7 Summary and future prospects

7.1 Summary

HIV infection is an epidemic with about 35 million infected peoples throughout the world. It is estimated that about 2 million people die each year with HIV/AIDS and new 2.5 million infections occur worldwide. Although HIV has been identified about 30 years ago, it is still a challenge to complete eradicate this virus from human body after infection. There are two main factors behind that- (1) high rate of mutations in its genome which lead to new variant and development of drug-resistant strains (2) high rate of replication, producing millions of virus particles in an infected person in a day. Though HAART therapy is effective, there is continuous emergence of HIV strains which are drug-resistant in nature. This trend shows that there is an urgent need for development of therapies which can target HIV at various points in its life cycle. HIV life cycle is dependent upon intricate network of protein-protein interactions among HIV-Human, Human-Human and HIV-HIV protein-protein interactions. There have been a number of studies which identified human proteins which are used by HIV in its life cycle. Also, a number of genomic screens have identified host factors which are recruited by HIV in process of its life cycle. Although, the number of factors overlapped with each other in these screens is very much less. This signs that a number of other factors have to be discovered which are essential from HIV survival. There are a number of factors which have been known non-essential for human but vital for HIV survival *e.g.* CCR5 is a chemokine receptor which is a 7-transmembrane GPCR family protein found essential for HIV but not for human. Similarly, there are numerous other host factors which yet to be discovered. In this study, we have made systematic attempt to know the human proteins which can directly interact with HIV proteins. The 'HIV-1 human protein interaction database' is a protein interactions repository at NCBI (http://www.ncbi.nlm.nih.gov/projects /RefSeq/HIVInteractions/), catalogued the major interactions among HIV-1 proteins and human proteins. There are two major categories of interactions: direct and indirect. In our study, we made an attempt to develop a prediction method which can predict the human proteins which can directly interact with HIV-1 proteins. There are mainly two models in this approach: (A) an overall model which predicts whether any human proteins will interact with HIV-1 proteins (B) Nine specific models for each HIV-1 protein, providing the probability of interaction of human protein with any of these nine HIV-1 proteins. We have used different types of input features to develop models for predicting HIV-Human protein interactions *e.g.*

BLAST, AAC, DPC, SAAC, Domains, Domains and various composition based hybrid models. It has been noticed that SAAC based model proved the best among composition based models, having accuracy 74.90% with MCC 0.54. In hybrid models, SAAC + domain based hybrid model was best with 75.07% accuracy and MCC 0.54. The prediction accuracy of each HIV-1 individual protein specific model varies as the input for each category was different. The performance accuracy varies from 75.83% to 86.17% for AAC, 72.26% to 82.32% for DPC, 74.89% to 85.38% in case of SAAC based SVM model. We have also developed model based on domain and hybrid model based on domains with AAC, DPC and SAAC based input features. It is assumed that such prediction method will help in identification of new human proteins which have the ability to interact with HIV proteins. For assisting the scientific community, a webserver 'HIVInt' has been developed and made freely available at 'www.imtech.res.in/raghava/HIVInt'.

It is well known that HIV uses CD4 receptor as its primary receptor and CCR5 or CXCR4 as its secondary receptor for successful invasion in human cells. This surface level protein-protein interaction has been discovered as a point to hinder HIV infection by blocking the CCR5 receptor by using CCR5 antagonist. Maraviroc, a CCR5 antagonist have been approved by FDA in 2007 for blocking CCR5 receptor and inhibiting HIV infection process. Genotypic as well as phenotypic methods have been developed to know the co-receptor used by HIV-1. The performance of genotypic methods is not satisfactory especially in case of CXCR4(X4)-tropic co-receptor usage prediction. It is well known that third variable loop of HIV gp120 protein is the main determinant of co-receptor usage. In our study, we have made a systematic attempt to develop an *in-silico* method for co-receptor usage prediction, based on the V3 loop sequences of HIV. We have taken the largest data set of R5- and X4-tropic V3 sequences and formed the positive and negative datasets, respectively. The third type of V3 sequences *i.e.* R5X4-tropic sequences was also included into the X4-tropic sequences *i.e.* negative (X4-tropic) dataset. We used support vector machine as machine learning technique to develop models based on various input features *e.g.* AAC, DPC, SAAC etc. It was found the BLAST was not efficient enough to discriminate between R5- and X4-tropic V3 sequences. It is found that the best composition based model predicted X4-tropic sequences with 80% accuracy, in order to enhance the prediction performance a hybrid model have been developed by using SAAC and BLAST. The hybrid approach found to be highly efficient in discriminating R5- and X4-tropic sequences with sensitivity 91.66%, specificity 81.77%, accuracy 89.19% and MCC 0.72. It is important to mention that besides predicting HIV-1 co-

for subtype B, it performed high accuracy when competed with other methods
:d for specific subtypes *e.g.* D, E, and AE. This is due to the factor that the dataset
HIVcoPred included all V3 sequences of all subtypes of HIV-1. A comparison of our
:d method 'HIVcoPred' has been done with other well known prediction methods for
receptor usage prediction. It is found that our method performed better or equally
comparison with the other developed methods of prediction of HIV-1 co-receptor
or service to scientific community, a webserver 'HIVcoPred' has been developed and
:ely available at 'www.imtech.res.in/raghava/HIVcoPred'.

scovery of CCR5 as HIV co-receptor and its subsequent blocking in successful HIV
ı lead to numerous chemical compounds to be designed as CCR5 antagonist. Lots of
ne derivatives, Pyrrolidines, Guanylhydrazone, Benzylpyrazole and Phenoxybenzyl
ves etc have been tried to develop CCR5 antagonists which may block the CCR5
· on human cells. In 2007, FDA approved Maraviroc (Selzentry) (belonging to
ropylamines, Tropane, Alkaloids and Alkaloid derivatives) to be used as CCR5
ist. Since then a number of other chemical compounds have been synthesized and
:o be used as anti-HIV compounds. In our study, we have also developed a
ational model to design better CCR5 antagonist by calculating the inhibitory activity
own compounds having ability to bind with CCR5. We took 91 chemical compounds
idine derivatives, originally used by Minghu Song (Song *et al.*, 2004). We calculated
criptors of these compounds and correlated these with physical property (pIC_{50}).
PaDEL software was used for descriptor calculation and a correlation coefficient (R)
:fficient of determination (R^2) of 0.92/0.84 respectively has been achieved.

ıave been a number of chemical compounds, peptides, and antibodies etc that have
:ed to block the initial infection by HIV. There are attachment inhibitors which block
eraction between gp120 and CD4 receptor, CCR5 antagonists which block the
:ion between V3 loop of gp120 and CCR5 receptor on human cell surface, and the
inhibitors which block the gp41-mediated cell membrane fusion of host and HIV cell
anes. Till date, there is no chemical entity have been designed to block the gp41-
ed cell membranes fusion. Enfuvirtide (T20) is the only FDA approved fusion
or, which is a 36 amino acid long peptide and block gp41-mediated cell membrane

Chapter 7 [Summary and Future prospects]

In our study, we have made a systematic attempt to predict the fusion inhibitors by calculating the *pEC50* and *pIC50* value of chemical compounds with potential to inhibit fusion process of HIV and human cells. We calculated the structural features in the form of descriptors and formed QSAR models. The model predicted the correlation value between the descriptors and the inhibitory values, as the highest R, R^2 were 0.85 and 0.72, respectively. Similarly, the best model developed for predicting the effective concentration achieved R and R^2 as 0.79 and 0.54, respectively. We have applied these QSAR models to known ZINC database and FDA approved compounds to know the possible compounds with their potential as potent fusion inhibitors. It can be said that such prediction models will help in discovering and designing of new effective fusion inhibitors.

HIV-1 is known as a notorious virus with continuous change in its genetic material to avoid drug susceptibility. Among the various option to check the virus replication in human body, siRNA based therapy have been designed as one of these methods. Since HIV life cycle is dependent upon numerous host factors, a number of siRNAs have been tried against human proteins to block HIV life cycle *e.g.* CD4, CCR5 etc. similarly, a number of siRNAs have been designed to block the expression of HIV gene *e.g. tat, rev, nef, vif* etc. In our study, we tried to predict and design the most promiscuous siRNAs which can block HIV gene expression. For designing the siRNAs, we have used published algorithm 'desiRM', which can predict complementary siRNA as well as mismatched siRNA against the target sequences. The input data is consists of full length of genomic sequences of nine HIV-1 genes, subtype B. We allowed the algorithm to predict siRNAs against the whole length of input sequences and selected only those siRNAs which have efficacy equal to or greater than 0.8. Among these siRNAs, top five siRNAs which targeted maximum number of sequences have been selected for each genomic region. It is found the most of the genomic sequences have been covered by these siRNAs, in some cases the coverage was less. In order to cover the whole input sequences, mutated version of these top five siRNAs have been designed and targeted the remaining sequences. It has been found that the siRNAs with their mutated version targeted most of the input sequences in each category of nine HIV-1 genomic regions. In this way, we can have a set of highly promiscuous siRNAs which can target HIV-1 genomic regions with efficacy equal to or greater than 0.8.

Future prospects

ı-protein interactions play important role in HIV life cycle. It has been discovered that ıg the expression of some proteins resulted in halting HIV life cycle. Though important rug designing point of view, number of human proteins which interact with HIV is yet liscovered. A number of studies are currently investigating the human proteins which t with HIV proteins. We have designed a method 'HIVInt' which would help in ying in human proteins physically interacting with HIV proteins. Since Maraviroc have pproved by FDA which block cellular receptor CCR5, any new technique or method may help in identification critical cellular factor capable of interacting with HIV will ıigh use. Our method 'HIVInt' is one such method, capable of predicting the human s directly interacting with HIV-1 proteins. We hope that such method will help in fast cation of cellular factors required by HIV and whose blockage will result in inhibition life cycle.

andatory to know the exact type of HIV strains before prescribing any coreceptor-ıist as a drug. Although a number of phenotypic as well as genotypic methods have esigned to determine the co-receptor usage, the CXCR4 usage prediction with high cy is still a challenge. Since determining co-receptor usage by phenotypic assay is a time consuming and expensive procedure, number of genotypic methods have been ıed to assist the fast characterization of HIV strains. In this thesis, a new HIV co-r usage prediction method has been designed. This method 'HIVcoPred' use a novel approach developed by using SAAC and BLAST, is highly efficient in determining eceptor usage by HIV of major subtypes *e.g.* B, C, D, AE etc. Since determining co-r usage is prerequisite for CCR5 antagonist usage, any tool or method which will help ate determination of co-receptor usage will be of high importance. Since our method Pred' have been developed on the largest dataset among all genotypic developed s, it is capable of predicting co-receptor usage of majority of subtypes. Hence it is a luable tool for HIV therapeutics and directly related with quick assessment of HIV patients for possible cure with co-receptor antagonist *i.e.* Maraviroc. It is important hat since there are a number of other factors which also play important role in co-usage determinations *e.g.* C4, V1/V2 region and gp41 region, more sophisticated s for determining co-receptor usage may be developed with integrated information of gions.

Chapter 7 [Summary and Future prospects]

The viral entry into human cells is a complex phenomenon involving multiple receptor and ligands *e.g.* gp120, V3 loop of gp120, gp41 proteins of HIV whereas CD4 as main receptor, CCR5 or CXCR4 as co-receptor are part of cellular machinery which involves in initial fusion process. A number of peptide, antibodies, chemical compounds has been designed to inhibit initial fusion process. We have designed QSAR models which can predict the inhibitory (pIC_{50}) value of any potential lead compound for blocking CCR5 expression and act as new CCR5 antagonist *e.g.* Maraviroc. Also we have designed QSAR models which predict the inhibitory (pIC_{50}) as well as effective concentration ($pEC50$) of chemical entities which have capabilities to act as potential fusion inhibitors. Since only a peptide based drug *i.e.* Enfuvirtide has been approved as fusion inhibitor, there is high necessity of chemical compounds which can block or prevent fusion process. It is anticipated that our study will helpful in identification of lead compounds having the capabilities to inhibit HIV fusion process.

HAART is available therapy to keep HIV RNA load below a level which give rise to AIDS. A number of therapies have been designed to inhibit HIV life cycle at various points including siRNA based therapy *i.e.* RNAi. In order to design highly effective siRNAs against HIV genome, we have used desiRM algorithm and predicted five most promiscuous siRNAs against each of the nine major genomic regions of HIV. Since HIV mutates very rapidly or mutates its target sequence quickly in order to avoid inhibition by any therapy, the siRNAs are designed which can target the largest number of sequences of HIV with high efficacy. Also, the mutated forms of these promiscuous siRNAs have been designed to target the sequences which were not targeted originally by five siRNAs. siRNAs based therapies have been designed earlier by various research groups but none of them have been cleared by FDA in order to inhibit HIV life cycle. We anticipate that our predicted siRNAs will be highly effective against HIV *in-vivo* and will result into inhibition of HIV cycle. As HAART uses 2-3 medicines in its course to prevent the arise of drug-resistant HIV strains, similarly multiple siRNAs based therapy (targeting viral or cellular factors essential for virus) have to designed to effectively inhibit HIV and developing new siRNA based treatment.