Structural determinants of selected stages of small regulatory RNA biogenesis in Arabidopsis thaliana

Agnieszka Mickiewicz-Piotrowska

Abstract

For plants, the regulation of gene expression is a process through which they can defend themselves against exogenous pathogens, control development processes and respond to changing environmental conditions. Small regulatory RNAs, which are elements of the pathway of post-transcriptional regulation of gene expression, are a precise and fast way to reduce the level of gene expression, which makes them a potential tool for research of molecular biology and application purposes on cultivated plants.

The biogenesis of small regulatory RNAs starts with long RNAs with double-stranded fragments in their structure. One of the DCL proteins cuts 20-24 nt duplexes from them, which are then unravelled in the microprocessor complex where the leading strand remains attached therefrom as a template for mRNA recognition. In the next stage, this complex attaches the mRNA and cuts it at the site of srRNA hybridization. Each of the DCL plant proteins cuts fragments of its own characteristics, fixed length, however, the structure of none of them has yet been experimentally determined. The structure of the miRNA:miRNA* duplex and the interaction between the miRNA guide strand and the target sequence in the mRNA seems to have a significant impact on the effectiveness of posttranscriptional gene expression regulation, however, the exact structural parameters of srRNAs determining their effectiveness are unknown. As part of this dissertation, the following research objectives have been achieved in order to learn about the structural conditions of srRNA plant biogenesis:

1. A number of features have been identified, both at the sequence and secondary structure level of the dsRNA. Two sets of data were investigated: one included miRNA: target duplexes, and the other, miRNA precursors, of which, after obtaining the secondary structure, miRNA duplexes: miRNA * duplexes extended by 10 nt before and after the miRNA

sequence were collected for further analysis. In this way, the reference thermodynamic profiles were obtained, which were later used as a parameter to verify potential amiRNAs.

2. A new algorithm was developed that uses the thermodynamic features of the native structures that make up the miRNAs in their precursors as well as interacting with the target transcript of the silenced gene. AmiRNA Designer – a bioinformatics tool to design sequences for silencing a specific gene was designed. The developed program was used to generate artificial miRNA sequences – amiRNAs – for exemplary genes. After designing each sequence, the program checks its compliance with the parameters of the thermodynamic profiles. A method of introducing modifications to the srRNA sequence was developed so that the thermodynamic details of the duplexes formed by the designed sequences would better fit the profiles. It runs according to the following plan

1. Obtaining complementary sequence to given TARGET=>amiRNA
2. Compare the thermodynamic details of the duplexes formed by the designed sequences with the thermodynamic profile of the TARGET:miRNA duplex
3. amiRNA sequence modifications, re-comparison as above
4. Obtaining the final sequence of amiRNA
5. Obtaining complementary sequence to amiRNA= >amiRNA*
6. Compare the thermodynamic details of the duplexes formed by the designed sequences with the thermodynamic profile of the miRNA:miRNA* duplex
7. amiRNA* sequence modifications, re-comparison as above
8. Obtaining the final sequence of amiRNA*

3. The spatial structure of the AtDCL4 protein in complex with its dsRNA substrate was modelled, determining its relationship between the structure of the enzyme and the length of generated dsRNA fragments. Using *in silico* methods, a few protein structural features were identified – RNA-binding catalytic sites and folds, domains and individual amino acid residues – responsible for dsRNA binding and directly impacted the formation of srRNA. In particular, comparative modelling (based on homology) and modelling of the protein-RNA complex, the method of normal mode analysis and molecular dynamics simulations of protein-RNA complexes were used. Modelling the protein structure based on scarce data – few known structures and their low sequence similarity to the modelled protein – turned out

to be a non-trivial task. Both fragments of the sequence without any equivalent that could constitute a structural template and those with little resemblance to known structures required more attention and additional efforts. The obtained models were submitted to a series of computer analyses verifying their reliability as high, as well as a comparative analysis combining the levels of sequence, structure and data from biochemical experiments.

As a result of the implementation of the objectives of this dissertation, five publications were created, both reviewing and describing the results of own research, as well as the AmiRNA Designer computer program, which aims to design silencing sequences specifically targeted at a specific gene, in particular for plants.