

**Title:** PFRED: A computational platform for siRNA and antisense oligonucleotides design.

**Authors:** Simone Sciabola<sup>1</sup>, Hualin Xi<sup>2</sup>, Dario Cruz<sup>1,8</sup>, Qing Cao<sup>3</sup>, Christine Lawrence<sup>4</sup>, Tianhong Zhang<sup>4</sup>, Sergio Rotstein<sup>4</sup>, Jason D Hughes<sup>5</sup>, Daniel R Caffrey<sup>6</sup>, and Robert Stanton<sup>7\*</sup>

<sup>1</sup> Medicinal Chemistry, Biogen, Cambridge, MA 02142, USA

<sup>2</sup> Regenta, Cambridge, MA, 02140, USA

<sup>3</sup> Medicinal Chemistry, Ra Pharmaceuticals, Cambridge, MA, 02140, USA

<sup>4</sup> Business Technology, Pfizer, Cambridge, MA 02140, USA

<sup>5</sup> Computational Biology, Foundation Medicine, Cambridge, MA, 02141, USA

<sup>6</sup> Department of Medicine, University of Massachusetts Medical School, Worcester, MA 01605, USA

<sup>7</sup> Simulation and Modeling Sciences, Pfizer, Cambridge, MA 02139, USA

<sup>8</sup> Chemical Engineering, Northeastern University, Boston, MA 02115, USA

\*To whom correspondence should be addressed. Tel: 1-617-551-3274, Email: [Robert.Stanton@pfizer.com](mailto:Robert.Stanton@pfizer.com)

## Abstract

PFRED a software application for the design, analysis, and visualization of antisense oligonucleotides and siRNA is described. The software provides an intuitive user-interface for scientists to design a library of siRNA or antisense oligonucleotides that target a specific gene of interest. Moreover, the tool facilitates the incorporation of various design criteria that have been shown to be important for stability and potency. PFRED has been made available as an open-source project so the code can be easily modified to address the future needs of the oligonucleotide research community. A compiled version is available for downloading at <https://github.com/pfred/pfred-gui/releases> as a java Jar file. The source code and the links for downloading the precompiled version can be found at <https://github.com/pfred>.

## Introduction

Over the last three decades small interfering ribonucleic acid (siRNA) and antisense oligonucleotides (ASO) have emerged as powerful tools to modulate the expression of genes in vivo and in vitro. Potential oligonucleotide therapeutics have been discovered and developed with mixed success[1,2]. For example, Mipomersen, Inotersen, Patisiran and Givlaari have all been approved for use in recent years. Mipomersen is a lipid lowering agent that targets the ApoB gene [3,4] which was eventually approved by the United States Food and Drug Administration (FDA) in 2013 and later discontinued due to adverse effects [5], Inotersen is an antisense compound targeting transthyretin (TTR) for the treatment of polyneuropathy caused by hereditary transthyretin-mediated amyloidosis [6], Patisiran was the first siRNA approved by FDA also targeting transthyretin-mediated amyloidosis [7], and Givlaari targets aminolevulinic acid synthase 1 (ALAS1) for the treatment of acute hepatic porphyria a rare genetic condition in which patients struggle to produce heme [8]. Furthermore, the approval of Nusinersen for the treatment of spinal muscular atrophy (SMA) in pediatric and adult patients represents a historical achievement for the SMA community and a success story for the field of nucleic acids based therapeutics [9].

However, the development of oligonucleotide therapeutics remains challenging as they often have unintended off-target effects and cause non-specific hepatic and renal toxicity. Additionally, it is exceptionally difficult to deliver oligonucleotides to most tissue types and organs. This can result in the need for tissue specific delivery systems and several siRNA compounds have advanced into development using this paradigm; these include liposomally liver targeted ALN-RSV01 which targets RSV through the nucleocapsid “N” gene [10], PF-4523655 for age related macular degeneration targeting the eye through direct injection [11] and galNac conjugates [12] such as the late stage ALN-AT3 [13] targeting antithrombin for the treatment of hemophilia. The recent FDA approval of Patisiran[14] represents a first-of-its-kind RNA interference (RNAi) therapeutic for the treatment of the polyneuropathy of hereditary transthyretin-mediated (hATTR) amyloidosis in adults and the first ever FDA approval of a siRNA treatment [7].

Designing antisense oligonucleotides (ASOs) and siRNA can be logistically challenging given the many competing design criteria that can be incorporated into the selection of tool or therapeutic sequences. Design considerations may include splice variants, cross-species targeting for validation, single nucleotide polymorphisms (SNPs), secondary structure, undesirable motifs (e.g. toxic, poly-A, or poly-G repeats), complementarity with off-target sequences, intron/exon boundaries, chemical modification pattern of nucleotides, and predicted activity. These factors need to be weighed according to the intended use of

the oligonucleotide. For example, compounds designed as in vitro tools only need to be active in a single species, whereas it is advantageous for therapeutic oligonucleotides to be active in model organisms as well as patients. A number of computational tools have been developed to address different aspects of the design process, including OptiRNAi [15], siExplorer [16], DSIR [17], i-Score [17,18], MysiRNA-Designer [19], siDirect [20,21], siRNA Selection Server [22], siRNArules [23] and siRNA-Finder [24]. Additionally, major vendors of siRNA reagents such as Dharmacon (siDESIGN Center) [25], GeneScript (Target Finder) [26] and Thermo Fisher (BLOCK-iT) [27] have web based design tools available.

In the present paper we describe PFRED (**P**Fizer **R**NAi **E**numeration and **D**esign tool), a client-server software system designed to assist with the entire oligonucleotide design process, starting with the specification of a target gene (Ensembl ID) and culminating in the design of siRNAs or RNase H-dependent antisense oligonucleotides. Sequences are chosen using bioinformatics algorithms built upon careful mining of the sequence-activity relationships found in public datasets as well as internal collections. The tool provides researchers with a user-friendly interface where the only required input is an accession number for the target gene and it returns a list of properties that are believed to contribute to the efficacy of an siRNA or ASO. These properties include human transcripts and cross-species homology, GC content, SNPs, intron-exon boundary, duplex thermodynamics, efficacy prediction score and off-target matches. An automated oligonucleotide selection procedure is available to quickly select one potential set of sequences with an appropriate property profile. The selection protocol can be customized by the user through changes of the selection cutoffs or the addition of alternate design parameters and algorithms.

## Materials and Methods

PFRED is implemented using a client-server software architecture. On the client-side, the graphical user interface (GUI) was developed in the Java programming language and is deployed to a user's desktop through Java Web Start, the user will need to have a recent version of the Java development kit setup in their system. Because of the cross-platform support of Java, PFRED can be run from multiple desktop environments such as Windows, Mac and Linux. The most CPU-intensive algorithms used in PFRED are implemented as Python and Perl scripts which are wrapped within a docker container that works as a web service hosted in an Amazon Web Service (AWS) instance and invoked remotely from the client through a RESTful API, also developed in Java. Beyond the performance improvements, the client-server architecture also greatly simplifies the integration of disparate design algorithms that often rely on heterogeneous third-party libraries or reference datasets. Calls to the server-side algorithms are made

through the PFRED java web service which uses the REST API that was developed using the Glassfish Jersey Java modules. New algorithms can be added to PFRED following the conventions established for the existing tool.

To facilitate the use of PFRED, users only need to download the latest GUI release and run it on their own computer; it is also possible for users to build their own private PFRED service by downloading the PFRED Docker repository in order to create and run the PFRED Service container following the documentation given at <https://github.com/pfred/pfred-docker>. This technique named containerization, facilitates cloud and server agnosticism, because the container automatically builds the needed environment for the PFRED back-end to work; users can also use this local service to test their own alternative algorithms or models, and deploy the PFRED service in a cloud environment like AWS, Azure, Google cloud, or on a local server. It is only required for the user to have Docker and Docker-compose setup in the system (see <https://docs.docker.com/compose/>). The current client-server architecture consists on an already built-in Docker Service within a publicly queryable AWS instance managed by Biogen, which offers a REST API for easy PFRED end-point access to all the back-end services (see <https://github.com/pfred/pfred-rest-service> for documentation on how to compile the REST source code and access the PFRED end-points using a browser). The PFRED GUI takes advantage of the RESTful API in order to access all the needed end-points for ASO and siRNA design (Schema 1 illustrates the PFRED architecture).

Figure 1 gives an example of the main PFRED interface showing the enumerated antisense sequences and data annotations related to these compounds. Oligonucleotide sequences and their properties can be easily visualized and analyzed in the main table. This main PFRED table provides most of the common spreadsheet functions such as alphanumeric sorting, column/row resizing, column coloring and conditional coloring. In addition, users can choose which properties are displayed in the table and customize the display order by drag-and-drop. To help users sift through large numbers of sequences, the main table supports filtering by either sequence patterns or property values. For example, a user can search for oligonucleotides that are conserved between human and mouse and have high predicted siRNA efficacy. Oligonucleotides can also be organized into different groups and then reviewed separately. Additional functions are also included to help users analyze their data. For example, the mathematical formula feature can be used to apply weights to multiple properties and summarize them into a single score for ranking and selection. The data can be exported as a comma separated file (CSV) and then imported into other applications for further analysis.

A key element of PFRED's functionality is its canonical representation of oligonucleotide structures. In a separate paper [28] we presented a monomer based notation language to represent complex biomolecule structures (the Hierarchical Editing Language for Macromolecules or HELM) and a desktop application named PME (Pfizer Macromolecule Editor) for their drawing and visualization, which has become an industry standard for biopolymer notations [29]. Figure 2A shows an example of HELM notation for a short antisense gapmer. As polymeric biomolecules, antisense oligonucleotides and siRNAs can be represented with HELM notation and different sequence visualization options can be created for specific purposes. Included within PFRED are three oligonucleotide representations, shown in Figure 2B-D. Figure 2B shows an explicit component or monomer view of the oligonucleotide breaking it down into sugars, bases and linkers. A sequence view is shown in Figure 2C, using a standard notation with RNA (capital) and DNA (lowercase) but emphasizing "chemical modifications" through the use of the pink dots. For example, the sequence shown in Figure 2C has a fully phosphorothioated backbone which is denoted by the pink dots between the base letters. It also has three locked nucleic acids (LNAs) at the 3' and 5' ends of the compound that are distinguished by the pink dots above the letters denoting these nucleosides. A final block representation is shown as Figure 2D. The block representation was created to emphasize patterns of chemical modifications to a structure. Base modifications are shown as pink dots below the blocks, sugars are colored depending on their structure (e.g. LNA-yellow, 2'OMe-Blue, etc.) and backbones other than phosphodiester are represented by pink connectors above the blocks.

The PFRED enumeration workflow is shown in Figure 3. Transcript IDs and sequences of a target for different ortholog species (human, rat, mouse, dog, chimpanzee and macaque) are retrieved from the public ENSEMBL database through the ENSEMBL REST API (please refer to <https://rest.ensembl.org/>) given a single transcript or gene ID. Users define a primary transcript sequence of their length of choice to be enumerated (Figure 3, steps 1-3). Each enumerated oligonucleotide is then aligned to the other selected transcripts including splice variants and ortholog transcripts using the fast memory-efficient short read aligner, Bowtie, against the pre-calculated Bowtie index covering all transcript sequences [30]. Alignment results are processed by a Python interface and the number of mismatches to each transcript is reported back to PFRED as properties of the oligonucleotides. Exon location and SNP information are also retrieved through the EnSEMBL APIs for each base position and mapped to each oligonucleotide as properties.

Pre-built bowtie indexes for all species are required for off-target searches. These were built by creating a Python interface that automatically downloads all transcripts (cDNA) and unspliced gene

sequences for any species in FASTA format using the ENSEMBL REST API which provides the advantage of always relying on the latest version of the ENSEMBL database. The FASTA files are written with transcript ID and Gene ID included in the sequence names. These FASTA files are used to build indexes for searching cDNA and gene off-target matches. The ENSEMBL IDs are used for post-processing the bowtie alignment results to calculate the number of genes that an oligonucleotide aligns to with 0, 1 and 2 mismatches, and to exclude the intended target from the off-target count. For siRNA, only the cDNA database is searched for both the guide and passenger strands. In contrast, for antisense design, only the sense sequences (or antisense sequence in reverse complementary direction) is searched against both the cDNA and unspliced gene sequence indexes for off-target hits. The number of cDNA off-target hits is then reported as an oligonucleotide property for 0, 1 and 2 mismatches.

Due to the cost of synthesizing and testing siRNAs and ASOs, rational design represents a critical step in any oligonucleotide experiment. A priori prediction of a compound's function is a key component of RNAi and ASO applications to obtain effective silencing of the desired gene. Early models were based on empirical rules for designing functional siRNA, obtained from the experimental results of ASO and RNAi screening campaigns. For RNAi sequence design, Tuschl [31,32] first and later Reynolds [33] reported a series of guidelines for the implementation of a “point-based” scoring scheme where the presence or absence of key features in an siRNA duplex will lead to an increase or decrease of the final functional score. Several other algorithms have emerged since then [16,17,34–39] based on more sophisticated scoring schemes which use sequence descriptors and supervised machine learning techniques to select siRNAs capable of inducing effective gene silencing in cell lines. In PFRED, an algorithm for predicting siRNA functionality has been implemented by using diverse sets of oligonucleotide descriptors combined with a support vector machine (SVM) algorithm. Details of the sequence descriptors and algorithm implementation were previously described elsewhere [40]. Models of ASO activity have also improved but can be highly dependent on the length of the ASO as well as its chemical modification. Most published models use either sequence motif based filters [41] or computational approaches including secondary structure prediction [42–44], thermodynamics [45,46], machine learning [47–49] and molecular dynamics simulation [50], but all have reported only limited success. Here we integrated both the thermodynamics parameters calculated using Oligowalk [51] and the sequence motifs reported in Matveeva [41]. Using these parameters, a scoring scheme was derived that penalized motifs correlated with toxicity or promiscuity, strong inter- and intra-molecular binding energy as well as weak binders to the RNA target. The model was derived and tested using data from AOBBase [52] including more than 500 ASOs against 46 targets. The thermodynamic calculations were based on unmodified DNA antisense oligonucleotides and

therefore of limited relevance to modifications that significantly increase the melting temperature ( $T_m$ ) of DNA/RNA duplex such as LNA.

Once all the descriptors have been calculated for each potential ASO or siRNA the design workflow becomes a process of selection. An example of the PFRED selection interface for siRNA design is shown in Figure 3, step 4. Oligonucleotide selections are made by first filtering out those compounds which align with: 1) known human SNPs, 2) sequence motifs associated with non-specific binding (such as polyG, polyT) and 3) low complexity sequences that have a high number of perfect off-target matches. An additional filter for ortholog transcript matches can be added depending on the experimental design requirement. Also, if predicted activities are available, they may be included as a selection criterion (models may depend on a specific oligonucleotide length or modification type). Other oligonucleotide properties such as number of off-target matches and diversity of location of each oligonucleotide along the gene sequence can be used to finalize the selection of desired compounds for a given experiment. The final selected compounds can then be exported in text format in preparation for synthesis or further analysis outside of the tool.

## Conclusions

We present here an informatics platform for the design of antisense and siRNA oligonucleotides. The system includes a basic set of design algorithms but is extensible such that experienced developers can add novel or proprietary descriptors. Workflows are included which walk researchers through the process of identifying a target sequence and choosing the design criteria to be applied in selection of candidate molecules. A series of oligonucleotide visualization and plotting algorithms are included along with spreadsheet functions allowing data import/export, formulas, and conditional formatting. Links to the source code and hosted versions can be found at <https://github.com/pfred>.

## Acknowledgements

The authors would like to thank Art Kreig, Matteo Di Tommaso and Enoch Huang for their support of this work. Dario Cruz acknowledges support from the Biogen Internship & Co-op program.

## References

1. Bennett CF, Swayze EE. RNA Targeting Therapeutics: Molecular Mechanisms of Antisense Oligonucleotides as a Therapeutic Platform. *Annual Review of Pharmacology and Toxicology*. 2010;50: 259–293. doi:10.1146/annurev.pharmtox.010909.105654
2. Watts JK, Corey DR. Clinical status of duplex RNA. *Bioorganic & Medicinal Chemistry Letters*. 2010;20: 3203–3207. doi:10.1016/j.bmcl.2010.03.109
3. Akdim F, Visser ME, Tribble DL, Baker BF, Stroes ESG, Yu R, et al. Effect of Mipomersen, an Apolipoprotein B Synthesis Inhibitor, on Low-Density Lipoprotein Cholesterol in Patients With Familial Hypercholesterolemia. *American Journal of Cardiology*. 2010;105: 1413–1419. doi:10.1016/j.amjcard.2010.01.003
4. Raal FJ, Santos RD, Blom DJ, Marais AD, Charng M-J, Cromwell WC, et al. Mipomersen, an apolipoprotein B synthesis inhibitor, for lowering of LDL cholesterol concentrations in patients with homozygous familial hypercholesterolaemia: a randomised, double-blind, placebo-controlled trial. *The Lancet*. 2010;375: 998–1006. doi:10.1016/S0140-6736(10)60284-X
5. Fogacci F, Ferri N, Toth PP, Ruscica M, Corsini A, Cicero AFG. Efficacy and Safety of Mipomersen: A Systematic Review and Meta-Analysis of Randomized Clinical Trials. *Drugs*. 2019;79: 751–766. doi:10.1007/s40265-019-01114-z
6. Benson MD, Waddington-Cruz M, Berk JL, Polydefkis M, Dyck PJ, Wang AK, et al. Inotersen Treatment for Patients with Hereditary Transthyretin Amyloidosis. *New England Journal of Medicine*. 2018;379: 22–31. doi:10.1056/NEJMoa1716793
7. Adams D, Gonzalez-Duarte A, O’Riordan WD, Yang C-C, Ueda M, Kristen AV, et al. Patisiran, an RNAi Therapeutic, for Hereditary Transthyretin Amyloidosis. *New England Journal of Medicine*. 2018;379: 11–21. doi:10.1056/NEJMoa1716153
8. Agarwal S, Simon AR, Goel V, Habtemariam BA, Clausen VA, Kim JB, et al. Pharmacokinetics and Pharmacodynamics of the Small Interfering Ribonucleic Acid, Givosiran, in Patients With Acute Hepatic Porphyria. *Clinical Pharmacology & Therapeutics*. n/a. doi:10.1002/cpt.1802
9. Prakash V. Spinraza—a rare disease success story. *Gene Ther*. 2017;24: 497–497. doi:10.1038/gt.2017.59
10. DeVincenzo J, Lambkin-Williams R, Wilkinson T, Cehelsky J, Nochur S, Walsh E, et al. A randomized, double-blind, placebo-controlled study of an RNAi-based therapy directed against respiratory syncytial virus. *PNAS*. 2010;107: 8800–8805. doi:10.1073/pnas.0912186107
11. Study Evaluating Efficacy and Safety of PF-04523655 Versus Laser in Subjects With Diabetic Macular Edema - Full Text View - ClinicalTrials.gov. [cited 6 Dec 2019]. Available: <https://clinicaltrials.gov/ct2/show/NCT00701181>
12. Efficient Synthesis and Biological Evaluation of 5'-GalNAc Conjugated Antisense Oligonucleotides. [cited 6 Dec 2019]. doi:10.1021/acs.bioconjchem.5b00265

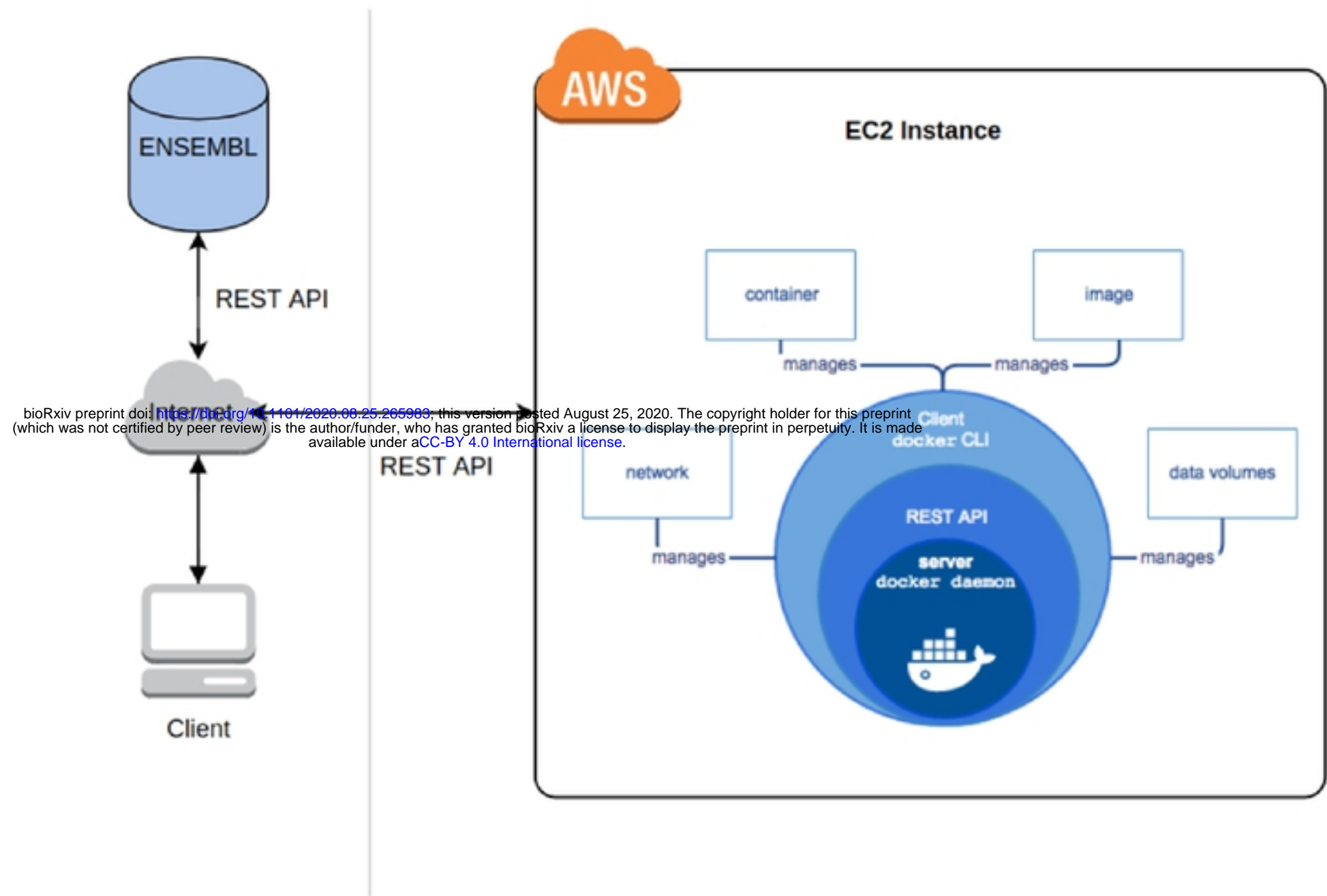


13. Sehgal A, Barros S, Ivanciu L, Cooley B, Qin J, Racie T, et al. An RNAi therapeutic targeting antithrombin to rebalance the coagulation system and promote hemostasis in hemophilia. *Nat Med.* 2015;21: 492–497. doi:10.1038/nm.3847
14. Hoy SM. Patisiran: First Global Approval. *Drugs.* 2018;78: 1625–1631. doi:10.1007/s40265-018-0983-6
15. Cui W, Ning J, Naik UP, Duncan MK. OptiRNAi, an RNAi design tool. *Computer Methods and Programs in Biomedicine.* 2004;75: 67–73. doi:10.1016/j.cmpb.2003.09.002
16. Katoh T, Suzuki T. Specific residues at every third position of siRNA shape its efficient RNAi activity. *Nucleic Acids Res.* 2007;35: e27. doi:10.1093/nar/gkl1120
17. Vert J-P, Foveau N, Lajaunie C, Vandenbrouck Y. An accurate and interpretable model for siRNA efficacy prediction. *BMC Bioinformatics.* 2006;7: 520. doi:10.1186/1471-2105-7-520
18. 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;35: e123. doi:10.1093/nar/gkm699
19. Mysara M, Garibaldi JM, ElHefnawi M. MysiRNA-Designer: A Workflow for Efficient siRNA Design. *PLOS ONE.* 2011;6: e25642. doi:10.1371/journal.pone.0025642
20. 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;32: W124–W129. doi:10.1093/nar/gkh442
21. Naito Y, Yoshimura J, Morishita S, Ui-Tei K. siDirect 2.0: updated software for designing functional siRNA with reduced seed-dependent off-target effect. *BMC Bioinformatics.* 2009;10: 392. doi:10.1186/1471-2105-10-392
22. Yuan B, Latek R, Hossbach M, Tuschl T, Lewitter F. siRNA Selection Server: an automated siRNA oligonucleotide prediction server. *Nucleic Acids Res.* 2004;32: W130–W134. doi:10.1093/nar/gkh366
23. Holen T. Efficient prediction of siRNAs with siRNARules 1.0: An open-source JAVA approach to siRNA algorithms. *RNA.* 2006;12: 1620–1625. doi:10.1261/rna.81006
24. 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;10. doi:10.3389/fpls.2019.01023
25. siDESIGN Center. Horizon Discovery; Available: <https://horizondiscovery.com/en/products/tools/siDESIGN-Center>
26. GenScript siRNA Target Finder. Genscript; Available: <https://www.genscript.com/tools/sirna-target-finder>
27. BLOCK-iT™ RNAi Designer. ThermoFisher Scientific; Available: <https://rnaidesigner.thermofisher.com/rnaiexpress/>

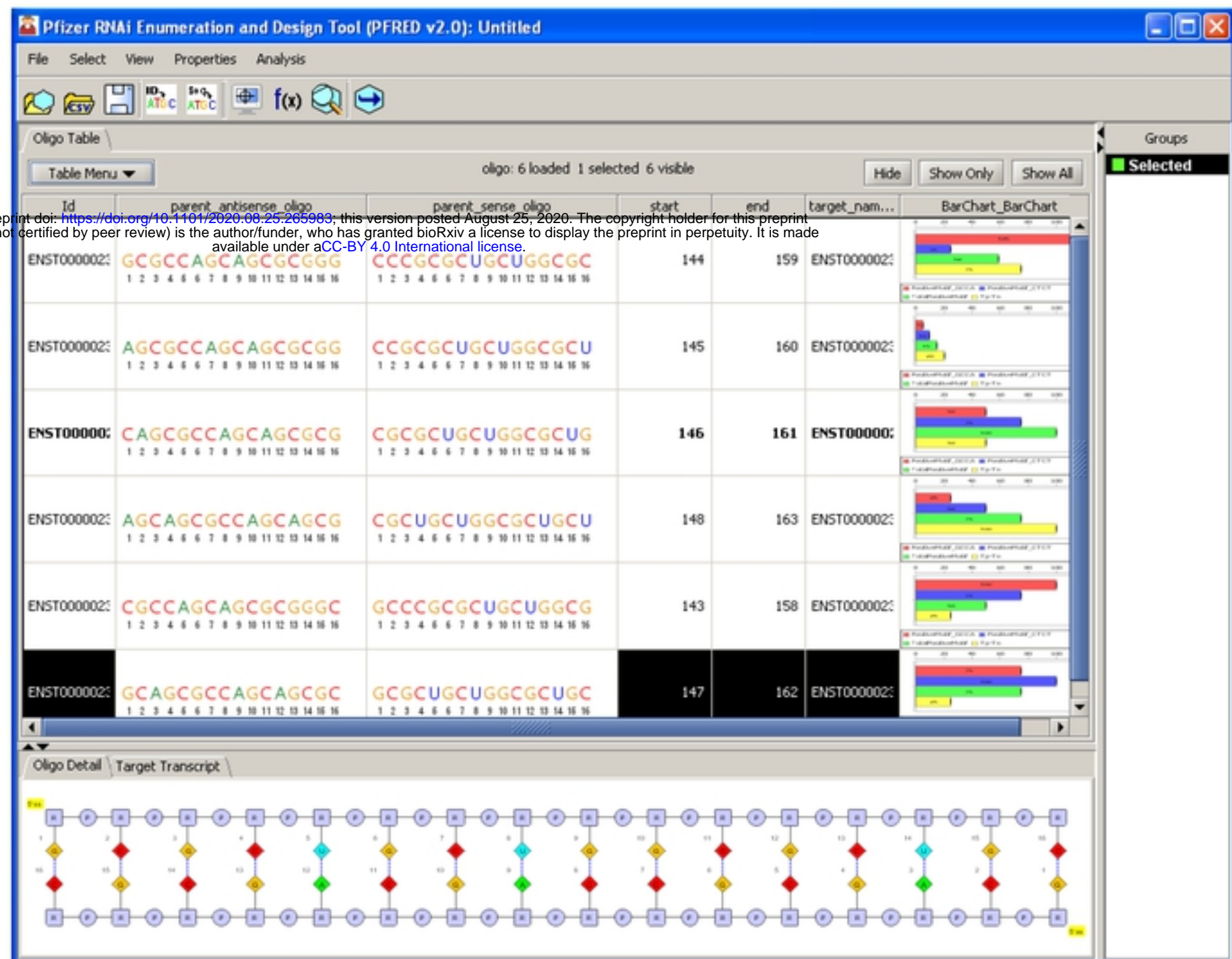
28. Zhang T, Li H, Xi H, Stanton RV, Rotstein SH. HELM: A Hierarchical Notation Language for Complex Biomolecule Structure Representation. *J Chem Inf Model.* 2012;52: 2796–2806. doi:10.1021/ci3001925
29. Milton J, Zhang T, Bellamy C, Swayze E, Hart C, Weisser M, et al. HELM Software for Biopolymers. *Journal of Chemical Information and Modeling.* 2017;57: 1233–1239. doi:10.1021/acs.jcim.6b00442
30. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology.* 2009;10: R25. doi:10.1186/gb-2009-10-3-r25
31. Elbashir SM, Lendeckel W, Tuschl T. RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev.* 2001;15: 188–200. doi:10.1101/gad.862301
32. Elbashir SM, Martinez J, Patkaniowska A, Lendeckel W, Tuschl T. Functional anatomy of siRNAs for mediating efficient RNAi in *Drosophila melanogaster* embryo lysate. *The EMBO Journal.* 2001;20: 6877–6888. doi:10.1093/emboj/20.23.6877
33. Reynolds A, Leake D, Boese Q, Scaringe S, Marshall WS, Khvorova A. Rational siRNA design for RNA interference. *Nat Biotechnol.* 2004;22: 326–330. doi:10.1038/nbt936
34. Sætrom P, Snøve O. A comparison of siRNA efficacy predictors. *Biochemical and Biophysical Research Communications.* 2004;321: 247–253. doi:10.1016/j.bbrc.2004.06.116
35. Teramoto R, Aoki M, Kimura T, Kanaoka M. Prediction of siRNA functionality using generalized string kernel and support vector machine. *FEBS Letters.* 2005;579: 2878–2882. doi:10.1016/j.febslet.2005.04.045
36. Huesken D, Lange J, Mickanin C, Weiler J, Asselbergs F, Warner J, et al. Design of a genome-wide siRNA library using an artificial neural network. *Nature Biotechnology.* 2005;23: 995–1001. doi:10.1038/nbt1118
37. Shabalina SA, Spiridonov AN, Ogurtsov AY. Computational models with thermodynamic and composition features improve siRNA design. *BMC Bioinformatics.* 2006;7: 65. doi:10.1186/1471-2105-7-65
38. Matveeva O, Nechipurenko Y, Rossi L, Moore B, Saetrom P, Ogurtsov AY, et al. Comparison of approaches for rational siRNA design leading to a new efficient and transparent method. *Nucleic Acids Research.* 2007;35: e63–e63. doi:10.1093/nar/gkm088
39. Hajeri PB, Singh SK. siRNAs: their potential as therapeutic agents – Part I. Designing of siRNAs. *Drug Discovery Today.* 2009;14: 851–858. doi:10.1016/j.drudis.2009.06.001
40. Sciabola S, Cao Q, Orozco M, Faustino I, Stanton RV. Improved nucleic acid descriptors for siRNA efficacy prediction. *Nucleic Acids Research.* 2013;41: 1383–1394. doi:10.1093/nar/gks1191
41. Matveeva OV. Identification of sequence motifs in oligonucleotides whose presence is correlated with antisense activity. *Nucleic Acids Research.* 2000;28: 2862–2865. doi:10.1093/nar/28.15.2862

42. Liao L, Li Z. Correlation between gene silencing activity and structural features of antisense oligodeoxynucleotides and target RNA. *In Silico Biol.* 2007;7: 527–534.
43. Far RK-K, Leppert J, Frank K, Sczakiel G. Technical Improvements in the Computational Target Search for Antisense Oligonucleotides. *Oligonucleotides.* 2005;15: 223–233. doi:10.1089/oli.2005.15.223
44. Shao Y, Wu Y, Chan CY, McDonough K, Ding Y. Rational design and rapid screening of antisense oligonucleotides for prokaryotic gene modulation. *Nucleic Acids Res.* 2006;34: 5660–5669. doi:10.1093/nar/gkl715
45. Matveeva OV. Thermodynamic calculations and statistical correlations for oligo-probes design. *Nucleic Acids Research.* 2003;31: 4211–4217. doi:10.1093/nar/gkg476
46. Lu ZJ, Mathews DH. OligoWalk: an online siRNA design tool utilizing hybridization thermodynamics. *Nucleic Acids Res.* 2008;36: W104–W108. doi:10.1093/nar/gkn250
47. Chalk AM, Sonnhammer ELL. Computational antisense oligo prediction with a neural network model. *Bioinformatics.* 2002;18: 1567–1575. doi:10.1093/bioinformatics/18.12.1567
48. Camps-Valls G, Chalk AM, Serrano-López AJ, Martín-Guerrero JD, Sonnhammer EL. Profiled support vector machines for antisense oligonucleotide efficacy prediction. *BMC Bioinformatics.* 2004;5: 135. doi:10.1186/1471-2105-5-135
49. Sætrom P. Predicting the efficacy of short oligonucleotides in antisense and RNAi experiments with boosted genetic programming. *Bioinformatics.* 2004;20: 3055–3063. doi:10.1093/bioinformatics/bth364
50. Noy A, Luque FJ, Orozco M. Theoretical Analysis of Antisense Duplexes: Determinants of the RNase H Susceptibility. *J Am Chem Soc.* 2008;130: 3486–3496. doi:10.1021/ja076734u
51. Giddings MC, Matveeva OV, Atkins JF, Gesteland RF. ODNBase-a web database for antisense oligonucleotide effectiveness studies. *Bioinformatics.* 2000;16: 843–844. doi:10.1093/bioinformatics/16.9.843
52. Bo X. AOBBase: a database for antisense oligonucleotides selection and design. *Nucleic Acids Research.* 2006;34: D664–D667. doi:10.1093/nar/gkj065

**Schema 1.** PFRED client-server architecture based on the AWS and Docker containerization.

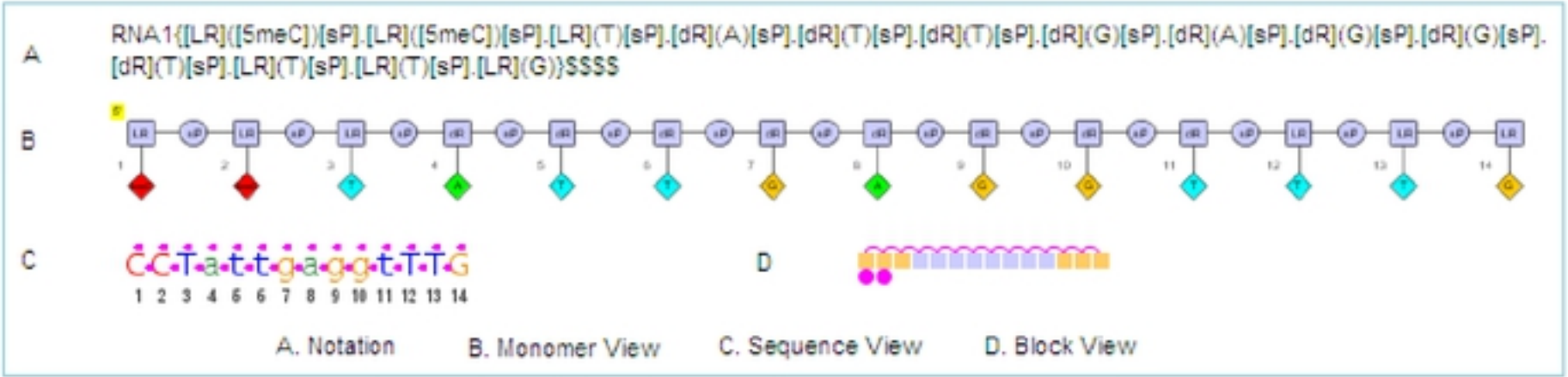


**Figure 1.** PFRED user interface. The main “Oligo Table” stores all the data generated through the oligo enumeration workflow (oligo sequences with their properties) and provides access to common spreadsheet functions. It is possible to assign selected oligo sequences to specific groups which can be shown and accessed under “Groups” on the right-side panel. More detailed information about the oligo sequence notation and the primary mRNA transcript can be visualized at the bottom of the user interface by enabling the “Oligo Detail” and “Target Transcript” panels.





**Figure 2.** Alternative representations of a short antisense gapmer sequence using the HELM notation (A) and different sequence visualization options (B-D).



**Figure 3.** Oligo Enumeration and Annotation workflow. The process of designing an ASO or siRNA in PFRED starts with the Advanced Oligo Enumerator (1) by providing the Ensembl ID of a target of interest and the list of ortholog species to design against. The Retrieve Ortholog Sequence button (2) uses the Ensembl REST API to communicate with the public Ensembl database to retrieve all the sequence information related to the gene of interest (splice variants and ortholog transcripts). At this point the user has the option to choose the Design type (siRNA design or AntiSense design) as well as additional oligo annotations (Off-target and Efficacy predictions) before triggering the Enumerate & Annotate button (3). Once all the descriptors have been calculated for each potential ASO or siRNA the design workflow becomes a process of selection (4).

bioRxiv preprint doi: <https://doi.org/10.1101/2020.08.25.265983>; this version posted August 25, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

