

Hairpin Database: Why and How?

Clark Jeffries

Research Professor

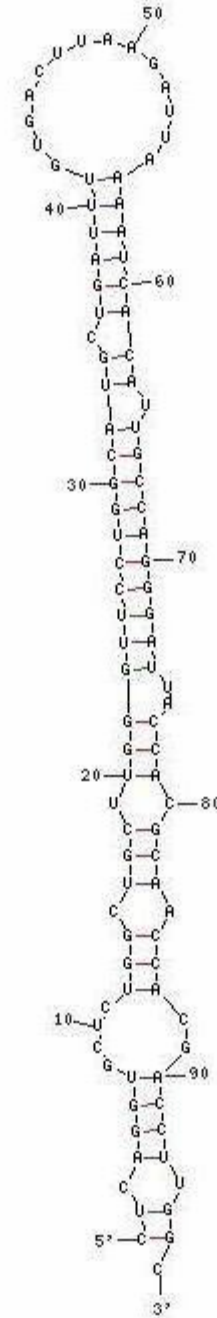
Renaissance Computing Institute and School of Pharmacy

University of North Carolina at Chapel Hill, United States

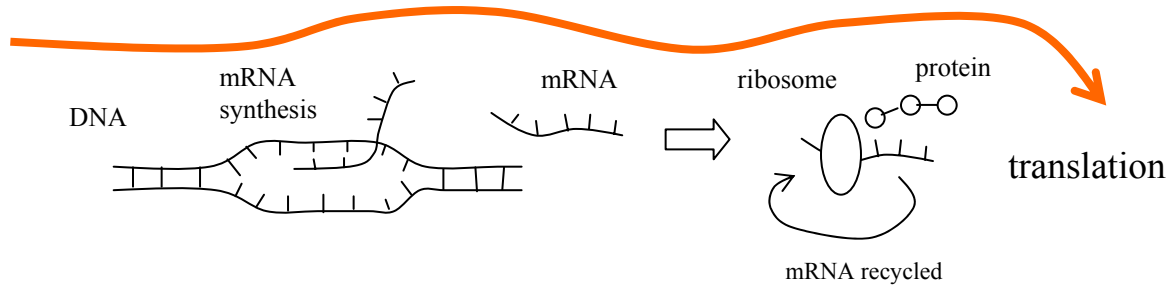


UNC
PHARMACY

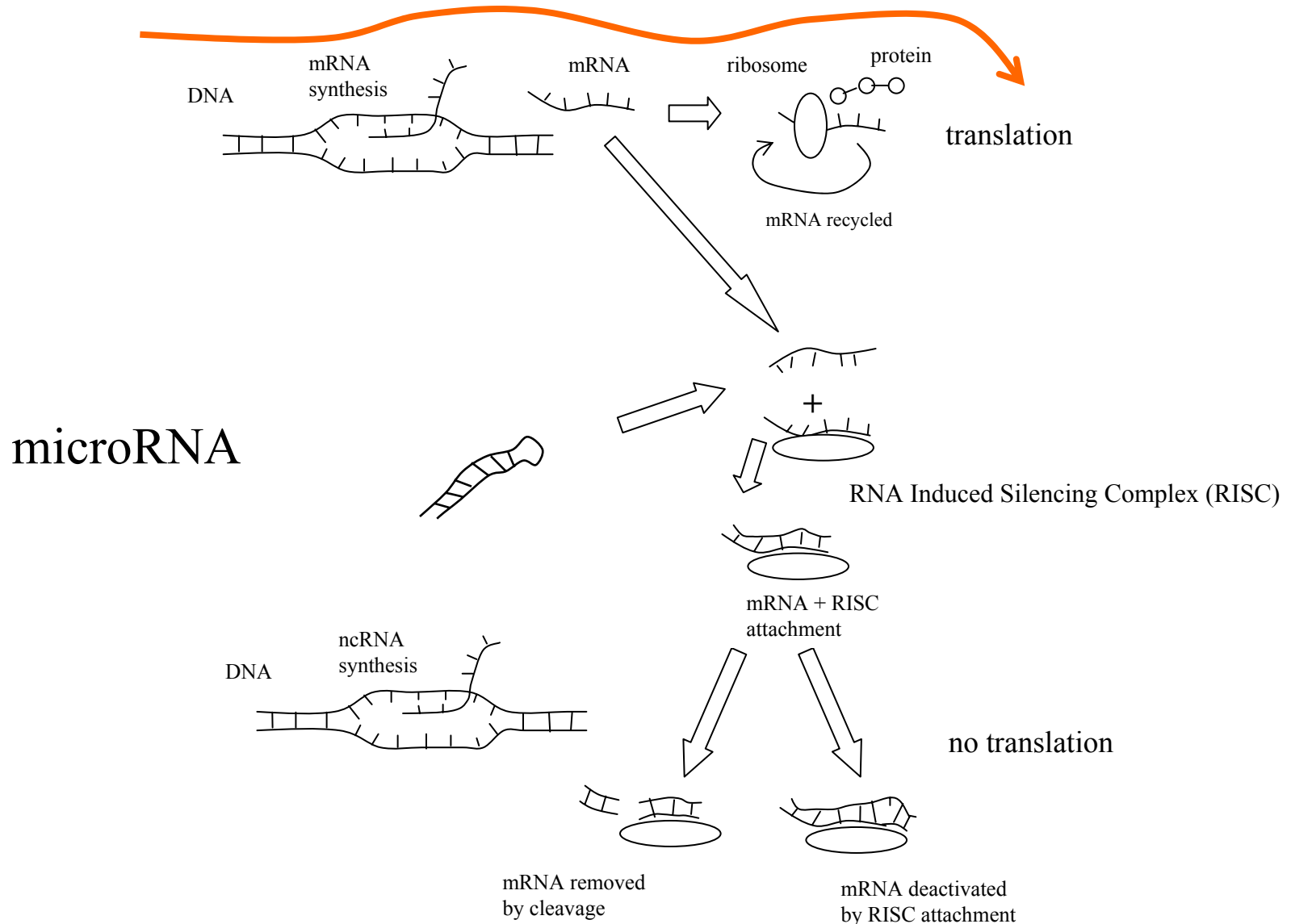
Why should a database be constructed that can be used as a search engine for genomic hairpins?



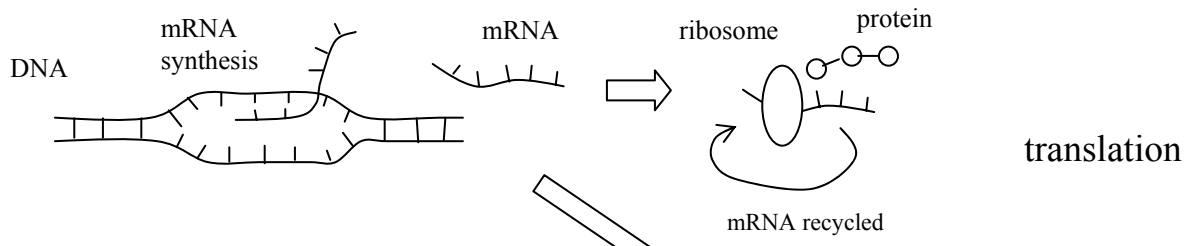
Central Dogma transcription and translation



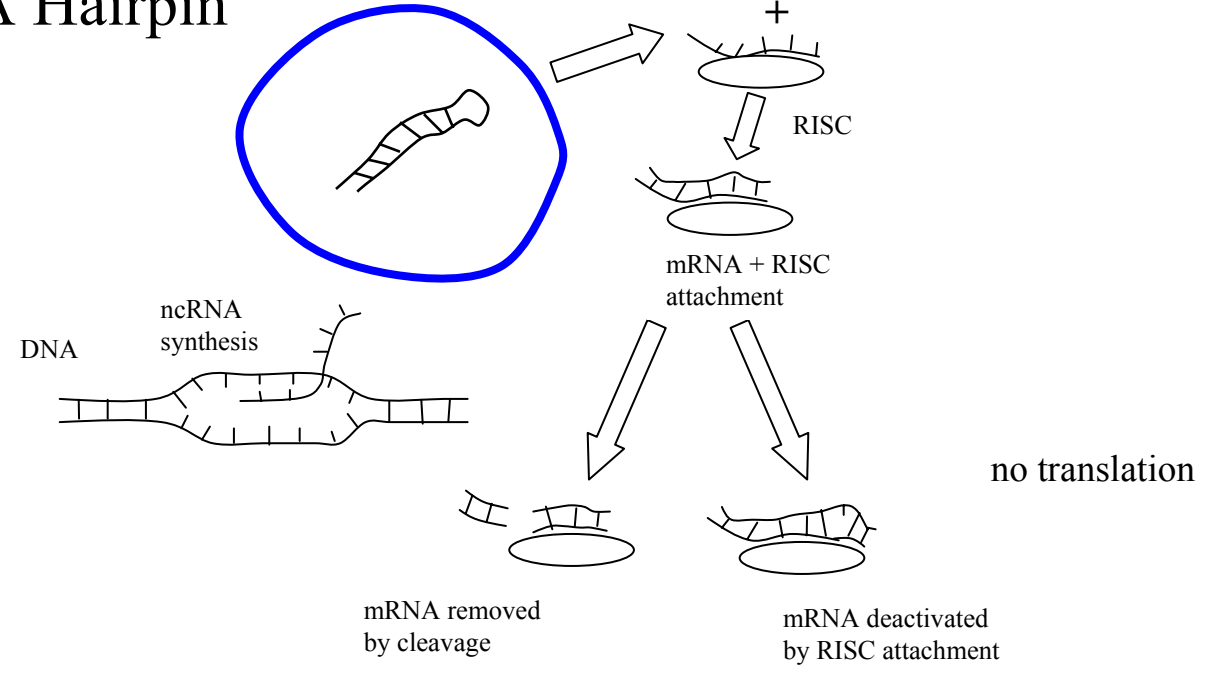
Central Dogma transcription and translation



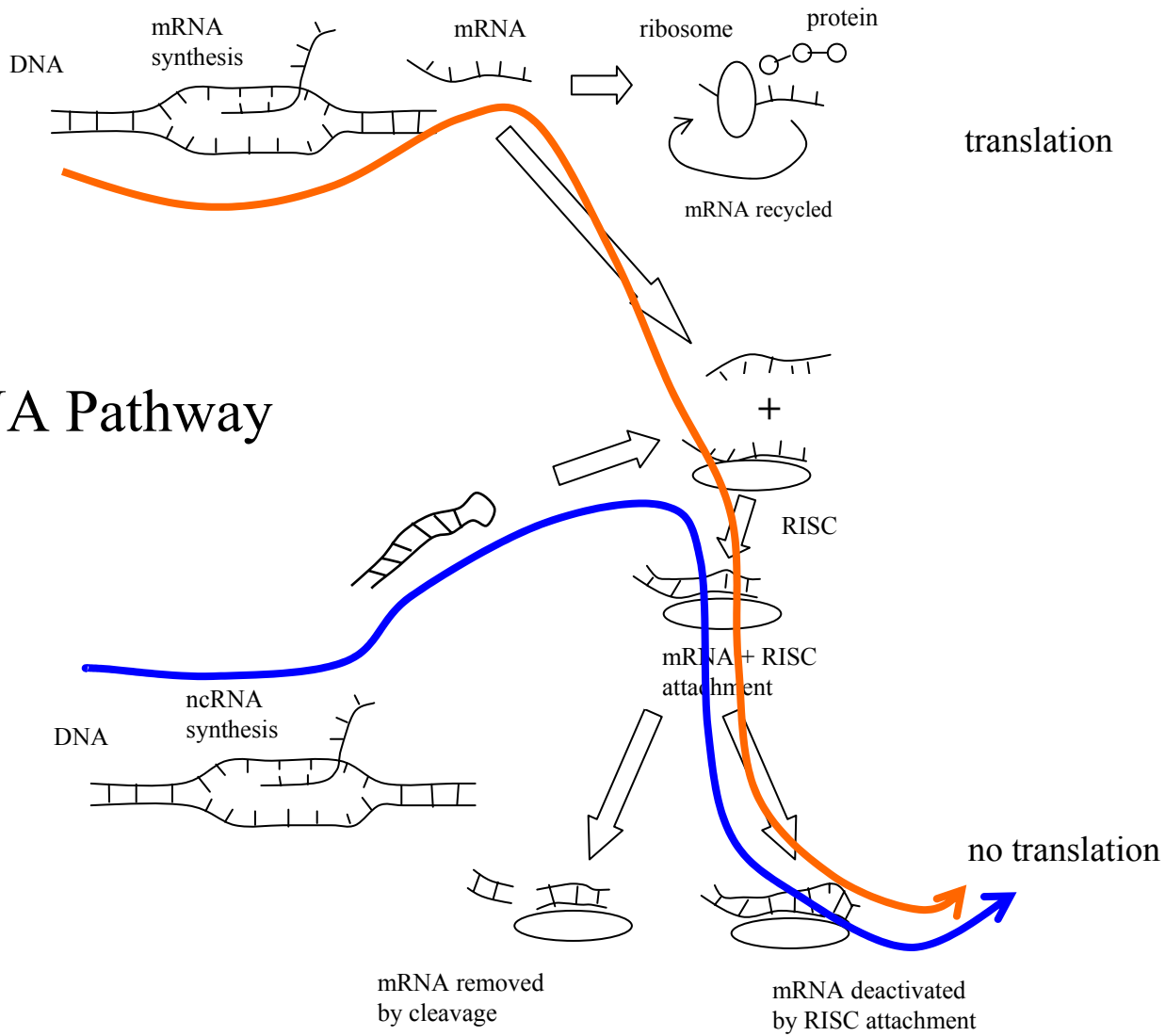
Science, special section, 2 September 2005



RNA Hairpin



RNA Pathway



Ambion: dsRNA ~200 nt on a side can be processed by DICER1 into a “cocktail” of dsRNAs about 20 nt on a side.

This has stimulated many researchers to suggest that dsRNA folded from noncoding RNA nearby reverse complement (revcom) might be processed by the miRNA pathway.

Data from the FANTOM3 project recently showed that 62% of the mouse genome is transcribed (Science 309, 2005, page 1559).

In particular, Alus and other repeats that lie in introns and near their revcoms might also be feedstock for the miRNA pathway.

The miRNA research focus to date has been largely on complementary alignments with the 3' UTR of an mRNA, but it appears just as reasonable to look for alignments on either 3' or 5' flank or indeed anywhere along the mRNA.

In fact, transcription factors can potently target introns of pre-mRNAs...why not RISCs?

The consequences for the ribosome and the RISC after cleavage or translational inhibition are not well understood, nor are the consequences of binding at different points on an mRNA and the roles or effects of the of flanking sequences.

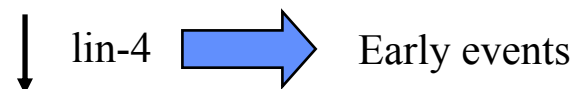
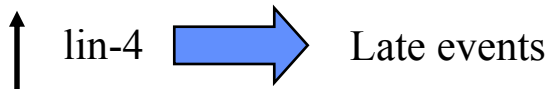
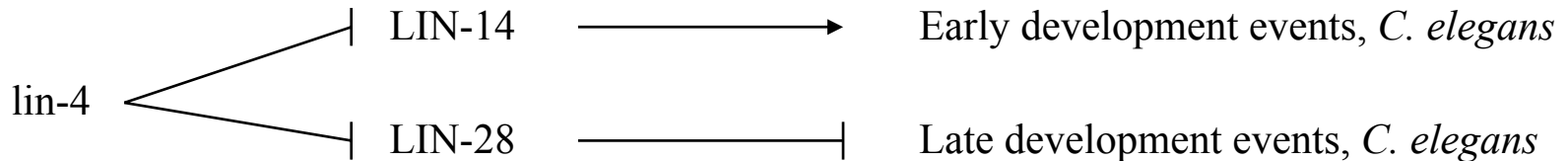
Examples of miRNA alignments (Ambros, Nature 431, 2004)

```

..nnnnnn      acca----      nnnnnn... 3'  LIN-14 3' UTR
      ucaca          cucagggga
      |||||         |||||
3' agugu          gagucccu
      gaacucca      u 5'          lin-4 miRNA
    
```

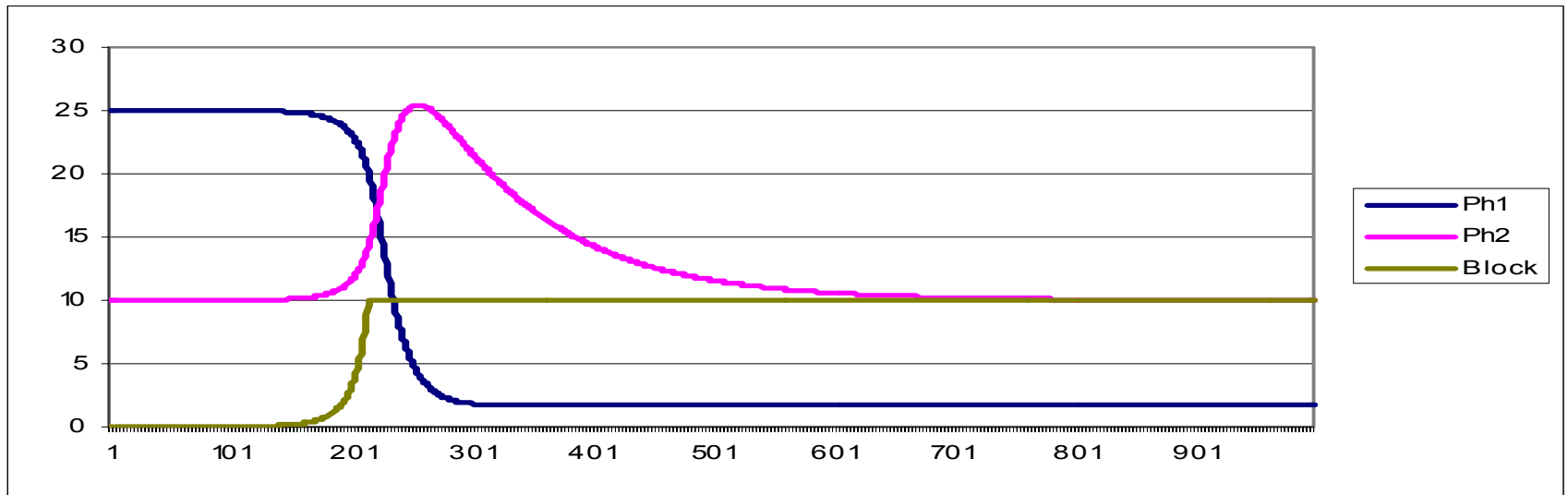
```

..nnnnnn      -----      nnnnnn... 3'  LIN-28 3' UTR
      uugcac          ucucagggga
      |||||         |||||
3' agugug          agagucccu
      aacucc        u 5'          lin-4 miRNA
    
```



Fire et al (Nature 391, 1998) demonstrated that “...only a few molecules” of dsRNA can cause a phenotypic toggle in *C. elegans*. Tuschl et al demonstrated the same in mice in 2000, and reversible toggling in mammal cells was later shown by several researchers (eg Gupta et al, PNAS 101, 2004).

A dynamical system model...



Chemically engineered oligonucleotides called “Antagomirs” (Krutzfeldt et al, Nature 2005) can efficiently and specifically silence miRNAs. In mice, miR-122 was reduced to undetectable levels for up to 23 days following a single intravenous injection.

Like the action of miRNAs on target sequences in mRNAs, the silencing of miR-122 used a complementary sequence.

Subtleties of Alignments: Reporter constructs in HeLa cells:

Successful translational inhibition (using luciferases in 3' UTR of LIN-41 sequences)
uugaa

```
gcacagccua      cuagguca in gene LIN-41
|||||          |||||
uguguuggau ... gauggagu in miRNA let-7
```

Unsuccessful translational inhibition

```
          cuugaa
gcacagccua      cuagguca in gene LIN-41
|||||          |||||
Uguguuggau ... gauggagu in miRNA let-7
```

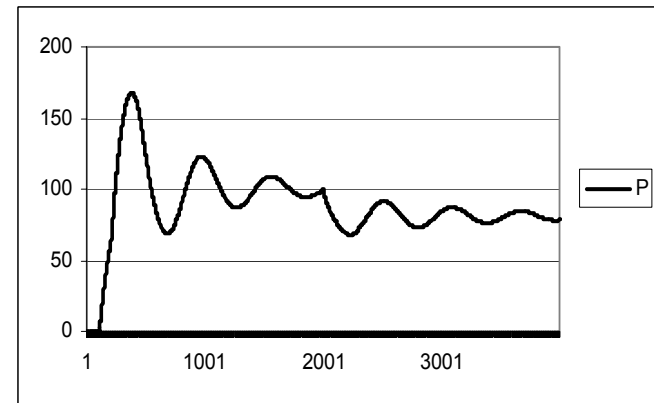


one base different

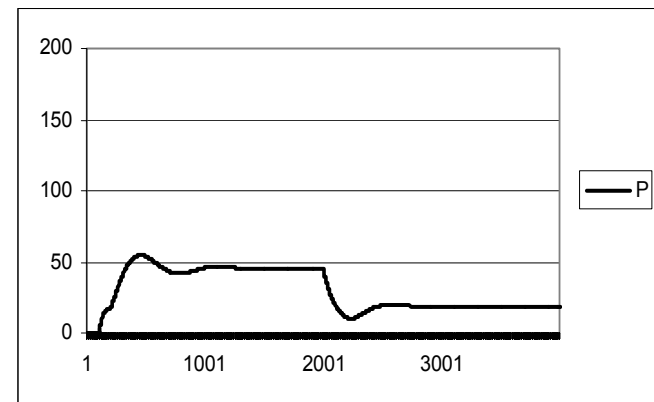
(Kiriakidou et al, Genes & Dev 18, 2004)

Why might nature use miRNAs?

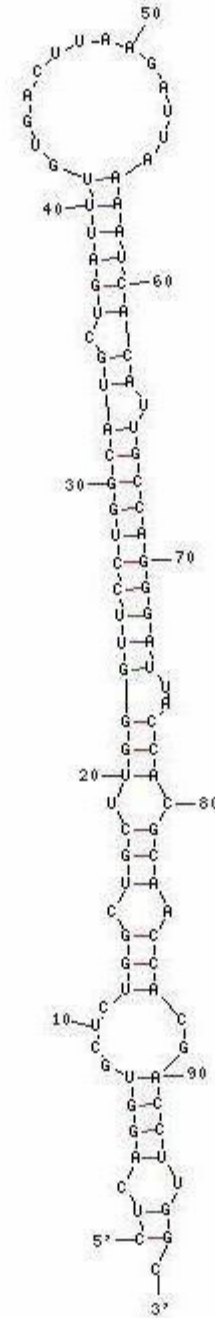
Protein levels in a model without Blocks and large time delays for feedback to transcription and translation.



The same model with a Block added that responds promptly to Target (mRNA) levels.



How should a database be constructed that can be used as a search engine for genomic hairpins?



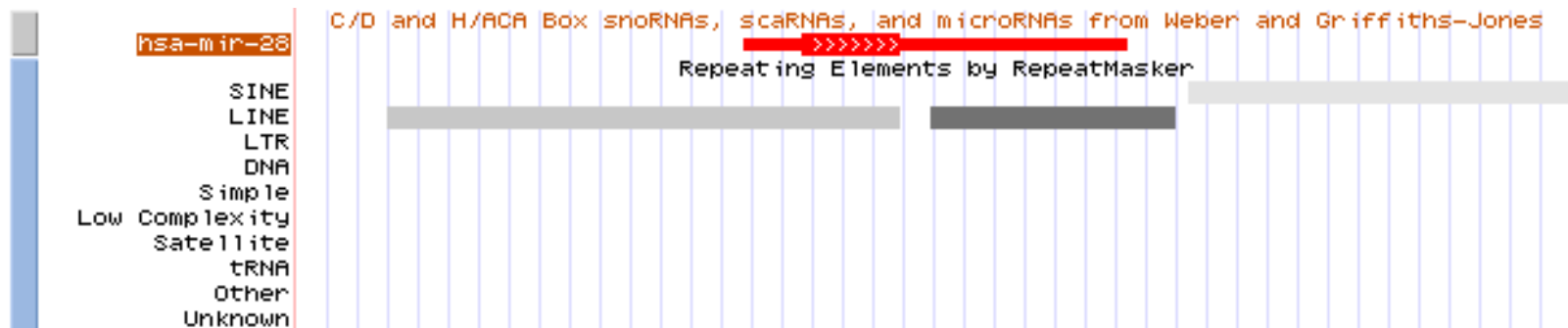
Wish list for hairpin database:

1. Improved algorithms for finding hairpins
2. Classification of hairpins by size, shape, and use of repeats
3. Further information on hairpins including chemical stability (such as $-dG/\#nts$ ratio and melting point)
4. Scoring system for evaluation of alternative secondary structures
5. Putative or known functional annotation
6. Target rules (such as exact complementary alignment of a subsequence with a region of the target mRNA)
7. Recognition of flanking sequences that mediate hairpin transcription
8. Hierarchies or clusters of hairpins, especially if known to be co-expressed

By the way, several miRNAs are known to overlap with repeats...

hsa-mir-330, 95, 151, 489, 9-1, 132, 28.

For example, hsa-mir-28 overlaps with two LINE L2 and almost overlaps with a MIRb (UCSC Genome Bioinformatics)

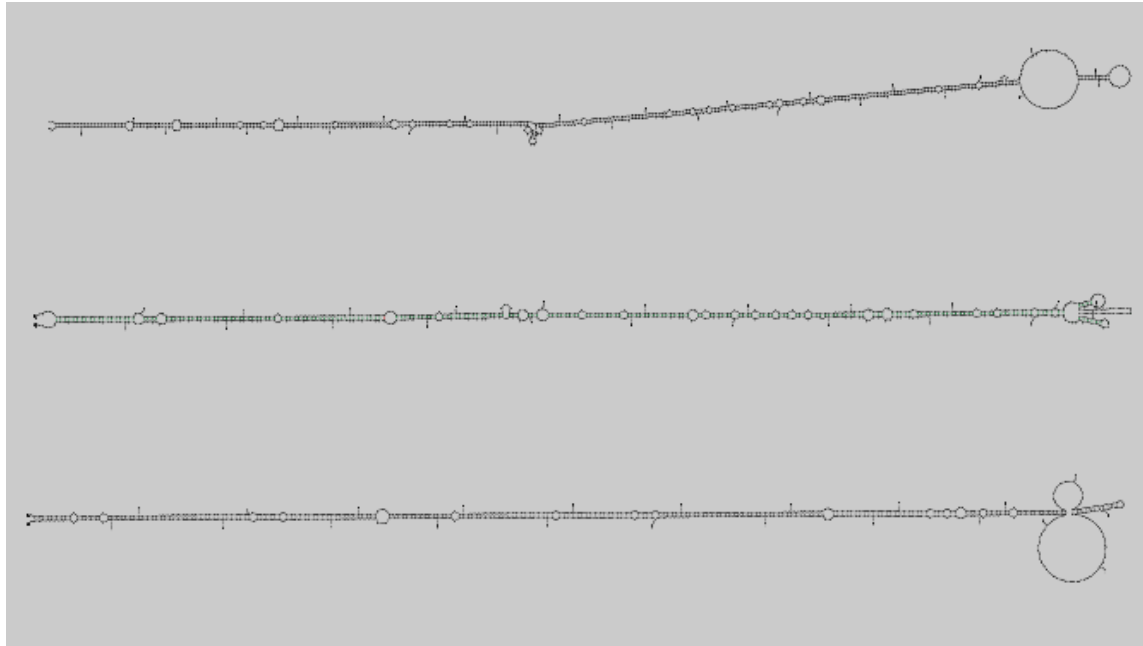


Hairpins and apoptosis...a connection?

Three hairpins seem to be possible byproducts from intronic splicing of the pre-mRNA of pro-apoptosis gene PAWR (5074, alias PAR-4). Here is one:

```
TTTTTTTTTTTTTTTTTTTTGAGACAGAGTCTCGCTTCTTCGCCCAGGCTGGAGTGCAATGGCGCTATCGT
AGCTCACTGCAACCTCCGCCTCCCGGGTTCAAGCAATTCCCCTGCCTCAGCCTCCCCAGTAGCTGGGAT
TACAGGCTAATTTTTGTATTTTTGTTAGAGATGGGGTTTCACCCTGTTGGCCAGGCTGGTCTCGAACTC
CTGACCTCAAGTGATACACCCACCTCGGCCTCCCAAAGTGCTGGGATTGCAGGCGTGAGTCATTGTGCC
CAGCCAAAATTGTACGCTCTTAAGGACAGATTTTTTTTCTTAGCTATATGGACTTAATCTTTATATTCTG
TAGTCTATATAATTATATACTTTAAAAAAAAAATACAGTGGCTGGCCGCAATGGCTCTCGCCTGTAATCC
CAGCACTTTGGGAGGCTGAGGCAGGTGGATCAGCTCAGTTCAGGAGTTCAAGACCATCCTGACCAATAT
GGTGAAACCCCGTCTCTAGTAAAATTACAAAATTAGCTGGGCCTGGTGGCATGCACCTGTAGTCCCAG
CTACTCGGGAGACTGAGGCAGGAGAATCACTTGAACCCGGGAGGCAGAGGTTGCAGTGAGCCAAGATTG
TGCCACTGCACTCCAGCCTGGGCAACAGAACGAGACTCCATCTCAAAAAAAAAAAAAAAAAAAAAA
```

Some dsRNA hairpins from one intron of pro-apoptosis gene PAWR



RNA Hairpins Located from intron 2-3 of PAWR. RNA folding engine predicts (mfold, M. Zuker) these three hairpins would form from intronic pre-mRNA for PAWR.

Functional suggestion:

Transcription of pro-apoptosis PAWR automatically yields a byproduct of three hairpins, each with Alu-like stems.

dsRNA from the hairpins might enter the miRNA pathway to produce RISCs that inhibit translation of many other genes, including anti-apoptosis gene BIRC4.

In this way, the initiation of apoptosis might be driven efficiently and irreversibly to its conclusion, the dissolution of a cell that is dangerously mutated or simply unneeded.

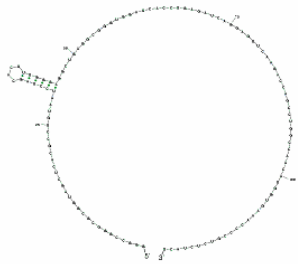
However, prediction of hairpins from RNA sequences is not easy...

Different strokes...

Different folds of

```
GGGCCGGGCGCGGTAGCTCACGCCTGTAATCCTAGCACTTTGGGAGGCTGAGGCGGGTGGATCACCTGAGATCAGGAGTTCAAGACCA  
GCCTGGCCAACATGGTGAAACCCCGTCTCTACT
```

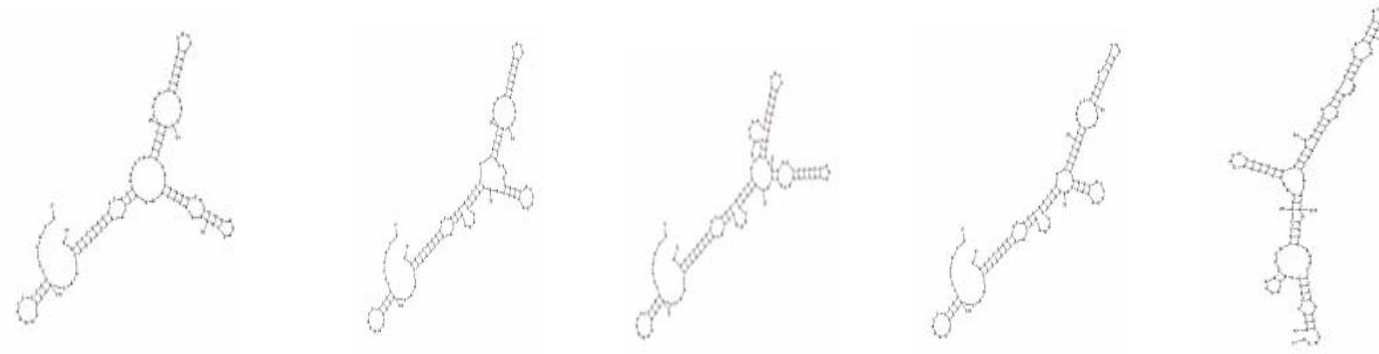
Srna from <http://sfold.wadsworth.org/srna.pl>
yields one shape:



Different predicted folds of

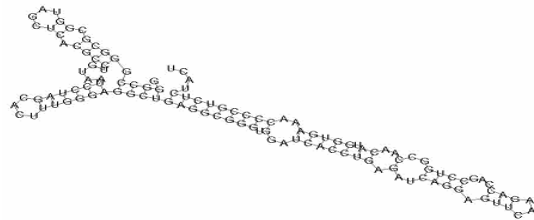
```
GGGCCGGGCGCGGTAGCTCACGCCTGTAATCCTAGCACTTTGGGAGGCTGAGGCGGGTGGATCACCTGAGATCAGGAGTT  
CAAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACT
```

mfold from <http://www.bioinfo.rpi.edu/applications/mfold/old/rna/form1.cgi>
yields five possible shapes:



Different predicted folds of
GGGCCGGGCGCGGTAGCTCACGCCTGTAATCCTAGCACTTTGGGAGGCTGAGGCGGGTGGATCACCTGAGATCAGGAGTTCAAGACCA
GCCTGGCCAACATGGTGAAACCCCGTCTCTACT

Vienna RNAfold <http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi>
yields one shape:



Tasks for hairpin database:

1. Close cooperation with RNA biochemists
2. Study of existing engines such as *einverted* (EMBOSS)
3. Application of repeat classification concepts
4. Where possible, inclusion of information from nucleic acid chaperon research
5. User documentation on effects of search parameter choices
6. Functional annotation, if possible
7. Target rules, if possible
8. Research on effects of flanking sequences that mediate hairpin transcription
9. User-friendly maps of locations, densities, clusters

CCEGA (NIH/NCRR):

The project described was supported by Grant Number 5-P20-RR020751-01-02 from the National Center for Research Resources. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health/National Center for Research Resources.



UNC
PHARMACY