12   Summary and Future prospects

12.1   Summary

Viruses are responsible for causing a variety of diseases. Some common diseases caused by them include AIDS, influenza, chickenpox, hepatitis, SARS and cold sores. The comparative capacity of viruses to produce infection is expressed in terms of virulence. Depending on the viral species, they have diverse mechanisms using which they generate disease in the host. Conventionally drugs and vaccines are being used in controlling the virus epidemics. However researchers have found that novel methods like antiviral RNAi therapy, virus inhibitory peptides and antiviral compounds have proved as effective approaches. During the past decade a lot of data has been generated in these fields as a result of experimentation by different workers.

Till date a handful of siRNA resources have been developed that spotlight on dealing with human and mammalian genes. However, viral siRNA databases that provide experimentally validated data are missing. We designed a manually curated resource, VIRsiRNAdb that harbors inclusive niceties of 1358 siRNAs directed against 42 important viruses. This resource also gives the users with amenities such as browsing, advance search, data submission, hyper linking to other resources and valuable siRNA investigation tools specially siTarAlign that aligns the siRNA with reference viral genomes to check target conservation. This resource is freely accessible at http://crdd.osdd.net/servers/virsirnadb.

Choosing efficient viral siRNA is a crucial rung in the creation of siRNA based antivirals. In spite of colossal prospectives, an algorithm that can predict viral siRNA effectiveness is not present. Besides, accuracy of the current general mammalian siRNA efficiency algorithms is not adequate for viral siRNAs. Hence, we came up with "VIRsiRNApred" a support vector machine based algorithm that can predict the effectiveness of viral siRNA. VIRsiRNApred is the only method that can predict efficiency of viral siRNAs and is developed employing experimentally confirmed viral siRNAs. This approach will be valuable in selecting highly potent viral siRNA to assist development of siRNA based antivirals. The web server can be accessed via http://crdd.osdd.net/servers/virsirnapred/.

Peptides have also shown potential to stall the Human immunodeficiency virus (HIV) lifecycle besides the antiretroviral drugs. We collected experimentally verified data on anti-HIV peptides and developed a resource namely, HIPdb that provides comprehensive details about HIV inhibiting peptides (HIPs). The HIPs target a variety of pathways in the life cycle of the virus e.g. attachment, replication, assembly etc. This resource furnishes details of 981 HIPs. Furthermore, physicochemical properties and predicted structure of the peptides are also integrated. HIPdb will be supportive in choosing valuable HIPs. Hence it can be helpful to researchers working in the field of peptide based drugs development against HIV. This web based resource is available via http://crdd.osdd.net/servers/hipdb.

Since numerous peptide drugs have been considered for clinical trials the need for antiviral peptide (AVP) resources is growing. In addition to anti-HIV peptides database, we also developed AVPdb, a comprehensive resource of experimentally established AVPs. The database has 2683 AVPs directed against more than 60 medically important viruses including HCV, HBV, SARS, Influenza, DENV, RSV, HSV etc. The resource also provides information on 624 modified AVPs in which a chemical group is linked to enhance their efficiency and stability. AVPdb caters information about: peptide sequence, its source, target virus along with its taxonomy, cell line, inhibition efficiency, target protein and assay along with references. AVPdb further provides predicted structure as well as physicochemical properties for each peptide. The resource has easy to use search/browse options and additional analysis tools. The web based resource is expected to assist researchers working for the improvement of peptide based antivirals. AVPdb can be accessed at http://crdd.osdd.net/servers/avpdb.

In spite of titanic significance of AVPs, a committed prediction resource was lacking. To cater this need we selected 1245 peptides that were directed against imperative human viruses like HIV, HCV, influenza etc. We used assorted peptide sequence attributes like amino acid composition, physicochemical properties, motifs and alignment for developing the algorithm. Support Vector Machine based physiochemical properties model showed 0.70 Matthew's Correlation Coefficient (MCC) with an accuracy of 85%. On independent data set the algorithm reached 0.71 MCC with an accuracy of 86%. The AVPpred web server will be helpful in choosing efficient AVPs. The algorithm is presented at http://crdd.osdd.net/servers/avppred.

AVP-IC$_{50}$Pred is a regression grounded IC$_{50}$ predictor created employing experimentally confirmed peptides.

Evaluation of peptide antiviral activity as IC$_{50}$ takes a lot of time and effort. Hence, we developed AVP-IC$_{50}$Pred, a regression grounded predictor of peptide antiviral activity in terms of IC$_{50}$. Unique experimentally checked peptides with quantitative efficacies from AVPdb and HIPdb were used for model development. Several MLTs were employed to develop all-inclusive prediction models. Imperative peptide attributes like amino acid compositions, binary patterns and selected physicochemical features were used in model development. This web server presented at http://crdd.osdd.net/servers/ic50avp will aid the researchers to select highly efficient AVPs.

Antiviral compounds (AVCs) are a category of antimicrobial drugs used specially for treating viral infections by inhibiting the development of a viral pathogen inside the host cell. We have developed Quantitative structure-activity relationship (QSAR) based models for predicting AVCs against deadly viruses like Human immunodeficiency virus (HIV), Hepatitis C virus (HCV), Hepatitis B virus (HBV), Human herpesvirus (HHV) and 26 others using publicly available experimental data from the ChEMBL bioactivity database. Support Vector Machine (SVM) models achieved a maximum Pearson Correlation Coefficient of 0.72, 0.74, 0.66, 0.68 and 0.71 during 10-fold cross-validation and 0.63, 0.65, 0.61, 0.64 and 0.67 on the independent validation sets. This algorithm will be helpful for virtual screening of AVCs and designing new molecules to target the viruses. User can also sketch the desired compound or generate analogs based on given building blocks and predict their inhibition on the viruses using tools provided on the web server. AVCpred is also hoped to assist the researchers in the identification of novel antiviral agents as well as save time and money for discovering a new drug before the synthesis and animal testing of chemicals. We have integrated these models in the AVCpred web server, freely available at http://crdd.osdd.net/servers/avcpred.

Human Immunodeficiency Virus (HIV) infections constantly claim a huge toll of human lives each year. Researchers have identified many inhibitors for treating HIV infections by targeting indispensable viral proteins. In the present study, we have developed Quantitative structure-activity relationship (QSAR) based models for predicting inhibitors against HIV proteins such as

reverse transcriptase, protease and integrase using publicly available experimental data from the ChEMBL bioactivity database. Support Vector Machine (SVM) models achieved a maximum Pearson Correlation Coefficient of 0.68, 0.76 and 0.72 during 10-fold cross-validation and 0.63, 0.60 and 0.65 on the independent validation sets. This algorithm will be helpful for virtual screening of inhibitors and designing new molecules to target the important HIV proteins and speed up the drug discovery process. HIVprotI web server is freely available at http://bioinfo.imtech.res.in/manojk/hivproti.

12.2   Future prospects

As rising number of papers are being reported in the region of viral RNAi and antiviral peptides, consequently, in prospect our focal precedence would be to bring up to date the current viral siRNA and AVP data and also include new viruses once apposite data is accessible. Also peptides targeting different strains/types of HIV is also one of the main future prospects for HIPdb. In case of VIRsiRNApred, presently the data for viral RNAi is composed of a large assortment of studies because of which the general accuracy of the model is affected. As more homogeneous siRNA data is provided, we shall update our algorithm to comply with with the new data to make the algorithm more useful. A virus specific siRNA design tool may also be developed once enough data is available. There is also a need to develop resources on chemically modified siRNAs. In addition valuable data that includes newly discovered unique AVPs for other viruses will necessitate to update our AVP prediction algorithms. Their performance can be improved further if a large homogeneous dataset is available in future. Also, algorithms to predict the IC_{50} value of virus specific peptides should be possible in the near upcoming. This will also encourage developing newer regression-based algorithms besides the existing classification-based methods in the area of peptide-based therapeutics. Prediction algorithms for chemically modified peptides is yet another future work. In case of AVCpred and HIVprotI, we used the pharmacological data from the ChEMBL resource for training/testing the models developed for general as well as specific viruses and HIV proteins. However, as more high throughput screening data on antiviral drugs becomes available, performance of the QSAR methods will be improved.