Genomes to Hits *In Silico* – Towards the discovery of antiviral inhibitors for NS1, NS3 and NS5 from Dengue

A Major Project dissertation submitted

In partial fulfillment of the requirement for the degree of

Master of Technology

In

Bioinformatics

Submitted by

Jai Nandini

(DTU/13/M.Tech/357)

Delhi Technological University, Delhi, India

Under the supervision of

Dr.Yasha Hasija

Department of Biotechnology
Delhi Technological University
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road,
Delhi-110042, INDIA

CERTIFICATE

This is to certify that the M. Tech. dissertation entitled **"Genomes to Hits *In Silico* – Towards the discovery of antiviral inhibitors for NS1, NS3 and NS5 from Dengue"**, submitted by **Jai Nandini (DTU/13/M.Tech/357)** in partial fulfilment of the requirement for the award of the degree of Master of Technology, Delhi Technological University (Formerly Delhi College of Engineering, University of Delhi), is an authentic record of the candidate's own work carried out by her under my guidance.

The information and data enclosed in this dissertation is original and has not been submitted elsewhere for honouring of any other degree.

Date:

Dr. Yasha Hasija
(Project Mentor)
Department of Bio-Technology
Delhi Technological University

Prof. D. Kumar
(Head of the Department)
Department of Bio-Technology
Delhi Technological University

DECLARATION

I, **Jai Nandini (DTU/13/MTECH/357)** declare that M. Tech. dissertation entitled **"Genomes to Hits *In Silico* – Towards the discovery of antiviral inhibitors for NS1, NS3 and NS5 from Dengue"**, submitted in partial fulfilment of the requirement for the award of the degree of Master of Technology, Delhi Technological University (Formerly Delhi College of Engineering, University of Delhi), is an authentic record of my own work carried out under the guidance of **Dr. Yasha Hasija**.

The information and data enclosed in this dissertation is original and has not been submitted elsewhere for honouring of any other degree.

Date: Name:

Place: Signature:

ACKNOWLEDGEMENT

I take this opportunity to express my sincere gratitude to all those who have been helpful in the successful completion of this project.

Assistant Professor Dr.Yasha Hasija, Deptt. Of Biotechnlogy, Delhi Technological University, my project guide, has guided me for the successful completion of this project. It is worth mentioning that she always provided the necessary guidance and support.

I would also like to present my sincere regards to my Head of Biotechnology Department Prof. D. Kumar, for his support and encouragement throughout the programme.

I am also thankful to all the faculty and staff members especially Mr. Chhail Bihari and Mr. Jitendra. I am also very thankful to my classmates for the support and motivation during this work. To all the named and many unnamed, my sincere thanks. Surely it is Almighty's grace to get things done fruitfully.

JAI NANDINI
2K13/BIO/07

CONTENTS

	TOPIC	PAGE NO
	LIST OF FIGURES	1
	LIST OF TABLES	2
1.	ABSTRACT	3
2.	INTRODUCTION	4
3.	REVIEW OF LITERATURE	5
	3.1 DENGUE FEVER	5
	3.1.1 DENGUE SYMPTOMS	5
	3.2 VIRAL STRUCTURE	6
	3.2.1 LIFE-CYCLE OF DENV	7
	3.3 THE DENV TARGET FOR ANTIVIRAL RESEARCH	8
	3.3.1 THE NON-STRUCTURAL PROTEINS	9
	3.4 GENOME ANOTATION	11
	3.5 PROTEIN TERTIARY STRUCTURE PREDICTION	13
	3.6 DRUG DESIGN	15
4.	METHODOLOGY	19
5.	RESULTS	27
6.	CONCLUSIONS	47
7.	DISCUSSION AND FUTURE PERSPECTIVE	48
8.	REFERENCES	49

LIST OF FIGURES

Figure	**Page No.**
Figure-1 Dengue virus gene structure	6
Figure-2 Excel sheet of gene information.	20
Figure-3 Homepage of Chemgenome	21
Figure-4 Homepage of Bhageerath	22
Figure-5 Homepage of Active site prediction	23
Figure-6 Homepage of RASPD	24
Figure-7 Homepage of ParDOCK	25
Figure-8 Docking Directly from AADS	26
Figure-9 Output of Chemgenome 2.0	27
Figure-10 Tabular view of Gene	28
Figure-11 Graphical view of gene	28
Figure-12 Output of Blastn	29
Figure-13 Output of Bhageerath-H	30
Figure-14 Output of AADS	31
Figure-15 cavity 1 information	32
Figure-16 Output of RASPD	33
Figure-17 Output of ParDOCK	34

LIST OF TABLES

Table	**Page No.**
Table-1 List of Tools Available for Gene Prediction	11
Table-2 List of Tools Available a for Protein Tertiary Structure a Prediction	13
Table-3 List of Software Available for Active Site Prediction	15
Table-4 list of Softwares for Drug Design	17
Table-5 Dengue virus serotype	20
Table-6 list of molecules used in docking for NS1, NS3 and NS5	34
Table-7 list of highly scored molecules	41
Table-8 Name, SMILES and Structure of hits	42

Genomes to Hits *In Silico* – Towards the discovery of antiviral inhibitors for NS1, NS3 and NS5 from Dengue

Jai Nandini

Delhi Technological University, Delhi, India

ABSTRACT

Dengue is an ignored disease which one is the cause of 25 to 100 million infections per year and 22000 deaths every year in the regions where it is common. Despite the great efforts invested in the research and study for dengue treatments, there is still no vaccine or antiviral drug has reached in the market, and the disease treatment and cure is limited to supportive care. Among all the strategies which are used in the search for the vaccine and for development of an antiviral drug against dengue, the most successful approach for treating the dengue is inhibiting the viral enzymes.

Making a truly effective dengue vaccine has proven difficult because of a phenomenon called antibody dependent enhancement. It is very difficult to create a vaccine which is used against all four type of dengue virus and can protect people In today's scenario biological information are rich but functional knowledge are poor, so challenges are grand. These challenges includes the functional annotation of the genes present in whole genome, detection of the druggable targets and calculation of three -dimensional structures of the druggable protein targets which can be done from the information of their amino acid sequences and arriving at lead compounds from the hits for these targets. We propose here a "Genome to Hits *In Silico*" strategy and illustrate it on Dengue virus (DENV). "Genome to hits" is a novel pathway incorporating a series of steps such as gene prediction, protein tertiary structure determination, active site identification, hit molecule generation, docking and scoring of hits to arrive at lead compounds. The current state of the art for each of the steps in the pathway is high-lighted and the feasibility of creating an automated genome to hits assembly line is discussed.

So best way to design antiviral is inhibiting the viral enzymes. NS1, has been associated with a role in protective immunity. NS3 is a protease and a helicase, whereas NS5 is the RNA polymerase in charge of viral RNA replication. NS3 (non-structural 3) and NS5 carry out all the enzymatic activities needed for polyprotein processing and genome replication.

INTRODUCTION

Antiviral drug discovery is becoming increasingly important due to the global threat of viral disease pandemics. Many members of the genus Flavivirus are significant human pathogens, among which dengue virus (DENV) alone poses a public health threat. Neither vaccine nor effective therapeutics is currently available for DENV. Development of a DENV vaccine has been challenging, because of the need to simultaneously immunize and induce a long-lasting protection against all four serotypes of DENV; an incompletely immunized individual may be sensitized to life-threatening dengue hemorrhagic fever or dengue shock syndrome. The challenges associated with vaccine development have underscored the importance of development of antiviral therapies for DENV and other flaviviruses.

The antiviral agents for a dengue can inhibit (*i*) viral entry;a (*ii*) the viral proteins; (*iii*) the viral capsid; (*iv*) the viral protease; (*v*) the viral helicase; (*vi*) the viral methyltransferase; (*vii*) the viral polymerase; and (*viii*) the host target. Inhibition of viral enzymes has been proven to be the most successful of antiviral approaches. (Canard *et al., 2009*)

Flaviviruses are small, enveloped RNA viruses which cause a variety of diseases into animals and man. The dengue virus particle is about 50 nm in diameter. The 10,723-nucleotide RNA genome encodes an uninterrupted open reading frame (ORF), directing the synthesis of a polyprotein precursor in the order NH2-C-prM-E-NS1 - NS2A-NS2B-NS3 -NS4A-NS4B-NS5-COOH, where C is the capsid protein, M is the membrane-associated protein, E is the envelope protein, and NS1 through NS5 are nonstructural proteins. NS1 is produced in large quantities during dengue viral replication, and it can be detected as early as the first day the patient experiences a fever. NS3 and NS5 are sufficiently conserved within the four serotypes. The flavivirus proteases, including NS2B-NS3 protease, are essential for viral replication and infectivity. NS3 is a multi-functional enzyme, acting as a protease for polyprotein processing. A number of studies have already revealed that the non-structural NS3 serine protease is required for the maturation of the viral polyprotein and thus is a promising target for the development of antiviral inhibitors. NS5 is the largest (104 kDa) and the most conserved protein in DENV. It is also a bifunctional enzyme with a methyltransferase domain and a RNA-dependent RNA polymerase. The best characterized DENV nonstructural proteins are NS3 and NS5, which are multifunctional proteins presenting several enzymatic activities. (Agnihotri *et al., 2012*)

The automation of genomes to hit molecules pathway poses several challenges. It involves, *inter alia*, (i) accurate genome annotation, (ii) identification of druggable target proteins, (iii) determination of 3-dimensional structures of protein targets, (iv) identification of hits for the target, (v) optimization of hits to lead molecules to realize high levels of affinity and selectivity to the target and low toxicity. It is important to mention that inhibition of virus enzymes is one of the most successful approaches for drug antiviral discovery.

REVIEW OF LITERATURE

3.1 Dengue fever

In the tropical countries the dengue virus comes as a major threat to the health. A typical mosquito transmits the dengue virus, still with this limitation there are many people get infected by this virus per year. Patients with severe dengue illnesses can be treated successfully if they are diagnosed as early as possible. There are four antigenically different serotypes of the Dengue virus.

Dengue virus -1
Dengue virus -2
Dengue virus -3
Dengue virus -4

People become immune to a particular type of dengue virus once they've had it, but can still get sick from the other types of dengue if exposed. Catching different types of dengue, even years apart, increase the risk of developing severe dengue.

3.1.1 Dengue symptoms

Dengue fever has many symptoms, and people usually experience a combination of symptoms. Common symptoms of dengue fever are:

- sudden fever and extreme tiredness
- intense headache (especially behind the eyes)
- muscle and joint pain
- loss of appetite
- vomiting, diarrhoea, abdominal pain
- a metallic taste in the mouth
- red or macular (small, flat red spots) rash occurs in half of cases
- minor bleeding from nose and gums(Oliveira *et al., 2014*)

An approach involves diagnosing dengue infections by detecting NS1, one of the seven nonstructural dengue proteins. NS1 is produced in large quantities during dengue viral replication, and it can be detected as early as the first day the patient experiences a fever.

3.2 Viral Structure

The dengue genome is a single stranded RNA molecule of positive polarity. However, the replicative form of dengue RNA is not a single linear molecule but rather, a cyclic or dimerized genome. This special genomic RNA organization proficient for replication carries many highly ordered secondary and tertiary structures ensuring proper regulation of dengue RNA synthesis. Most of these RNA structures are located in the 5' and 3' untranslated regions.

The dengue virus particle is about 50 nm in diameter. The 10,723-nucleotide RNA genome encodes an uninterrupted open reading frame (ORF), directing the synthesis of a polyprotein precursor in the order NH2-C-prM-E-NS1- NS2A -NS2B-NS3-NS4A -NS4B-NS5-COOH, where C is the capsid protein, M is the membrane-associated protein, E is the envelope protein, and NS1 through NS5 are nonstructural proteins (figure-1). The long open reading frame encoding a large polyprotein is flanked in 5' and 3' by untranslated regions (UTRs). The latter carry a number of cis-acting signals (stem loops, conserved sequences, …) required for viral replication, and possibly RNA capping. There are complementary sequences in these UTRs that are thought to be responsible for cyclization of the genome, which is essential for replication. (Aleshin *et al., 2007*)

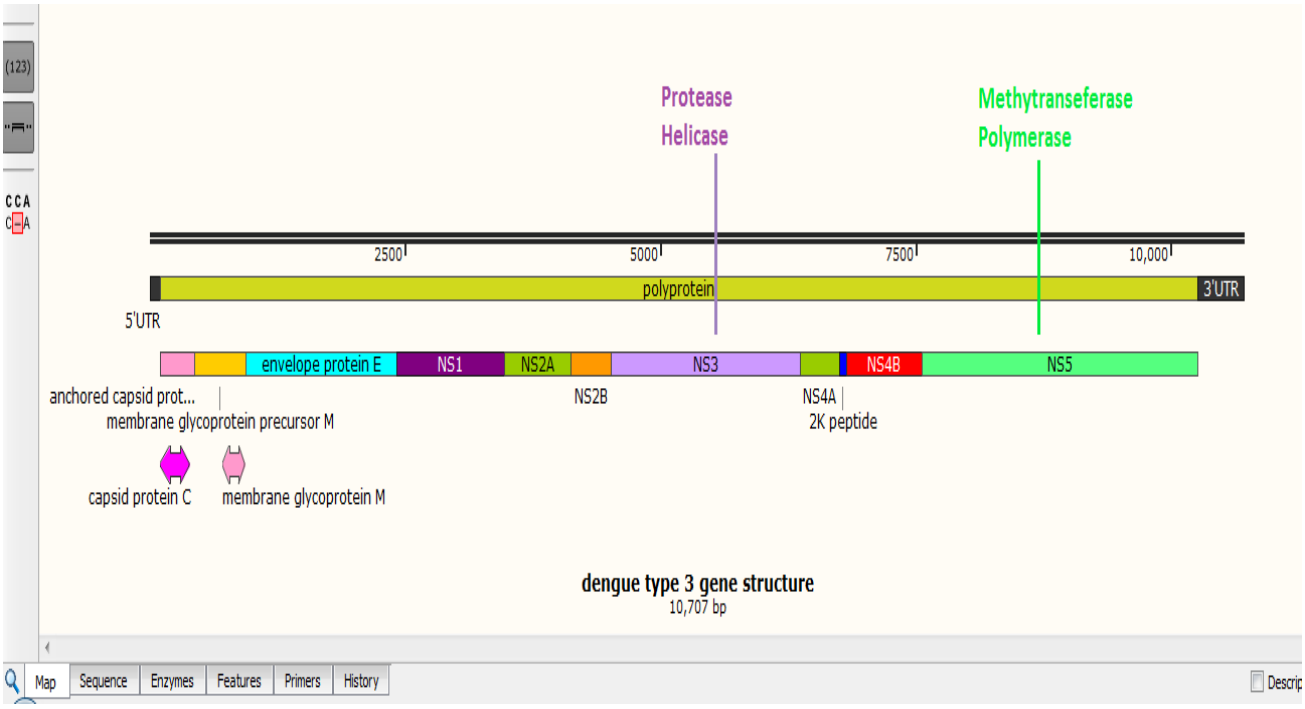


Figure-1: Dengue virus gene structure

There are four serotypes of Dengue virus. Here, a serotype is a group of viruses classified together based on their antigens on the surface of the virus. These four subtypes are different strains of dengue virus that have 60-80% homology between each other. The major difference for humans lies in subtle differences in the surface proteins of the different dengue subtypes. Infection induces long-life protection against the infecting serotype, but it gives only a short time cross protective immunity against the other types. The first infection cause mostly minor disease, but secondary infections has been reported to cause severe diseases (DHF or DSS) in both children and adults. This phenomenon is called Antibody-Dependent Enhancement. (Halstead *et al., 2003*)

3.2.1 Life-Cycle of DENV

The life cycle of dengue involves endocytosis via a cell surface receptor. The virus uncoats intracellularly via a specific process. In the infectious form of the virus, the envelope protein lays flat on the surface of the virus, forming a smooth coat with icosahedral symmetry. However, when the virus is carried into the cell and into lysozomes, the acidic environment causes the protein to snap into a different shape, assembling into trimeric spike. Several hydrophobic amino acids at the tip of this spike insert into the lysozomal membrane and cause the virus membrane to fuse with lysozome. This releases the RNA into the cell and infection starts. **(Alcaraz *et al., 2010*)**

The DENV RNA genome is in the infected cell translated by the host ribosomes. The resulting polyprotein is subsequently cleaved by cellular and viral proteases at specific recognition sites. The viral nonstructural proteins use a negative-sense intermediate to replicate the positive-sense RNA genome, which then associates with capsid protein and is packaged into individual virions. Replication of all positive -stranded RNA viruses occurs in close association with virus-induced intracellular membrane structures. DENV also induces such extensive rearrangements of intracellular membranes, called replication complex (RC). These RCs seem to contain viral proteins, viral RNA and host cell factors. The subsequently formed immature virions are assembled by budding of newly formed nucleocapsids into the lumen of the endoplasmic reticulum (ER), thereby acquiring a lipid bilayer envelope with the structural proteins prM and E. The virions mature during transport through the acidic trans -Golgi network, where the prM proteins stabilize the E proteins to prevent conformational changes. Before release of the virions from the host cell, the maturation process is completed when prM is cleaved into a soluble peptide and virion-associated M by the cellular protease furin. Outside the cell, the virus particles encounter a neutral pH, which promotes dissociation of the pr peptides from the virus particles and generates mature, infectious virions. At this point the cycle repeats itself. But researchers at the University of North Carolina have discovered a new target for human antibodies that could hold the key to a vaccine for the world's most widespread mosquito-borne disease.

3.3 The DENV targets for antiviral research

Making a truly effective dengue vaccine has proven difficult because of a phenomenon called antibody dependent enhancement. People infected with one type of dengue usually develop a natural immune response that rids the body of the virus and prevents a repeat infection of that same virus type. But if those people are infected with a second type of dengue, the virus is enhanced because of that first immune response. The result can be severe dengue hemorrhagic fever, which can be deadly. (Halstead *et al., 2003*)

The best characterized DENV nonstructural proteins are NS3 and NS5, which are multifunctional proteins presenting several enzymatic activities. These proteins are the most conserved ones in all four dengue virus serotypes. Therefore, NS3 and NS5 are considered attractive targets for the search and development of dengue antiviral chemotherapeutics. It is important to mention that inhibition of virus enzymes is one of the most successful approaches for drug antiviral discovery. (Canard *et al., 2009*)

A dengue vaccine has proven difficult to develop, in part because there are four major subtypes of dengue virus, each with slightly different viral proteins. Many researchers currently believe that the deadly dengue hemorrhagic disease is caused when a person is infected with one subtype, and then infected later by a second subtype. The antibodies, and immunity, gained from the first infection appear to assist with the infection by the second subtype, instead of providing a general immunity to all subtypes. This means that an effective vaccine will have to stimulate protective antibodies against all four types at once, a feat that has not yet been achieved.

Dengue virus also makes several proteins that create new viruses once it is inside a cell. Two of the major ones are multifunctional proteins with several enzymes strung together.One is NS5 contains a methyltransferase and a polymerase, and the other one is NS3 contains a protease and a helicase. Each of these enzymes performs a different part of the life cycle. The polymerase builds new RNA strands based on the viral RNA, the helicase helps to separate these strands, and the methyltransferase adds methyl groups to the end of them, protecting the RNA strands and coaxing the cell's ribosomes to create viral proteins based on them. The viral proteins are created in one long polyprotein chain, which is finally clipped into the functional units by the protease. (Agnihotri *et al., 2012*)

Proteolysis yields ten proteins, the three structural proteins (C, prM, and E) and the seven nonstructural (NS) proteins involved in genome replication and capping (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5). Some of these NS proteins also participate in pathogenesis and counteract the innate immunity of the host cell. Replication of the viral genome does not occur freely in the cytoplasm. Instead, there is an extensive intracellular membrane re-arrangement in

the infected cell, with various observable cell substructures containing most NS proteins organized along the virus replication cycle. (Lescar *et al., 2008*)

3.3.1 The Non-Structural proteins

Out of the seven NS1-to-NS5, here we consider NS1, NS3 and NS5 so far as drug targets, not only because they are essential to virus growth but also because they exhibit enzyme activity, which is a plus regarding drug screening. The role and structure of NS1 is unknown. It is a soluble protein detected very early during infection. NS2a, NS2b, NS4a, and NS4B are membrane –associated proteins believed to anchor and regulate the replication complex during the virus life-cycle.

Presently, there are no drugs against dengue in the clinic. Therefore, what we call a validated target is only derived by analogy to other viruses for which drugs have proven to be effective. Herpes and HIV have been the drug-design founding viruses. Now for DENV, HCV is fulfilling this role. Recent data presented regarding a nucleoside analogue, although toxic, have shown that the NS5 protein is a validated target. The validated targets are thus the RNA-dependent RNA polymerase and, by analogy to HCV, the protease. More recently, the HCV NS5A protein seems to emerge as a very interesting target, but NS5A does not have an equivalent in dengue, so far. Conversely, the dengue RNA genome is capped, and the DENV genome encodes most of its own RNA capping machinery. This is not the case for HCV, which does not rely on RNA capping for gene expression. The RNA capping enzymes of dengue so far await validation in an animal model both at the level of efficiency, and toxicity. Indeed, it is not known if the abundance ofv cellular MTases will cause a specificity problem, i.e., if it will be possible to design an anti-dengue MTase inhibitor having non-significant toxicity effect through co-lateral inhibition of host cell MTases. (Goodsell *et al., 2008*)

• **NS1** The Flavivirus nonstructural protein 1 (NS1) is a conserved, membrane-associated and secreted glycoprotein with replication and immune evasion functions. Secreted NS1 is a hexameric, barrel -shaped lipoprotein that can bind back to the plasma membrane of cells. Antibodies targeting cell surface-associated NS1 can be protective in vivo in a manner dependent on Fc effector functions.

• **NS3** carries two functional domains, a N-terminal serine protease (~170 aa), (69 kDa) and a Cterminus helicase/RNA triphosphatase (~440 aa). The protease domain is inactive alone, and needs the presence of 40 aa of NS2b bound to form a protease active site. The NS2b/NS3 protease has been the first dengue protein target actively used in drug design programs. Two complementary approaches have been followed to discover antivirals based on the inhibition of this enzyme. The first approach has been the screening of a large chemical library and the second approach has been to design peptidomimetics The NS3 helicase domain is also an

interesting target because it contains features unique to flaviviruses, such as Domain III. (Agnihotri *et al., 2012*)

The viral protease is responsible for cleavage at a number of sites, including NS2A-NS2B, NS2B-NS3, NS3-NS4A, and NS4B-NS5; it also cleaves just upstream of the signal sequences at the C-prM and NS4A- NS4B junctions, within NS2A, and within NS3 itself. Later studies showed that expression of NS3 alone did not lead to production of an active protease; however, including a portion of NS2B with NS3 led to full proteolytic activity.(Lescar *et al., 2008*)

Viral proteases are proven antiviral targets, as evidenced by the clinical availability of ten human immunodeficiency virus 1 (HIV-1) protease inhibitors and the hepatitis C virus (HCV) protease inhibitors. Thus, it is plausible that a protease inhibitor for dengue virus would be efficacious in the clinic.(Noble *et al., 2010*)

The important role of NS3 for the replication of the dengue virus makes this protein an interesting site for viral inhibiting.

• **NS5** is the largest and most conserved and most conserved dengue protein. The NS5 is about 900 amino acids long and comprises a methyltransferase domain at its N terminus and an RNA-dependent RNA polymerase domain at its C-terminal end. Both enzymatic activities from attractive targets for antiviral development.

The domain RNA-dependent RNA polymerase (RdRp) is responsible for the replication of the positive-strand RNA genome in an asymmetric and semi -conservative process in which the antigenome is only present in a double-stranded RNA replication intermediate. In DENV infections, the NS5 protein is primarily localized within the nucleus. However, not all flavivirus RdRps localize to the nucleus.(Lim *et al., 2015*)

Viral MTases are involved in the mRNA capping process, transferring a methyl group from the cofactor *S*-adenosyl-L-methionine (AdoMet) to the N7 atom of the cap guanine and onto the 2'OH group of the ribose moiety of the first RNA nucleotide. In the genus *Flavivirus*, both (guanine -N7)-methyltransferase (N7MTase) and (nucleoside -2'-O-)- methyltransferase (2'OMTase) activities have been associated with the N-terminal domain of the viral NS5 protein.

The enzymes codified by NS5 show they have important functions in the replication of the virus, playing a key role and suggesting NS5 as a potential antiviral target**.** (Lim *et al., 2015*)

3.4 Genome Annotation

In recent times genomes and proteome related databases are growing very fast at an accelerated pace, so these are the times for bioinformatician, drug designer and computational biologist. That's why in this post genomic era there is heigh expectation for successful and fast treatment of diseases.

The computational genome annotation can play a vital role in finding potential therapeutic target molecules for pathogens. In the present research scenario, it is a big challenge to carry out the structural and functional annotation of the whole genome sequence or the translated ORFs (open reading frames). These annotations can be used in comparative genomics, pathway reconstruction and particularly in drug design. Genome annotation is the process of exploring biological/ functional information from sequences. It is done by following two main steps: (i) identification of distinct, potentially functional elements on the genome, a process called gene prediction in the context of identification of protein coding regions and (ii) assignment of biological function to these elements (genes or proteins). Automated annotation tools provide a faster computational annotation as compared to manual annotation (curation) which involves human expertise. Ideally, these approaches coexist and complement each other in the same annotation pipeline. (Jayaram *et al., 2012*)**.** The basic level of annotation involves finding genes and isolating the protein coding sequences from non- coding sequences. A variety of computational approaches have been developed to permit scientists to view and share genome annotations (Table-1).

Table-1: List of Tools Available for Gene Prediction

	TOOL	LINK	METHOD
1.	FGENESH	http://linux1.softberry.com/all.htm	*Ab initio*
2.	GeneID	http://www1.imim.es/geneid.html	*Ab initio*
3.	GeneMark	http://exon.gatech.edu/GeneMark/gmchoice.html	*Ab initio*
4.	GeneMark.hmm	http://exon.gatech.edu/hmmchoice.html	*Ab initio*
5.	GeneWise	http://www.ebi.ac.uk/Tools/Wise2/	Homology
6.	GENSCAN	http://genes.mit.edu/GENSCAN.html	*Ab initio*
7.	Glimmer	http://www.tigr.org/software/glimmer/	*Ab initio*
8.	GlimmerHMM	http://www.cbcb.umd.edu/software/glimmerhmm/	*Ab initio*
9.	GRAILEXP	http://compbio.ornl.gov/grailexp	*Ab initio*
10	GENVIEW	http://zeus2.itb.cnr.it/~webgene/wwwgene.html	*Ab initio*
11	GenSeqer	http://bioinformatics.iastate.edu/cgi-bin/gs.cgi	Homology

12.	PRODIGAL	http://prodigal.ornl.gov/	Homology
13.	MORGAN	http://www.cbcb.umd.edu/~salzberg/morgan.html	*Ab initio*
14.	PredictGenes	http://mendel.ethz.ch:8080/Server/subsection3_1_8.html	Homology
15.	MZEF	http://rulai.cshl.edu/software/index1.htm	*Ab initio*
16.	Rosetta	http://crossspecies.lcs.mit.edu	Homology
17.	EuGéne	http://eugene.toulouse.inra.fr/	*Ab initio*
18.	PROCRUSTES	http://www.riethoven.org/BioInformer/newsletter/archives/2/procrustes.html	Homology
19.	Xpound	http://mobyle.pasteur.fr/cgi-bin/portal.py?#forms::xpound	*Ab initio*
20.	Chemgenome	http://www.scfbio-iitd.res.in/chemgenome/chemgenome3.jsp	*Ab initio*
21.	Augustus	http://augustus.gobics.de/	*Ab initio*
22.	Genome Threader	http://www.genomethreader.org/	Homology
23.	HMMgene	http://www.cbs.dtu.dk/services/HMMgene/	*Ab initio*
24.	GeneFinder	http://people.virginia.edu/~wc9c/genefinder/	*Ab initio*
25.	EGPRED	http://www.imtech.res.in/raghava/egpred/	*Ab initio*
26.	mGene	http://mgene.org/web	*Ab initio*

Most of the available computational methods are knowledge -based and adopt techniques like Hidden Markov Models or machine learning methods. The accuracies of these models are limited by the availability of data on experimentally validated genes, and as typically seen in newly sequenced genomes, can lead to suboptimal levels of prediction. Ab initio methods originating in physico- chemical properties of DNA can help overcome the limitations of knowledge-based methods. Generally for annotation purposes, homologous sequences in protein sequence databases are searched. The state of the art tool for such database searches is PSI-BLAST (Position Specific Iterated Basic Local Alignment Search Tool) . The performance of PSI-BLAST and other database search tools to identify homologs of a given query in a sequence database has been measured by others . However these benchmarks do not suffice the requirements in genome annotation. Our efforts are aimed at eliminating the limitations of PSI BLAST in correctly annotating protein coding sequences in genomes by using ab initio approach. (Soni *et al., 2013*)

3.5 Protein Tertiary Structure Prediction

Despite the large genome sizes, only a fraction of the genome codes for proteins, particularly in higher organisms. For example, Human genome is ~3300Mb with ~20,500 protein coding genes and ~3000 non -protein coding RNA genes which is about 2% of the genome. The increasing gap between sequences and structures makes it simply impossible to solve the structures of all existing proteins experimentally. The knowledge of the 3D structure of a protein can usher in tools for structure based drug discovery. Thus, a reliable computational method for protein tertiary structure prediction is desperately needed. Various computational methodologies have been developed for the prediction o f tertiary structures of proteins over the past few years. These include (a) comparative modeling, (b) fold recognition or threading, (c) *ab initio* or *de novo* methods. Comparative modeling and fold recognition methods are database driven and their prediction accuracies depend on the sequence similarities realized in known structures. These methods are extremely popular, reliable and fast for protein tertiary structure prediction when a close sequence homolog exists in the database. *Ab initio* or *de novo* methods are used for predicting structure of protein sequences with no close structural homologs. (Jayaram *et al., 2012*) Homology modeling and fold recognition methods utilize the information derived from structures solved previously via x -ray and NMR methods. This method is effective, popular, reliable and fast for protein tertiary structure prediction when a close sequence homolog exists in the structural repositories. Several protein structure prediction tools are available in the public domain (Table-2).

Table-2: List of Tools Available for Protein Tertiary Structure Prediction

	TOOL	LINK	METHOD
1.	CPHModels 3.0	http://www.cbs.dtu.dk/services/CPHmodels/	Protein homology modeling server
2.	SWISS-MODEL	http://swissmodel.expasy.org/SWISS-MODEL.html	A fully automated protein structure homology-modeling server
3.	Modeller	http://salilab.org/modeller/	Program for protein structure modeling by satisfaction of spatial Restraints
4.	3D-JIGSAW	http://3djigsaw.com/	Server to build three-dimensional models for proteins based on homologues of known structure

5.	PSIPRED	http://bioinf.cs.ucl.ac.uk/psipred/	A combination of methods such as sequence alignment with structure based scoring functions and neural network based jury system to calculate final score for the alignment
6.	3D-PSSM	http://www.sbg.bio.ic.ac.uk/~3dpssm/index2.html	Threading approach using 1D and 3D profiles coupled with secondary structure and solvation potential
7.	ROBETTA	http://robetta.bakerlab.org	*De novo* Automated structure prediction analysis tool used to infer protein structural information from protein sequence data
8.	PROTINFO	http://protinfo.compbio.washington.edu/	*De novo* protein structure prediction web server utilizing simulated annealing for generation and different scoring functions for selection of final five conformers
9.	SCRATCH	http://scratch.proteomics.ics.uci.edu/	Protein structure and structural features prediction server which utilizes recursive neural networks, evolutionary information, fragment libraries and energy
10.	I-TASSER	http://zhanglab.ccmb.med.umich.edu/I-TASSER/	Predicts protein 3D structures based on threading approach
11.	BHAGEERATH	http://www.scfbio-iitd.res.in/bhageerath/index.jsp	Energy based methodology for narrowing down the search space of small globular proteins
12.	BHAGEERATH-H	http://www.scfbioiitd.res.in/bhageerath/bhageerath_h.jsp	A Homology *ab-initio* Hybrid Web server for Protein Tertiary Structure Prediction

To make biological sense out of large volumes of sequence data, it is necessary to compare the protein sequences with those proteins that have been already characterized biochemically. To design drug molecules, structural annotation plays an important role. Structural genomics (SG) efforts facilitate such comparisons by determining the structures for a large number of protein sequences, but most SG targets have not been functionally characterized. It is already known that accurate functional details of a protein can neither be inferred from its sequence alone nor from sequence comparisons with other proteins whose structures and functions are known but only from its own native structure. Several efforts are being made to unravel the physico -chemical basis of protein structures and to establish some fundamental rules of protein folding. Despite the successes, protein tertiary structure prediction still remains a grand challenge - an unsolved

problem in computational biochemistry. *Ab initio* or *de novo* methods are frequently employed for predicting tertiary structures of proteins by incorporating the basic physical principles, irrespective of the availability of structural homologs. The knowledge of tertiary structures of proteins serves as a basis for structure-based drug design. (Noble *et al., 2010*)

3.6 Drug Design

The automation of genomes to hit molecules pathway poses several challenges. It involves, *inter alia*, (i) accurate genome annotation , (ii) identification of druggable target proteins , (iii) determination of 3-dimensional structures of protein targets, (iv) identification of hits for the target, (v) optimization of hits to lead molecules to realize high levels of affinity and selectivity to the target and low toxicity.

The ability of biomolecules to bind to their substrates in a highly specific manner is an important characteristic in many biological processes. One of the major challenges for the CADD (Computer Aided Drug Design) techniques is to achieve this specificity i.e . specific binding of a small molecule to the biomolecular target *in vivo* at minimum cost and time, while maintaining novelty of the scaffolds and proper ADMET profiles. The two main steps in computer based drug design protocols are target identification and lead optimization . The knowledge of the structure of a biomolecule, such as a protein / nucleic acid assists in understanding the molecular level mechanism of action of drugs and lead optimization. (Kalra *et al., 2001*)

Structure of the protein target molecule can be determined either experimentally or computationally (as discussed above). The next step after target identification and structure determination is the detection of the ligand binding site. Most of the experimentally determined structures have some information on ligand binding site . In the absence of such information, detection of ligand binding site is required. Design of small molecules in structure based drug discovery requires knowledge of the binding pocket on the protein which upon blockade results in loss of function. Experimental information on protein active sites and function loss are useful In the absence of any experimental information, one could identify all potential binding sites on the protein from the structural information (Table 3).

Table-3: List of Software Available for Active Site Prediction

	TOOL	LINK	METHOD
1.	SitesIdentify	http://www.manchester.ac.uk/bioinformatics/sitesidentify/	Sequence and geometry based
2.	PAR-3D	http://sunserver.cdfd.org.in:8080/protease/PAR_3D/index.html	Structure based
3.	FUZZY-OIL-DROP	http://www.bioinformatics.cm-	Fuzzy oil drop

		uj.krakow.pl/activesite/	model
4.	CASTp	http://cast.engr.uic.edu	Structure based
5.	Pocket-Finder	http://www.modelling.leeds.ac.uk/pocketfinder/	Energy based
6.	Q-site finder	http://www.modelling.leeds.ac.uk/qsitefinder/	Energy based
7.	PASS	http://www.ccl.net/cca/software/UNIX/pass/overview.shtml	Structure based
8.	SURFNET	http://www.biochem.ucl.ac.uk/~roman/surfnet/surfnet.html	Structure based
9.	LIGSITE	http://projects.biotec.tu-dresden.de/pocket/	Based on Connolly surface
10.	VOIDOO	http://xray.bmc.uu.se/usf/voidoo.html	Structure based
11.	LiGandFit	http://www.phenix-online.org/documentation/ligandfit.htm	Structure based
12.	Active site prediction	http://www.scfbio-iitd.res.in/dock/ActiveSite.jsp	Structure based
13.	AADS	http://www.scfbio-iitd.res.in/dock/ActiveSite_new.jsp	Structure based
14.	Fpocket	http://fpocket.sourceforge.net/	Based on Voronoi tessellation
15.	Pocket Picker	http://gecco.org.chemie.uni-frankfurt.de/pocketpicker/index.html	
16.	IsoCleft	http://bcb.med.usherbrooke.ca/isocleftfinder.php	graph-matching-based method
17.	metaPocket	http://sysbio.zju.edu.cn/metapocket/	Structure based
18.	LIGSITE	http://gopubmed2.biotec.tu-dresden.de/cgi-bin/index.php	Structure based
19.	GHECOM	http://strcomp.protein.osaka-u.ac.jp/ghecom/	Structure based
20.	ConCavity	http://compbio.cs.princeton.edu/concavity/	Structure based
21.	POCASA	http://altair.sci.hokudai.ac.jp/g6/Research/POCASA_e.html	Structure based

The next step in CADD protocol is lead generation and optimization. A candidate molecule can either be sketched using publicly available drawing tools or searched from the small molecule libraries such as NRDBSM which embeds Lipinski's rules or the one developed by us more recently consisting of one million small molecules (http://www.scfbioiitd. res.in/software/nrdbsm/moleculesearchnew .jsp). For a given biomolecular target, these databases of small molecules can be scanned for identifying hit molecules based on their physico -chemical parameters and functional groups. Assessment of the candidate molecules is performed by calculating the binding energies with the target, one of the major bottlenecks being

the computational time. The calculation of binding energy of a small ligand with a protein target using docking and scoring methods can take minutes, which approximates to around 12000 to 15000 hours for a dataset of one million molecules. There are many softwares and tool which can be used for the docking and drug design (Table 4). (Mukherjee *et al., 2013*)

Table-4: A list of Softwares for Drug Design

	TOOL	LINK	METHOD
1.	Discovery studio	http://accelrys.com/products/discovery-studio/structurebased-design.html	Molecular modeling and *de novo* drug design
2.	Sybyl	http://www.tripos.com/	Computational software for drug discovery
3.	Bio-Suite	http://www.staff.ncl.ac.uk/p.dean/Biosuite/body_biosuite.html	Tool for Drug Design, structural analysis and simulations
4.	Molecular Operating Environment (MOE)	http://www.chemcomp.com/	Structure-based drug design, molecular modeling and simulations
5.	Glide	https://www.schrodinger.com/products/14/5	Ligand-receptor docking
6.	Autodock	Ligand-receptor docking	Protein-ligand docking
7.	DOCK	http://dock.compbio.ucsf.edu/	Protein-ligand docking
8.	*Sanjeevini*	http://www.scfbio-iitd.res.in/sanjeevini/sanjeevini.jsp	A complete software suite for structure-based drug design
9.	ArgusLab	http://www.arguslab.com/arguslab.com/ArgusLab.html	Ligand-receptor docking
10.	eHITS	http://www.simbiosys.ca/ehits/index.html	Ligand-receptor docking
11.	FlexX	http://www.biosolveit.de/FlexX/	Ligand-receptor docking
12.	FLIPDock	http://flipdock.scripps.edu/	Ligand-receptor docking

13.	FRED	http://www.eyesopen.com/oedocking	Ligand-receptor docking
14.	GOLD	http://www.ccdc.cam.ac.uk/products/life_sciences/gold/	Protein-ligand docking
15.	ICM-Docking	http://www.molsoft.com/docking.html	Protein-ligand docking
16.	PLANTS	http://www.tcd.uni-konstanz.de/research/plants.php	Protein-ligand docking
17.	Surflex	http://www.biopharmics.com/	Protein-ligand docking

METHODOLOGY

The "Genome to hit" assembly line *in silico* for Dengue is given in a flowchart format below.

1. Whole genome sequences of all 4 serotype of Dengue virus is retrived from
http://www.ncbi.nlm.nih.gov/nuccore/NC_001477.1 (Type1)
http://www.ncbi.nlm.nih.gov/nuccore/NC_001474.2 (type2)
http://www.ncbi.nlm.nih.gov/nuccore/NC_002640.1(Type3)
http://www.ncbi.nlm.nih.gov/nuccore/NC_001475.2 (type4)

2. Genes are predicted which are then translated into protein sequences using Chemgenome 3.0
http://www.scfbio-iitd.res.in/chemgenome/chemgenome3.jsp

3. Perform blastn of these gene's sequences with Human genomic + transcript database.
http://blast.ncbi.nlm.nih.gov/Blast.cgi

4. These polyprotein sequences are splitted w.r.t. literature and results are processed for tertiary structure prediction by Bhageerath-H
http://www.scfbio-iitd.res.in/bhageerath/bhageerath_h.jsp

5. Modeled structure are examined for identification of potential active sites by active site finder AADS
http://www.scfbio-iitd.res.in/dock/ActiveSite_new.jsp

6. A million compound library of small molecules is screened against binding sites using RASPD
http://www.scfbio-iitd.res.in/software/drugdesign/raspd.jsp

7. The screened molecules are docked, scored and optimized iteratively using ParDock
http://www.scfbio-iitd.res.in/dock/pardock.jsp

8. Hits ready to be synthesized and tested in laboratory and optimized iteratively.

Above mentioned steps are described below.

1. As all we know that there are 4 serotypes of Dengue virus. To download the whole sequences of these serotype of dengue we go to the NCBI site http://www.ncbi.nlm.nih.gov and download the all four serotypes with help of NCBI ID NC_001477.1, NC_001474.2, NC_002640.1, NC_001475.2 for type-1, type-2, type-3, type- 4 respectively.

Table- 5: Dengue virus serotype

s.no.	Dengue virus	Nucleotide	GC %	NCBI ID	Gene Start	Gene End
1	Type-1	10,735	46.70	NC_001477.1	95	10273
2	Type-2	10,723	45.80	NC_001474.2	97	10272
3	Type-3	10,707	46.70	NC_002640.1	95	10267
4	Type-4	10,649	47.70	NC_001475.2	102	10265

In table-5 basic information about all serotypes of Dengue virus is written like nucleotide length, GC content, their NCBI ID and the starting point and the ending point of the gene.

gene ID	start	end	Product	Protein ID	nt sequen	PDB ID	protein se	aa start	aa end	AA count	comments
type 3	1	94	5' UTR		AGTTGTTA						
5075727	95	10267	polyprote	YP_00162	ATGAACA		MNNQRKI	1	3390	3390	
	95	436	anchored	YP_00153	ATGAACA	NA	MNNQRKI	1	114	114	5..114 = Flavivirus capsid prote
	95	394	capsid p	YP_00153	ATGAACA	NA	MNNQRKI	1	100	100	
	437	934	membrane	YP_00153	TTCCACTT	NA	FHLTSRDG	115	280	166	119..204= Flavivirus polyprotei
	710	934	membrane	YP_00153	ATCAGTG	3J6S	SVALAPH	206	280	75	206..280 = Flavivirus envelope
	935	2413	envelope	YP_00153	ATGAGAT	3J6S, 1UZ	MRCVGVG	281	773	493	281..574 = 583..672 = 677..773 =
	2414	3469	nonstruc	YP_00153	GACATGG	NA	DMGCVIN	774	1125	352	1,491..1,6 1,663..1,792 = DEAD
	3470	4123	nonstruc	YP_00153	GGAAGTG	NA	GSGKVDN	1126	1343	218	
	4124	4513	nonstruc	YP_00153	AGCTGGC	NA	SWPLNEG	1344	1473	130	
	4514	6370	nonstruc	YP_00153	TCCGGCG	NA	SGVLWDV	1474	2092	619	
	6371	6751	nonstruc	YP_00153	TCAATCG	NA	SIALDLVT	2093	2219	127	
	6752	6820	2K pepti	YP_00153	ACCCCCC	NA	TPQDNQL	2220	2242	23	presumably plays a role in men
	6821	7564	nonstruc	YP_00153	AATGAAA	NA	NEMGLLE	2243	2490	248	
	7565	10267	nonstruc	YP_00153	GGAACAG	4V0Q	GTGSQGE	2491	3390	900	2,545..2,7 2,740..3,386= Flavivi
	10268	10707	UTR		ACGTAGG						
type 1											
	2420	3475	ns1	NP_722461.1			DSGCVINWKGRELKCGSGIFVTNEVHTWTEQYKFQADSPKRLSAAIGKAWEEGVC				
	4520	6376	ns3	NP_722463.1			SGVLWDTPSPPEVERAVLDDGIYRILQRGLLGRSQVGVGVFQEGVFHTMWHVTRGAVLMYQGKR				
	7574	10270	ns5	NP_722465.1			GTGAQGETLGEKWKRQLNQLSKSEFNTYKRSGIIEVDRSEAKEGLKRGETTKHAVSRGTAKLRW				
type2											
	2422	3477	ns1	NP_739584.2			DSGCVVSWKNKELKCGSGIFITDNVHTWTEQYKFQPESPSKLASAIQKAHEEGICGIRSVTRLE				
	4522	6375	ns3	NP_739587.2			AGVLWDVPSPPPMGKAELEDGAYRIKQKGILGYSQIGAGVYKEGTFHTMWHVTRGAVLMHKGKR				
	7570	10269	ns5	NP_739590.2			GTGNIGETLGEKWKSRLNALGKSEFQIYKKSGIQEVDRTLAKEGIKRGETDHHAVSRGSAKLRW				
type4											
	2424	3479	ns1	NP_740318.1			DMGCVASWSGKELKCGSGIFVVDNVHTWTEQYKFQPESPARLASAILNAHKDGVCGIRSTTRLE				
	4524	6377	ns3	NP_740321.1			SGALWDVPSPAATKKAALSEGVYRIMQRGLFGKTQVGVGIHMEGVFHTMWHVTRGSVICHETGR				
	7563	10262	ns5	NP_740325.1			GTGTTGETLGEKWKRQLNSLDRKEFEEYKRSGILEVDRTEAKSALKDGSKIKHAVSRGSSKIRW				

Figure-2 Excel sheet of gene information.

In figure-2 all the essential information about the Dengue Virus gene is available like the start and end point of nucleotide of genes and their protein IDs, their nucleotide sequences and their protein sequences.Making of this sheet will be very useful to go ahead in this project.

2. After that we perform blastn of all the gene's sequences with the Human genomic + transcript database to check that whether the sequences of viral genes is similar to human genomic data + transcript data or not.

3. After downloading the whole genome sequences of all the 4 serotypes next step is the finding the genes, protein coding region. So we go to the home page of Chemgenome. Paste the genome in FASTA format. We run chemgenome with default parameters. If you want to specify, you can specify additional parameters Specify email address optionally to get the result mailed in inbox. The output generated will be both in tabular and graphical representation.

An alternative to the knowledge based methods for gene prediction is an *ab initio* model ChemGenome 2.0 (figure-3) is an *ab initio* method for gene prediction in prokaryotes. It examines physicochemical properties and geometrical structures of codons to structural and thermodynamic properties of DNA sequences, more specifically , the hydrogen bond, the stacking and the solvation energies computed from molecular dynamics generated structures (Soni *et al., 2013*)

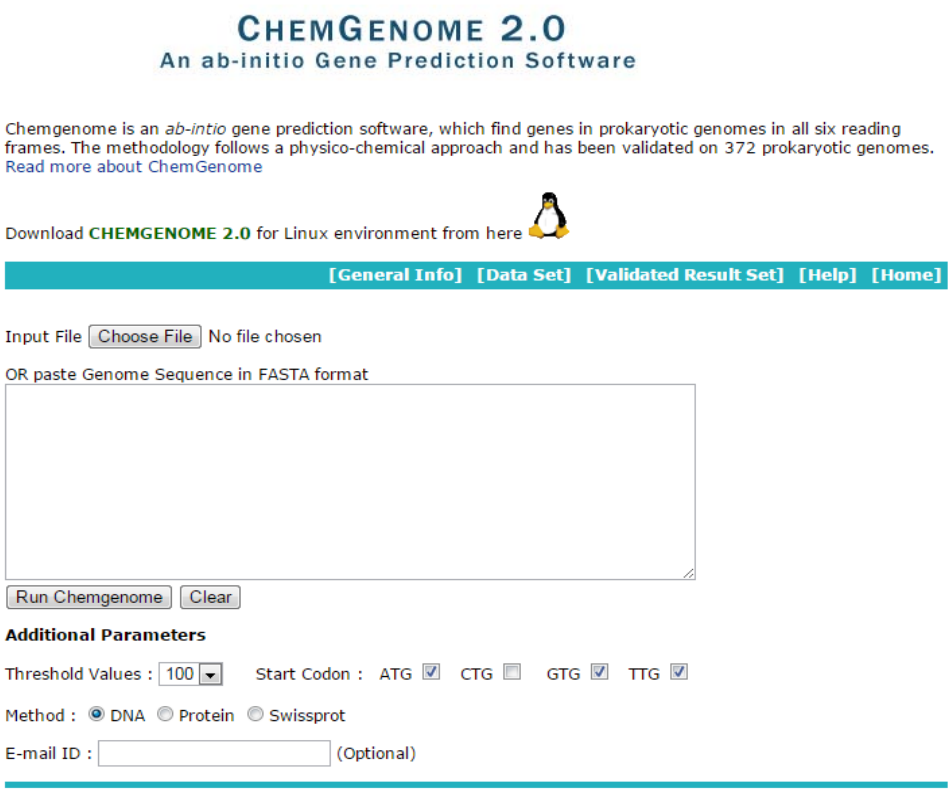


Figure-3: Homepage of Chemgenome

4. Third step is predicting 3- D structure of viral enzymes. Bhageerath is an energy based software suite for predicting tertiary structures of small globular proteins. The protocol comprises eight different modules which use physicochemical properties of proteins and *ab initio* methodology to predict five candidates for the native from the input query sequence. Bhageerath -H (figure-4) is a homology *ab initio* hybrid server for protein tertiary structure prediction. The protocol identifies regions having local sequence similarity with database to generate 3D fragments which are patched with *ab initio* modeled fragments to put together complete structure of proteins. For sequences with available sequence homologs, Bhageerath - H software predicts a structure within 5 Å RMSD from the native. It identifies regions which show local sequencesimilarity in respect to sequences in RCSB (protein data bank) to generate 3D fragments which are patched with *ab initio* modeled fragments to generate complete structures of the proteins. The knowledge of tertiary structures of proteins serves as a basis for structure-based drug design. (Bhushan *et al., 2006*)

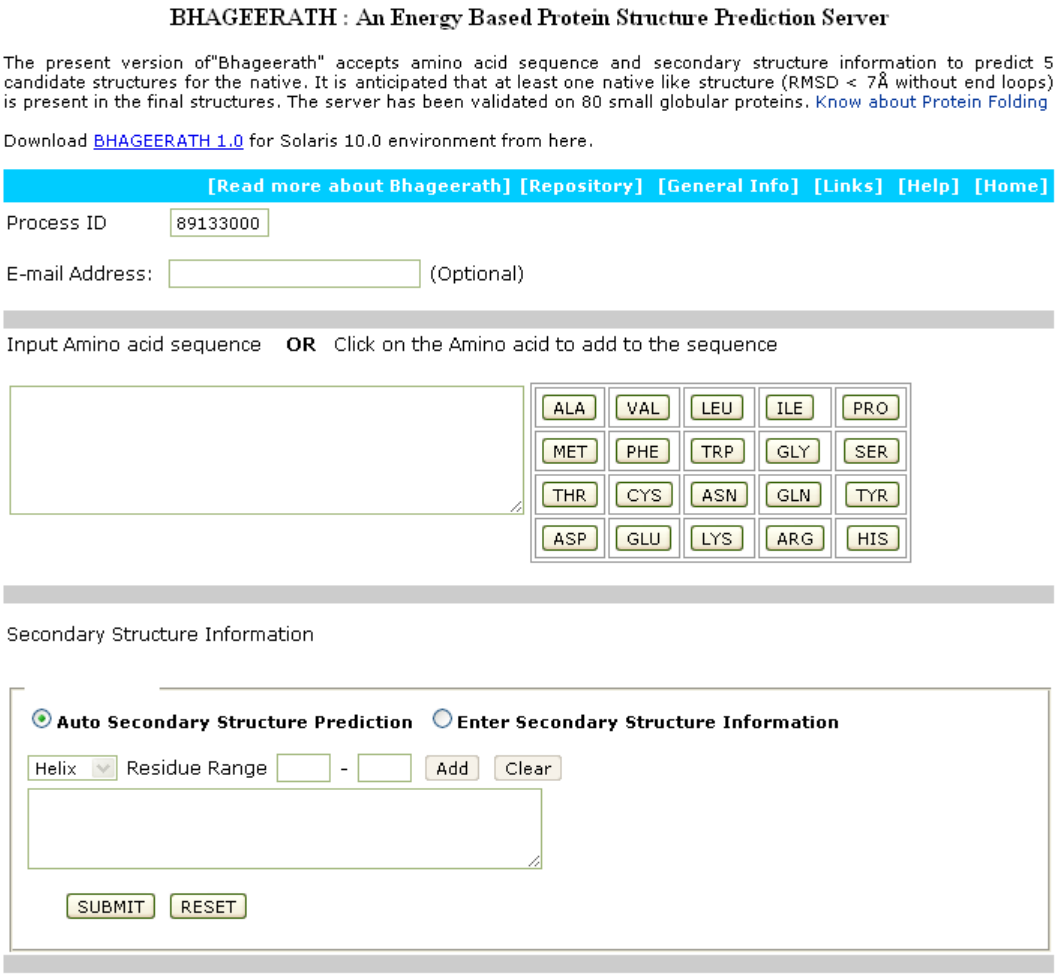


Figure-4: Homepage of Bhageerath

5. After predicted the 3-D structure of proteins, next steps is to finding the active sites present in the viral enzymes. The active site finder identifies all cavities in a protein and scores them based on the physicochemical properties of functional groups lining the cavities in the protein. Methodology for an automated identification of ten potential binding pockets which are expected to bracket the true "active site" (binding pocket) from AADS (figure-5). Top ten cavity points identified are then submitted for an automated docking of an input ligand/candidate molecule. The docking protocol uses an all atom energy based Monte Carlo method. (Singh *et al., 2011*)

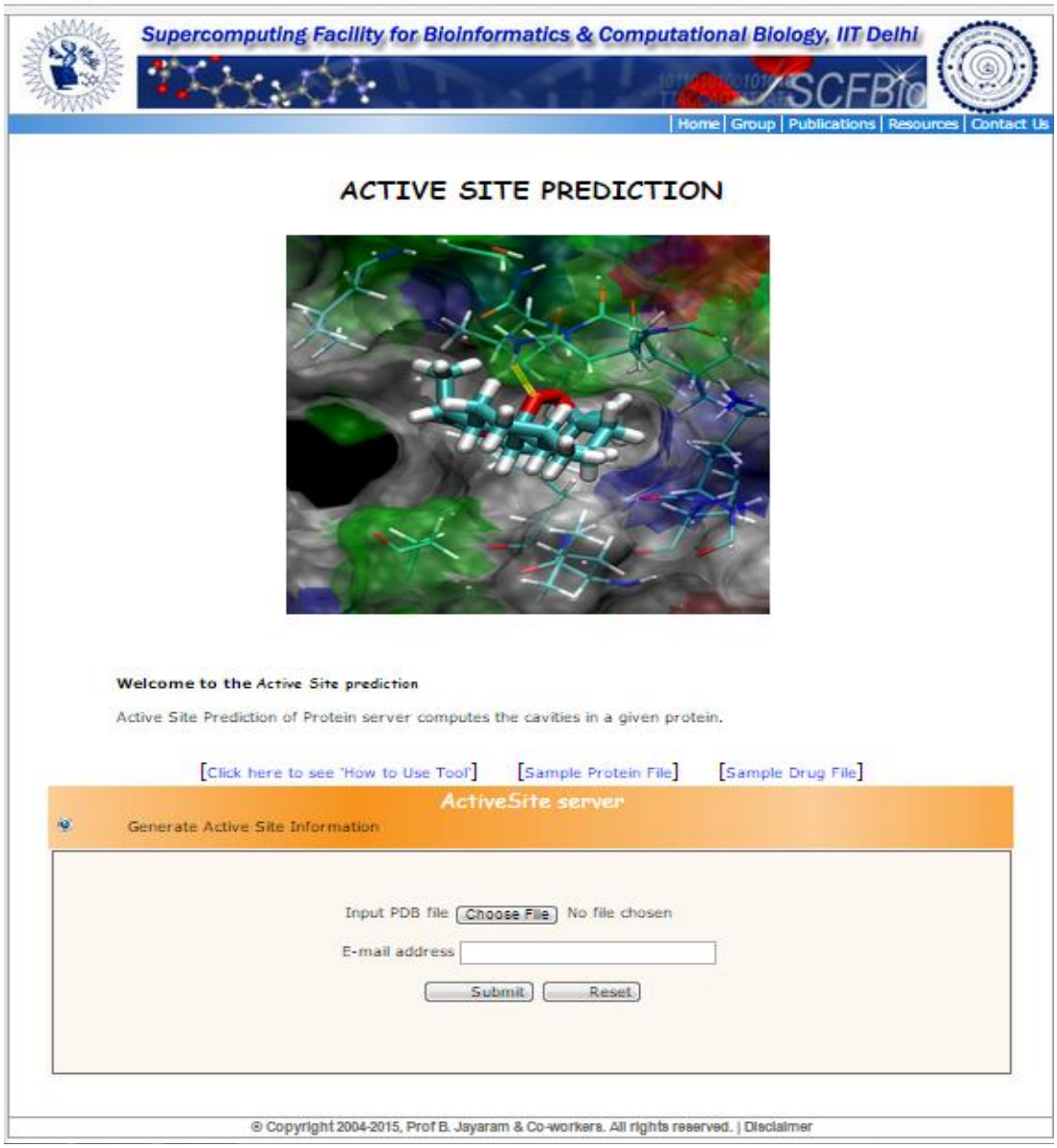


Figure-5: Homepage of Active site prediction

6. Now next step is RASPD (A Rapid Scoring Methodology Based on Physico-Chemical Descriptors of Million Molecules)by using various molecular descriptors of candidate drug like molecules/active site of proteins for computation of binding free energy of target protein with significantly reduced time and cost (figure-6). This is done by predicting the highly possible candidates for binding from a large data set (millions) and using a reduced set of likely candidate molecules to achieve the same result as done in traditional techniques. For checking the protein active site information and determining the binding affinity of you drug molecule to the target protein, click on RASPD button and then on Submit button in the upcoming pop up window. (Mukherjee *et al., 2013*)

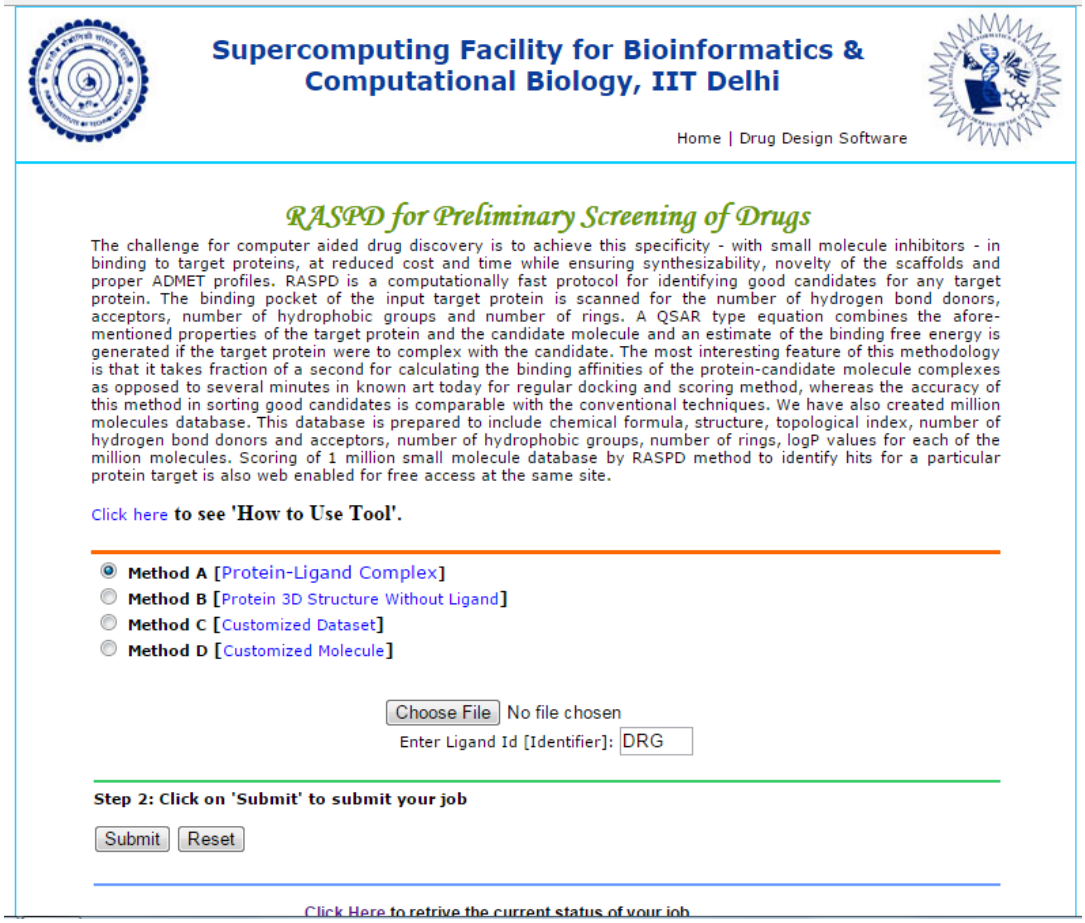


Figure-6: Homepage of RASPD

7. Docking can be performed in two ways.
   - Docking and Scoring with Case 1**:** After We have selected the drug and target protein-ligand complex by screening them through Lipinski's filter and RASPD respectively, go to the ParDOCk home page (figure-7). On doing so, in a pop-up window be appear and asked to give the total formal charge of the chosen drug molecule along with e -mail id and then submit the job. Four docked structures along with their predicted binding affinity values will be sent back via e-mail. (Latha *et al., 2006*)

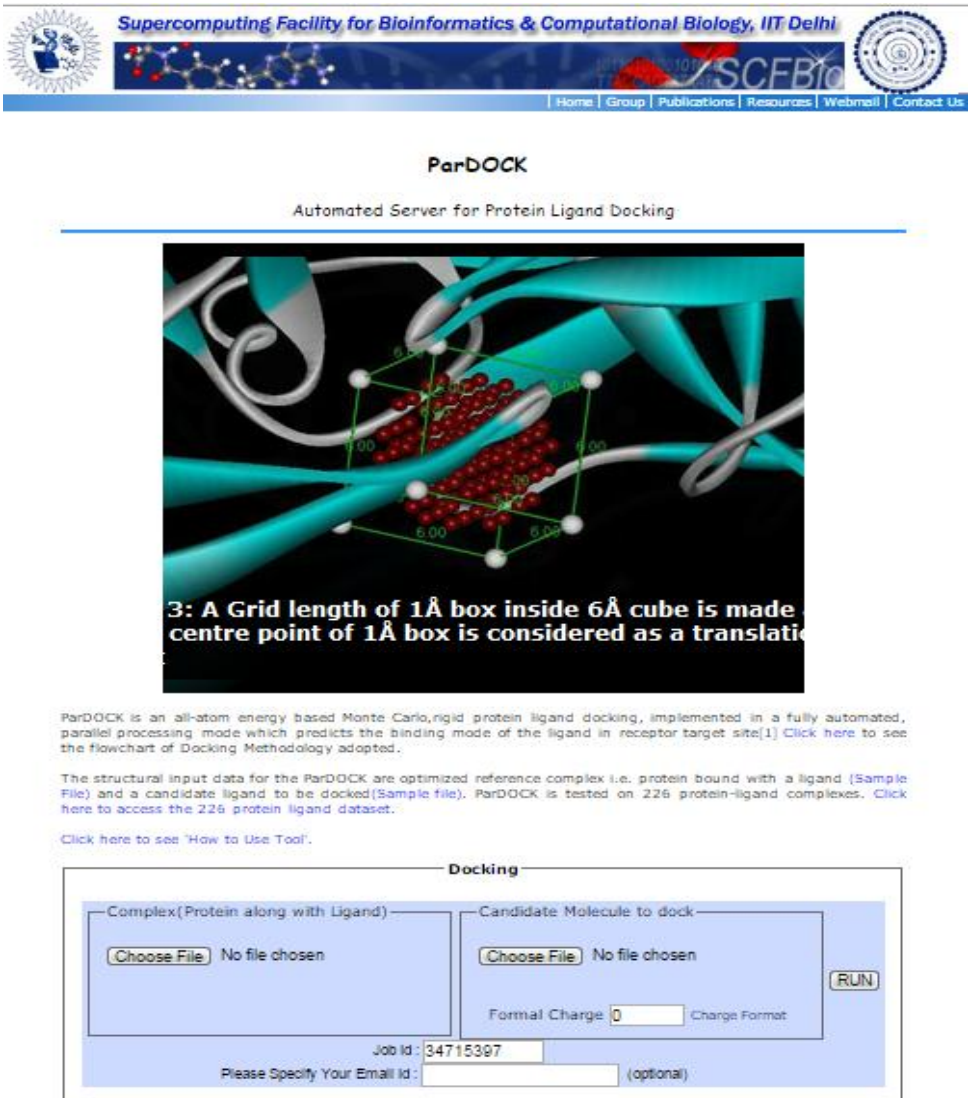


Figure-7: Homepage of ParDOCK

- Docking and Scoring with Case 2: When the target protein has a Multiple Binding Site. AADS (Active Site identification followed by docking and scoring) methodology predicts binding sites in a protein. The top ten binding sites reported by the above algorithm have an accuracy of 100%. The top ten binding sites predicted by the above algorithm are submitted for an automated docking and scoring. The candidate drug molecule is docked at all the ten binding sites predicted by the above algorithm and the docked structures are finally scored using an in house developed scoring function christened BAPPL. Finally ten docked structures along with their predicted binding free energy values are reported back to the user via e-mail. (Latha *et al., 2006*)

  After obtaining the cavity points through AADS, click on the Click here for Docking button. On doing so, in a pop-up window we'll be asked to give the total formal charge of the chosen drug molecule along with e-mail id and then submit the job (figure-8). Ten docked structures along with their predicted binding affinity values will be sent back via e-mail.

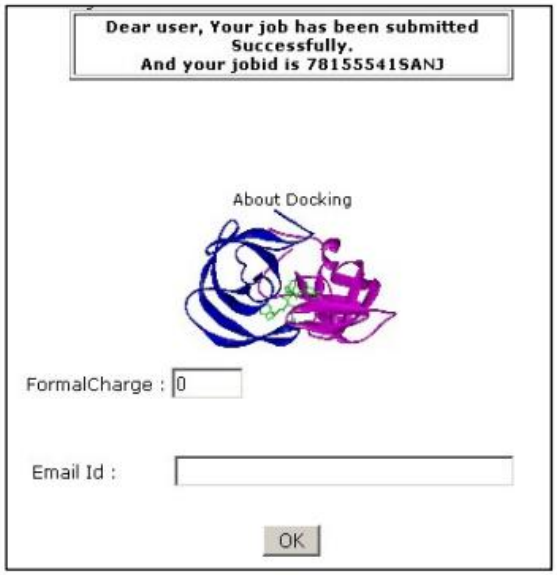


Figure-8: Docking Directly from AADS

All steps which are mentioned above will be repeated for all the genomes of four serotypes of Dengue virus.

RESULTS

ChemGenome 2.0 output

There is one protein coding regions in all four serotype of Dengue virus. Gene predicted by the ChemGenome 2.0 software encodes for Dengue virus polypeptide.

Dengue virus	Gene Start	Gene End
Type-1	95	10273
Type-2	97	10272
Type-3	95	10267
Type-4	102	10265

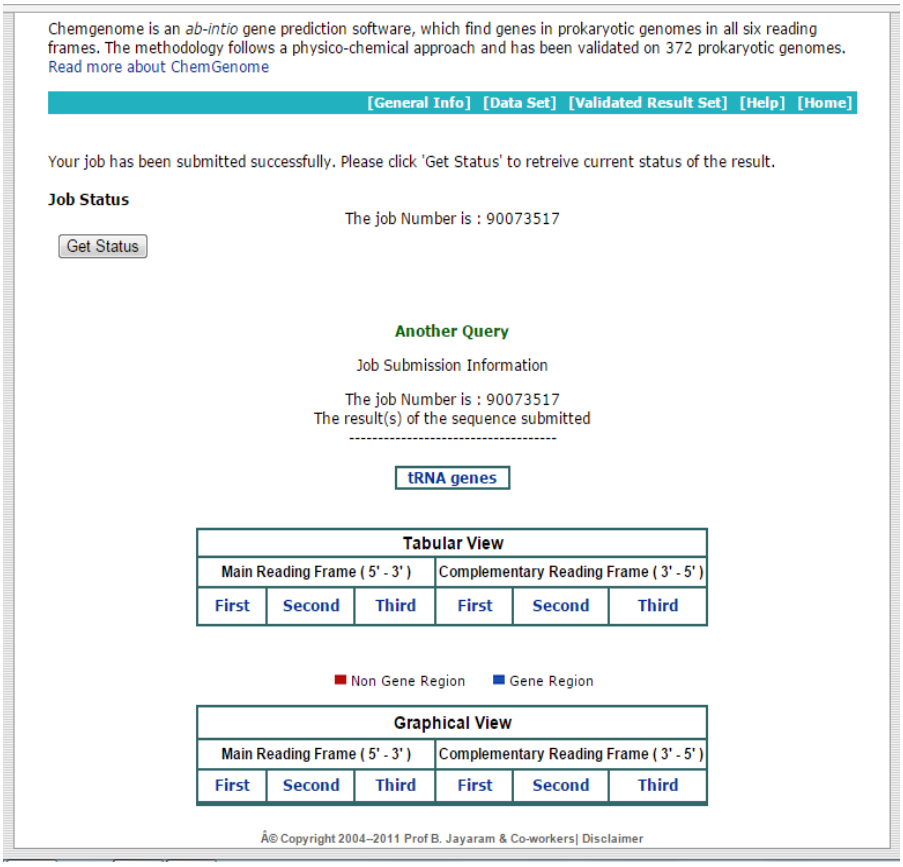


Figure-9: Output of Chemgenome 2.0

Output shows the both tabular view and graphical view of the genome to show the genes presents in the genome (figure-9). It shows the result from all six reading frames

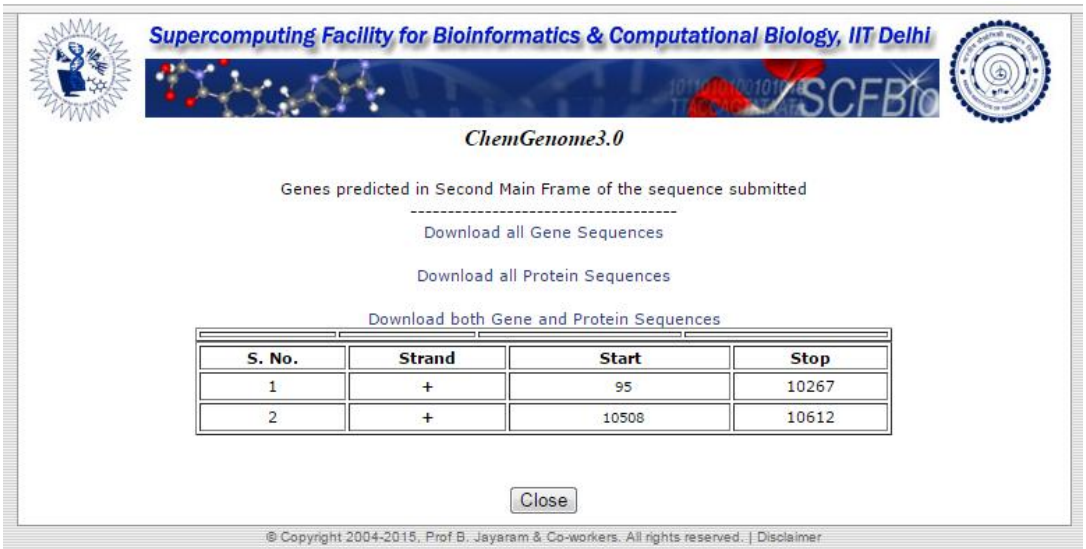


Figure-10: Tabular view of Gene

Figure- 10 showing the result of second reading frame in tabular form. And figure-11 showing the result of the same reading frame but in graphical representation, red regions are non-coding region and blue regions are coding regions in genome of Dengue Type 3 virus.

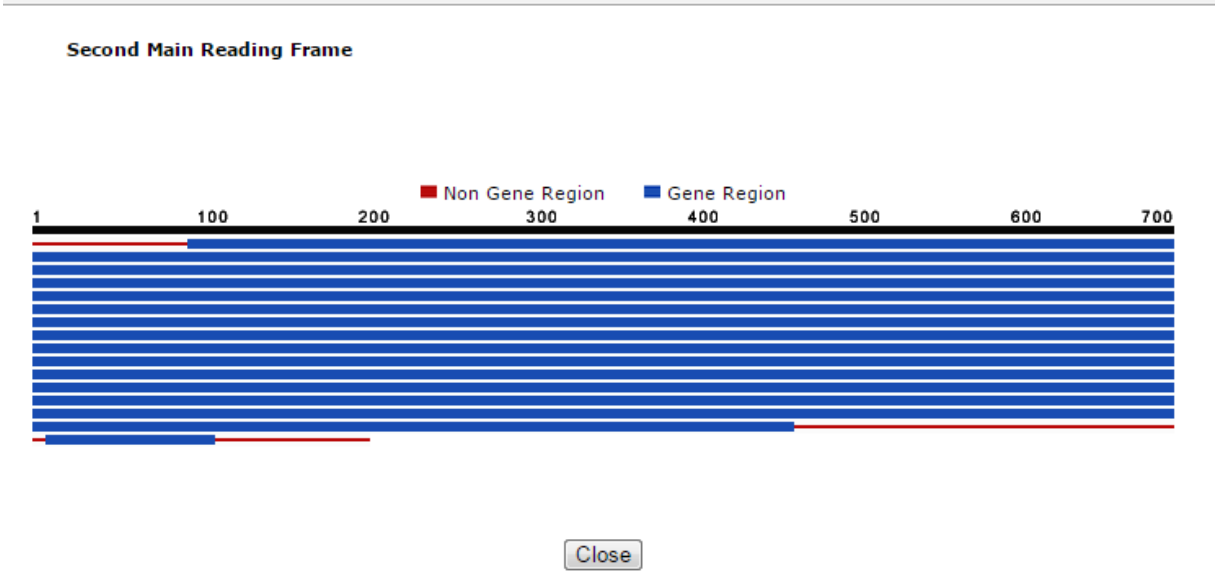


Figure-11: Graphical view of gene

The gene sequence is translated into amino acid sequence using DNA converter (http://www.scfbio-iitd.res.in/chemgenome/genetranslator.jsp), which is then used as an input to the *Bhageerath*-H (or *Bhageerath*) for 3D structure prediction.

Blast Output

In the output of blastn of genes of Dengue virus with Human genomic + transcript database, there was no similarity between the sequences of the genes of viral genome and the transcript of Human (figure-12). It means that in humans there is no such protein which is similar to viral protein. On the basis of the output of the blastn we can assume that the hits which we are going to produce against the viral enzymes are not going to bind with human proteins. It means hits against NS1, NS3 and NS5 will not bind to human proteins and will not block the activity of human proteins.

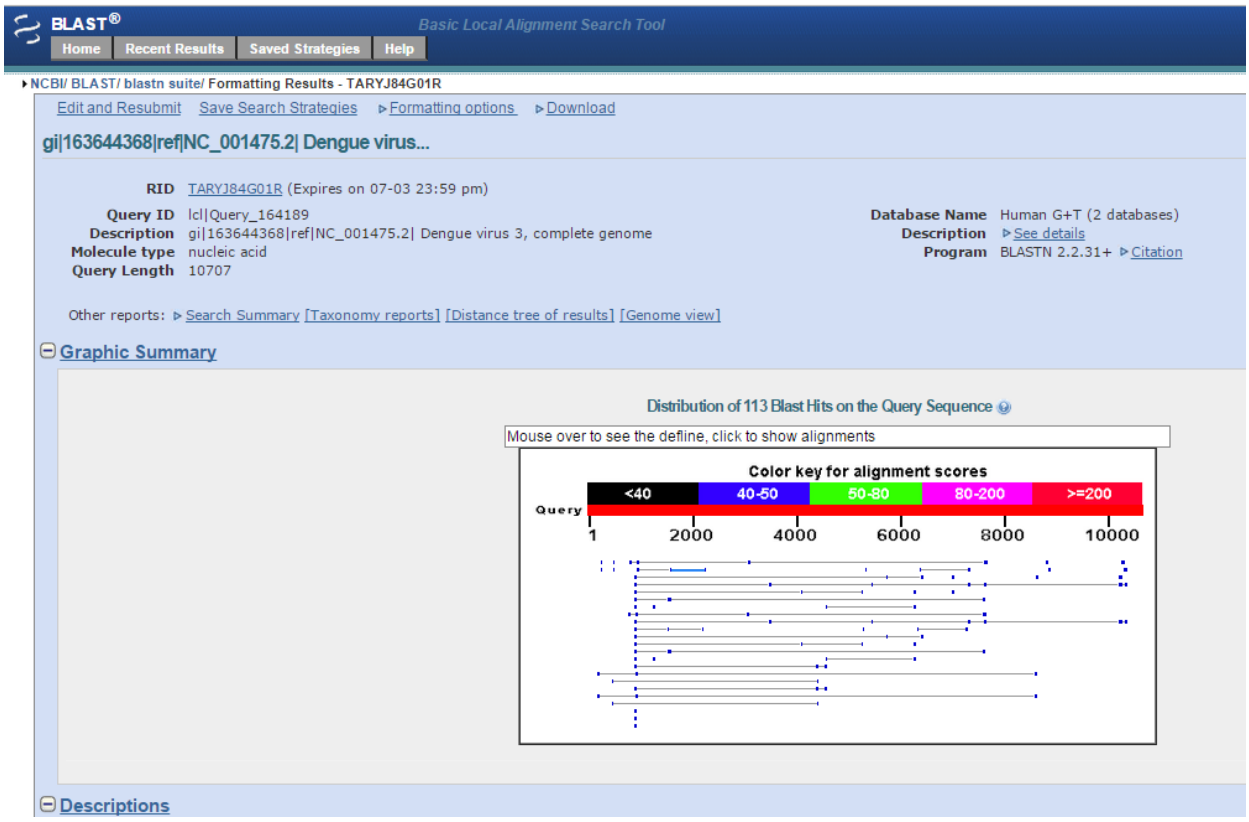


Figure-12: Output of Blastn

Bhageerath-H Output

Proteins with experimentally known (X-ray/NMR) 3D structures can be used directly for active site detection and docking studies. In the absence of a known structure, computational softwares like Bhageerath-H or Bhageerath can be used. For the purposes of illustration, we have given the amino acid sequence of the protein with known crystal structure for Bhageerath-H prediction.

Top five energy ranked structures are shown in figure-13 as Bhageerath-H output.

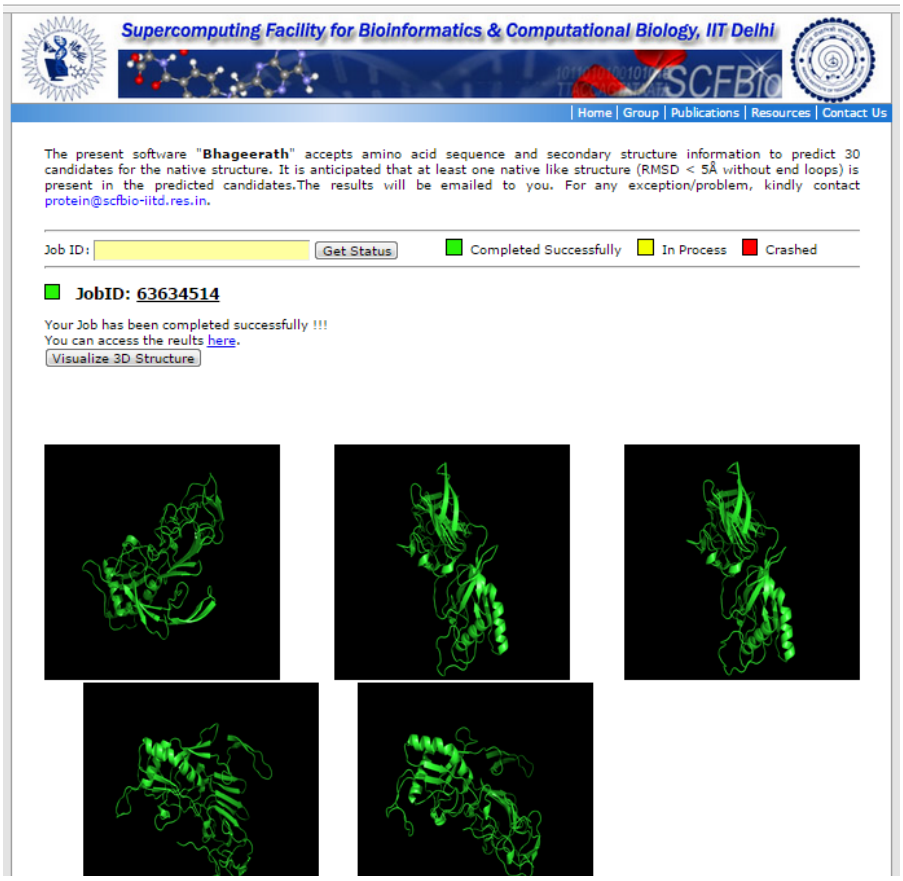


Figure-13: Output of Bhageerath-H

AADS Output

All the structures modeled by Bhageerath- H could be considered for active site identification. Here we used Model 1 as an input to the Active Site Finder. There will be top ten potential binding pockets which are expected to bracket the true "active site" (binding pocket) shown in figure- 14.

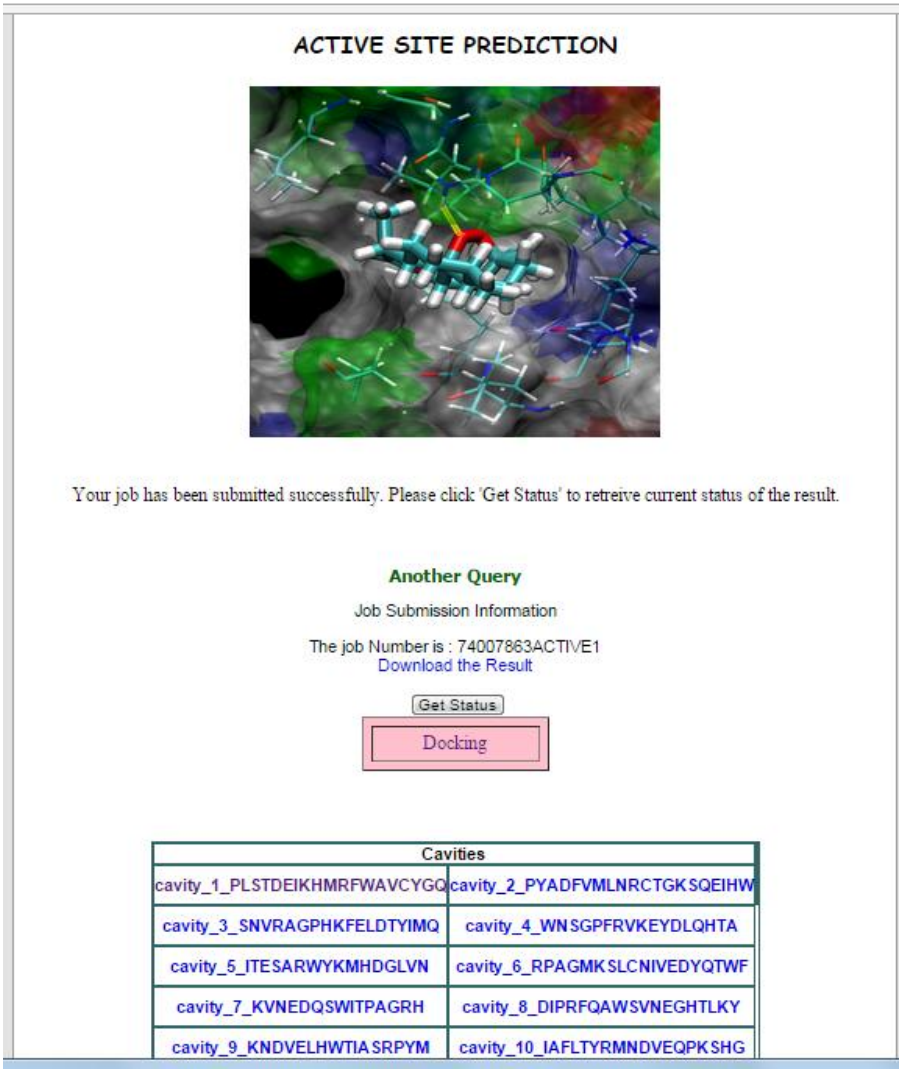


Figure-14: Output of AADS

File Name = cavity_1_VATPEWNFHSIGMKLDRQY
Cavity point 4.006 -26.561 -20.696
Volume of the Cavity in Angstrom cube = 878

1	VAL	537	H	10.193	-26.949	-26.733
2	ALA	186	O	-8.515	-27.000	-25.695
3	ALA	186	C	-7.918	-27.764	-24.933
4	ALA	196	2HB	-7.754	-25.056	-18.231
5	VAL	21	1HG2	-8.300	-23.385	-22.548
6	THR	22	H	-7.928	-20.145	-21.709
7	PRO	250	2HG	-7.502	-34.547	-17.599
8	ALA	536	3HB	-7.067	-25.706	-28.807
9	VAL	21	2HG2	-7.020	-24.588	-22.836
10	VAL	21	CG2	-7.231	-23.526	-22.710
11	THR	22	O	-6.980	-21.196	-19.183
12	GLU	24	H	-7.095	-20.836	-15.829
13	TRP	201	HE3	-5.708	-24.520	-16.802
14	VAL	21	3HG2	-6.681	-23.146	-21.849
15	TRP	201	HZ3	-5.732	-23.403	-18.929
16	VAL	370	3HG1	-6.585	-22.134	-30.470
17	THR	22	3HG2	-6.143	-16.959	-21.889
18	ASN	255	CB	-4.932	-38.835	-17.686
19	ASN	255	2HB	-5.464	-37.998	-18.138
20	VAL	188	3HG1	-4.917	-30.355	-29.158
21	ASN	23	HA	-5.582	-21.575	-17.214
22	ASN	23	C	-5.456	-19.685	-16.145
23	ASN	23	O	-4.789	-18.687	-15.872
24	PHE	20	HE2	-5.035	-15.386	-23.005
25	THR	27	2HG2	-5.025	-15.746	-14.505
26	ASN	255	H	-4.185	-38.850	-20.116
27	HIE	254	2HB	-4.078	-36.597	-21.877
28	VAL	188	N	-4.027	-28.512	-27.559
29	VAL	188	H	-4.503	-28.048	-28.319
30	ALA	187	C	-4.556	-28.477	-26.345
31	THR	27	1HG2	-4.203	-17.233	-13.975
32	THR	27	CG2	-4.274	-16.155	-13.829
33	ASN	255	N	-3.381	-38.603	-19.557
34	ASN	255	CA	-3.460	-38.791	-18.105
35	ASN	255	HA	-2.975	-37.965	-17.586
36	HIE	254	CB	-3.011	-36.679	-22.085
37	ALA	193	2HB	-2.851	-32.709	-24.285
38	VAL	188	O	-2.742	-28.345	-30.097
39	SER	534	CB	-3.015	-27.291	-33.263
40	SER	534	2HB	-3.297	-26.818	-32.323

Figure-15: cavity 1 information

In above figure-15 the information about the cavity which was predicted by the AADS like which amino acids are presents are in the cavity, what is the cavity point and volume of the cavity.

RASPD output

The predicted binding sites could also be used to derive this information which is then used as input for RASPD calculation with an average cut off binding affinity to limit the number of candidates. A list of molecules were selected with good binding energy from one million molecule database corresponding to the top 10 predicted binding sites. (Mukherjee *et al., 2013*) Figure- 16, RASPD output gives library of molecules for the all top ten cavities of the protein.

Cavity-1			
-16.8	ZINC19857639		
-16.5	ZINC12050585		
-16.4	ZINC02831975		
-16.4	ZINC09669021		
-16.4	ZINC11913294		
-16.4	ZINC12576410		
-16.3	ZINC16480347		
-16.2	ZINC00680516		
-16.2	ZINC02836173		
-16.2	ZINC03143011		
-16.2	ZINC03877717		
-16.2	ZINC22064237		
-16.1	ZINC12419773		
-16.1	ZINC19835705		
-16	ZINC01139950		
-16	ZINC09233632		
-16	ZINC11790331		
-16	ZINC11974393		
-15.9	ZINC01799612		
-15.9	ZINC02109070		
-15.9	ZINC02760064		
-15.9	ZINC02942316		
-15.9	ZINC08769779		
-15.9	ZINC11912462		
-15.9	ZINC13081002		
-15.9	ZINC14885566		
-15.9	ZINC15005335		

Figure-16: Output of RASPD

ParDOCK Output

Out of these molecules which are extracted from RASPD, 100 molecules for each Non structural protein chosen randomly are given as input to ParDOCK for atomic level binding energy calculations. Common molecules from the RASPD output for the all four serotype given to ParDOCK for a better hits (figure-17). ParDOCK gives the output of the Ligand -protein docking result with binding affinity of the molecule to the all top 10 active sites of the predicted protein structure.

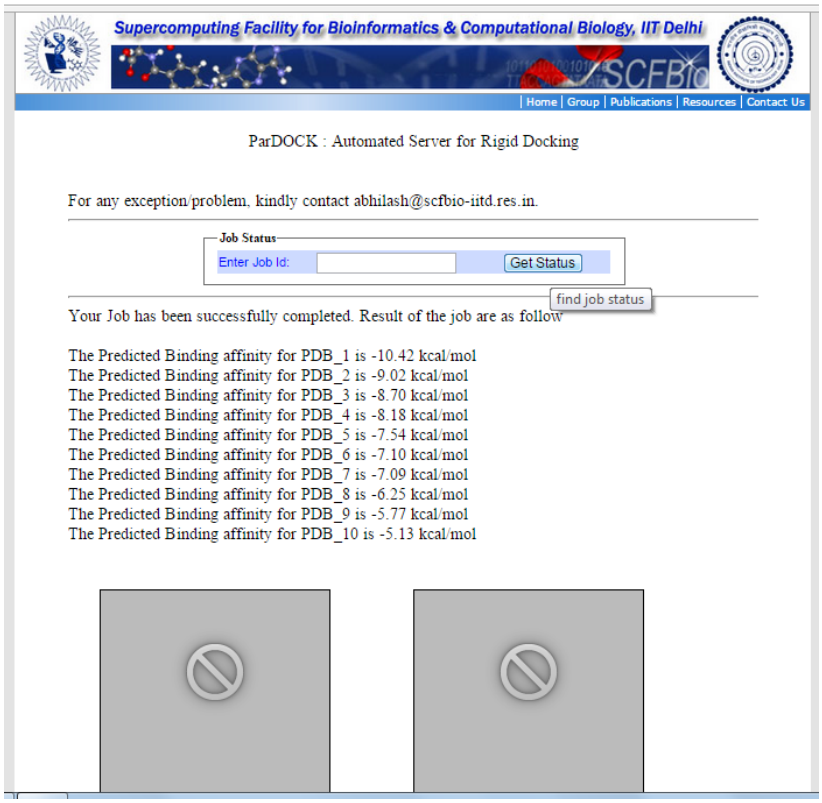


Figure-17: Output of ParDOCK

There are all values of the binding affinity of the molecules with NS1, NS3 and NS5 of all four serotypes of Dengue virus given in table-6.

Table- 6: list of molecules used in docking for NS1, NS3 and NS5

Protein	Drug	Type3	Type 1	Type 2	type 4
ns1	ZINC19857639	-8.4			
	ZINC12050585	-8.7			

	ZINC09669021	-9.02	-7.6	-8	-8.5
	ZINC03143011	-7.5			
	ZINC12419773	-8.1			
	ZINC01139950	-10.42	-7.2	-7.8	-9.3
	ZINC02108812	-7.3			
	ZINC03005330	-8.6			
	ZINC08769779	-7.4			
	ZINC11912462	-8.1			
	ZINC13081002	-6.8			
	ZINC14885566	-8			
	ZINC02872334	-7.5			
	ZINC19817993	-8			
	ZINC19839187	-9.1	-7.4	-6.8	-9.5
	ZINC20608356	-7.6			
	ZINC14731992	-7.1			
	ZINC14878716	-8.7	-7.6	-8	-9.3
	ZINC14960061	-7			
	ZINC14971028	-8.4			
	ZINC14976848	-8.2			
	ZINC14976913	-9	-7.8	-7.5	-9.1
	ZINC14980992	-7.5			
	ZINC14981585	-8.7	-7.6	-7.5	-8.7
	ZINC14983543	-8.8	-7		
	ZINC14988652	-8.3	-7.2	-7	-8.1
	ZINC14990396	-7.6			
	ZINC14994511	-7.5			
	ZINC14996254	-9.1	-8.9	-9	-9.7
	ZINC14996886	-7.4			
	ZINC15003471	-6			
	ZINC15004922	-8.7	-7.2	-7.8	-8.7
	ZINC15005051	-9.7	-8.9	-8.9	-10
	ZINC15005205	-8.3			
	ZINC15006729	-8.5	-8.1	-8	-10.2
	ZINC15006917	-7.9			
	ZINC15007810	-7.7			
	ZINC15009246	-8.4	-7.2	-8.1	-9.6
	ZINC15064125	-8.1			
	ZINC15071932	-7.8			
	ZINC15636531	-8.6	-7.2	-8.6	-9.3
	ZINC16193598	-6.1			
	ZINC16430580	-7.9			

	ZINC19885609	-10.2	-7.6	-8.3	-9.3
	ZINC20118037	-8			
	ZINC20213083	-8			
	ZINC20352944	-8.1			
	ZINC20394103	-7.2			
	ZINC11966255	-8.4	-7.4	-8.5	-9.3
	ZINC11972241	-7.8			
	ZINC01139950	8.9	-8	-7.9	-9.5
ns3	ZINC03143011	-6.5			
	ZINC02880085	-6.8			
	ZINC16671375	-7.7			
	ZINC16667348	-6.4			
	ZINC03143011	-8.3	-6.6		
	ZINC02880085	-7.6			
	ZINC09554231	-6.7			
	ZINC16671375	-6.6			
	ZINC01139952	-8.4	-9.7	10.3	-10.1
	ZINC19805326	-8.3	-9	-9.3	-8.7
	ZINC02836173	-8.5	-7.9		-8.2
	ZINC08892130	-7.8			
	ZINC19797529	-8.3	-8.5		-8.8
	ZINC11913294	-8			
	ZINC13081002	-7.5			
	ZINC04122853	-8.2	-8.1		-9
	ZINC12213012	-7.9			
	ZINC12419770	-8.4	-8.2		-7.9
	ZINC19835705	-8.2	-8.5		-8.6
	ZINC02851420	-8			
	ZINC16248648	-6.9			
	ZINC01799612	-6.7			
	ZINC04020431	-7.6			
	ZINC11912462	-7.8			
	ZINC13322128	-8.6	-8.6		-8.9
	ZINC13424926				
	ZINC16193214	-7.5			
	ZINC19857639	-9.6	-9.6	-8.4	-9.1
	ZINC05221544	-7.9			
	ZINC02985323	-8.6	-8.7		-9
	ZINC11908856	-7.8			
	ZINC11974393	-8.5	-9.6	-8.2	-8.9
	ZINC12492235	-8.4	-7.7		

	ZINC01903996	-7.4			
	ZINC02744150	-8.4	-8.5		
	ZINC12243318	-7.4			
	ZINC16480347	-7.7			
	ZINC00808592	-8.4	-8.8	-8.7	.8.2
	ZINC03877718	-6.9			
	ZINC22033431	-7.6			
	ZINC11790367	-8.9	-9.7	-8.9	-9.3
	ZINC11869608	-7.8			
	ZINC14882872	-6.7			
	ZINC16671239	-7.9			
	ZINC00671474	-9.3	-9.4	-10.1	-9.8
	ZINC06498003	-9.5	-9	-9.2	-8.8
	ZINC12157057	-9	-9.2	-9.9	-9.8
	ZINC08917967	-8.3	-8.5		
	ZINC09226916	-7.8			
	ZINC11882026	-9.2	-8.2	-9.8	-8.4
	ZINC14885566	-8.8	-8.5		
	ZINC15005338	-8.8	-8.3		
	ZINC20451377	-9	-8.3	-8.4	-8.7
	ZINC02942316	-7.9			
	ZINC09191993	-7.9			
	ZINC12205684	-8.1			
	ZINC12243261	-7.4			
	ZINC16286107	-7.8			
	ZINC20214547	-6.8			
	ZINC20601870	-9.8	-9.7	-9	-9.3
	ZINC01158015	-8.8	-8.6		
	ZINC01794178	-8			
	ZINC02760064	-8.7	-9.4	-8.8	-8.7
	ZINC11783262	-9.8	-9.9	-9.7	-9.9
	ZINC11784161	-8.8	-8.5	-9.8	-9.4
	ZINC12576410				
	ZINC19811543	-8.6	-8.6		
	ZINC20508667	-7.8			
	ZINC02745710	-8.1			
	ZINC08769779	-9.2	-9.5	-8.6	-9.6
	ZINC11840986	-8.3			
	ZINC11841107	-8.6	-8.6		
	ZINC12247323	-8.6	-7.9		
	ZINC12247658	-9.8	-9	-9.8	-9.9

	ZINC14994048	-7.9			
	ZINC16275934	-9.2	-9.1	-7.7	-8
	ZINC17097190	-5.6			
	ZINC19758600	-9.3	-8.8	-8	-8.9
	ZINC19884539	-7.8			
	ZINC01109335	-9	-9.5	-9.8	-10
	ZINC01807072	-8.4			
	ZINC02843658	-8.2			
	ZINC08846618	-8.8			
	ZINC11786333	-8.8	-8.9	-9.2	-9.4
	ZINC11871377	-8			
	ZINC11963881	-8.7			
	ZINC11980941	-8.3			
	ZINC11982148	-8.6			
	ZINC12333572	-8.4			
	ZINC14879312	-8.5			
	ZINC15823058	-9.9	-9.7	-10.6	-9.9
	ZINC17195094	-7.8			
	ZINC19114279	-7.8			
	ZINC19842927	-8.6			
	ZINC20562945	-7.9			
	ZINC22147048	-8			
	ZINC02895958	-7.4			
	ZINC02952200	-8.9	-9.5	-7.3	9.4
	ZINC02985067	-8.8	-8.8	-9.1	-8.7
	ZINC03656658	-8.3			
	ZINC04113772	-8.7			
	ZINC04122191	-9.6	10.3	-9.3	-9.8
ns5	ZINC16667348	-6.5			
	ZINC08935093	-6.8			
	ZINC03143011	-9.2	-9.4	-9.2	-9.7
	ZINC00675736	-8.7	-7.3	-8.4	
	ZINC09554231	-6.9			
	ZINC11882026	-7.7			
	ZINC02811191	-8.8	-9.6	-9.5	-9.6
	ZINC11787305	-8.5			
	ZINC19811543	-9.8	-9.5	-9	-8.9
	ZINC20388027	-8.2			
	ZINC20494835	-7.6			
	ZINC08769779	-8.3			
	ZINC12576410				

	ZINC32603782	-7.8			
	ZINC09191993	-8			
	ZINC09261062	-7.5			
	ZINC12242758	-8.4	-9.6	-9.7	-8.5
	ZINC12243261	-6.4			
	ZINC12247658	-8.3			
	ZINC13081002	-7.3			
	ZINC14885570	-9.3	-9	-9.3	-8.9
	ZINC19805326	-8.3			
	ZINC20562945	-8.6			
	ZINC22147048	-6.8			
	ZINC09309891	-9.1	-10.4	-9.8	-9.7
	ZINC11783262	-8.4			
	ZINC11784200	-8.2			
	ZINC11911425				
	ZINC12034968	-9.2	-10	-9.9	-9.5
	ZINC12242780	-8.1			
	ZINC19114279	-7.9			
	ZINC20601870	-8.6	-8.7	-9.1	-9.3
	ZINC01299985	-8.9	-8.6	-9.6	-8.9
	ZINC01794178	-7.6			
	ZINC01807072	-9.3	-9.4	-8.4	-9
	ZINC02109070	-7.3			
	ZINC02745710	-8.2			
	ZINC02952200	-7.7			
	ZINC04113772	-7			
	ZINC08442648	-7.4			
	ZINC11786333	-8.4			
	ZINC11788825	-8.8	-9.7	-8.9	-8.8
	ZINC11963881	-9.2	-9.6	-9.8	-9.5
	ZINC12247323	-7.9			
	ZINC12419770	-7.1			
	ZINC14976814	-7.8			
	ZINC15014145	-8.2	-9.1	-8.4	-8.6
	ZINC16115868	-6.3			
	ZINC17195094	-6.9			
	ZINC19758600	-8.7	-9.2	-8.8	-9.2
	ZINC19842927	-8.4			
	ZINC20451377	-8.3			
	ZINC00793735	-9.2	-8.8	-9	-9.1
	ZINC02800990	-8.5	-9.6	-9.4	-8.8

	ZINC02851420	-7.5			
	ZINC02895958	-8.3			
	ZINC04620640	-7.5			
	ZINC11787010	-8.8	-8.1		
	ZINC12343898	-8.5			
	ZINC12455413	-8.1			
	ZINC14982226	9.1	-9.6	-9.8	-10.1
	ZINC15008014	-9.4	-9.5	-9.7	-9.3
	ZINC15722724	-6.2			
	ZINC16275934	-8.2			
	ZINC19790803	-9.1	-9.5	-9.4	-9.7
	ZINC19830686	-7.2			
	ZINC20136483	-9.1	-9.2	-9.5	-8.9
	ZINC20294685	-6.8			
	ZINC20919473	-8.1			
	ZINC20999348	-9	-8.5	-8.2	-8.7
	ZINC01080544	-7.9			
	ZINC01139952	-8.4			
	ZINC01863017	7.5			
	ZINC01917379	8.8	-9.1	-8.5	-8.2
	ZINC02812284	-7.3			
	ZINC02895117	-8.9	-8.8	-9.1	-8.4
	ZINC04122191	-7.9			
	ZINC08442648	7.5			
	ZINC08843102	-7			
	ZINC09445558	-8.1			
	ZINC11791105	-8.2			
	ZINC11912369	-8			
	ZINC11980933	-8.8	-9.6	-9.5	-9.3
	ZINC12190465	-8.2			
	ZINC12243664				
	ZINC14978739	-8.6	-8.3	-8.9	-9.2
	ZINC14990472	-8.2			
	ZINC15857222	-7.5			
	ZINC19831804	-8.3			
	ZINC00990870	-8.2			
	ZINC01978122	6.1			
	ZINC02424886	-8.5	-8.8	-9	-8.7
	ZINC02846676	-9	-8.5	-8.8	-8.5
	ZINC02860582	-8.3			
	ZINC02861847	-8.7	-8.1	-8.6	-9.3

	ZINC02986228	-7.4			
	ZINC03656658	-8			
	ZINC08772915	-7.1			
	ZINC09041389	-8			

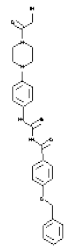
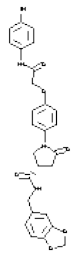
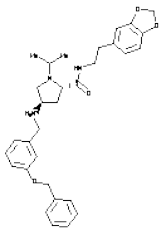
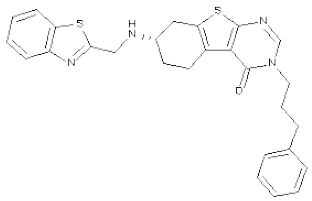
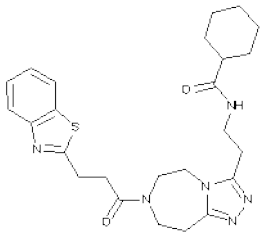
After all this we screen best 10 molecules having best binding affinity for NS1, NS3 and NS5 each. This is given in table-7.

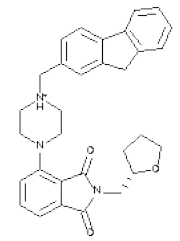
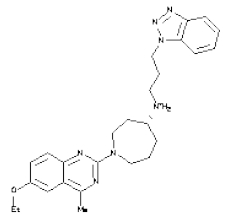
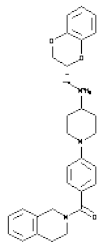
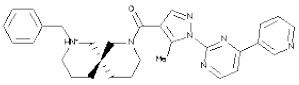
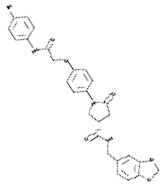
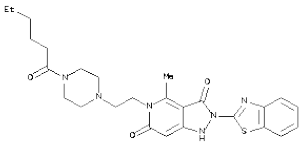
Table-7: list of highly scored molecules

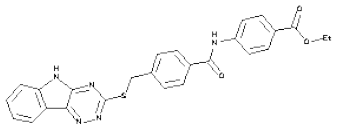
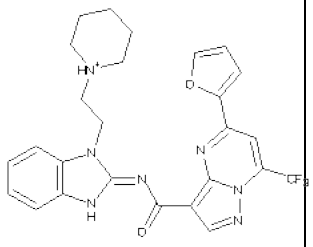
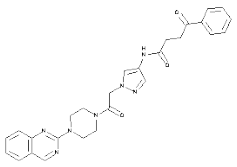
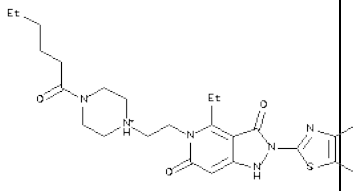
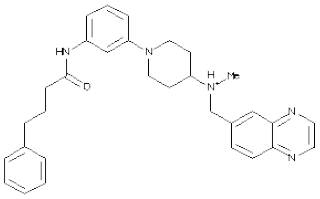
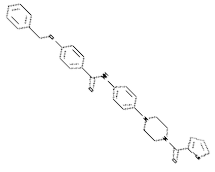
Protein	Drug	Type3	Type 1	Type 2	Type 4
NS1	ZINC09669021	-9.02	-7.6	-8	-8.5
	ZINC01139950	-10.42	-7.2	-7.8	-9.3
	ZINC19839187	-9.1	-7.4	-6.8	-9.5
	ZINC14878716	-8.7	-7.6	-8	-9.3
	ZINC14976913	-9	-7.8	-7.5	-9.1
	ZINC14996254	-9.1	-8.9	-9	-9.7
	ZINC15004922	-8.7	-7.2	-7.8	-8.7
	ZINC15005051	-9.7	-8.9	-8.9	-10
	ZINC19885609	-10.2	-7.6	-8.3	-9.3
	ZINC01139950	8.9	-8	-7.9	-9.5
NS3	ZINC19857639	-9.6	-9.6	-8.4	-9.1
	ZINC00671474	-9.3	-9.4	-10.1	-9.8
	ZINC06498003	-9.5	-9	-9.2	-8.8
	ZINC20601870	-9.8	-9.7	-9	-9.3
	ZINC11783262	-9.8	-9.9	-9.7	-9.9
	ZINC08769779	-9.2	-9.5	-8.6	-9.6
	ZINC12247658	-9.8	-9	-9.8	-9.9
	ZINC01109335	-9	-9.5	-9.8	-10
	ZINC15823058	-9.9	-9.7	-10.6	-9.9
	ZINC04122191	-9.6	10.3	-9.3	-9.8
NS5	ZINC03143011	-9.2	-9.4	-9.2	-9.7
	ZINC19811543	-9.8	-9.5	-9	-8.9
	ZINC09309891	-9.1	-10.4	-9.8	-9.7
	ZINC12034968	-9.2	-10	-9.9	-9.5
	ZINC01807072	-9.3	-9.4	-8.4	-9
	ZINC11963881	-9.2	-9.6	-9.8	-9.5
	ZINC00793735	-9.2	-8.8	-9	-9.1
	ZINC14982226	9.1	-9.6	-9.8	-10.1
	ZINC15008014	-9.4	-9.5	-9.7	-9.3
	ZINC19790803	-9.1	-9.5	-9.4	-9.7

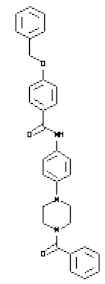
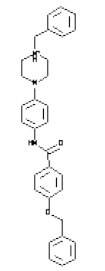
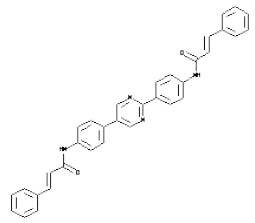
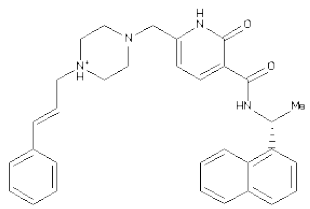
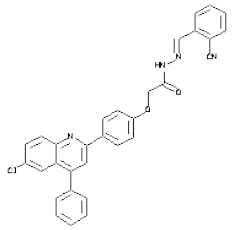
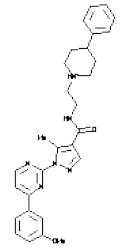
In table-8 popular name, SMILES and structure of these 30 screened molecules are written.

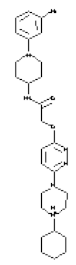
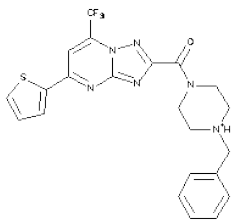
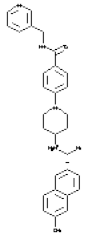
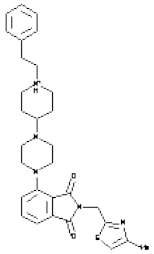
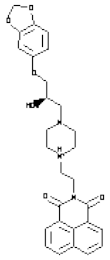
Table-8: Name, SMILES and Structure of hits

	Drug	Name	SMILES	Structure
NS1	ZINC09669021	4-benzyloxy-N-[[4-(4-butanoylpiperazin-1-yl)phenyl]thiocarbamoyl]benzamide	CCCC(=O)N1CCN(CC1)c2ccc(cc2)NC(=S)NC(=O)c3ccc(cc3)OCc4ccccc4	
	ZINC01139950	(3S)-1-[4-[2-(4-ethylanilino)-2-keto-ethoxy]phenyl]-5-keto-N-piperonyl-pyrrolidine-3-carboxamide	CCc1ccc(cc1)NC(=O)COc2ccc(cc2)N3C[C@H](CC3=O)C(=O)NCc4ccc5c(c4)OCO5	
	ZINC19839187	(2S,4R)-N-[2-(1,3-benzodioxol-5-yl)ethyl]-4-[(3-benzyloxyphenyl)methylamino]-1-isopropyl-pyrrolidine	CC(C)N1C[C@@H](C[C@H]1C(=O)NCCc2ccc3c(c2)OCO3)[NH2+]Cc4cccc(c4)OCc5ccccc5	
	ZINC14878716	(7S)-7-(1,3-benzothiazol-2-ylmethylamino)-3-(3-phenylpropyl)-5,6,7,8-tetrahydrobenzothiopheno[3,2-e]	c1ccc(cc1)CCCn2cnc3c(c2=O)c4c(s3)C[C@H](CC4)NCc5nc6ccccc6s5	
	ZINC14976913	N-[2-[7-[3-(1,3-benzothiazol-2-yl)propanoyl]-5,6,8,9-tetrahydro-[1,2,4]triazolo[4,5-d][1,4]diazepin-	c1ccc2c(c1)nc(s2)CCC(=O)N3CCc4nnc(n4CC3)CCNC(=O)C5CCCCC5	

	ZINC14996254	4-[4-(9H-fluoren-2-ylmethyl)piperazin-1-yl]-2-[[(2S)-tetrahydrofuran-2-yl]methyl]isoindoline-1,3-dio	c1ccc-2c(c1)Cc3c2ccc(c3)C[NH+]4CCN(CC4)c5cccc6c5C(=O)N(C6=O)C[C@@H]7CCCO7	
	ZINC15004922	(4R)-N-[3-(benzotriazol-1-yl)propyl]-1-(6-ethoxy-4-methyl-quinazolin-2-yl)azepan-4-amine	CCOc1ccc2c(c1)c(nc(n2)N3CCC[C@H](CC3)[NH2+]CCCn4c5ccccc5nn4)C	
	ZINC15005051	[4-[4-[[(3R)-2,3-dihydro-1,4-benzodioxin-3-yl]methylamino]-1-piperidyl]phenyl]-(3,4-dihydro-1H-isoqu	c1ccc2c(c1)CCN(C2)C(=O)c3ccc(cc3)N4CCC(CC4)[NH2+]C[C@@H]5COc6ccccc6O5	
	ZINC19885609	[(6R)-8-benzyl-4,8-diazaspiro[5.5]undecan-4-yl]-[5-methyl-1-[4-(3-pyridyl)pyrimidin-2-yl]pyrazol-4-y	Cc1c(cnn1c2nccc(n2)c3cccnc3)C(=O)N4CCC[C@]5(C4)CCC[NH+](C5)Cc6ccccc6	
	ZINC01139950	(3S)-1-[4-[2-(4-ethylanilino)-2-keto-ethoxy]phenyl]-5-keto-N-piperonyl-pyrrolidine-3-carboxamide	CCc1ccc(cc1)NC(=O)COc2ccc(cc2)N3C[C@H](CC3=O)C(=O)NCc4ccc5c(c4)OCO5	
NS3	ZINC19857639	2-(1,3-benzothiazol-2-yl)-5-[2-(4-hexanoylpiperazin-1-yl)ethyl]-4-methyl-1H-pyrazolo[5,4-d]pyridine-	CCCCCC(=O)N1CCN(CC1)CCn2c(c3c(cc2=O)[nH]n(c3=O)c4nc5ccccc5s4)C	

	ZINC06498003	ethyl 4-({4-[(5H-[1,2,4]triazino[5,6-b]indol-3-ylsulfanyl)methyl]benzoyl}amino)benzoate	CCOC(=O)c1ccc(cc1)NC(=O)c2ccc(cc2)CSc3nc4c(c5ccccc5[nH]4)nn3	
	ZINC20601870	5-(2-furyl)-N-[1-[2-(1-piperidyl)ethyl]benzimidazol-2-yl]-7-(trifluoromethyl)pyrazolo[1,5-a]pyrimidi	c1ccc2c(c1)[nH]/c(=N\C(=O)c3cnn4c3nc(cc4C(F)(F)F)c5ccco5)/n2CC[NH+]6CCCCC6	
	ZINC11783262	4-oxo-N-[1-[2-oxo-2-(4-quinazolin-2-ylpiperazin-1-yl)ethyl]pyrazol-4-yl]-4-phenyl-butanamide	c1ccc(cc1)C(=O)CCC(=O)Nc2cnn(c2)CC(=O)N3CCN(CC3)c4ncc5ccccc5n4	
	ZINC08769779	8-benzothiazol-2-yl-5-ethyl-4-[2-(4-hexanoylpiperazin-1-yl)ethyl]-4,8,9-triazabicyclo[4.3.0]nona-1,5	CCCCCC(=O)N1CC[NH+](CC1)CCn2c(=O)cc3c(c2CC)c(=O)n([nH]3)c4nc5ccccc5s4	
	ZINC12247658	N-[3-[4-(methyl-(quinoxalin-6-ylmethyl)amino)-1-piperidyl]phenyl]-4-phenyl-butanamide	C[NH+](Cc1ccc2c(c1)nccn2)C3CCN(CC3)c4cccc(c4)NC(=O)CCCc5ccccc5	
	ZINC01109335	4-benzoxy-N-[4-[4-(2-thenoyl)piperazino]phenyl]benzamide	c1ccc(cc1)COc2ccc(cc2)C(=O)Nc3ccc(cc3)N4CCN(CC4)C(=O)c5cccs5	

	ZINC15823058	N-[4-(4-benzoylpiperazin-1-yl)phenyl]-4-benzyloxy-benzamide	c1ccc(cc1)COc2ccc(cc2)C(=O)Nc3ccc(cc3)N4CCN(CC4)C(=O)c5ccccc5	
	ZINC04122191	4-benzyloxy-N-[4-(4-benzylpiperazin-1-yl)phenyl]-benzamide	c1ccc(cc1)C[NH+]2CCN(CC2)c3ccc(cc3)NC(=O)c4ccc(cc4)OCc5ccccc5	
NS5	ZINC03143011	(E)-3-phenyl-N-[4-[2-[4-[[(E)-3-phenylacryloyl]amino]phenyl]pyrimidin-5-yl]phenyl]acrylamide	c1ccc(cc1)/C=C/C(=O)Nc2ccc(cc2)c3cnc(nc3)c4ccc(cc4)NC(=O)/C=C/c5ccccc5	
	ZINC19811543	6-[(4-cinnamylpiperazin-1-yl)methyl]-N-[(1R)-1-(1-naphthyl)ethyl]-2-oxo-1H-pyridine-3-carboxamide	C[C@H](c1cccc2c1cccc2)NC(=O)c3ccc([nH]c3=O)CN4CC[NH+](CC4)C/C=C/c5ccccc5	
	ZINC09309891	2-[4-(6-chloro-4-phenyl-2-quinolyl)phenoxy]-N-[(2-cyanophenyl)methyleneamino]acetamide	c1ccc(cc1)c2cc(nc3c2cc(cc3)Cl)c4ccc(cc4)OCC(=O)N/N=C/c5ccccc5C#N	
	ZINC12034968	1-[4-(3-methoxyphenyl)pyrimidin-2-yl]-5-methyl-N-[2-(4-phenyl-1-piperidyl)ethyl]pyrazole-4-carboxami	Cc1c(cnn1c2nccc(n2)c3cccc(c3)OC)C(=O)NCC[NH+]4CCC(CC4)c5ccccc5	

	ZINC11963881	2-[6-(4-cyclohexylpiperazin-1-yl)pyridazin-3-yl]oxy-N-[1-(m-tolyl)-4-piperidyl]acetamide	Cc1cccc(c1)N2CCC(CC2)NC(=O)COc3ccc(nn3)N4CC[NH+](CC4)C5CCCCC5	
	ZINC00793735	(4-benzylpiperazin-1-yl)-[3-(2-thienyl)-5-(trifluoromethyl)-2,6,7,9-tetrazabicyclo[4.3.0]nona-2,4,7,	c1ccc(cc1)C[NH+]2CCN(CC2)C(=O)c3nc4nc(cc(n4n3)C(F)(F)F)c5cccs5	
	ZINC14982226	4-[4-[[(1R)-1-(6-methoxy-2-naphthyl)ethyl]amino]-1-piperidyl]-N-(3-pyridylmethyl)benzamide	C[C@H](c1ccc2cc(ccc2c1)OC)[NH2+]C3CCN(CC3)c4ccc(cc4)C(=O)NCc5cccnc5	
	ZINC15008014	2-[(4-methylthiazol-2-yl)methyl]-4-[4-(1-phenethyl-4-piperidyl)piperazin-1-yl]isoindoline-1,3-dione	Cc1csc(n1)CN2C(=O)c3cccc(c3C2=O)N4CCN(CC4)C5CC[NH+](CC5)CCc6ccccc6	
	ZINC19790803	2-[4-[(2R)-3-(1,3-benzodioxol-5-yloxy)-2-hydroxy-propyl]piperazin-1-yl]ethylBLAHdione	c1cc2cccc3c2c(c1)C(=O)N(C3=O)CC[NH+]4CCN(CC4)C[C@H](COc5ccc6c(c5)OCO6)O	

These molecules may prove to be potential identity in modulating disease manifestation for all the serotypes of Dengue virus.

CONCLUSIONS

Post-genomic research era encompasses many diverse aspects of modern science. The "Genome to hits" pathway described here symbolizes the emergence of an integrated technology to address specific health issues, and more specifically provides a novel and rapid approach to identifying new and potent hit molecules from genomic information.

Inhibiting the viral enzymes is the best way to design an antiviral. NS1 has been associated with a role in protective immunity. NS3 is a protease and a helicase; whereas NS5 is the RNA polymerase in charge of viral RNA replication.NS3 (non-structural 3) and NS5 carry out all the enzymatic activities needed for polyprotein processing and genome replication.

Two general strategies can be pursued for any antiviral therapy. The first is to inhibit viral targets. The second is to inhibit host targets. Inhibition of targets in the host continues to be risky since it is not yet known what all the mechanisms of infection and disease development in the host are, particularly in cases of dengue hemorrhagic fever. Another point is that the action of the drug should be well known to minimize side effects.

In this case, inhibition of viral targets is a good way. The best characterized DENV proteins are nonstructural proteins NS3 and NS5, which are multifunctional proteins presenting several enzymatic activities. These proteins are the most conserved ones in all four dengue virus serotypes. It is a good point to create an efficient drug that acts in all four serotypes. Checking the similarities between the genes of viral genome with Human transcripts shows that there is no such similarity found. It means hits against NS1, NS3 and NS5 will not bind to human proteins and will not block the activity of human proteins.

Although there are still no specific vaccines or chemotherapeutic regimes for prevention and treatment for flavivirus based diseases like dengue, hemorrhagic fever, encephalitis etc. In recent years there has been substantial progress in our understanding the life cycle of flavivirus, the various stages of which represent potential targets for the development of novel antiviral drugs. NS3 and NS5 protein are particularly interesting molecular target for antiviral compounds because of it central role in all the viral life cycle of flaviviruses. The development of such drugs requires a more informed structure- based drug discovery program. We hope that the lead molecules generated from this structure based drug designing of NS1, NS3 and NS5 protein would be helpful in identifying structurally diverse compounds with desired biological activity for the successful treatment of Dengue.

DISCUSSION AND FUTURE PERSPECTIVE

The wealth of information available from experimental host pathogen interaction studies invites computational biologists to develop databases and newer computational methods to advance further focused experimentation. Consequently, bioinformatics is rapidly evolving into independent fields addressing specific problems in interpreting (i) genomic sequences, (ii) protein sequences and 3D-structures, as well as (iii) transcriptome and macromolecular interaction data. It is thus increasingly difficult for the biologist to choose the computational approaches that perform best in inhibiting the growth of pathogen in the host. Genome to hit technology is used in this project to Dengue virus. Genome to hit assembly line is a culmination of several recent advances in computational chemistry and computational biology implemented in a high performance computing environment. At least three areas for further improvement can be immediately identified : (i) development of algorithms for cleavage of polyproteins , (ii) algorithms for identification of druggable protein targets , (iii) improved accuracies in tertiary structure prediction of nonstructural proteins, (iv) development of methods for determining tertiary structures of structural proteins and (v) identification of hit molecules with reduced toxicities. This protocol should ultimately result in an accelerated emergence of new methods for treating infectious diseases. Similarly, metabolic disorders can also be accessed via the "Genome to Hit" pathway. (Noble *et al., 2010*)

Even with forthcoming advances, a drug for clinical trials has not yet been obtained. A greater understanding of the viral replicative cycle in the host and the variations that occur in different serotypes will be necessary for the choice and improvement of a suitable drug. A challenging aspect in the search for potent, Selective antiviral drugs that interfere with multifunctional NS3 and NS5 is the design of appropriate assays for druggable sites that are relevant for viral replication *in vivo*.

Although challenging, recent research has generated optimistic results and has encouraged researchers to develop an effective therapy against the dengue virus in the near future . This protocol can be used for the other structural and non-structural viral enzymes to identify the hits for Dengue virus.

REFERENCES

1. Agnihotri, S; Narula, R; Joshi, K; Rana, S; Singh, M. (2012). In silico modeling of ligand molecule for non structural 3 (NS3) protein target of flaviviruses. Bioinformation.;8(3):123-7. (Agnihotri *et al., 2012*)

2. Aleshin, A; Shiryaev, S; Strongin, A; Liddington, R.. (2007). Structural evidence for regulation and specificity of flaviviral proteases and evolution of the Flaviviridae fold. Protein Sci. 16 (5), 795–806. (Aleshin *et al., 2007*)

3. Alcaraz, SL; Monroy, MY; Rosa María, R (2010). Insights into dengue virus genome replication. Future Virology Vol. 5, No. 5, Pages 575-592 . (Alcaraz *et al., 2010*)

4. Bartenschlager, R; Miller, S. (2008). Molecular aspects of Dengue virus replication. Future Microbiol. Apr;3(2):155-65. (Bartenschlager *et al., 2008*)

5. Bhushan, K; Jayaram, B; Shenoy, SR. (2006). Bhageerath: an energy based web enabled computer software suite for limiting the search space of tertiary structures of small globular proteins. Nucleic Acids Res; 34: 6195-204. (Bhushan *et al., 2006*)

6. Brooks, A.J; Johansson, M; John, A.V; Xu, Y; Jans, D.A; Vasudevan, S.G. (2002). The interdomain region of dengue NS5 protein that binds to the viral helicase NS3 contains independently functional importin beta 1 and importin alpha/beta recognized nuclear localization signals. J. Biol. Chem. 277 (39), 36399–36407. (Brooks *et al., 2002*)

7. Canard, B. (2009). Antiviral Research and Development Against Dengue Virus. (Canard *et al., 2009*)

8. Dhingra, P; Mishra, A; Kaushik, R; Jayaram, B; Mukherjee, G; Singh, A; Shekhar, S. (2014). Bhageerath-H: A homology/ab initio hybrid server for predicting tertiary structures of monomeric soluble proteins. BMC Bioinformatics. 15(Suppl 16):S7. (Dhingra *et al., 2014*)

9. Goodsell, DS. (2008). Molecule of the month: Dengue virus (Goodsell *et al., 2008*)

10. Halstead, S.B. (2003). Neutralization and antibody-dependent enhancement of dengue viruses. Adv. Virus Res. 60, 421–467. (Halstead *et al., 2003*)

11. Jayaram, B; Dhingra, P; Mukherjee, G; Perumal, V. (2012). Genomes to Hits: The Emerging Assembly Line In Silico. Proceedings of the Ranbaxy Science Foundation 17th Annual Symposium on "New Frontiers in Drug Design, Discovery and Development', Chapter 3, 13-35. (Jayaram *et al., 2012*)

12. Kalra P, Reddy TV, Jayaram B(2001). Free Energy Component Analysis for Drug Design: A Case Study of HIV-1 Protease.Inhibitor Binding. J Med Chem; 44: 4325-38. (Kalra *et al., 2001*)

13. Latha, N; Jain, T; Sharma, P; Jayaram, B; Gandhimathi, A; Pandey VS.(2006). Sanjeevini: A comprehensive active site directed lead design software. Indian J Chem A 2; 45: 1834-37. (Latha *et al., 2006*)

14. Lescar, J; Luo , D; Xu, T; Sampath , A; Lim , SP,;Canard , B; Vasudevan, SG. (2008). Towards the design of antiviral inhibitors against flaviviruses: The case for the multifunctional NS3 protein from Dengue virus as a target. Antiviral Res.;80(2):94-101. (Lescar *et al., 2008*)

15. Lim, SP; Noble, CG; Shi, PY. (2015). The dengue virus NS5 protein as a target for drug discovery. Antiviral Res;119:57-67. (Lim *et al., 2015*)

16. Luo, D; Vasudevan, SG; Lescar, J. (2015). The flavivirus NS2B–NS3 protease–helicase as a target for antiviral drug development. Antiviral Res.;118:148-158. (Luo *et al., 2015*)

17. Marijke, MF; Alen, F. (2012). Dengue Virus Entry as Target for Antiviral Therapy. Journal of Tropical Medicine. (Agnihotri *et al., 2012*)

18. Mathur, A; Shekhar, S; Jayaram, B; Singh, T; Mukherjee, G; Shekhar, V.(2012). Sanjeevini: A freely accessible web-server for target directed lead molecule discovery. BMC Bioinformatics; 13: S7. (Mathur *et al., 2012*)

19. Mukherjee G, Jayaram B. (2013). A rapid scoring methodology based on physico-chemical descriptors of small molecules (RASPD) for identifying hits against a protein target. Phys Chem Chem Phys. (Mukherjee *et al., 2013*)

20. Noble, CG; Chen , YL; Dong , H; Gu, F; Lim , SP; Schul, W; Wang, QY; Shi, PY. (2010). Strategies for development of dengue virus inhibitors. Antiviral Res. Mar;85(3):450-62. (Noble *et al., 2010*)

21. Oliveira, AS; Silva, ML ; Oliveira, FC; Silva, F; Teixeira, RR ;Paula,SO. (2014). NS3 and NS5 Proteins: Important Targets for Anti-Dengue Drug Design. J. Braz. Chem. Soc., Vol. 25, No. 10, 1759-1769. (Oliveira *et al., 2014*)

22. Shenoy SR, Jayaram B.(2013) .Proteins: Sequence to Structure and Function-Current Status. Curr Protein Pept Sci; 11: 498-514. (Shenoy *et al., 2013*)

23. Singh, T; Jayaram, B; Biswas, D. (2011). AADS - An Automated Active Site Identification, Docking, and Scoring Protocol for Protein Targets Based on Physicochemical Descriptors. J. Chem. Inf. Model. 51 (10), pp 2515–2527. (Singh *et al., 2011*)

24. Soni, A; Khushhali, M; Pandey,KM; Ray, P; Jayaram, B. (2013)Genomes to Hits In Silico - A Country Path Today, A Highway Tomorrow: A Case Study of Chikungunya. Current Pharmaceutical Design, 19, 4687-4700. (Soni *et al., 2013*)

25. Seema; Jain, SK. (2005). MOLECULAR MECHANISM OF PATHOGENESIS OF DENGUE VIRUS : ENTRY AND FUSION WITH TARGET CELL. Indian Journal of Clinical Biochemistry, 20 (2) 92-103. (Seema *et al., 2005*)

26. Tomlinso, SM; Malmstrom, RD; Watowich, SJ.(2009). New Approaches to Structure-Based Discovery of Dengue Protease Inhibitors. Infectious Disorders - Drug Targets, 9, 327-343. (Tomlinso *et al., 2009*)

27. Xie, X; Zou, J; Wang, QY; Shi, PY. (2015). Targeting dengue virus NS4B protein for drug discovery. Antiviral Res.;118:39-45. (Xie *et al., 2015*)