

# The Sfold Project: Computational tools for regulatory RNA studies



The American Cancer Society estimates that 12 million people are affected by cancer worldwide. As **Dr Ye Ding** explains, a software platform called Sfold is helping scientists to a better understanding of RNAs, which could potentially lead to new treatments

**EFFICIENT GENE REGULATION** methods are of paramount importance to the functional studies of genes and genetic products in humans and model organisms, as well as in infectious pathogens. 'The Sfold Project', an NSF and NIH funded initiative based at the Wadsworth

Center and led by Dr Ye Ding, intends to provide tools for studies on regulatory RNAs including microRNAs that may have implications for the better understanding and treatment of cancer.

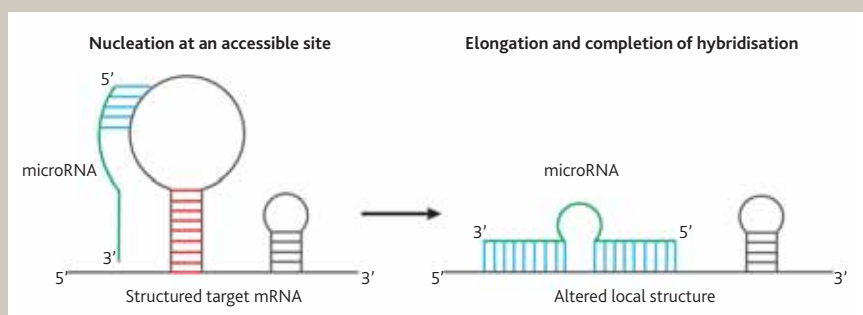
MicroRNA targeting is a very exciting field, with significant potential to treat human

diseases. The Sfold Project team based at the Wadsworth Center has been developing state-of-the-art tools for genome-wide target identification of microRNAs. As Ding explains, microRNAs (miRNAs) are small non-coding RNAs of ~22 nucleotides that repress translation of target messenger RNAs (mRNAs) in multi-cellular eukaryotes. "RNA molecules are involved in some of the most fundamental processes in cells," he outlines. "The functions of many RNA molecules can be understood from their structures. However, elucidation of RNA structure by experimental means is usually difficult and time consuming." In order to facilitate the identification and validation of new therapeutic targets and agents for the treatment of human diseases, a better understanding of gene function is indispensable; RNA-targeting antisense nucleic acids provide a valuable alternative to gene knockout and mutagenesis which is inefficient and complex.

The main objectives of Ding's work are to develop novel algorithms and methods for improved prediction of RNA higher order structures, to hone the methods and tools for rational and efficient design of antisense nucleic acids, and to predict microRNA targets. With these goals in mind, the core of the project is the creation and maintenance of Sfold, a unique nucleic acid folding and design software which has already met with encouraging results, as Ding is keen to point out: "Sfold has been licensed for commercial use, because RNAi is an important tool for drug target validation and functional genomics," he remarks, before going on to emphasise the wider significance of the programme for genetic scientists: "I received many emails from users indicating that the software is valuable for their research."



**PROCESSING OF TARGET MRNA SEQUENCE BY SFOLD FOR SELECTION OF EFFECTIVE TARGET SITES FOR POTENT GENE SILENCING BY RNAI**



**A TWO-STEP HYBRIDISATION MODEL: NUCLEATION AT AN ACCESSIBLE TARGET SITE, FOLLOWED BY HYBRID ELONGATION TO DISRUPT LOCAL TARGET SECONDARY STRUCTURE AND FORM THE COMPLETE MICRORNA-TARGET DUPLEX**

this in mind, a collaborative project with the Jun Lu lab at Yale University and a collaboration with the Anke van den Berg lab at University Medical Center Groningen in the Netherlands have been planned. These joint projects would help to develop statistical methods for improving target predictions and elucidating regulatory networks for microRNAs associated with cancer, while incorporating new tools from these studies into Sfold, and making these tools freely available to the scientific community. Looking to the future, Ding is clear that there is much to be optimistic about in terms of RNA based treatment of diseases: "There will be further advancement with delivery approaches for getting siRNAs to specific target cells, tissues or organs," he says. "There will also be exciting development of RNAi-based therapeutics and diagnostics for cancer, viral infectious diseases and other diseases. It will not be surprising if the first FDA-approved RNAi drug emerges within a decade."

Encouragingly, the molecular biology component of the project has already yielded some positive results. Initial findings have led to new novel statistical algorithms for improved RNA secondary structure prediction, and the creation of tools for the rational design of antisense oligos, hammerhead ribozymes and siRNAs. In their collaboration with the Victor Ambros lab, Ding's team have furthered their understanding of structure-based two stage microRNA-target hybridisation which they have utilised to create models and develop their Sfold software further, leading to better interpretation of worm microRNA targeting data.

The Sfold Project is currently awaiting the results of a pending NIH grant renewal application

which, if successful, would enable the team to advance the development of computational methods and design tools to address both the potency and specificity of RNAi triggers. Ding's laboratory emphasises the desire to broaden the understanding of microRNAs in cancer, and with

**For both exogenous and endogenous RNA-targeting nucleic acids that include antisense oligos, trans-cleaving ribozymes, siRNAs and microRNAs, target structural accessibility is important for successful target binding for inducing gene down-regulation**

## Harnessing RNA in cancer treatment

Dr Ye Ding tells International Innovation about his exciting research into the regulatory targets of nematode microRNAs, which could provide a greater insight into the function of RNA molecules

**What are microRNAs, and how do they regulate developmental and physiological processes? Do they have the same function in humans, as in animals?**

MicroRNAs are endogenous RNA molecules of 21 – 23 nucleotides in length. Since they are not translated into proteins, they are an abundant class of non-coding RNAs. They can regulate developmental and physiological processes, mainly by inhibiting the expression of the genes essential for these processes. Although thousands of microRNAs have been discovered to date, the functions of most microRNAs have not been established by experimental studies. Many microRNAs are conserved between human and animals. If one uses conservation as the sole criterion, then these microRNAs can have similar functions between human and animals. However, there

are many human-specific microRNAs that presumably have human-specific functions. For human and animals, microRNAs typically are partially complementary to target mRNAs, thus microRNAs serve as the guide for target recognition by a complex similar to that for RNAi. Successful target binding induces translational repression or microRNA destabilisation.

**Given that microRNAs are associated with human cancers, can they be harnessed for cancer treatment? Can you alter gene expression by duplicating and implanting the resistant genes?**

The association between microRNAs and diseases has presented opportunities for development and commercialisation of microRNA-based diagnostics and

therapeutics. Several companies are working on commercialising the potential for microRNAs to serve as cancer biomarkers as many studies have shown that changes in microRNA expression profiles are linked directly to some aspects of cancer's biological processes. Therefore, controlling microRNA expression may influence cancer development. Gene expression can be altered by gene amplification or implantation, but with technical difficulties. Silencing genes by RNAi or inhibiting microRNA activities can be effective alternatives for controlling specific gene expression. The activities of microRNAs can be specifically inhibited by antagomirs, small synthetic RNAs that are complementary to specific microRNAs. It is believed that an antagomir can irreversibly bind the microRNA so that the microRNA is no longer available for binding to its target

mRNA to induce translational repression or target destabilisation. Thus, antagomirs may represent a therapeutic strategy for silencing microRNAs in disease.

**Further to the above, what effect will this gene manipulation have to the virulence of diseases? Could such techniques preclude pathogens from attacking cells in the first place?**

Virulence of pathogens can be related to microRNAs. For example, the development of immunological cells in response to a pathogen can be influenced by microRNAs. It has been suggested that microRNAs can have a key role in regulating viral pathogenesis. Manipulation of microRNA genes can possibly ameliorate pathogen virulence. A pathogen can be precluded from attacking host cells through the silencing of pathogenic genes or host genes essential for pathogen-host interactions. For example, a recent study published by Science showed that the Hepatitis C virus infection in primates can be suppressed by the silencing of microRNA-122.

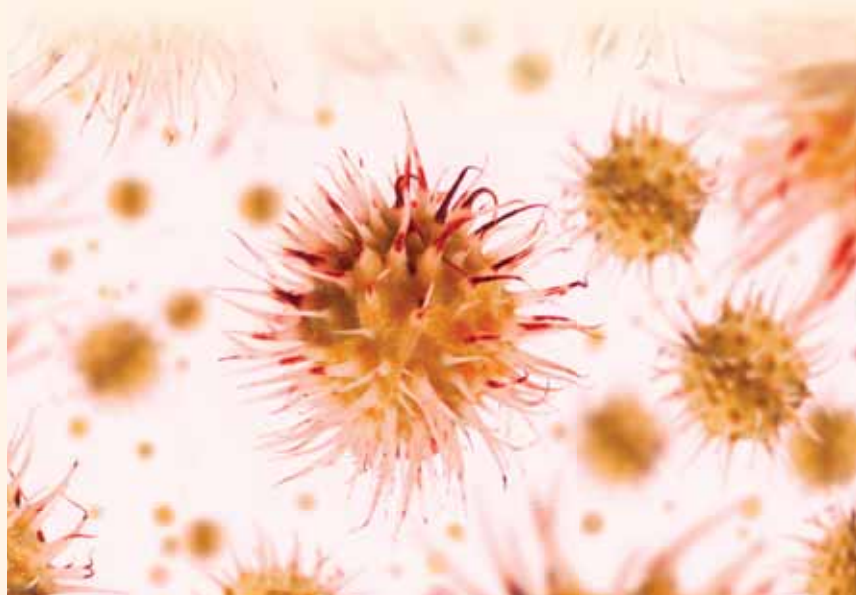
**In comparison with previous computational methods, how much success have you had in identifying microRNA targets by incorporating Sfold target structure predictions?**

In a collaborative study, Dang Long, I and the Victor Ambros lab investigated the role of target secondary structures on the efficacy of repression by microRNAs, by employing the Sfold program for predicting target structures. We developed a novel structured-based paradigm for modelling the interaction between a microRNA and a target. The model performed surprisingly

well on available worm target data, and was validated by experimental testing in worm. Two key parameters of the model were further validated by an analysis of a large worm microRNA target data set from an immunoprecipitation study. In comparison with previous methods, the model was particularly effective in accounting for certain false positive predictions made by those methods. Based on the model and other additional parameters, my lab has recently developed a prediction method. The results so far suggest this method offers a substantial improvement over existing methods.

**Have your studies allowed your lab to successfully develop a nematode microRNA target database including validated and predicted targets? How will this knowledge manifest itself in future computational biology research?**

Similar to our finding for exogenous RNA-targeting nucleic acids, a significant discovery is that target structural accessibility is also important for target recognition by microRNAs. We also found support for the preference of seed matches. We have developed a user-friendly version of a target database for worm and fly and will update the database with predictions from our recent method and make it available to the scientific community. We hope that such a database will provide useful information for experimental scientist to test computational predictions. Feedback experimental data will be helpful for us to improve our prediction methods by data analysis and further predictive modelling development.



## Rational Design Tools for Antisense Nucleic Acids

Dr Ye Ding explains the techniques employed to design tools for engineering of antisense nucleic acids, the likely impact of this research, and the work still to be done

**Can you outline the project's focus on RNA-targeting nucleic acids including antisense oligonucleotides, trans-cleaving ribozymes and short interfering RNAs, and the role that they play in gene regulation? What methods have you adopted to identify and manipulate the mechanisms that cause gene down-regulation?**

RNA-targeting nucleic acids can bind target RNA by complementary base pairing to induce translation inhibition or target cleavage. Although the mechanisms are not completely understood, there are known differences for different types of nucleic acids. Target cleavage by RNase H is the predominant pathway of inhibition by antisense oligos. Target cleavage by hammerhead ribozymes is magnesium dependent. Gene silencing by RNAi mediated by siRNAs involves argonaute proteins as the catalytic components of the RNA-induced silencing complex (RISC). While studies of mechanisms have been pursued by other labs through biochemical and structural biology approaches, our focus has been on the effects of predicted secondary structure features and sequence features on the efficacy of antisense nucleic acids, for the optimisation of computational design of these molecules. Significance of target secondary structure also provides a mechanistic insight into the functionality of RNA-targeting nucleic acids.

**What techniques did you use to develop a methodology for the efficient design of ribozymes, and with what success?**

A hammerhead ribozyme is an RNA molecule with its own structure. We



## THE SFOLD PROJECT

## OBJECTIVES

- To develop novel algorithms and methods for improved prediction of RNA higher order structures
- To develop methods and tools for the rational and efficient design of antisense nucleic acids for high-throughput gene down-regulation
- To develop methods and tools for improved prediction of microRNA targets
- To develop and maintain Sfold, a unique nucleic acid folding and design software
- To develop and maintain a comprehensively annotated, web-accessible microRNA target database

## FUNDING

National Science Foundation (NSF): \$2 million; National Institute of Health (NIH): \$1.2 million

## CONTACT

Dr Ye Ding

Wadsworth Center  
New York State Department of Health  
Center for Medical Science  
150 New Scotland Avenue  
Albany NY 12208

T 1 518 486 1719

F 1 518 402 4623

E yding@wadsworth.org

**DR YE DING** is a research scientist and the Principal Investigator of the Sfold Project team. He is interested in the field of Computational and Structural Biology and has a PhD from the Carnegie Mellon University.



employed Sfold for both the target and the ribozyme, and introduced a number of structural and thermodynamic parameters for modelling ribozyme target hybridisation. In collaboration with the Erasmus Schneider lab at Wadsworth Center, we designed and tested a series of hammerhead ribozymes to target the mRNA of the breast cancer resistance protein (BCRP). We found two significant parameters that are predictive of ribozyme activity, and several rules for preferred folding features for the ribozyme itself. These findings suggest rules for the structure-based rational design of *trans*-cleaving hammerhead ribozymes in gene knockdown studies.

**How close is the data you have gleaned to finding application in the engineering of antisense nucleic acids? Can you specify the particular areas of healthcare where these could make an impact?**

The tools we have developed from our studies have been incorporated into Sfold. These tools have been used by other scientists for the design and engineering of antisense nucleic acids in their gene modulation studies. Effective design tools can be valuable for the development of nucleic acid based therapeutics and diagnostics.

**Can you explain the potential of siRNAs to revolutionise biological science? How may these be synthetically constructed and engineered in mammalian gene knockdown studies, and how will your**

**computational methods play a part in the design process of RNAi triggers?**

RNAi has become the method of choice for mammalian gene knockdown studies. Synthesised siRNAs are available from a number of RNAi reagent providers. RNAi triggers can also be stably expressed in cells through lentiviral vectors. The Sfold application module for RNAi design has heavy usage. So scientists have been using the Sfold server for designing RNAi triggers.

**In terms of your the key objective what have been your main findings to date, and what questions remain unanswered?**

I and my computational collaborators Charles Lawrence and Chi Yu Chan successfully developed novel methods for improved RNA secondary structure predictions. These methods serve as the basis for Sfold. As a part of Sfold development, we studied antisense oligos for prokaryotic gene inhibition in an collaboration with the Kathleen McDonough lab at Wadsworth Center. My lab and the Igor Roninson lab of Ordway Research Institute have successfully collaborated on analysis of RNAi data. From these collaborative studies and the study on ribozymes, we found that targeting structural accessibility is the most important determinant for the function of these antisense nucleic acids. Our recent focus has been on RNAi. How can we further improve potency by computational design? What are key factors for RNAi off-target effects? These are the questions we would like to answer.

**Wadsworth Center**  
New York State Department of Health

