Sfold Tools for microRNA Target Prediction

William Rennie[1], Shaveta Kanoria[1], Jun Lu[2] and Ye Ding[1*]

*Correspondence:

Email: ye.ding@health.ny.gov; Phone: (518) 486 1719; Fax: (518) 474 3183

[1]Wadsworth Center, New York State Department of Health, 150 New Scotland Avenue, Albany, NY 12208, USA

[2] Department of Genetics and Yale Stem Cell Center, Yale University, New Haven, CT 06520, USA

Abstract

Computational prediction of miRNA binding sites on target mRNAs facilitates experimental investigation of miRNA functions. In this chapter, we describe the command-line use of STarMir, an application module of the Sfold RNA package. STarMir performs miRNA binding site predictions for target RNAs. Output data include comprehensive sequence, thermodynamic and target structure features, a logistic probability measuring confidence for each predicted site. and a quantitative score. Based on a large 3' UTR assay dataset, this score provides a quantitative measure of the overall regulatory effects of both seed and seedless sites on the target. Sfold is

now on Wikipedia at https://en.wikipedia.org/wiki/Sfold. STarMir and the rest of the Sfold package can be downloaded from GitHub at https://github.com/Ding-RNA-Lab/Sfold.

Keywords

miRNA, CLIP, target mRNA, RNA secondary structure, miRNA binding site, efficacy, quantitative score

Running head: Sfold microRNA target prediction

1. Introduction

MicroRNAs (miRNAs) are a class of naturally occurring, small non-coding RNAs (ncRNAs) of ~22 nucleotides (nt) in length, found in plants, animals and some viruses [1,2] A mature miRNA recognizes a target by guiding the RNA-induced silencing complex (RISC) to hybridize with partially complementary sequences typically in the 3' untranslated regions (3' UTRs) of the target mRNAs. miRNA binding induces translational repression and/or mRNA degradation of the target mRNA [1,2]. miRNA mediated gene regulation is extensive, as one miRNA may regulate hundreds of targets, whereas an individual mRNA can be targeted by multiple miRNAs [3]. miRNAs play important roles in various biological processes such as development, differentiation, apoptosis and proliferation [2,4]. Moreover, dysregulation in miRNA activity has been associated with numerous human diseases [5,6].

Most existing algorithms for miRNA target prediction are primarily based on the seed rule [7]. With the development of the cross-linking immunoprecipitation (CLIP) technique [8], it has become possible to identify short Argonaute (AGO) crosslinked sequences containing miRNA binding sites. CLIP involves UV irradiation of tissues, organisms, or cells to covalently

crosslink miRNA targets to the AGO proteins (the catalytic component of the RISC complex). The crosslinked RNAs are shortened by partial digestion with RNase to ~50 nt and further amplified by RT-PCR. The shortened RNA fragments are then sequenced to identify AGO tags containing miRNA binding sites on the target mRNAs.

Several CLIP studies have been published, including HITS-CLIP for the mouse brain [8], PAR-CLIP in human cell lines [9], variants of PAR-CLIP [10], and a study in *C. elegans* [11]. These CLIP studies provide a genome wide map of miRNA target interactions by generating short target fragments containing miRNA binding sites. The high throughput data from the CLIP studies were successfully utilized in the development of logistic models for improved miRNA binding site predictions [12]. These models are based on a comprehensive list of sequence, thermodynamic and target structure features found to be enriched for miRNA binding sites identified by CLIP. They were validated through intra-dataset, inter-dataset as well as cross-species validations [12].

The models have been implemented into the STarMir application module of the Sfold RNA package, which predicts miRNA binding sites on a target mRNA [13]. The use of STarMir through the Sfold web server was described in an earlier chapter [14]. In this chapter, we describe the command-line use of STarMir after installing the Sfold package on a local computer.

2. Materials

Sfold is now on Wikipedia at https://en.wikipedia.org/wiki/Sfold. **Figure 1** shows the core of the Sfold Wikipedia page. STarMir, along with the rest of Sfold package, can be downloaded from GitHub at https://github.com/Ding-RNA-Lab/Sfold. **Figure 2** is a screen shot of the files and folders in the Sfold GitHub repository.

3. Methods

STarMir now provides a quantitative score to assess the combined regulatory effects of multiple seed and seedless sites. This score, developed based on an high-throughput luciferase reporter data set that includes 461 miRNAs and 11 3′ UTRs for a total of 4994 miRNA:3′ UTR pairs, provides a quantitative measure of the overall regulatory effects of both seed and seedless sites of one miRNA on a given target 3′ UTR [15].

3.1 Operating system and Github Sfold package download

For running on a user's local computer, STarMir is only available for the Linux operating system. It runs it on the Ubuntu distribution, version 18.04 or newer. Unbuntu's tutorial for installation is available at https://ubuntu.com/tutorials/install-ubuntu-desktop#1-overview.

The zip archive for STarMir can be downloaded from https://github.com/Ding-RNA-Lab/Sfold, by selecting the green "Code" button. Experienced GitHub users can also "fork" the archive. The zip file should be saved in a directory where the user plans to install the archive. The zip file can be extracted using an unzip utility tool. After unzipping, a directory called "Sfold-main" is created containing the components of the Sfold package including the STarMir program.

3.2 Installation of other required software and utilities

3.2.1 RNAhybrid

RNAhybrid [16] is used by STarMir to create a set of candidate binding sites. It can be downloaded from https://bibiserv.cebitec.uni-bielefeld.de/rnahybrid. Download and installation instructions are available on this site.

3.2.2 R

R is a statistical package (https://www.r-project.org) used by STarMir for executing the prediction models. R is usually available on most Linux systems and can be downloaded and installed using the Linux distributions package manager. The installed executable must be globally accessible on the host computer.

3.2.3 Perl

The bulk of the STarMir code was written in Perl. Version 5 or newer is required. Perl is usually pre-installed on any Linux system. Two Perl modules used by STarMir, Bio::Seq and Bio::SeqIO, must be installed, typically using CPAN. The main Bioperl installation page provides recommendations and links to the main CPAN page. The URL is https://github.com/bioperl/bioperl-live/blob/master/README.md. The CPAN package is included with the Perl installation.

3.2.4 Shell

STarMir can be run on the command line within a console window. It was developed using bash and thus should run with any shell.

3.3 Configuration

3.3.1 *Sfold*

To configure sfold, the user must enter the '`Sfold-main/bin`' directory. This directory contains the Sfold executable and a document called '`Running_Sfold`', which provides instructions for configuring and running Sfold. The document also explains how to use the testing utility to confirm that the Sfold package is installed correctly.

3.3.2 STarMir

STarMir does not have an automated configuration utility. The user must manually edit a

few parameter files located in the 'Sfold-main/STarMir' directory. The README file in that directory contains relevant information.

The file 'starmir_param.pl' needs to be edited. The user needs to set the $RNA_bindir path to the location of the RNAhybrid executable. Typing "which RNAhybrid" on the command line will display the bin directory of the RNAhybrid executable. Only the directory (path) is needed, not the program name.

The user must also set the path to disruptEn, a program that calculates the free energy required to open a local structure. This program is part of the Sfold package, and the binary can be found in the Sfold-main/bin directory. To ensure proper functionality, use the full path, such as: "/home/williamrennie/development/Sfold_main/bin/". The final slash is important and must be included. Additionally, the line "$SFOLDBIN="/home/bill/Desktop/Sfold-main/bin/" must be set to the same Sfold-main/bin directory specified in the previous above.

The steps above complete the installation and configuration of the programs required for STarMir.

3.4 Procedures for executing STarMir

3.4.1 Executing Sfold

Sfold must be run first for predicting target structures. Running Sfold is straightforward, and the default parameters are sufficient for STarMir to use predicted structures. It has many configuration options which can be viewed by running Sfold without any arguments. Sfold should always be executed using the full path to the executable, for example "/home/bill/Sfold-main/sfold".

- The input to Sfold is a file containing a single RNA sequence in FASTA format.

• Sfold produces a directory of output files, some of which are required by STarMir. The user shall select the directory, which need to be passed to STarMir.

• An example command to execute Sfold is "`/home/williamrennie/Sfold-main/bin/sfold -o myoutputdir myseq.fasta`"

• Running Sfold can take anywhere from a few minutes to a few hours depending on both the computational power of the host computer and the length of the target sequence (see **Note 1**).

3.4.2 Executing STarMir

• STarMir is a system of Perl scripts and helper application that predict and rank miRNA binding sites on a target mRNA.

• The main command script is located in the distribution's `Sfold main/STarMir` directory of the distribution. The user can run the program through that script.

• The script MUST be executed from the directory that contains it. Otherwise, it may not find all the necessary supporting scripts and may not run correctly.

• The main command script "`starmir_research.pl`", requires nine arguments. Arguments that point to files MUST give the full path to the file (absolute path). The code has no facility for deducing the path to the file. The arguments, in the order they appear, are the following:

3.4.2.1 Required arguments (in order)

1. <working directory>: The directory where output files will be stored. By default, the intermediate output files will also be saved here. This argument must end with a backslash (\) to set an absolute path to the directory.

2. <miRNA file>: A file containing one or more miRNA sequences in FASTA format (absolute path).

3. <mRNA file>: A file containing a single mRNA sequence in FASTA format (absolute path).

4. <sfold output directory>: The directory containing Sfold results for the mRNA sequence can be found (absolute path, must end with a forward slash (/).

5. <target species>: The species name used by RNAhybrid. Must be one of fly, human or worm (these are only species supported by RNAhybrid).

6. <model species>: The species used by the STarMir prediction model. It must be one of human, mouse, or worm, STarMir predicts miRNA binding sites using models built for human (Homo sapiens), mouse (*Mus musculus*) and worm (*Caenorhabdites elegans*). These models were trained on V-CLIP data for human [9], HITS-CLIP data for mouse [8] and ALG-1 CLIP data for worm [11]. The human and mouse models were cross-validated and can be broadly used for other species [12].

7. <CDS start>: The start of the coding region.

8. <CDS end>: The end of the coding region.

3.4.2.2 Optional argument

An optional ninth argument can be either 1 or 0. Setting this argument to 1 deletes all the intermediate files, whereas the default (0) preserves the intermediate files.

The command line, which must be run in the same directory as the scripts is:

```
<path to script directory>/starmir_research.pl <working
directory> <miRNA file> <mRNA file> <sfold output
directory> <target species> <model species> <CDS start>
<CDS end> <optional 1 for deleting intermediate files>
```

For best practice, use absolute path names for all input files and directories. Below is an example of STarMir command line run with intermediate files deleted:

```
./starmir_research.pl Data/myseq mirnas.fasta myseq.fasta
~/runs/sfoldDataMyseq/ human human 234 874 1
```

3.2 Output

A successful local installation and execution of STarMir will generate separate final output files for miRNA binding sites in the 5′ UTR, the coding region and the 3′ UTR, and for both seed and seedless sites (**Table 1)**. These output files are prefixed with 'Final-'. If no prediction is made for a specific site, e.g.., in the case of the lack of a single seed, the corresponding file will not be generated.

Examples of seed and seedless output files are presented in **Figure 3**. For the seed output file, several seed-specific features (e.g., Seed_Access) are provided (see **Figure 3A**, and **Table 2**). In each file, the binding sites are listed in the descending order of their logistic probabilities. In addition, a file containing miRNA:target hybrid conformations in text format is generated (**Table 1**, **Figure 4**).

Publication quality hybrid diagrams in PDF format can also be produced using the "`create_PDF.pl`" script, located in the STarMir subdirectory of the Sfold GitHub depository. The input file for this script is "`Total-En-Hyb-Fil-`*`mRNA id`*`.out`" file (where "mRNA

id" corresponds to the target name). This file will be located in the output directory specified during the execution of STarMir. The resulting PDF file will be named "`SiteX.pdf`", where X refers to the specific site number provided for the `create_PDF.pl` script. For example, the command "`./create_pdf.pl ../../OutputDir2/TotalEn-Hyb-Fil-NM_017589.4.out 3`" will generate the PDF diagram for site 3, saved as "`Site3.pdf`" in the same directory in which the command was executed. Examples of diagrams for both a seed site and a seedless site are shown in **Figure 5**, with the miRNA seed highlighted.

For each binding site, STarMir provides a comprehensive list of site features. Several are unique to STarMir: structure-based free-energy measures, a logistic probability, and a score for the miRNA:target pair (**Table 2**). The logistic probability is a measure of confidence for a predicted site [17] . The score is a measure of predicted regulatory efficacy of the miRNA on the target, based on a linear combination of the contributions from both seed and seedless binding sites [15]. Examples of these scores are given in **Table 3**. Notable, although 8-mer sites are often considered the most effective among all types of sites, they do not necessarily ensure high scores which are predictive of effective regulation. On the other hand, large numbers of seedless sites can influence the combined score, potentially leading to a strong regulatory impact.

4. Notes

1. The Sfold/STarMir calculations arte both CPU and memory intensive. For example, the Sfold web server utilizes a ThinkStation P700 Tower equipped with two Xeon E5-2650 v3 processors (2.3 GHz), 128 GB of memory, and an 8 TB SATA disk. Typical calculation times are as follows: three minutes for 500 nts, five minutes for 1,000 nts, 30 minutes for 2,000 nts, two hours for 3,000 nts, five hours for 4,000 nts and nine hours for 5,000 nts.

2. The logistic probability is a measure of confidence in a predicted binding site, calculated using models developed from crosslinking immunoprecipitation (CLIP) studies. While it is challenging to directly assign a specific level of contribution to the overall functional efficacy based on the logistic probability of a given binding site, we recommend considering a logistic probability 0.75 or higher for further experimental investigation.

3. The quantitative score was developed using a large reporter expression dataset. This dataset indicated that when the score is 0.09 or higher (≥ 0.09), 54% of the targets were effectively down-regulated.

Acknowledgment

This work is supported in part by the National Institutes of Health (grant GM138856 to J. L. and Y.D.).

5. References

1. Ambros V (2004) The functions of animal microRNAs. Nature 431 (7006):350-355

2. Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. In: Cell, vol 116. vol 2. pp 281-297

3. Friedman RC, Farh KK, Burge CB, Bartel DP (2009) Most mammalian mRNAs are conserved targets of microRNAs. Genome Res 19 (1):92-105. doi:10.1101/gr.082701.108

4. Harfe BD (2005) MicroRNAs in vertebrate development. In: Curr Opin Genet Dev, vol 15. vol 4. pp 410-415. doi:10.1016/j.gde.2005.06.012

5. Esau CC, Monia BP (2007) Therapeutic potential for microRNAs. In: Adv Drug Deliv Rev, vol 59. vol 2-3. pp 101-114. doi:10.1016/j.addr.2007.03.007

6. Erson AE, Petty EM (2008) MicroRNAs in development and disease. In: Clin Genet, vol 74. vol 4. pp 296-306. doi:10.1111/j.1399-0004.2008.01076.x

7. Lewis BP, Burge CB, Bartel DP (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell 120 (1):15-20

8. Chi SW, Zang JB, Mele A, Darnell RB (2009) Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. Nature 460 (7254):479-486. doi:10.1038/nature08170

9. Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M, Jr., Jungkamp AC, Munschauer M, Ulrich A, Wardle GS, Dewell S, Zavolan M, Tuschl T (2010) Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell 141 (1):129-141. doi:10.1016/j.cell.2010.03.009

10. Kishore S, Jaskiewicz L, Burger L, Hausser J, Khorshid M, Zavolan M (2011) A quantitative analysis of CLIP methods for identifying binding sites of RNA-binding proteins. Nat Methods 8 (7):559-564. doi:10.1038/nmeth.1608

11. Zisoulis DG, Lovci MT, Wilbert ML, Hutt KR, Liang TY, Pasquinelli AE, Yeo GW (2010) Comprehensive discovery of endogenous Argonaute binding sites in Caenorhabditis elegans. In: Nat Struct Mol Biol, vol 17. vol 2. pp 173-179. doi:10.1038/nsmb.1745

12. Liu C, Mallick B, Long D, Rennie WA, Wolenc A, Carmack CS, Ding Y (2013) CLIP-based prediction of mammalian microRNA binding sites. In: Nucleic Acids Res, vol 41. vol 14. p e138. doi:10.1093/nar/gkt435

13. Rennie W, Liu C, Carmack CS, Wolenc A, Kanoria S, Lu J, Long D, Ding Y (2014) STarMir: a web server for prediction of microRNA binding sites. Nucleic Acids Res 42 (Web Server issue):W114-118. doi:10.1093/nar/gku376

14. Rennie W, Kanoria S, Liu C, Carmack CS, Lu J, Ding Y (2019) Sfold Tools for MicroRNA Target Prediction. Methods Mol Biol 1970:31-42. doi:10.1007/978-1-4939-9207-2_3

15. Mastriano S, Kanoria S, Rennie W, Liu C, Li D, Cheng J, Ding Y, Lu J (2024) High-Throughput Quantification of miRNA-3'-Untranslated-Region Regulatory Effects. bioRxiv. doi:10.1101/2024.12.05.626985

16. Rehmsmeier M, Steffen P, Hochsmann M, Giegerich R (2004) Fast and effective prediction of microRNA/target duplexes. RNA 10 (10):1507-1517. doi:10.1261/rna.5248604

17. Liu C, Mallick B, Long D, Rennie WA, Wolenc A, Carmack CS, Ding Y (2013) CLIP-based prediction of mammalian microRNA binding sites. Nucleic Acids Res 41 (14):e138. doi:10.1093/nar/gkt435

18. Bartel DP (2009) MicroRNAs: target recognition and regulatory functions. Cell 136 (2):215-233. doi:10.1016/j.cell.2009.01.002

19. Long D, Lee R, Williams P, Chan CY, Ambros V, Ding Y (2007) Potent effect of target structure on microRNA function. Nat Struct Mol Biol 14 (4):287-294. doi:10.1038/nsmb1226

Figure legends

Figure 1. The core section of the Sfold Wikipedia page.

Figure 2. The files and folders in Sfold GitHub repository.

Figure 3 An example of the seed site output file that provide several seed specific features (**3A**); and an example of the output file for seedless sites (**3B**).

Figure 4. Hybridization diagrams in text format for the first five binding sites predicted for hsa-miR-501-3p on target NM_017589.4 (gene BTG4).

Figure 5. Examples of high-quality hybridization diagrams, one for a hsa-miR-501-3p seed site on NM_017589.4 (**5A**), one for a seedless site (**5B**). Note that the seed region of the miRNA is highlighted in red.

Tables

Table 1. Final prediction files from STarMir

File name	File content description
Final-3pUTR_seed_sites.txt	Features and predictions for 3′ UTR seed sites
Final-3pUTR_seedless_sites.txt	Features and predictions for 3′ UTR seedless sites
Final-CDS_seed_sites.txt	Features and predictions for CDS seed sites
Final-CDS_seedless_sites.txt	Features and predictions for CDS seedless sites
Final-5pUTR_seed_sites.txt	Features and predictions for 5′ UTR seed sites
Final-5pUTR_seedless_sites.txt	Features and predictions for 5′ UTR seedless sites
conformation_all_sites.out	Predicted miRNA-target conformations for all sites

Table 2 Sequence, structural features and predictions for target sites from STarMir	
Site ID	Predicted sites are sequentially numbered along the target sequence
Target	Accession number of the target mRNA
miRNA	Name of the microRNA (miRNA)
Target_Len	Length of the target
Site_Position	Start and end position of the target region (site) predicted to be bound by miRNA
Seed_Position	Start and end position of the target sub-region complementary to the miRNA seed (i.e. positions 2-7/8 of the miRNA)
Seed_Type	6mer, offset 6mer, 7mer-A1, 7mer-m8, and 8mer seed sites [18]
Site_Access	A measure of structural accessibility as computed by the average probability of a nucleotide being single-stranded (i.e., unpaired) for the nucleotides in the predicted binding site
Seed_Access	A measure of structural accessibility as computed by the average of single-stranded probabilities of the nucleotides in the target sub-region complementary to the miRNA seed
Upstream_Access (# nt)	A measure of structural accessibility as computed by the average of single-stranded probabilities for the block of nucleotides upstream of the predicted binding site (# is the block size)
Dwstream_Access (# nt)	A measure of structural accessibility as computed by the average of single-stranded probabilities for the block of nucleotides downstream of the predicted binding site (# is the block size)
Upstream_AU (# nt)	Percentage of AU for the block of nucleotides upstream of the binding site (# is the block size)
Dwstream_AU (# nt)	Percentage of AU for the block of nucleotides downstream of the binding site (# is the block size)
Site_Location	Relative starting location of the predicted binding site along the length of the sequence (e.g., for 3′ UTR, 0 indicates the 5′ end of the UTR, and 1 corresponds to the 3′ end)
3′_bp	Presence of contiguous Watson Crick base pairing for miRNA nucleotide positions 12-17 (sites with 3′_bp are also called 3′ compensatory/supplementary sites) [18]
dG_hybrid = ΔG_{hybrid}	A measure of stability for miRNA:target hybrid as computed by RNAhybrid [16]
dG_nucl = ΔG_{nucl}	A measure of the potential of nucleation for miRNA:target hybridization [19]
dG_total = ΔG_{total}	A measure of the total energy change of the hybridization [19]
LogitProb	Probability of the site being an miRNA binding site as predicted by nonlinear logistic model [17]
Score	A quantitative measure of the overall regulatory effects of both seed and seedless sites on the target [15]

Table 3. Examples of quantitative scores with number of different types of binding sites

Transcript name	microRNA name	Number of 8mer sites	Number of 7mer sites	Number of 6mer sites	Number of seedless sites	Score
BTG4	hsa-miR-501-3p	0	1	0	22	0.040742190
hKLF4	hsa-mir-153-2	0	0	2	13	0.037326758
hDICER	hsa-let-7c	0	2	3	319	0.470307676
mBMI1	hsa-mir-154	1	0	4	77	0.144362275
hOCT4	hsa-mir-335	1	0	0	19	0.021329511
hLIN28	hsa-mir-9-2	0	1	6	196	0.282444367

Sfold is a software program developed to predict probable RNA secondary structures through structure ensemble sampling and centroid predictions[1][2] with a focus on assessment of RNA target accessibility,[3] for major applications to the rational design of siRNAs[4] in the suppression of gene expressions, and to the identification of targets for regulatory RNAs particularly microRNAs.[5][6]

Original author(s)	Ye Ding and Charles E. Lawrence
Developer(s)	Dang Long and Chaochun Liu (application modeling); Clarence Chan, Adam Wolenc, William A. Rennie and Charles S. Carmack (software development)
Initial release	1 April 2003; 21 years ago
Repository	github.com/Ding-RNA-Lab/Sfold
Operating system	Linux
Website	www.healthresearch.org/sfold-software-for-sirna/

Development [edit]

The core RNA secondary structure prediction algorithm is based on rigorous statistical (stochastic) sampling of Boltzmann ensemble of RNA secondary structures, enabling statistical characterization of any local structural features of potential interest to experimental investigators. In a review on nucleic acid structure and prediction,[7] the potential of structure sampling described in a prototype algorithm[8] was highlighted. With the publication of the mature algorithms for Sfold,[1][2] the sampling approach became the focus of a review[9] Both the sampling approach and the centroid predictions were discussed in a comprehensive review.[10] As an application module of the Sfold package, the STarMir program[11] has been widely used for its capability in modeling target accessibility.[6] STarMir was described in an independent study on microRNA target prediction[12] STarMir predictions have been used in an attempt to derive improved predictions.[13] Predictions by Sfold have led to new biological insights.[14] The novel ideas of ensemble sampling and centroids have been adopted by others not only for RNA problems, but also for other fundamental problems in computational biology and genomics.[15][16][17][18][19]

An implementation of stochastic sampling has been included in two widely used RNA software packages, RNA Structure[20] and the ViennaRNA Package,[21] which are also based on the Turner RNA thermodynamic parameters.[22] Sfold was featured on a *Nucleic Acids Research* cover,[23] and was highlighted in *Science* NetWatch.[24] The underlying novel model for STarMir[11] was featured in the Cell Biology section of *Nature Research Highlights*.[25]

Distribution [edit]

Sfold runs on Linux, and is freely available to the scientific community for non-commercial applications, and is available under license for commercial applications. Both the source code and the executables are available at GitHub.

External links [edit]

- Sfold GitHub repository
- Sfold commercial licensing

Figure 1

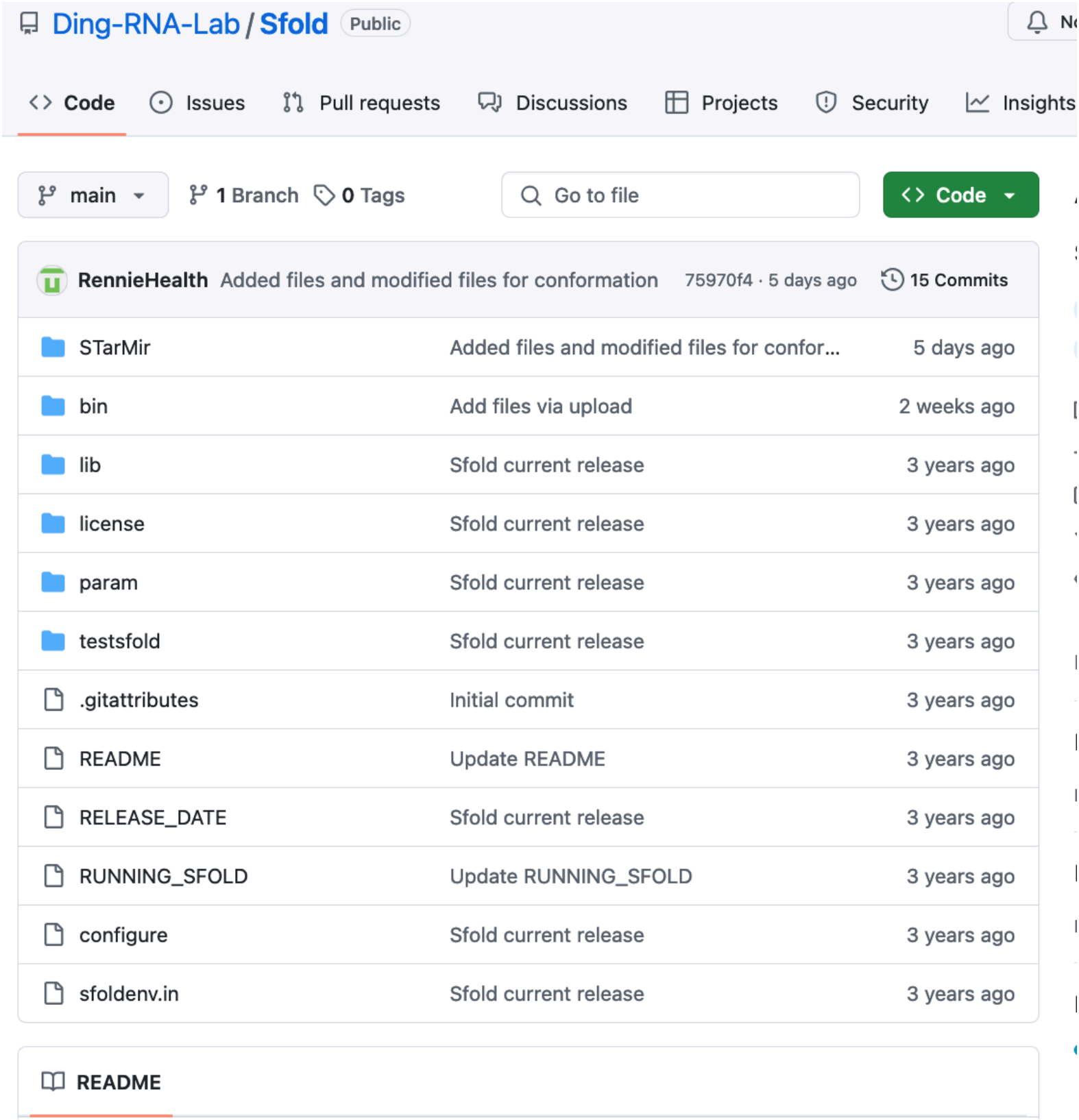

Ding-RNA-Lab / Sfold
Public
Code
Issues
Pull requests
Discussions
Projects
Security
Insights
main
1 Branch
0 Tags
Go to file
Code
RennieHealth Added files and modified files for conformation 75970f4 · 5 days ago 15 Commits
STarMir Added files and modified files for confor... 5 days ago
bin Add files via upload 2 weeks ago
lib Sfold current release 3 years ago
license Sfold current release 3 years ago
param Sfold current release 3 years ago
testsfold Sfold current release 3 years ago
.gitattributes Initial commit 3 years ago
README Update README 3 years ago
RELEASE_DATE Sfold current release 3 years ago
RUNNING_SFOLD Update RUNNING_SFOLD 3 years ago
configure Sfold current release 3 years ago
sfoldenv.in Sfold current release 3 years ago
README

Figure 2

a

```
"SiteID"        "Target"        "miRNA" "Target_Len"    "Site_Position"
"Seed_Position" "Seed_Type*"    "3'_bp" "dG_hybrid"     "dG_nucl"
"dG_total"      "Site_Access"   "Seed_Access"   "Upstream_Access(5nt)"
"Dwstream_Access(5nt)"  "Upstream_AU(5nt)"      "Dwstream_AU(5nt)"
"Upstream_Access(10nt)" "Dwstream_Access(10nt)" "Upstream_AU(10nt)"
"Dwstream_AU(10nt)"     "Upstream_Access(15nt)" "Dwstream_Access(15nt)"
"Upstream_AU(15nt)"     "Dwstream_AU(15nt)"     "Upstream_Access(20nt)"
"Dwstream_Access(20nt)" "Upstream_AU(20nt)"     "Dwstream_AU(20nt)"
"Upstream_Access(25nt)" "Dwstream_Access(25nt)" "Upstream_AU(25nt)"
"Dwstream_AU(25nt)"     "Upstream_Access(30nt)" "Dwstream_Access(30nt)"
"Upstream_AU(30nt)"     "Dwstream_AU(30nt)"     "Site_Location"
"LogitProb"     "Upstream_Access(30nt)" "Dwstream_Access(30nt)"
"Upstream_AU(20nt)"     "Dwstream_AU(30nt)"     Score
8       NM_017589.4     hsa-miR-501-3p  1050    383-411 405-411 7mer-m8
1       -23.1   -8.413  -0.137  0.357   0.283   0.033   0.099   0.6
0.4     0.119   0.46    0.5     0.6     0.096   0.634   0.467   0.667
0.14    0.55    0.45    0.65    0.297   0.526   0.48    0.72    0.407
0.476   0.533   0.667   0.224702380952381       0.655441321236195
0.407   0.476   0.45    0.667   0.0407421894466
```

b

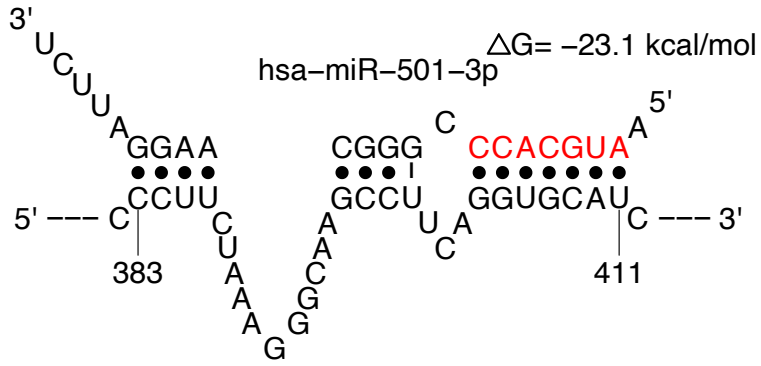
```
"SiteID"        "Target"        "miRNA" "Target_Len"    "Site_Position" "3'_bp"
"dG_hybrid"     "dG_nucl"       "dG_total"      "Site_Access"
"Upstream_Access(5nt)"  "Dwstream_Access(5nt)"  "Upstream_AU(5nt)"
"Dwstream_AU(5nt)"      "Upstream_Access(10nt)" "Dwstream_Access(10nt)"
"Upstream_AU(10nt)"     "Dwstream_AU(10nt)"     "Upstream_Access(15nt)"
"Dwstream_Access(15nt)" "Upstream_AU(15nt)"     "Dwstream_AU(15nt)"
"Upstream_Access(20nt)" "Dwstream_Access(20nt)" "Upstream_AU(20nt)"
"Dwstream_AU(20nt)"     "Upstream_Access(25nt)" "Dwstream_Access(25nt)"
"Upstream_AU(25nt)"     "Dwstream_AU(25nt)"     "Upstream_Access(30nt)"
"Dwstream_Access(30nt)" "Upstream_AU(30nt)"     "Dwstream_AU(30nt)"
"Site_Location" "LogitProb"     "Upstream_Access(10nt)" "Dwstream_Access(30nt)"
"Upstream_AU(30nt)"     "Dwstream_AU(30nt)"     Score
20      NM_017589.4     hsa-miR-501-3p  1050    906-928 0       -17.6   -0.224
-7.354  0.379   0.636   0.499   0.8     0.2     0.481   0.379   0.9     0.4
0.483   0.379   0.8     0.467   0.441   0.317   0.65    0.45    0.411   0.349
0.56    0.44    0.366   0.363   0.567   0.433   0.0136986301369863
0.528740278449575       0.481   0.363   0.567   0.433   0.0407421894466
22      NM_017589.4,    hsa-miR-501-3p  1050    996-1008        0       -17
-0.115  -2.25   0.266   0.416   0.222   0.6     0.8     0.51    0.115   0.6
0.6     0.452   0.182   0.6     0.533   0.352   0.227   0.55    0.6     0.329
0.34    0.52    0.68    0.36    0.444   0.5     0.7     0.63013698630137
0.568329262359457       0.51    0.444   0.5     0.7     0.0407421894466
23      NM_017589.4,    hsa-miR-501-3p  1050    1011-1022       0       -15.4
-0.371  5.848   0.091   0.237   0.377   0.6     0.8     0.199   0.541   0.5
0.9     0.302   0.678   0.6     0.867   0.33    0.749   0.6     0.85    0.385
0.788   0.6     0.8     0.377   0.819   0.6     0.8     0.732876712328767
0.569609983804448       0.199   0.819   0.6     0.8     0.0407421894466
21      NM_017589.4,    hsa-miR-501-3p  1050    949-981 0       -20.4   -0.167
3.465   0.344   0.129   0.461   0.4     0.4     0.255   0.492   0.5     0.5
0.256   0.512   0.533   0.6     0.317   0.466   0.45    0.65    0.317   0.403
0.48    0.56    0.29    0.376   0.533   0.6     0.308219178082192
0.434462062811596       0.255   0.376   0.533   0.6     0.0407421894466
```

Figure 3

Figure 4

```
Target name   = NM_017589.4

Site 1 -- hsa-miR-501-3p

    5'->3'         A   GG  GAGAGAAA  G   U       G
    Target      3   GGA  CU         GC CUG GG GCG    29
                    |||  ||         || ||| || |||
    miRNA      22   UCU  GG         CG GGC CC CGU    1
    3'->5'             UA  AA               A   AA


Site 2 -- hsa-miR-501-3p

    5'->3'         U   CG    CUGAAGG    CCAA   A
    Target     25   GGG  GUCC        GCCU    GGG        51
                    |||  ||||        ||||    |||
    miRNA      22   UCU  UAGG        CGGG    CCC        1
    3'->5'                 AA                ACGUAA


Site 3 -- hsa-miR-501-3p

    5'->3'           N    C  AA    C
    Target     30     GUCC UG   GGG        41
                      |||| ||   |||
    miRNA      22     UAGG AC   CCC        1
    3'->5'         UCU    A  GGG   ACGUAA


Site 4 -- hsa-miR-501-3p

    5'->3'         N   CG   A AAGA   A    N
    Target     52   AGA  CCU G    UGG G GC     72
                    |||  ||| |    ||| | ||
    miRNA      22   UCU  GGA C    GCC C CG     1
    3'->5'             UA   A GG       A  UAA


Site 5 -- hsa-miR-501-3p

    5'->3'         N   CG   AGAAGAUGGAG    A
    Target     52   AGA  CCU           GCCC           74
                    |||  |||           ||||
    miRNA      22   UCU  GGA           CGGG           1
    3'->5'             UA   A               CCCACGUAA
```

a


ΔG= −23.1 kcal/mol
hsa−miR−501−3p
3'
5'
5' −−− C
C −−− 3'
383
411
NM_017589.4

b

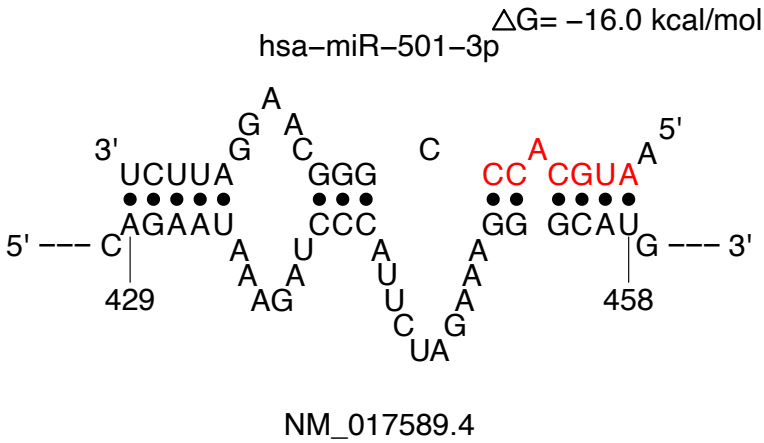

ΔG= −16.0 kcal/mol
hsa−miR−501−3p
3'
5'
5' −−− C
G −−− 3'
429
458
NM_017589.4

Figure 5