In Silico Evidence of the Relationship Between miRNAs and siRNAs

L. Montanucci1, P. Fariselli*,1, P.L. Martelli1, I. Rossi1,2 and R. Casadio1

1Biocomputing Group, Department of Biology, via Irnerio 42, University of Bologna, Italy
2BioDec Srl, Via Calzavecchio 20/2, I-40033 Casalecchio di Reno (BO), Italy

Abstract: Both short interfering RNAs (siRNAs) and microRNAs (miRNAs) mediate the repression of specific sequences of mRNA through the RNA interference pathway. In the last years several experiments have supported the hypothesis that siRNAs and miRNAs may be functionally interchangeable, at least in cultured cells. In this work we verify that this hypothesis is also supported by a computational evidence. We show that a method specifically trained to predict the activity of the exogenous siRNAs assigns a high silencing level to experimentally determined human miRNAs. This result not only supports the idea of the biological interchangeability of siRNAs and miRNAs but indicates that it is possible to use computational tools developed using synthetic small interference RNAs to study endogenous miRNAs.

INTRODUCTION

RNA interference (RNAi), first discovered in the nematode Caenorhabditis elegans [1], is an evolutionarily conserved endogenous pathway that mediates negative post-transcriptional gene regulation in eukaryotic cells [2-4]. The sequence-specific mediators of this regulatory pathway are small RNA molecules (~23 nt), that bind complementary regions of messenger transcripts and induce highly specific gene silencing [5].

Small RNAs belong to two major classes: microRNAs (miRNAs) and short interfering RNAs (siRNAs). MiRNAs originate from short hairpins (called microRNA precursors) endogenously produced through a multi-step process that starts in the nucleus [6]. On the contrary siRNAs can be either endogenous or exogenous and derive from double stranded RNA molecules (dsRNAs). miRNAs and siRNAs seem to differ only in their biogenesis [7-10].

Long dsRNAs and the miRNA hairpin-shaped precursors are initially processed by the cytoplasmic ribonuclease III-like enzyme Dicer which cleaves them producing dsRNAs of about 22 bp long with characteristic 2nt overhangs at the 3’ terminus. Then they are incorporated, as single stranded RNA, into a ribonucleoprotein complex named RNA-induced silencing complex (RISC). RISC identifies target genes through an antisense complementarity between the mi/siRNA and the mRNA. When RISC binds the target messenger, mRNA can be inhibited and later reprocessed or degraded, depending on the degree of pair-matching [11, 12].

To study the effect of siRNAs on the gene expression, synthetic siRNAs are generally synthesized with perfect base complementarity to their target. Notwithstanding the perfect complementarity, different siRNAs have different knock-down efficacies and it is not completely known what determines the high effectiveness and specificity of some molecules or the weak functionality of others [13].

Similarly to siRNAs, plant microRNA were recognized to function through near-perfect base pairing with their targets [14-16]. On the contrary it was discovered that animal miRNAs/miRNA duplexes display only a partial base pair complementarity, since they can form complex secondary structures [17]. In particular a short sequence of 6 to 8 consecutive bases, located at the 5’ of the mature miRNA, was found to be determinant in animal target recognition [18-21].

Summarizing, while animal miRNAs carry out their function through a partial complementarity, by construction synthetic siRNAs hybridize with perfect base pair matching to target mRNAs.

Many computational analyses have been carried and predictors based on different features have been developed to infer siRNAs effectiveness. Some were based on the siRNA sequence alone [22], others added thermodynamic properties [23-25], secondary structure information [26-28], compositional biases [29] or position specific properties [30-33].

Despite much knowledge has been achieved in the last few years, many aspects of the silencing mechanism are still not fully understood, including the reasons for the high specificity of the silencing mechanism and the exact relationship between small interference and micro RNAs functioning [14, 17, 34].

Recently, Huesken and coworkers [35] made available a conspicuous experimental data set on which they implemented a powerful predictor of siRNA activity. Using these data we show that a method specifically trained to predict the activity of synthetic siRNA, assigns a high silencing level to experimentally determined human miRNAs. This result not only supports the idea of siRNAs and miRNAs equivalence but indicates that it is possible to use computational tools developed using synthetic small interference RNAs to investigate endogenous miRNAs.

MATERIALS AND METHODOLOGY

Databases

The data set of siRNAs is taken from Novartis [35] and consists of 2431 siRNA sequences whose inhibitory activi-
The evaluation of the performance of the prediction methods was carried using the Pearson correlation coefficient, defined as:

\[ r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\left[ \sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2 \right]^{1/2}} \]

**RESULTS**

**Tuning the Predictor**

In order to evaluate the accuracy of the different neural networks we adopted a 10-fold cross validation procedure. We then split the data set in 10 disjoint subsets and we evaluated the accuracy summing up the results obtained for the ten different testing sets. Other types of splitting such as the original division in two sets previously adopted [35] did not change the results. The results are reported in Table 1, where it is evident that an encoding that takes into account all the 21 siRNA nucleotides provides a better accuracy. Furthermore, as previously noticed with other methods, the inclusion of dipeptide and tripeptides information helps in improving the correlation with the experimental data.

The neural networks previously designed by the authors of the experimental data set [35] used only the sequence information and the reported Pearson correlation coefficient for the test set was 0.66, which is similar to the one obtained with our cross-validation procedure using the same encoding (0.65 in Table 1).

The predictions on the data set using our best performing method are also plotted in Fig. (1).

**Table 1. Pearson Accuracy of on the Huesken DB**

<table>
<thead>
<tr>
<th></th>
<th>Pearson Correlation Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Seq.</td>
</tr>
<tr>
<td>19-nt</td>
<td>0.63</td>
</tr>
<tr>
<td>21-nt</td>
<td>0.65</td>
</tr>
</tbody>
</table>

**siRNA and miRNA Comparision**

Using our best performing predictor of Table 1 we tested the hypothesis of chemical-functional equivalence between siRNAs and endogenous miRNAs in silico. In particular, even if the computational method was trained to identify active siRNAs, it assigns high scores also to the endogenous miRNAs. However, the two types of small RNAs maintain a subtle difference: the designed siRNAs have a perfect base-matching with the corresponding miRNAs, while the miRNAs extracted from the data have only a partial base pairing.

Nonetheless, the results reported in Fig. (2), confirm the hypothesis, since the neural-network predicts high inhibition activities for most of the experimental miRNAs. This is true for both human microRNAs derived from TarBase and miRBase databases (Fig. 2). In Fig. (2), where we report the distribution of the predicted efficacies for miRNAs, it is evident that the vast majority of miRNAs are predicted with an inhibitory level higher than 0.7. Moreover, the random selection of the human cDNA sequences behaves as expected, showing a normal-like distribution of the predicted activity with median value around 0.7 (Fig. 2). It is worth noticing
Fig. (1). Correlation between observed and predicted inhibitory activities for the test set of 249 siRNAs. The Pearson correlation value is 0.68.

Fig. (2). Distribution of predicted inhibitory activities for the two miRNA data sets of (from TarBase: dotted line; from miRBase: gray line) compared to the random selection of cDNA candidates (black solid line).
that the distribution of the inhibitory activity values of the siRNAs experimentally obtained [35] has a similar shape and the same median value. This findings not only support the idea of siRNAs and miRNAs equivalence but indicate that it is possible to adopt computational tools developed using synthetic small interference RNAs to investigate endogenous miRNAs.

CONCLUSION

In this paper we developed a new method for predicting the inhibitory activity of small interference RNAs starting from their sequence. Our results indicate that an input encoding that takes into account di-peptides and tri-peptides information achieves higher accuracy with respect to simpler input encoding based only on the sequence representation [35]. Furthermore, with a computational approach we showed that siRNAs and miRNAs can be functionally interchangeable. Although several experimental evidences support this hypothesis, to the best of our knowledge this is the first time that a computational method confirms the relationships between siRNAs and miRNAs.

ACKNOWLEDGEMENTS

This work was supported by the following grants: LIBI-Laboratorio Internazionale di BioInformatica and the Biosapiens Network of Excellence project within the European Union’s VI Framework Programme (contract number LSHG-CT-2003-503265), delivered to RC. IR acknowledges the EU FP6 Specific Targeted Research Project TargetHerpes (LSHG-CT-2006-037517) for support.

REFERENCES

In Silico Evidence of the Relationship Between miRNAs and siRNAs


Received: February 25, 2007 Revised: April 2, 2008 Accepted: April 5, 2008

© Montanucci et al.; Licensee Bentham Open.
This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/license/by/2.5/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.