Novel and Simple Simulation Method to Design and Development of Antisense Template

Devendra Vilas Deo*[1], Dr. Nawaj Shaikh [2]

1. Medgenome Labs Inc. Ltd, Bangalore, India; Datar Cancer Genetics Limited, Nashik, India.

2. Society for Health Allied Research and Education, India (SHARE-INDIA).

Corresponding Author: Devendra Vilas Deo, Department of Medical Oncology and Hematology, Amrita Institute of Medical Sciences and Research Centre, Amrita Vishwa Vidyapeetham, Kochi - 682041, Kerala, India.

Abstract

Antisense technology is emerging as potential therapeutics against lethal infections. Antisense-mRNA complex inhibits the protein translation of pathogens and hence it is used for treatment. To make these specific antisenses, currently, there is various online tools/ software available. Based on the previous online tools for antisense and literature, the difficulties for designing antisense templates, finding high conserved regions from large number of long sequences, by taking all those factors in consideration, we proposed new innovative offline target simulation methods i.e. Deletion of the unwanted region from viral sequence alignment (DURVA) and Most frequent region (MFR) for designing and developing antisense template from a large number of a long sequence or genomic data.

To evaluate our offline tool, we used the long genomic sequence of the current pandemic virus-SARS-CoV-2 for simulation. Initially, we hypothesized that DURVA-MFR would find stable regions from large annotated sequencing data. As per Ho et.al. observation for antisense designing and development, we designed a couple of algorithms and python scripts to process the data of approximately 30kbp sequence length and 1Gb file size in a short turnaround time. The steps involved were as: 1) Simplifying the whole genome sequence in a single line; 2) Deletion of the unwanted region from Virus sequence alignment (DURVA); 3) Identification of the Most frequent antisense target region (MFR) and 4) Designing and development of antisense template. This simulation method is identifying the most frequent regions between 20-30bp long, with GC count≥10 compare to other online tools. The targets identified in our study were highly identical with a large population and similar with a high number of remaining sequences. In addition, designed antisense sequences were stable and each sequence is having tighter binding with targets. After studying each parameters, here we suggested that our offline method would be one of the first which helpful for finding the best antisense against all present and upcoming lethal infections. The initial design of this logic was published in Indian Patent Office Journal No. 08/2021 with Application number 202121005964A.

Keywords- DURVA-MFR, Antisense/ASO, Coronaviridae, NCBIvirus.

Citation: Devendra Vilas Deo, "Novel and Simple Simulation Method to Design and Development of Antisense Template"
MAR Oncology Volume 5 Issue 5
www.medicalandresearch.com (pg. 2)

Simple Summary

The antisense development is state of the art for modern therapeutics. There are a number of online softwares and open sources for designing of antisense templates. But all other tools did not consider the frequency as a major factor in designing antisense. Secondly; all sources except our simulation approach, does not process large size files or long sequences. Therefore, we designed an offline innovative simulation method that deletes the unwanted region from targeted sequences and stores the data which are fulfilled antisense criteria. This article explained all information about how our new approach is best for designing antisense template against SARS-CoV-2 and may applicable to many other lethal infectious viruses etc.

Introduction

Modern medicine is curiously looking toward antisense technology. Antisense is a short template around 20-30bp and it binds complementary to host mRNA. Antisense-mRNA complex inhibits protein translation. Thus; antisense involves in the regulation of gene expressions. This is a modern approach to inhibit viral protein translation in case of viral infection.(1-4)

We consider new and popular pandemic virus now days for our study evaluation and simulation purpose and further leads to evaluate the efficiency and accuracy of our self- derived methodology for long and complex genomic sequences which was not easy to define conserved and stable targets because of long and complex genomic sequence structure. The Coronavirus belongs to family Coronaviridae and genera Betacoronavirus. This is positive single-stranded genomic RNA virus about 29Kbp long and act as mRNA within the host cell. Generally, the strains of the virus attacks on respiratory and other systems and it binds to Angiotensin- converting enzyme 2 (ACE-II enzyme) receptors present on respiratory system cells and multiply within the host cells. This persistent infection may cause lungs damage.(5) This second reason was to choose SARS-CoV-2 in addition with assess the accuracy and effectiveness of our self- designed method.

In case of COVID-19 as describe in Figure 1, either antisense would bind to genomic RNA templates or discontinuous transcripts of virus which would inhibit the viral protein translation and virions reproduction ultimately. In order to regulate target gene expression, antisense ought to reach disease-associated tissues and cross cell membranes. This is in part facilitated by the manipulation of their chemical structure, which makes oligonucleotides also stronger and safer with a lower chance to have side effects on host immune

Citation: Devendra Vilas Deo, "Novel and Simple Simulation Method to Design and Development of Antisense Template"
MAR Oncology Volume 5 Issue 5
www.medicalandresearch.com (pg. 3)

system(1–4). Apart from this, the most important factor for RNA virus is that we need highly stable and frequent region where your antisense would bind(6-8). Based on this idea, we decided to count frequency of each plausible target via most frequent region (MFR). High the MFR, high the chances of complementory binding.
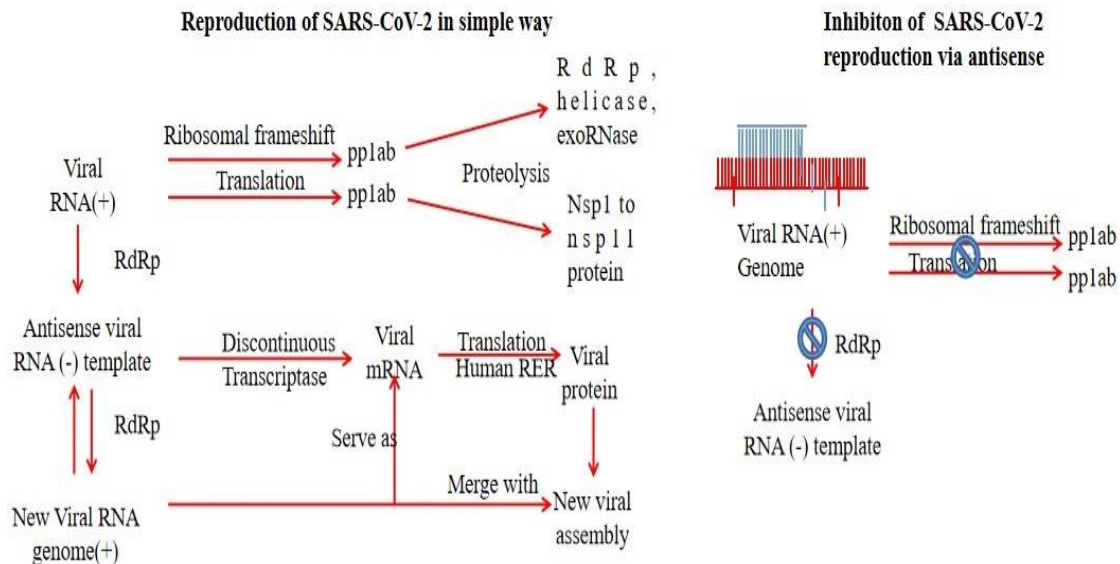

Figure 1 : The schematic diagram is showing reproduction of SARS-CoV-2 in simple way(Left side) and how antisense could be helpful to inhibit the SARS-CoV-2 reproduction via antisense blocking. Antisense binds complementarily to inhibit translation and negative template synthesis via RdRp protein.

Based on literature survey, the number of algorithm or method were available to find conserved region or homology. For example, Clustal Omega multiple sequence alignment tool(9,10) and NCBI BLAST(11). This tool mostly uses for finding conserved regions. But this tool has certain limitation i.e. it can process only 4000 sequences or 4MB size file(9,10). The number of open source software around us to design antisense and siRNA for all such as OptiRNAi(12), siExplorer(13), DSIR(14), i-Score(8), MysiRNA-Designer(15), siDirect(16), siRNA Selection Server(17), siRNA-Finder(18), PFRED(19) and OligoWalk(20). These all are web-based design tools. These algorithms predict and aid in designing/development of best antisense but not considers frequency as major finding criteria which we used in addition to design DURVA and MFR method along with offline mode. Hence, we hypothesized that this new approach would be helpful to find out highly conserved region from SARS-CoV-2 RNA sequences and complementary antisense. To prove this hypothesis, we designed a script of DURVA and

Citation: Devendra Vilas Deo, "Novel and Simple Simulation Method to Design and Development of Antisense Template"
MAR Oncology Volume 5 Issue 5
www.medicalandresearch.com (pg. 4)

MFR method using Python programming language. Our first objective was to find conserved regions which are meeting Ho et. al(4) criteria for antisense designing. This antisense used to evaluate for ASO target sites and finally to compare our tool with other existing methods for performance.

Based on our study, we understood that DURVA - MFR can process more than ten thousand (nearly 30kbps) long sequences and can calculate most frequent region from 1Gb text file size, which is to be most innovative and first time. At the end of our study, we concluded that our defined ASO-target sites were more than 88% identical with SARS-Cov-2 data. This could be helpful to understand how our method will be useful in future to find most frequents and highly conserved target site from large sample size and implementation in other lethal viral infection too.

Methods and Tools

Based on our short piolet study for finding most frequent region, we modified our python transcripts (Supplementary data of Python scripts are not given here to) process large sample size in short turnaround time. These transcripts is very easy and can work on Linux and Windows operating systems. To design and run the python scripts we used IDLE Python3.9 version on Microsoft windows11. The DURVA MFR method works as followed:

Simplifying whole genome sequence in single line

To find most frequent target region, we assembled all our Coronavirus related Sequence.FASTA data in single file. To check the efficiency of python script, data file was converted in whole genome in single line format which was downloaded sequences from NCBI Virus database (https://www.ncbi.nlm.nih.gov/). We had downloaded total 13875 SARS-CoV-2 RNA sequences from NCBI Virus database(21) (https://www.ncbi.nlm.nih.gov/). Out of the total downloaded sequences, 4144 were Asian sequences, 2024 were European sequences, 2266 were African sequences, 2427 were Kiwis (New Zealand) samples, 1185 were North American samples and 1829 were South American samples. We removed newline character (\n) from all these samples for making DURVA method simple. Newline character (\n) works same as 'Enter' in our text file.

Citation: Devendra Vilas Deo, "Novel and Simple Simulation Method to Design and Development of Antisense Template"
MAR Oncology Volume 5 Issue 5
www.medicalandresearch.com (pg. 5)

Deletion of unwanted region from virus sequence alignment (DURVA)

In this step, we studied total 13875 SARS-CoV-2 RNA sequences (https://www.ncbi.nlm.nih.gov/) to delete unwanted region from all these sequences according to Ho et. al(4) antisense guidelines.

Most frequent antisense target region (MFR)

In piolet study, we went through processing high number of samples in short interval. So, we improvised MFR python script (not given here) to find most frequent region irrespective of comparing known reference antisense target. Herein, we modified Severance logic for frequency. After improvisation, we included both complete and partial genome sequences for the study. we calculated frequency of targets from all these total 13875 SARS-CoV-2 RNA sequences to find most frequent target regions.

Designing and development of antisense sequence

As per Watson-crick DNA structural model, the antisense would bind complementary to mRNA. And designed antisense would bind to target sites which we found as outcome of MFR method. The short and very frequent sequence would be helpful to design and development of antisense. Thus; the antisense would have complementary nucleotides according to Watson-Crick DNA model. And on the principle of this assumption, we wrote our python scripts for developing antisense sequence from 5'à3' direction.

Initially, we designed antisense sequences based on all MFR data. Continental and world-wide level Where we found the best antisense from previous SARS-CoV-2 complete and partial genome sequences for piolet study. Finally, we modified our result by selecting loci of target in genome of SARS-CoV-2. After improvisation, we included both complete and partial genome sequences for the study and get positive results shown in 3.Result section

Data analysis

The frequency of each target was calculated via MFR method and based on MFR outcomes, we designed antisense by applying reverse complementary python code for each target. Similarity or homology between multiple sequences and targets were calculated via Jalview software (Clustal Omega) (9,10) and NCBI BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi)(11). The statistical data collected for parameters such as frequency, similarity, free energy and temperature etc. Data analysis and graph plots were

generated via GraphPad Prism 8.0.1 version. The hybridization thermodynamic were evaluated via online Oligowalk (siRNA web server)(20) (https://rna.urmc.rochester.edu/cgi-bin/server_exe/oligowalk/oligowalk_form.cgi) and bimolecular secondary structure between antisense and target were predicted via RNA structure webserver(22) (https://rna.urmc.rochester.edu/RNAstructureWeb/Servers/bifold/bifold.html) etc.

Results

Table 1. Why DURVA and MFR?

	ClustalOmega	**NCBIBLAST**	**OligoWalk**	**DURVA-MFR**
Advantage	Finds conserve And similar Regions from Sequences from4000 sequencesOr 4Mb file.	Shows most Similar Sequences with Detail information About homology	Predicts siRNA having19bp Length and hybridization thermodynamics	Finds most frequent Antisense target Based on Chan et.al. Study from large Number of sequences Irrespective of Length
Limitation	Online tool,dataFile should be upto 4MB and sequence length should be within 1Kb.	Shows result up To 100 sequences	The siRNA Prediction based on single input target	Only counts target if Sequence is identical.
Turn around time	More than 120 minutes	More than 30minutes	Take 15 -20minute	Maximum 3 minutes
Conclusion	Helpful to find Similarity from limited samples.	Not helpful toFind target sequence.	There is no Significance for stability in multiplesequences.	New approach to Design antisense based on frequency and finding high conserved regions

Table 1. Why DURVA and MFR? Comparative analysis of Clustal omega(9,10), NCBI BLAST(11), OligoWalk(20) and DURVA-MFR. The comparison between tools based on advantage, limitation, turnaround time of tools or method. We also defined the prime use of tools in particular aspect for design and development of antisense.

Citation: Devendra Vilas Deo, "Novel and Simple Simulation Method to Design and Development of Antisense Template"
MAR Oncology Volume 5 Issue 5
www.medicalandresearch.com (pg. 7)

In our study we are working on novel and simple simulation technique which is effective and producing data in short time. For processing the samples we used IDLE Python3.9 versions on Microsoft windows11 to write transcripts for DURVA MFR.Workflow of DURVA and MFR as follow as shown in Figure 2.
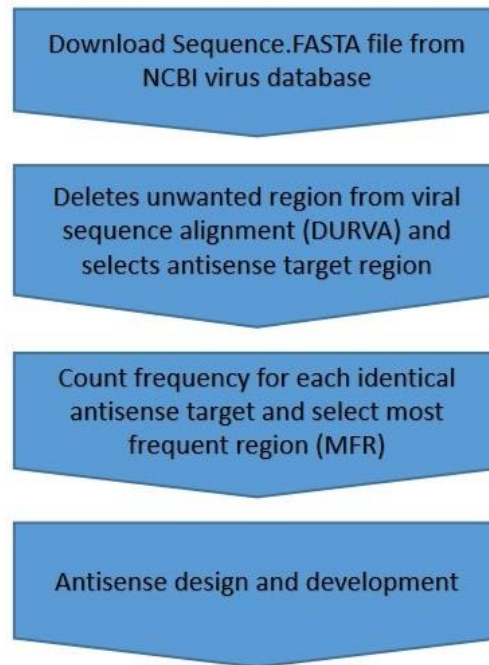

Figure 2: This is simplest block diagram to understant how DURVA and MFR will helpful to simulate targets and antisense sequence design for lethal infection. The detail about process were explained in method and tools section.

Simplifying whole genome sequence in single line by removing newline character

In piolet study, we checked the efficiency of this transcript to convert whole genome in single line format for high number and long sequences in very few seconds. After removing newline character from each end of total 13875 complete or partial sequences made further process easy. Each line without new line character serves as one word which contains equal number of characters or letters same as the length of each sequence.

Citation: Devendra Vilas Deo, "Novel and Simple Simulation Method to Design and Development of Antisense Template"
MAR Oncology Volume 5 Issue 5
www.medicalandresearch.com **(**pg. 8**)**

For example,

Sample Sequence:

ATGCATGCATGCATG

CATGCATGCATGC

Sample sequence after removing newline character: ATGCATGCATGCATGCATGCATGCATGC

Deletion of unwanted region from viral sequence alignment(DURVA)

After removing newline character (\n) from whole genome sequence and simplifying the genome sequence study, it eased our primary aid to delete unwanted region from all virus sequences according to Ho et. al(4)criteria. We studied total 13875 SARS-CoV-2 RNA sequences (SARS-Cov-2 targets.txt). We deleted unwanted region from all these sequences to find most frequent target region. The DURVA extracted and stored 100 to 4000 targets from each sample sequence. To evaluate speed of DURVA, we aggregated all continental data in single file and ran the program file. We concluded that the DURVA shortlisted and stored desirable target sequences (Supplementary file: DURVA.txt) within few minutes for high number of sample sequences. This proved the efficiency, accuracy and shorten the time length for DURVA method.

Most frequent antisense target region (MFR)

This was most important and challenging step in our critical thinking which could narrate our all results and accuracy to find most frequent target region. In preliminary study, we matched our selected sequences with respect to reference sequence (NC_045512.2) https://www.ncbi.nlm.nih.gov/)(11,21). That previous strategy for finding frequent region was lengthy and not working properly and harmful to desktop functioning. Although, some of the targets were not identical with reference antisense target sequences. So; MFR did not work properly and accurate in those cases. Our improvised MFR python script used to find most frequent region by using Severance logic. This improved method found most frequent antisense target from DURVA data (Table2).

Citation: Devendra Vilas Deo, "Novel and Simple Simulation Method to Design and Development of Antisense Template"
MAR Oncology Volume 5 Issue 5
www.medicalandresearch.com (pg. 9)

Targets	Sense strand	Old Frequency	New frequency
Target1	5'~AACAGGCCACAAAUAGGCGU~3'	99.05%	89.90%
Target2	5'~ACCACCUGGUACUGGUAAGAGU~3'	98.57%	88.68%
Target3	5'~AUUGCUUGGUACACGGAACGU~3'	99.63%	90.12%
Target4	5'~AUUGCUUGGUACACGGAACGUUCU~3'	99.60%	90.12%
Target5	5'~CCAGGGACCACCUGGUACUGG~3'	98.48%	88.69%
Target6	5'~GCCGCUGUUGAUGCACUAUG~3'	99.50%	88.80%
Target7	5'~GCCGCUGUUGAUGCACUAUGU~3'	99.48%	88.80%
Target8	5'~GCUGGUGGCACUACUGAAAU~3'	99.37%	51.46%
Target9	5'~GGACCACCUGGUACUGGUAA~3'	98.58%	88.68%
Target10	5'~UCCAGGGACCACCUGGUACU~3'	98.79%	88.69%
Target11	5'~UGCUUGGUACACGGAACGUUCU~3'	99.64%	90.12%
Target12	5'~UUGCUUGGUACACGGAACGUUCUG~3'	99.61%	89.54%
Target13	5'~GGUACUGGUCAGGCAAUAAC~3'	99.43%	88.94%

Table 2- Comparison of modified MFR python scripts between new and older frequencies

Table 2. The data represents most frequent targets discovered from SARS-CoV-2 genomic RNA sequences. These thirteen targets discovered from a)n=19572 SARS-CoV-2 sequences for old frequency and b)n=13875SARS-CoV-2 sequences for new frequency.*Please check DURVA MFR.xlxs

In this improvised process we need not to compare unknown target sequence with known reference antisense target sequences (NC_045512.2) https://www.ncbi.nlm.nih.gov/)(11,21). This would not only find specific or known targets, but helpful to find highly stable region throughout the studied sequences irrespective of loci. However; the advantage of this improvisation was: a) it reduced turnaround time and b) increased our accuracy. Also; the dependency of other tools such as Clustal Omega and NCBI BLAST became lesser than previous strategy. After improvisation, we included both complete and partial genome sequences and compared old data with new results to evaluate our hypothesis. And it showed that the decrease in frequency percentage as we seen in South American population in pilot study.(Supplementary Data file: DURVA MFR.xlxs) The reason behind that the decrease in frequency because a)some of the target region were belonged to other gene or b)contain unnecessary characters instead of nucleotides i.e. NNNNNN. However, the small change in previous python script and strategy lead us to remarkable results as shown (Table2).

Citation: Devendra Vilas Deo, "Novel and Simple Simulation Method to Design and Development of Antisense Template"
MAR Oncology Volume 5 Issue 5
www.medicalandresearch.com (pg. 10)

Designing and development of antisense sequence

As per Watson-crick DNA model, the antisense would bind complementary to mRNA. And developed antisense would bind to target site which we discovered as outcome of MFR method. The short and very frequent sequence would be helpful to design and development of antisense that were acquired to fight against SARS-CoV-2 virus.

Target	Antisense Sequence	Frequency%	Target Loci
Target1	5'~TCCAGTTTGCCCTGGAGCGAT~3'	94.70%	22733...22753
Target2	5'~TTCCAGTTTGCCCTGGAGC~3'	95.34%	22736...22754
Target3	5'~TTCCAGTTTGCCCTGGAGCGATTTGTCTG~3'	94.54%	22726...22754
Target4	5'~GTGTCGTGGATCACGGACAGCATC~3'	89.54%	23225...23249
Target5	5'~CACCCCTCGACATCGAAGCCAA~3'	90.05%	18248...18269
Target6	5'~CAGCAGCGTACAACCAAGC~3'	89.83%	10619...10637
Target7	5'~ACACCAGCTACGGTGCGAGCTCT~3'	89.22%	15046...15068
Target8	5'~TAGCTCCTCTAGTGGCGGCTATTG~3'	89.30%	15122...15145
Target9	5'~TTCCAGTTTGCCCTGGAGCGATTTGTCTG~3'	94.54%	22726...22754
Target10	5'~GACAACACGATGCACCACCAAAGG~3'	90.09%	13173...13196
Target11	5'~CCAGTACCAGGTGGTCCCTGGAGTGTAGA~3'	88.47%	17017...17045
Target12	5'~TCAACAGCGGCATGAGAGCAAGCTGT~3'	88.80%	17104...17129
Target13	5'~GTGCATCAACAGCGGCATGAG~3'	88.81%	17114...17134
Target14	5'~CACAGCGTCCTAGATGGTGTCCAGC~3'	86.61%	26919...26943
Target15	5'~CAGACAAAUCGCUCCAGGGCAAACUGGAA~3'	94.54%	22726...22754
Target16	5`-AGCCGAGTGACAGCCACACAG-3`	85.50%	33...53

Table 3-Antisense sequence designed based on modified method

Table 3. These are top 16 designed antisense sequences (including Target16 is previous potential antisense) with frequency and target loci on genomic RNA of SARS-CoV-2.

Designed and developed antisense sequences and their target (Table 3) was verified by homology were evaluated using NCBI BLAST(11) (https://blast.ncbi.nlm.nih.gov/Blast.cgi), similarity via Jalview Clustal omega multiple alignment tool(9,10) (https://www.ebi.ac.uk/Tools/msa/clustalo/) and stability and hybridisation thermodynamics of most frequent targets OligoWalk tools(20,22) (https://rna.urmc.rochester.edu/cgi-bin/server_exe/oligowalk/oligowalk_form.cgi) respectively.

Citation: Devendra Vilas Deo, "Novel and Simple Simulation Method to Design and Development of Antisense Template"
MAR Oncology Volume 5 Issue 5
www.medicalandresearch.com (pg. 11)

DURVA-MFR and other tools

However, the major drawback of this tool is: if any error in sequences due to improper sequencing or mutation, it does not able to find conserve regions from more sequences or complete genome of SARS-CoV-2 virus in such cases. As some of the sequences did not contain 100% identical fragment or complete fragment with respective to reference fragment after MFR run. Therefore; due to this limitation of MFR in such cases,we chose Jalview (Clustal-Omega Multiple alignment tool) (EBI) (9,10) and NCBI BLAST(11).These tools helped us to find similarity between partial or semi-conserved fragments and most frequent region within non-identical antisense-target sequences. This helped to assimilate relation between DURVA-MFR and Clustal omega. Clustal-Omega Multiple alignment tool (9,10) results showed that most frequent regions were highly similar with few genomic RNA sequences and BLAST(11) results had shown all short template having E value<0.006 against SARS-CoV-2. The Supplementary file:Target.xls contains data for partially or semi- conserved with these 16 targets. The statistical data analysis was done in column and grouped table type and nested graph type for data analysis to generate graph1, graph2 respectively via GraphPad Prism 8.0.1 version. The graph 1 and graph 2 (Supplementary file: Data graph.pptx) defined the similarity in terms of PID (9,10) and comparative analysis between identical and similarity in terms of percentage respectively. As per our observations in terms of PID, Target1, Target3, Target4, Target9 and Target15 were identical within most frequent genomic RNA sequences and most similar within few RNA sequences too.

And here we compare oligoWalk(20) data for parameters such as stability and efficacy of our discovered antisense against previously known anstisense. In short, hybridisation thermodynamic and probabilities of efficient siRNA. Here we compared our discovered antisense with similar siRNA predicted via oligoWalk(20). The similar siRNAs are having efficient siRNA probabilities more than 0.55. Six antisense complementary against Target4, Target6, Target7, Target10 and Target11 were having efficient siRNA probabilities more than 0.8. Similar siRNA sequence had overall free energy (ΔGoverall) in the range of -17.4 kcal/mol to -28.1 kcal/mol and mean -22.04 kcal/mol. The free energy required for duplex formation (ΔGDuplex) in the range of -32.7 kcal/mol to -37.7 kcal/mol and mean -34.25 kcal/mol. The temperature (Tm-Dup) required for duplex formation in the range of 82.40Cto 91.10C and mean 85.740C. The free energy to break-target (ΔGtarget) in the range of -4.7 kcal/mol to -14.7 kcal/mol and mean -10.19 kcal/mol. The free energy of intra oligo (ΔGintra-oligomer) in the range of -0.1 kcal/mol to -3.3 kcal/mol and mean -1.00 kcal/mol and inter oligo(ΔGinter-oligomer) in the range of -10.6 kcal/mol to -15 kcal/mol

Citation: Devendra Vilas Deo, "Novel and Simple Simulation Method to Design and Development of Antisense Template"
MAR Oncology Volume 5 Issue 5
www.medicalandresearch.com (pg. 12)

and mean -12.04 kcal/mol. And free energy between 5' and 3' ends of predicted functional siRNA (ΔGEnd_diff) in the range of -0.73 kcal/mol to 2.32 kcal/mol and mean -1.44 kcal/mol.

However, we compared our result with one of potential therapeutic ASO tested against new SARS strains within in vitro experiments. The synthetic ASO sequence;. 5`- AGCCGAGTGACAGCCACACAG-3` and sense strand sequence: 5`- CTGTGTGGCTGTCACTCGGCT-3` with respective to our studied parameters. After our conclusion, we could state that their discovered target was 85.5% frequent within our SARS- CoV-2 genomic RNA sequences data. Neverthless; we noticed within both experiments after the BLAST of most frequent region that 5'~UGCUUGGUACACGGAACGUUCU~3' also belonged to ORF1ab gene of SARS-CoV-2 as same as 5`-CTGTGTGGCTGTCACTCGGCT- 3`. These both studies suggested ORF1ab was most stable gene within new strains of corona virus. Moreover we had calculated hybridisation thermodynamic and probabilities of efficient siRNA for known potential antisense(1). The similar siRNA sequences; 5`- AGCAUGCAGCCGAGUGACA-3` with respective to known potential antisense sequence 5`-AGCCGAGUGACAGCCACACAG-3` were having efficient siRNA probabilities 0.33, ΔGoverall were 22.5 kcal/mol, ΔGDuplex were -37.9 kcal/mol, Tm-Dup was 90.80C, ΔGtarget were -11.2 kcal/mol, ΔGintra-oligomer were -1.4 kcal/mol, ΔGinter-oligomer were -16.6 kcal/mol and ΔGEnd_diff was 0.03kcal/mol etc.

By comparing oligowalk siRNA web server results(20) for paramters such as melting temperature, free energies such as overall, interoligo and intraoligo and break target etc. for previous known antisense and designed antisense via DURVA and MFR suggested that our discovered antisense were having more efficient siRNA probabilities.. Six similar antisenses discovered through oligowalk against Target4, Target6, Target8, Target10, Target13 and Target14 were having more negative ΔGoverall meant high tighter binding against genomic RNA and transcribed RNA template of virus. One similar antisense discovered through oligowalk against Target8 was having ΔGDuplex and Tm-Dup values near to known potential antisense indicated both were equally more stable and required high melting temperature to hybridise duplex formation. Eight similar antisense discovered through oligowalk against Target1, Target2, Target3, Target5, Target7, Target9, Target11 and Target12 were having more negative ΔGtarget Value meant that these antisense templates will be less accessible for siRNA binding. Four similar antisense template discovered through oligowalk against Target8, Target11, Target13 and Target14 were having more negative ΔGintra-oligomer values meant these were more self-stable structure comparative to known potential antisense template. All similar discovered template discovered through oligowalk were having more positive ΔGEnd_diff. These meant that all similar functional siRNA will be shown an

unstable 5'end. Contrary, 5`- AGCAUGCAGCCGAGUGACA-3` was more stable dimerization comparative to our discovered antisense templates.

The multiple variable table type and column graph type plotted via GraphPad Prism 8.0.1 version for data analysis of graph3. Graph3 is correlation between parameters such as frequency, similarity, melting temperature required for duplex formation and free energies such as ΔGoverall, ΔGDuplex, ΔGtarget, ΔGintra-oligomer, ΔGinter-oligomer and ΔGEnd_diff. Overall data suggested that DURVA and MFR method of finding antisense work efficiently and accurately and can be helpful for designing and development of antisense.

Also; we studied predicted bimolecular secondary structure derived from RNA structure web server (https://rna.urmc.rochester.edu/cgi-bin/server_exe/oligowalk/oligowalk_form.cgi)(22). And we noticed all promising RNAstrucure bifold results for DURVA and MFR outcomes. The highest value of energy for previously studied high-scored AON and TR_3 was -31.5(24) while lowest value of energy for our frequent antisense and target15 was -37.2. But except target15, remaining targets and antisense were binding 100%. (Supplementary Data file: DURVA MFR.xlxs)

Discussion

Designing ASOs and siRNA can be logically challenging for selection of tool or therapeutic sequences. There are number of online open-source software around us to design antisense and siRNA such as OptiRNAi(12), siExplorer(13), DSIR(14), i-Score(8), MysiRNA-Designer(15), siDirect(16), siRNA Selection Server(17), siRNA- Finder(18), PFRED(19) and OligoWalk(20). Recently developed PFizer RNAi Enumeration and Design (PFRED) tool provides an user-interface client server to predicts a library of siRNA or antisense oligonucleotides that target a specific gene of interest. Specification of this tools was it starts with target gene (Ensembl ID) and culminating in the design of siRNAs or antisense oligonucleotides. Sequences selection worked using bioinformatics algorithms built upon careful mining of the sequence-activity relationships found in public datasets as well as internal collections. However, this online tool difficult to download for new user and requires docker or docker compose setup(19). OptiRNAi a computational tool based on Elbashir et al. prediction(12), siExplorer developed for calculating siRNA activity and its binding affinity against endogenous human genes and provides the 3-nt periodicity unveiling siRNA functionality(13), DSIR prdiction for active siRNA is based on a linear model combining particular nucleotides at given positions and specific motifs on the siRNA guide-strand,

Citation: Devendra Vilas Deo, "Novel and Simple Simulation Method to Design and Development of Antisense Template"
MAR Oncology Volume 5 Issue 5
www.medicalandresearch.com (pg. 14)

including 2-nt overhangs at the 3' end(14), siDirect 2.0 updated version predicts efficient siRNA candidate with minimal off-target effect based melting temperature Tm< 21.5°C, seed-target duplex between 2-8 positioned nucleotides from the 5' end of the siRNA guide strand and its target mRNA followed by the elimination of unrelated transcripts(16). However, we noticed lot of similarity between MysiRNA-Designer and our method. MysiRNA- Designer online source for windows is difficult for installing and this tool filters siRNA based on MysiRNA score i.e. siRNA efficacy prediction score for ranking the designed siRNAs and finding best targets sites once siRNA designed(15). While our method is inverse and finding stable target for designing antisense template.

Overall, all previous siRNA or RNAi designing tools were dependent on homology and thermodynamic parameters assessment for predicting efficient antisense template, While our method is not predicting antisense template and does not requires homology assessment once antisense designed. However; in our paper we have shown direct comparison with tools such as NCBI BLAST(11), Jalview (Clustal- Omega Multiple alignment tool) (EBI) (9,10) and OligoWalk(20) which are helpful for finding homology i.e. E-value, P-value, percentage of identity and hybridisation themodynamics etc. for cross-checking predicted targets site and binding efficiency and generally used by all online open sources directly or indirectly.

In short, DURVA and MFR is novel and simple approach for designing antisense because Jalview (Clustal- Omega Multiple alignment tool) (EBI) (9,10) did not deal with very large numbers (tens of thousands) of DNA/RNA or protein sequences due to it use of them BED algorithm for calculating guide trees. DURVA-MFR can be alternative and very simple method to find most identical and probable conserve region within maximum number of sequences by ignoring transcriptional errors. Similar to Clustal omega Multiple sequence alignment, T-coffee and Lalign(9,10) are used to find local alignment not to design target as per antisense requirement. Also; these tools can process if your input file contains 500 sequences or having size of1Mb. While DURVA and MFR can study more than ten of thousand sequences and up to 1Gb file. This is more accurate and less challenging to design therapeutic ASO without adding available features related to homology. We are already familiar with NCBI BLAST(11) and Oligowalk(20) web server tools for homology and antisense prediction. But, we also discussed limitation and turn around time of these tools in Table1. While the advantage of DURVA- MFR method over the finding efficient and high conserved region without need of BLAST(11) for homology and OligoWalk(20) for prediction of antisense. Because the target discovered from high number of sample size or annoted sequences of same species which were available at NCBI virus database (https://www.ncbi.nlm.nih.gov/)(11,21). And the Oligowalk SiRNA web server(20) predicts the efficient

Citation: Devendra Vilas Deo, "Novel and Simple Simulation Method to Design and Development of Antisense Template"
MAR Oncology Volume 5 Issue 5
www.medicalandresearch.com (pg. 15)

siRNA based on single target input with limit while DURVA-MFR method can define most efficient antisense based on frequency of target analysed from large number of long sequences.

The regression analysis with few studies related to potential ASO designed via high- score from oligonucleotide properties calculator tool. Our analysis proved the success rate of our newly innovated method. Goryachev and co-workers did comparative analysis of old and new SARS strains. Based on analysis they tested synthetic ASO i.e. 5`- AGCCGAGTGACAGCCACACAG-3` within in vitro model(1). The stability of their antisense template with respective to our findings was very lesser as comparison. The previous discovered sense strand was 85.5% frequent within our SARS-CoV-2 genomic RNA sequences data (Graph 2). The previous verified ASO target site was less than 86% frequent with SARS-Cov-2 data while our defined all ASO target sites are more than 86.61% frequent with SARS-Cov-2 data. However, both studies indicate that ORF1ab is most stable gene within new strains of corona virus. This completely states that combination of DURVA and MFR not only efficient to find most frequent region but also helpful for designing potential ASO. This analysis would be helpful for upcoming clinical trials against this worldwide pandemic. Moreover, H. Chubuk study for promising strategy against SARS-CoV-2 defined AON against 5'UTR, 3'UTR and start codon based on high-score of parameter described by Aartsma-Rus et al. He also proposed that targeting viral genome could be helpful to control viral replication and infection(24). However,we calculated the frequency of high-scored ASO predicted by Chubuk et al via DURVA-MFR method. (Supplementary Data file: DURVA MFR.xlxs) We concluded that frequency of high-scored ASO were between 47-87%.

Recently, Yan et al developed structure based ASO to inhibit SARS-CoV-2 replication(2). Their approach would be helpful to cross-check the binding of our designed antisense in 3D space at molecular level.And also helpful to find most useful ASO based drugs against disease like SARS-CoV-2. However, their study was focused on structure in stead of stability of genomic regions. And structure could be varying at molecular level which could be limitation of their critical thinking. However, we predicted bimolecular secondary structure via RNA structure web server (https://rna.urmc.rochester.edu/cgi-

bin/server_exe/oligowalk/oligowalk_form.cgi)(21). And we noticed promising RNA strucure bifold results for DURVA and MFR outcomes. The highest value of energy for previously studied high-scored AON and TR_3 was -31.5(24) while lowest value of energy for our frequent antisense and target15 was -37.2. This could be helpful to understand binding efficiency of our discovered antisense.

Citation: Devendra Vilas Deo, "Novel and Simple Simulation Method to Design and Development of Antisense Template"
MAR Oncology Volume 5 Issue 5
www.medicalandresearch.com (pg. 16)

In our preliminary study, we faced difficulties to find more appropriate outcomes. So; later we improvised MFR python script to find most frequent region irrespective of comparing known reference antisense target by modifying Severance frequency strategy(23). This improved method found most frequent antisense target from DURVA data. In this modified version, we need not to compare unknown target sequence with known reference sequence (NC_045512.2) (https://www.ncbi.nlm.nih.gov/)(11,21) antisense target sequence. This not only found certain most frequent target also helpful in finding highly stable region through out the studied sequences irrespective of loci. In Asian samples, we found that more than 99.28% frequent in all outcomes and the one short target sequence was having high frequency about 99.87%. Contrary, Indian study showed decrease in frequency percentage as we seen earlier. There we found only few target sequence with more than 95% frequency. The reason behind the decrease in frequency was either some of the target region were belonged to the gene or having the inappropriate nucleotides (e.g. NNNNNN) at same loci or position. After improvisation, we included both complete and partial genome sequences for the study. However; the benefit of this improvisation was that it does not saved our time only but increased our accuracy too and can process up to 1Gb data. Moreover, the dependency of other tools became lesser comparative to previous strategy. Although, it did not alter all previous results.

Overall our study suggested that DURVA-MFR is novel and simple simulation method for designing and development of antisense sequence. Based on correspondence data, we understood that DURVA - MFR able to process more than ten thousand (nearly 30kbps) long sequences and can assess most frequent region from 1Gb text file size, which is to be most innovative and first time. At the end of our study, we concluded that our defined ASO- target sites were more than 88% identical with SARS-Cov-2 data. This could be helpful to understand how our method will be useful in future to find most frequents and highly conserved target site from large sample size and implementation in other lethal viral infection too.

5.   Future scope of DURVA and MFR method:

Apart from antisense designing and developing DURVA-MFR method can be used in finding most stable homologues region between two or more sequences. Using these strategies, we could find similarity

Citation: Devendra Vilas Deo, "Novel and Simple Simulation Method to Design and Development of Antisense Template"
MAR Oncology Volume 5 Issue 5
www.medicalandresearch.com (pg. 17)

between paralogue and orthologue sequences. These algorithms have ability to cut short sequences in serial. This serial sequences can be compared with remaining other sequence and can evaluate most identical frequent region within two different species or similar species. This method helpful in finding most stable and identical region within mutating virus such as SARS-CoV-2, HIV and influenza viruses. Using these two algorithms we can design primers and probes for various biotechnological and genetic engineering aids. This method will not only efficient in genomic but useful for proteomics too. This method will mostly use for vaccinology in case of mutating viral strains such as HIV and Corona viruses etc. In future, these methods will be game changer and most powerful logic to design most efficient, accurate, user friendly, time saving and safe antisense to cure lethal infections.

Conclusion

In this study, we studied all types of parameter as per our hypothetical consideration and proved that our self-designed simulation technique i.e. DURVA and MFR methods can find highly identical and stable antisense targets and design antisense template from large number of sequences and large file size up to 1Gb. These methods are novel and simple way for designing and development of the antisense against all types of lethal infections.

References

1. Goryachev AN, Kalantarov SA, Severova AG, Goryacheva AS. Potential Opportunity of Anti-sense Therapy of COVID-19 on an in Vitro Model. bioRxiv [Internet]. 2020 Jan 1;2020.11.02.363598. Available from: http://biorxiv.org/content/early/2020/11/03/2020.11.02.363598.abstract

2. Li Y, Garcia G, Arumugaswami V, Guo F. Structure-based design of antisense oligonucleotides that inhibit SARS-CoV-2 replication. bioRxiv [Internet]. 2021 Jan 1;2021.08.23.457434. Available from: http://biorxiv.org/content/early/2021/08/24/2021.08.23.457434.abstract

3. Chan JH, Lim S, Wong WF. ANTISENSE OLIGONUCLEOTIDES: FROM DESIGN TO THERAPEUTIC APPLICATION. Clin Exp Pharmacol Physiol. 2006 May;33(5–6).

4. Ho SP, Britton DH, Stone BA, Behrens DL, Leffet LM, Hobbs FW, et al. Potent antisense oligonucleotides to the human multidrug resistance-1 mRNA are rationally selected by mapping RNA-accessible sites with oligonucleotide libraries. Nucleic Acids Res. 1996 May;24(11):1901–7.

Citation: Devendra Vilas Deo, "Novel and Simple Simulation Method to Design and Development of Antisense Template"
MAR Oncology Volume 5 Issue 5
www.medicalandresearch.com (pg. 18)

5. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nat Microbiol. 2020 Apr 2;5(5).

6. Fakhr E, Zare F, Teimoori-Toolabi L. Precise and efficient siRNA design: a key point in competent gene silencing. Cancer Gene Ther. 2016 Apr 18;23(5).

7. Sharma VK, Watts JK. Oligonucleotide therapeutics: chemistry, delivery and clinical progress. Future Med Chem. 2015 Oct;7(16).

8. Ichihara M, Murakumo Y, Masuda A, Matsuura T, Asai N, Jijiwa M, et al. Thermodynamic instability of siRNA duplex is a prerequisite for dependable prediction of siRNA activities. Nucleic Acids Res. 2007 Sep;35(18).

9. Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res [Internet]. 2019 Jul;47(W1):W636—W641. Available from: https://europepmc.org/articles/PMC6602479

10. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics [Internet]. 2009 May 1;25(9):1189–91. Available from: https://doi.org/10.1093/bioinformatics/btp033

11. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990 Oct;215(3):403–10.

12. Cui W, Ning J, Naik UP, Duncan MK. OptiRNAi, an RNAi design tool. Comput Methods Programs Biomed. 2004 Jul;75(1).

13. Katoh T, Suzuki T. Specific residues at every third position of siRNA shape its efficient RNAi activity. Nucleic Acids Res. 2007 Feb;35(5).

14. Filhol O, Ciais D, Lajaunie C, Charbonnier P, Foveau N, Vert J-P, et al. DSIR: Assessing the Design of Highly Potent siRNA by Testing a Set of Cancer-Relevant Target Genes. PLoS One. 2012 Oct 30;7(11).

15. Mysara M, Garibaldi JM, ElHefnawi M. MysiRNA-Designer: A Workflow for Efficient siRNA Design. PLoS One. 2011 Oct 26;6(11).

16. Naito Y, Yamada T, Ui-Tei K, Morishita S, Saigo K. siDirect: highly effective, target specific siRNA design software for mammalian RNA interference. Nucleic Acids Res. 2004 Jul 1;32(Web Server).

Citation: Devendra Vilas Deo, "Novel and Simple Simulation Method to Design and Development of Antisense Template"
MAR Oncology Volume 5 Issue 5
www.medicalandresearch.com (pg. 19)

17. Yuan B, Latek R, Hossbach M, Tuschl T, Lewitter F. siRNA Selection Server: an automated siRNA oligonucleotide prediction server. Nucleic Acids Res. 2004 Jul;32(Web Server issue):W130-4.

18. Lück S, Kreszies T, Strickert M, Schweizer P, Kuhlmann M, Douchkov D. siRNA-Finder (si-Fi) Software for RNAi-Target Design and Off-Target Prediction. Front Plant Sci. 2019 Aug 15;10.

19. Sciabola S, Xi H, Cruz D, Cao Q, Lawrence C, Zhang T, et al. PFRED: A computational platform for siRNA and antisense oligonucleotides design. PLoS One. 2021 Jan 22;16(1).

20. Lu ZJ, Mathews DH. OligoWalk: an online siRNA design tool utilizing hybridization thermodynamics. Nucleic Acids Res. 2008 Jul;36(Web Server issue):W104-8.

21. NCBI Virus [Internet]. Available from: https://www.ncbi.nlm.nih.gov/

22. Reuter JS, Mathews DH. RNAstructure: software for RNA secondary structure prediction and analysis. BMC Bioinformatics [Internet]. 2010;11(1):129. Available from: https://doi.org/10.1186/1471-2105-11-129

23. Biology M, Biology M. A promising strategy against SARS-CoV-2 infected patients: Antisense therapy Hasan Cubuk 1 * 1. 2.

24. Severance CR. Python for Everybody Trinket. 2018;245. Available from: https://books.trinket.io/pfe/index.html

Citation: Devendra Vilas Deo, "Novel and Simple Simulation Method to Design and Development of Antisense Template"
MAR Oncology Volume 5 Issue 5
www.medicalandresearch.com (pg. 20)