online.)



HNADOCK



HNADDOCK is a web server that implement specific DNA/RNA – DNA/RNA interaction function to its docking method

Figure 5. Comparison between the crystal structure (blue and red) and HNADOCK server prediction (green and yellow) for two RNA–RNA docking examples: (A) structure input (target code: 1KD5; ranked #1, IRMSD = 1.98 Å); (B) sequence input (target code: 1KIS; ranked #4, IRMSD = 2.39 Å),

Pipeline Predictor Short ncRNA



Figure 2.2. Flow chart for the annotation of short structured ncRNAs. The figure is an updated version from the published one in Bompfünewerer Consortium et al. [2007]. RNA families in the sense of the Rfam database are predominantly defined by sequence homology, while RNA classes are defined via functional and/or structural similarities that may or may not be the consequence of common ancestry. Computational RNA prediction is the key to a pile of subsequent analyses which coherently contribute to accurate RNA annotation and, in the long run, steadily improve our understandings of vitally important RNA-mediated cellular processes.

Clusters of Complex Structure



Figure 2.13. Cluster of complex structures. Structure-based clustering of RNAz hits with evidence for transcription by Pol III identifies a group of Y-shaped, potentially related putative ncRNAs. Abbreviations: N: number of sequences in cluster. MPI: mean pairwise identity of multiple alignment. SCI: structure conservation index.

Motivation: the World's Pandemic Time Cohort



https://www.sciencedirect.com/science/article/pii/S2319417020300445



COVID-19 Publications

from Our Group



 Parikesit, A. A., & Nurdiansyah, R. (2020). Drug Repurposing Option for COVID-19 with Structural Bioinformatics of Chemical Interactions Approach. Cermin Dunia Kedokteran, 47(3), 222–226. Retrieved from

http://kalbemed.com/DesktopModules/EasyDNNNews/DocumentDownload.ashx?portalid=0&mo duleid=471&articleid=1274&documentid=1804

- **Parikesit, A. A**. (2020). Protein Domain Annotations of the SARS-CoV-2 Proteomics as a Blue-Print for Mapping the Features for Drug and Vaccine Designs. Jurnal Matematika dan Sains ITB. Bandung https://jms.fmipa.itb.ac.id/index.php/jms/article/view/823
- Muhammad Ansori, A. N., Dhea Kharisma, V., Sabilil Muttaqin, S., Antonius, Y., & Parikesit, A. A. (2020). Genetic Variant of SARS-CoV-2 Isolates in Indonesia: Spike Glycoprotein Gene. Journal of Pure and Applied Microbiology, 14(suppl 1), 971–978. <u>https://doi.org/10.22207/JPAM.14.SPL1.35</u>
- **Parikesit, A. A.,** & Nurdiansyah, R. (2020). The Predicted Structure for the Anti-Sense siRNA of the RNA Polymerase Enzyme (RdRp) gene of the SARS-CoV-2. Berita Biologi LIPI, 19(1), 97-108. https://e-journal.biologi.lipi.go.id/index.php/berita_biologi/article/view/3849
- **Parikesit, A. A.**, Ratnasari, N. R. P., & Anurogo, D. (2020). Application of Artificial Intelligence-Based Computation in the Health Sciences to Ward off the COVID-19 Pandemic. International Journal of Human and Health Sciences (IJHHS), 5(2), 177. <u>https://doi.org/10.31344/ijhhs.v5i2.256</u>
- Adisurja, G. P & Parikesit, A. A. Virtual Screening of the Flavonoids Compounds with the SARS-CoV-2 3C-like Protease as the Lead Compounds for the COVID-19 (*in press* at Coronaviruses, Benthamsciences) https://www.eurekaselect.com/191637/article



BACKGROUND



- COVID-19 disease is caused by SARS CoV-2
- The designated 'ground zero' is Wuhan, China.
- Current standing of COVID-19 pandemic is more than 481 millions Infected patients, and more than 6 million death. Mostly in the US.
 Significant infection and death in Indonesia (per 1th of April 2022, WHO COVID-19 Dashboard).

SARS-COV2 is...

- RNA virus, generally it will have faster mutation rate than DNA virus or bacteria
- The capability of zoonotic infection make it easier to mutate as well
- Zoonotic? Meaning they are transmitted between animals and people.



Figure 1. Schematic diagram of coronavirus structure.

Scheme of SARS-CoV-2 Infection Cycle



https://www.oatext.com/pharmacological-approaches-to-the-treatment-of-covid-19-

patients.php#gsc.tab=0

SARS-CoV-2 Genomics Sequences from Indonesian Patients



SARS-CoV-2 Genome



The SARS-CoV-2 has a ~29.9 kilobase positive-sense RNA genome that contains as many as 29 open reading frames. Though the exact number of functional proteins remains to be established, there are at least 16 nonstructural proteins (nsp), four structural proteins, and at least six or seven accessory proteins.

https://www.genetex.com/MarketingMaterial/Index/SARS-CoV-

2 Genome and Proteome
S Gene

- Infectivity or virulence of SARS-CoV-2 virus is mainly catered by its spike protein in the viral surface
- It plays important role for viral penetration to the host cell, by facilitating attachment to the ACE2 receptor
- Thus, it is logical in the sense of rational drug design that the SARS-CoV-2 spike protein should be inhibited to ward off the viral infection

Objective of this research

 In this regard, combining transcriptomics and **CADD** approaches could be a viable solution to design SARS-CoV-2 drug, especially to provide anti-sense inhibitor to the mRNA expression of a gene. Thus, the objective of this research is to design transcriptomics-based drug candidate with bioinformatics pipeline to block the expression the mRNA of the SARS-CoV-2 S gene with siRNA.

2D dan 3D RNA PREDICTION METHOD

 The seach for S gene sequences from various localities was conducted with NCBI website:

https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/sars-cov-2

- The multiple sequence alignment using MUSCLE algorithms was applied to extract the sequence of the conserve region. All the S gene sequences were employed for constructing the phylogenetic tree with MEGAX
- The ncRNA or mRNA FASTA data were uploaded to the Vienna RNA package at http://rna.tbi.univie.ac.at. The respective tools were employed sequentially: *RNAfold, Barrier server, RNAup* dan *RNAxs*. They are useful to analyze the 2D annotation within the thermodynamics and kinetics sphere, and to determine the structure and the function of the ncRNA/mRNA.

2D dan 3D RNA PREDICTION METHOD

 The iFOLDRNA version 2 server (https://dokhlab.med.psu.edu/ifoldrna/) was utilized as well for de novo 3D structure prediction. In order to observe the chemical interaction of the siRNA and the mRNA, the HNAdock application for transcriptomics lead was employed. (http:// huanglab.phys.hust.edu.cn/hnadock/). Lastly, the docking result was visualized using the UCSF Chimera software (https:// www.cgl.ucsf.edu/chimera/)

Result of the siRNA design annotation



siRNA	1
Worst Rank	33
Position	187
Access 8nt	0.3012
Access 16nt	0.1726
Assymetry (S)	0.7500
Assymetry (E)	0.6638
Self Folding	1.0000
Free End	1.0000
Target Seq.	ACTTTCC TTTACAATCATA
siRNA Seq.	TATGATT GTAAAGGAAAGT
BLAST	NCBI BLAST

Figure 2: The RNAxs output for the S gene SARS-CoV-2 siRNA



The 3D structures of the conserved mRNA *S* gene and the siRNA



Figure 5: the 3D visualization of the S gene SARS-CoV-2 a) mRNA b) siRNA

The Molecular Docking Visualization of the mRNA and siRNA of the S gene



Figure 6: The docking result of the SARS-CoV-2 S gene mRNA and siRNA complex a) Complex visualization with HNADOCK b) Complex visualization with UCSF Chimera

b)

a)

Transcriptomics-based Drugs (Then)

Table 1

Anti-miRNA and siRNA/shRNA Therapeutics in Clinical Trials

Company	Drug	Delivery route	Target	Vehicle	Disease	Phase	Status
Santaris	SPC3649 (LNA)	SC	miR-122	Naked LNA	HCV	IIa	Ongoing
Opko Health	Bevasiranib	IVT	VEGF	Naked siRNA	AMD/DME	Ш	Terminated
Allergan/Sirna	AGN-745	IVT	VEGF-R1	Naked siRNA	AMD	II	Terminated
Quark/Pfizer	PF-655	IVT	RTP801	Naked siRNA	AMD/DME	Π	Completed
Quark Pharma	QPI- 1007	IVT	Caspase 2	Naked siRNA	NAION	Ι	Ongoing
TransDerm/IPCC	TD101	Intralesional injection	KRT6A(N171K)	Naked siRNA	Pachyonychia Congenita	Ib	Completed
Sylentis	SYL040012	Ophthalmic drops	ADRB2	Naked siRNA	Intraocular Pressure	п	Ongoing
Sylentis	SYL1001	Ophthalmic drops	TRPV1	Naked siRNA	Dry eye syndrome	Ι	Ongoing
ZaBeCor	ExcellairTM	Inhalation	Syk kinase	unknown	Asthma	П	Ongoing
Alnylam/Cubist	ALN-RSV01	Nebulization or intransal	RSV Nucleocapsid	Naked siRNA	RSV	IIb	Ongoing
Marina Biotech	CEQ508	Oral	Beta catenin	tkRNAi in E. Coli	FAP/ colon cancer	Ι	Ongoing
Silenseed Ltd	siG12D LODER	EUS biopsy needle	KRASG12D	LODER polymer	PDAC	Ι	Ongoing
Tekmira	ТКМ-АроВ	IV	Аро В	SNALP	Hypercholesterolemia	Ι	Terminated
Tekmira	TKM-PLK1	IV	PLK1	SNALP	Solid tumors	Ι	Ongoing
Alnylam/Tekmira	ALN-VSP02	IV	KSP and VEGF	SNALP	Solid tumors	Ι	Completed
Alnylam	ALN-TTR01	IV	TTR	SNALP	TTR-mediated amyloidosis (ATTR)	Ι	Ongoing
University Duisburg	Bcr-Abl siRNA	IV	Bcr-Abl	Anionic liposome	CML	Ι	Completed
Silence Therapeutics	Atu027	IV	PKN3	siRNA-lipoplex	Advanced solid cancer	Ι	Ongoing
Quark Pharma	I5NP	IV	P53	Naked siRNA	AKI and DGF	Π	Ongoing
Calando Pharma	CALAA-01	IV	RRM2	Cyclodextrin nanoparticle, TF, and PEG	Solid tumors	I	Ongoing
Gradalis Inc.	FANG vaccine	<i>Ex vivo</i> IV	Furin and GM-CSF	Electroporation	Solid tumors	П	Ongoing
Duke University	iPsiRNA	Ex vivo intradermal injection	LMP2, LMP7, MECL1	Transfection	Metastatic melanoma	Ι	Ongoing
City of Hope/Benitec	Tat/Rev shRNA	Ex vivo transplant	HIV Tat and Rev	Lentivirus	HIV	0	Ongoing

(Burnet and Rossi, 2012)

Transcriptomics-based Drugs (Now)

miRNA-Based Therapeutics							
Company	Name	Therapeutic Agent	Delivery System	Target Disease	Stage in Drug Development Pipeline		
Santaris Pharma/Roche	Miravirsen	AntimiR-122	LNA antagomiR	Hepatitis C; Chronic hepatitis C	Phase II clinical trials (NCT02452814; NCT2508090)		
Regulus Therapeutics	RG-101	AntimiR-122	GaLNAc-conjugated antagomiR	Chronic hepatitis C	Phase II clinical trials (discontinued)		
	RG-125	AntimiR-103/107	GaLNAc-conjugated antagomiR	Diabetic non-alcoholic steatohepatitis	Phase II (discontinued)		
	RG-012	AntimiR-21	NA	Hereditary nephritis	Phase II (NCT02855268)		
	RGLS4326	AntimiR-17	NA	Autosomal dominant polycystic kidney disease	Phase I (on hold)		
miRagen Therapeutics	MRG-106	AntimiR-155	LNA-modified antisense inhibitor	CTCL mycosis fungoides subtype; CLL; DLBCL; ATLL	Phase II (NCT03713320; NCT03837457); Phase I (NCT02580552)		
	MRG-107	AntimiR-155	NA	ALS; cardiac disorders; retinal disorders	Pre-Clinical		
-	MRG-110	AntimiR-92	LNA antagomiR	Wounds	Phase I (NCT03603431)		
	MRG-201	miR-29 mimic	Cholesterol-conjugated miRNA duplex	Keloid; fibrosis	Phase II (NCT03601052); Phase I (NCT02603224)		
EnGeneIC	MesomiR-1	miR-16 mimic	EnGeneIC Dream Vector	Malignant pleural mesothelioma; non-small-cell lung cancer	Phase I (NCT02369198)		
Mirna Therapeutics Inc.	MRX-34	miR-34 mimic	dsRNA liposomal nanoparticle	Solid tumours; haematological malignancies	Phase 1 (terminated)		

Table 1. miRNA-based therapeutics in clinical trials.

LNA: locked nucleic acid; GaLNAc: N-acetylgalactosamine; CTCL: cutaneous T cell lymphoma; ATLL: adult T-cell leukaemia lymphoma; CLL: chronic lymphocytic leukaemia; DLBCL: diffuse large B-cell lymphoma [activated B-cell (ABC) subtype]; ALS: amyloid lateral sclerosis.

(Bajan and Hutvagner, 2020)

Conclusion

 It is concluded that the both 2D and 3D designs of the siRNA lead and mRNA biomarker for S gene could be elucidated with in silico-based approach. Thus, the docking result indicates that there is a possibility that the docking between both the siRNA and mRNA biomarkers could happen in the computational manner.



THANKS GOES TO

- THE RESEARCH AND COMMUNITY SERVICE DEPARTMENT (LPPM) OF THE INDONESIA INTERNATIONAL INSTITUTE FOR LIFE SCIENCES
- NVIDIA CORP. GPU GRANT
- Indonesian Society for Bioinformatics and Biodiversity (ISBB/MABBI)



References

- Ahmed, Sabeena, Mohammad Mahbubul Karim, Allen G. Ross, Mohammad Sharif Hossain, John D. Clemens, Mariya Kibtiya Sumiya, Ching Swe Phru, et al. 2021. "A Five-Day Course of Ivermectin for the Treatment of COVID-19 May Reduce the Duration of Illness." International Journal of Infectious Diseases 103, no. February: 214–16. <u>https://doi.org/10.1016/j.ijid.2020.11.191.</u>
- Bajan, Sarah, and Gyorgy Hutvagner. 2020. "RNA-Based Therapeutics: From Antisense Oligonucleotides to MiRNAs." *Cells* 9, no. 1: 137. https://doi.org/10.3390/cells9010137.
- Burnett, John C, and John J Rossi. "RNA-Based Therapeutics: Current Progress and Future Prospects." *Chemistry & Biology* 19, no. 1 (January 27, 2012): 60–71. doi:10.1016/j.chembiol.2011.12.008.
- Callaway, E. (2020). The race for coronavirus vaccines: a graphical guide. Nature, 580(7805), 576–577. <u>https://doi.org/10.1038/d41586-020-01221-y</u>
- Caly, Leon, Julian D. Druce, Mike G. Catton, David A. Jans, and Kylie M. Wagstaff. 2020. "The FDA-Approved Drug Ivermectin Inhibits the Replication of SARS-CoV-2 in Vitro." *Antiviral Research* 178, no. June: 104787. https://doi.org/10.1016/j.antiviral.2020.104787.
- Durbin, R., Eddy, S. R., Krogh, A., & Mitchison, G. (1998). Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids. Analysis. <u>https://doi.org/10.1017/CB09780511790492</u>
- Gao J, Tian Z and Yang X 2020 Breakthrough: Chloroquine phosphate has shown apparent efficacy in treatment of COVID-19 associated pneumonia in clinical studies. Biosci. Trends 2020.01047
- Lu H 2020 Drug treatment options for the 2019-new coronavirus (2019-nCoV). Biosci. Trends 2020.01020
- Parikesit A.2020. How Bioinformatics bolster the fight against the coronavirus. <u>https://www.thejakartapost.com/academia/2020/02/14/how-bioinformatics-bolster-the-fight-against-the-coronavirus.html</u>
- Rose, Dominic. "The Long and the Short of Computational ncRNA Prediction," January 6, 2010. http://www.qucosa.de/recherche/frontdoor/?tx_slubopus4frontend%5Bid%5D=6215.
- Zhang, D., Wu, K., Zhang, X., Deng, S., & Peng, B. (2020). In silico screening of Chinese herbal medicines with the potential to directly inhibit 2019 novel coronavirus. Journal of Integrative Medicine, 18(2), 152–158. <u>https://doi.org/10.1016/j.joim.2020.02.005</u>
- Mattick, J. S., & Rinn, J. L. (2015). Discovery and annotation of long noncoding RNAs. Nature Structural & Molecular Biology, 22(1), 5–7. https://doi.org/10.1038/nsmb.2942Mattick, J. S. (2005). The functional genomics of noncoding RNA. Science, 309(5740), 1527–1528. https://doi.org/10.1126/science.1117806
- Mattick, J. S. (2005). The functional genomics of noncoding RNA. Science, 309(5740), 1527–1528. https://doi.org/10.1126/science.1117806

Further Correspondence?

- Email : <u>arli.parikesit@i3l.ac.id</u>
- Twitter Account
- : https://twitter.com/arli_ap
- Telegram
 Account
- : http://www.telegram.me/arliap

Line ID

: @arliap

