Chapter 10. Summary

10.1 Summary

RNA interference (RNAi) is one of the important discoveries (1998) in the field of biology, which coincided with the accomplishment of human whole genome sequencing project (2002). With RNAi providing an option for selective silencing of the gene products (mRNA), the other project provided us with sequencing information encoded in the genome and their products. The simultaneous occurrence of these two events created the way for many scientific labs to explore and blend them and within no time RNAi technique was used for gene knockdown assays. Subsequently, it became one of the routine molecular biology procedures. The technique of RNAi was further implemented for application as a potential therapeutic. In most of the cases, the abnormalities that result from protein expression (from mRNAs) could now be silenced. Also, as the antisense oligonucleotides (Fomivirsen) were already approved by the US-FDA (Food and Drug Administration) for medicinal use, the field of siRNA based gene silencing flourished exponentially. As for now, the siRNA dependent silencing has become a well-known molecular biology technique for single or genome wide screens in functional genomics. Though, for its therapeutic application, there are certain voids that need to be addressed first. These include their efficacy, cell specific delivery, off-target effects, toxicity, nuclease susceptibility and others. Different types of chemical modifications are engineered on either or both of the siRNA strands to overcome these limitations.

Till now, most of the computational work has been focused on the development of resources on normal siRNAs, but for now, almost normal (non chemically modified) siRNAs are outdated. Especially, for pre-clinical and clinical trials the siRNAs are chemically modified in one or another way to enhance their properties. On literature search, we do not find any bioinformatic resource for chemically modified siRNAs. So we developed siRNAmod, a first resource of experimentally tested chemically modified siRNAs with more than 5000 siRNAs sequences and more than 120 chemical modifications. These modifications were performed on all the three basic components (sugar, nitrogenous base and phosphate) of the building blocks (nucleosides) of siRNA. Also, the modifications are present for on both strands and can be more than one type. In case of more than one type on modification, different permutations and combinations are performed to check their effect. Detailed

information on the siRNAs, chemical modifications, experimental procedures etc. is provided. It is free for scientific use and is available on the URL: http://crdd.osdd.net/servers/sirnamod.

Modifying the normal siRNA with chemical modifications helps to overcome various limitations but it may have the effect on its knockdown efficiency. No algorithm was present in literature for predicting the efficacy of the chemically modified siRNAs. We curated the data and selected top 31 modifications and their efficacy values and developed the SMEpred platform. We used various features of these modified siRNAs including the information of chemical modifications and developed various models based on Support Vector Machines (SVM) technique. We integrated the three models in the web server that predict the efficacy of the chemically modified siRNAs. The mononucleotide composition based model provided best performance (PCC) of 0.80 in 10-fold cross validation and similar performance on independent validation dataset. We also added two other models that provide the positional information of these modifications on the antisense strand, as this information is important in terms of defining the siRNA activity. The web server is available on the following web-address: http://bioinfo.imtech.res.in/manojk/smepred.

Toxicity presents another important aspect of chemically modified (as well as normal) siRNAs that may arise due to chemical moiety or off-targets of basic nucleotide sequence. We collected data on chemically modified siRNAs and developed TOXsiRNA web server that predicts the toxicity of the 22 different chemically modified siRNAs. We employed four different machine-learning techniques (Linear Regression (LR), Artificial Neural Networks (ANN), k-Nearest Neighbor (KNN) and Support Vector Machines (SVM)) for model development and found that SVM performed best. We selected three models, for final web server implementation. We have also integrated the other tools for prediction of the siRNA efficacies, off targets etc. in the web server, which is available at the following link: http://bioinfo.imtech.res.in/manojk/toxsirna.

Delivery has been another important issue for the use of siRNAs as therapeutic and various types of methods are applied for the same. However, peptides are already in the market as delivery vehicle or as therapeutic. Furthermore, the peptides have

shown capacity on their own to overcome all the hurdles in the delivery of the siRNA, right from beginning (condensation with RNA) up to the release of siRNAs in the cytoplasm. For this endeavor, we developed PEPsRNA, a web server for peptide based siRNA delivery with manually curated, analyzed data from research articles and patents. It included about 225 different types of peptides and more than 2250 associated siRNA sequences in it. In addition, we have also provided the information on the conjugate molecules (ranging from metals like gold particles to large polymers) that were added to enhance the delivery of peptide-vehicle. Detailed information with analysis of the peptides, siRNAs as well as conjugate molecules is provided. The resource is available on the following URL: http://bioinfo.imtech.res.in/manojk/pepsirna.

To design and develop the efficient repertoire of the siRNAs is a bit tedious job especially for the viruses for which we need high-level biosafety facilities and live viruses to deal with. We used our in-house web server VIRsiRNApred and dual luciferase assay system to quickly design and check the efficacy of the siRNAs *in vitro*. We chose dengue virus (because of its global disease burden) and selected eleven siRNAs for its five different regions (two siRNAs for each genome segment but three for envelope region). The regions range from one end to another end of the virus namely 5'-Untranslated region, Capsid, Pre-Membrane, Envelop and 3'-Untranslated region. Furthermore, we selected siRNAs with efficacy values ranging from very high to low. Two siRNAs for the envelope gene were already present in the literature and experimentally evaluated using the dengue viral system. This was to provide an external validation to our *in vitro* dual luciferase assay model. In addition, we also designed high efficacy four siRNAs, two for each of luciferase genes. We performed the *in vitro* experiments and most of the siRNAs achieved efficacy as per the predicted values (except for two). The two envelope siRNAs also showed similar performances in dual luciferases assay system as reported in the earlier virus based study. We also tested the pooling of the siRNAs for each gene and their performance was almost equal to respective constituent siRNAs when the two had similar efficacies. In case of high and low siRNA efficacy pooling, the effect of knockdown was lowered. The viability tests for each siRNA showed their non-toxic behavior with respect to control (scrambled siRNA).

Another interesting phenomenon associated with the small double stranded RNA is gene expression induction. In this process the small double stranded RNA, around 21 nucleotides long, guided to the promoter of the gene leads to its induced expression. The small RNAs are called saRNAs for small activating RNAs as opposite to siRNAs and the phenomena called as RNA activation as opposite to RNA interference. The field is recently discovered but is related to the small RNA and various molecules are in pre-clinical and clinical trials for various human pathologies. Because of its relevance to RNAi and siRNA, we developed a web resource (saRNAdb) for the same with detailed information on the small activating RNA sequences, their analysis and related experimental information. Furthermore, we collected the information on the proteins that are involved in the process (although the mechanism is not clearly defined yet) and extracted their interacting partners and provided the information on the web server. It is also freely available for public use on the following URL: http://bioinfo.imtech.res.in/manojk/sarna.

10. 2 Future implications

The field of small RNA based therapeutics is growing exponentially by day and night and a lot of experimental information is accumulating on its different aspects. So our main focus will be to keep the information up to date and also increase the domain of our computational resources in the future. Furthermore, in case of the prediction systems, we will try to increase the number of chemical modifications and incorporate more features of these chemical moieties in predicting their efficacies and toxicities when engineered on siRNAs. In case of the delivery methods, we will incorporate all other types of vehicles used for small RNA delivery. Likewise, we will update and incorporate the information on the small activating RNAs and its various fields too. Finally, as we have generated a set of effective siRNAs for dengue virus, we will try to move them to the next level and test them in *in vivo* using live viruses. We will also increase our domain for other viruses and develop effective siRNAs against them too that will be useful in R&D as well as in the development of potential therapeutics.