

Genomes to Hits *In Silico* - A Country Path Today, A Highway Tomorrow: A Case Study of Chikungunya

Anjali Soni^{1,2}, Khushhali M. Pandey³, Pratima Ray⁴ and B. Jayaram^{1,2,5,*}

¹Department of Chemistry & ²Supercomputing Facility for Bioinformatics & Computational Biology, Indian Institute of Technology, Hauz Khas, New Delhi-110016, India; ³Department of Chemical Engineering & Biotechnology, Maulana Azad National Institute of Technology, Bhopal-462051, India; ⁴Department of Pediatrics, All India Institute of Medical Sciences, New Delhi-110029, India; ⁵Kusuma School of Biological Sciences, Indian Institute of Technology, Hauz Khas, New Delhi-110016, India

Abstract: These are exciting times for bioinformaticians, computational biologists and drug designers with the genome and proteome sequences and related structural databases growing at an accelerated pace. The post-genomic era has triggered high expectations for a rapid and successful treatment of diseases. However, in this biological information rich and functional knowledge poor scenario, the challenges are indeed grand, no less than the assembly of the genome of the whole organism. These include functional annotation of genes, identification of druggable targets, prediction of three-dimensional structures of protein targets from their amino acid sequences, arriving at lead compounds for these targets followed by a transition from bench to bedside. We propose here a "Genome to Hits *In Silico*" strategy (called *Dhanvantari*) and illustrate it on Chikungunya virus (CHIKV). "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.

Keywords: Genome annotation, protein folding, docking and scoring, lead molecule, CHIKV.

1. INTRODUCTION

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. Here, we describe the progresses achieved in each of the above areas, the conceivability of a "Genome to hits" assembly line *in silico* (Fig. 1) and illustrate the approach with chikungunya virus (CHIKV).

2. BACKGROUND

We describe here the science and the software behind "Genome to Hits" assembly line which comprises six steps (Fig. 1), classifiable into three major areas of research viz. (a) genome annotation (steps 1 and 2), (b) protein tertiary structure prediction (step 3) and (c) structure based drug design (steps 4 to 6). Information available on chikungunya virus, which is taken up as an illustrative case in this study is summarized in the subsection (d).

(a). *Genome Annotation.* 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 (Table 1). 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. 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 2). 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.

Table 1. Some Typical Features Considered During Genome Annotation

Genome Annotation	
<i>Structural annotation:</i> identification of genomic elements	<i>Functional annotation:</i> assigning biological information to genomic elements
<ul style="list-style-type: none"> • ORFs and their localization • gene structure • coding regions • regulatory motifs 	<ul style="list-style-type: none"> • biochemical function • biological function • involved in regulation and interactions • control of expression

Generally for annotation purposes, homologous sequences in protein sequence databases are searched. The state of the art tool for

*Address correspondence to this author at the Supercomputing Facility for Bioinformatics & Computational Biology, Indian Institute of Technology, Hauz Khas, New Delhi-110016, India; Tel: 91-11-2659 1505; 91-11-2659 6786; Fax: 91-11-2658 2037; Emails: bjayaram@chemistry.iitd.ac.in; website: www.scfbio-iitd.res.in

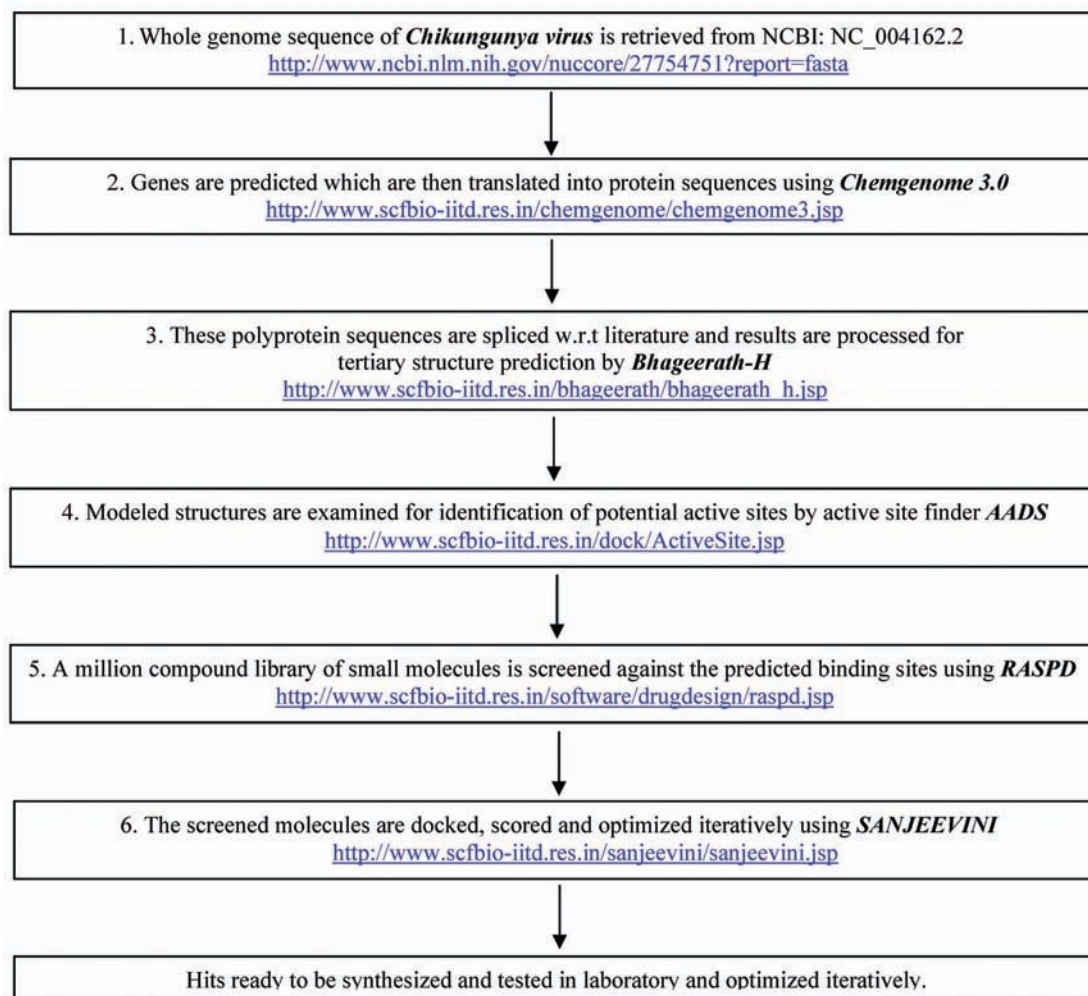


Fig. (1). Flow diagram illustrating the steps involved in *Dhanvantari* pathway to achieve hit molecules from genomic information.

such database searches is PSI-BLAST (Position Specific Iterated Basic Local Alignment Search Tool) [1, 2]. 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 [3]. 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. Physico-chemical properties such as hydrogen bonding, stacking, solvation etc. show clear signatures of the functional destiny of DNA sequences [4-8], which has formed the basis of *Chemgenome*. In the present study, we have used *Chemgenome*, the SCFBio tool (<http://www.scfbio-iitd.res.in/chemgenome/chemgenome3.jsp>) to produce and interpret structural annotations for the viral genome of *Chikungunya virus*.

(b). *Protein tertiary structure prediction*. The genome annotation is followed by protein annotation at structural, functional and at genomic scale which is essential for routine work in biology and for any systematic approach to the modeling of biological systems. To bridge the expanding sequence-structure gap, many computational approaches are becoming available which assign structure to a novel protein from its amino acid sequence. A plethora of automated methods to predict protein structure have been developed based on a variety of approaches. These include (a) homology modeling, (b) fold recognition or threading, (c) *ab initio* or *de novo* methods. 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 3). 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 [9-11].

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 [11, 12-26].

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. In this study, *Bhageerath* and *Bhageerath-H* servers are employed for protein structure prediction. *Bhageerath* is an energy based software suite for predicting tertiary structures of small globular proteins, available at <http://www.scfbio->

Table 2. List of Tools Available for Gene Prediction

Sl. No.	Softwares	URLs	Methodology
1.	FGENESH	http://linux1.softberry.com/all.htm	<i>Ab initio</i>
2.	GeneID	http://www1.imim.es/geneid.html	<i>Ab initio</i>
3.	GeneMark	http://exon.gatech.edu/GeneMark/gmchoice.html	<i>Ab initio</i>
4.	GeneMark.hmm	http://exon.gatech.edu/hmmchoice.html	<i>Ab initio</i>
5.	GeneWise	http://www.ebi.ac.uk/Tools/Wise2/	Homology
6.	GENSCAN	http://genes.mit.edu/GENSCAN.html	<i>Ab initio</i>
7.	Glimmer	http://www.tigr.org/software/glimmer/	<i>Ab initio</i>
8.	GlimmerHMM	http://www.cbcb.umd.edu/software/glimmerhmm/	<i>Ab initio</i>
9.	GRAILEXP	http://compbio.ornl.gov/grailexp	<i>Ab initio</i>
10.	GENVIEW	http://zeus2.itb.cnr.it/~webgene/wwwgene.html	<i>Ab initio</i>
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	<i>Ab initio</i>
14.	PredictGenes	http://mendel.ethz.ch:8080/Server/subsection3_1_8.html	Homology
15.	MZEF	http://rulai.cshl.edu/software/index1.htm	<i>Ab initio</i>
16.	Rosetta	http://crossspecies.lcs.mit.edu	Homology
17.	EuGéne	http://eugene.toulouse.inra.fr/	<i>Ab initio</i>
18.	PROCRUSTES	http://www.riethoven.org/BioInformer/newsletter/archives/2/procrustes.html	Homology
19.	Xpound	http://mobyline.pasteur.fr/cgi-bin/portal.py?#forms::xpound	<i>Ab initio</i>
20.	Chemgenome	http://www.scfbio-iitd.res.in/chemgenome/chemgenome3.jsp	<i>Ab initio</i>
21.	Augustus	http://augustus.gobics.de/	<i>Ab initio</i>
22.	Genome Threader	http://www.genomethreader.org/	Homology
23.	HMMgene	http://www.cbs.dtu.dk/services/HMMgene/	<i>Ab initio</i>
24.	GeneFinder	http://people.virginia.edu/~wc9c/genefinder/	<i>Ab initio</i>
25.	EGPRED	http://www.imtech.res.in/raghava/egpred/	<i>Ab initio</i>
26.	mGene	http://mgene.org/web	<i>Ab initio</i>

iitd.res.in/bhageerath/index.jsp [12, 27]. It predicts five candidates for the native, from the input query sequence. *Bhageerath-H* [28] is a hybrid (homology + *ab initio*) server for protein tertiary structure prediction [29, 30]. It identifies regions which show local sequence similarity 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. This server again predicts the best five energetically favorable structures, which are expected to be close to the native. The knowledge of tertiary structures of proteins serves as a basis for structure-based drug design.

(c). *Structure based drug design*. 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 4). In this study we use, AADS (http://www.scfbio-iitd.res.in/dock/ActiveSite_new.jsp) methodology for an automated identification of ten potential binding pockets which are expected to bracket the true “active site” (binding pocket). AADS requires the 3D structure of the target protein and detects the top 10 potential binding sites with 100% accuracy in capturing the actual binding (active) site.

Once the binding pockets on proteins are identified, libraries of small molecules are screened against these sites to identify a few hit molecules using software such as RASPD (<http://www.scfbio-iitd.res.in/software/drugdesign/rasped.jsp>). RASPD protocol is designed in the spirit of structure-based drug design approach but with a rapid turnover rate. RASPD screens small molecule databases against the active sites based on physicochemical descriptors or in general the set of Lipinski parameters such as hydrogen bond donors, hydrogen bond acceptors, molar refractivity, Wiener index and volume for the protein and drug and also the functional groups

Table 3. List of Tools Available for Protein Tertiary Structure Prediction

Sl. No	Softwares	URLs	Description
1.	CPHModels3.0	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
5.	3D-JIGSAW	http://3djigsaw.com/	Server to build three-dimensional models for proteins based on homologues of known structure
6.	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
7.	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
8.	ROBETTA	http://robeta.bakerlab.org	<i>De novo</i> Automated structure prediction analysis tool used to infer protein structural information from protein sequence data
9.	PROTINFO	http://protinfo.compbio.washington.edu/	<i>De novo</i> protein structure prediction web server utilizing simulated annealing for generation and different scoring functions for selection of final five conformers
10.	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
11.	I-TASSER	http://zhanglab.ccmb.med.umich.edu/I-TASSER/	Predicts protein 3D structures based on threading approach
12.	BHAGEERATH	http://www.scfbio-iitd.res.in/bhageerath/index.jsp	Energy based methodology for narrowing down the search space of small globular proteins
13.	BHAGEERATH-H	http://www.scfbio-iitd.res.in/bhageerath/bhageerath_h.jsp	A Homology <i>ab-initio</i> Hybrid Web server for Protein Tertiary Structure Prediction

[31-33]. The most interesting feature of RASPD is that it generates a set of hit molecules based on the complementarities of the aforementioned properties with a certain cutoff binding affinity bypassing the conventional docking and scoring strategies, which reduces the search time significantly. The libraries incorporated in RASPD are a million compound library of small molecules and a natural product library. The users can also sketch molecules of their choice or use a non-redundant dataset of small molecules NRDBSM [34] (<http://www.scfbio-iitd.res.in/software/nrdbsm/index.jsp>) and submit them for RASPD screening.

The screening is followed by atomic level docking and scoring strategies (Table 5) such as *Sanjeevini* (<http://www.scfbio-iitd.res.in/sanjeevini/sanjeevini.jsp>) to identify a few candidates which could be pursued as leads for experimental synthesis and validation [35, 36]. ParDOCK module of *Sanjeevini* is an all-atom energy based Monte Carlo algorithm for protein-ligand docking. It involves the positioning of ligands optimally with best configuration in the protein binding site and scores them based on their interaction energies. This utility is freely accessible at <http://www.scfbio-iitd.res.in/dock/pardock.jsp> [37]. ParDOCK uses BAPPL scoring function [38] for atomic level scoring of non-metallo protein ligand complexes and in ranking them accurately with their estimated free energies. BAPPL is again freely accessible at <http://www.scfbio-iitd.res.in/software/drugdesign/bappl.jsp>. The accuracy of this scoring function in predicting binding free energy is high with ± 1.02 kcal/mol average error and a correlation coefficient

of 0.92 between the predicted and experimental binding energies for 161 protein-ligand complexes. An extended version of BAPPL, i.e. BAPPL-Z can be used for the prediction of binding energies of the complexes having zinc metal ion in their active sites. BAPPL-Z utility is accessible at <http://www.scfbio-iitd.res.in/software/drugdesign/bapplz.jsp> [39]. All these tools are collectively gathered in *Sanjeevini* software, which is a complete drug design software suite, freely accessible at (<http://www.scfbio-iitd.res.in/sanjeevini/sanjeevini.jsp>) [34, 40-47]. Thus, the assessment of candidate molecules is done based on their binding energies and the molecules identified as good binders to the target are considered further for synthesis and testing.

(d). *Chikungunya Virus*. Chikungunya fever (CHIK) is a mosquito (*Aedes aegypti*) borne devastating disease caused by Chikungunya virus (CHIKV), an alphavirus belonging to the family *Togaviridae*. It is one of the most important re-emerging infectious diseases in Africa and Asia with sporadic intervals and is responsible for significant global impact on public health problems [48-62]. CHIKV is listed as a category C pathogen in 2008 by National Institute of Allergy and Infectious Diseases (NIAID) and as a biosafety level 3 (BSL3) pathogen [50, 63-66]. CHIKV causes debilitating and prolonged arthralgic syndrome incapacitating the affected population for longer periods. CHIKV is usually found in tropics but has widespread across the globe in recent years due to a range of transmission vectors, globalization and climatic changes [67-111]. The 'Chikungunya' word has originated from the Ma-

Table 4. List of Software Available for Active Site Prediction

S.No	Software	URL	Description
1	SitesIdentify	http://www.manchester.ac.uk/bioinformatics/sitesidentify/	Sequence and geometry based
2	PAR-3D	http://sunserver.cdfd.org.in:8080/tease/3D/index.html	Structure based
3	FUZZY-OIL-DROP	http://www.bioinformatics.cm-uj.krakow.pl/activesite/	Fuzzy oil drop 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 ^{CSC}	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://bcf.med.usherbrooke.ca/isocleftfinder.php	graph-matching-based method
17	metaPocket	http://sysbio.zju.edu.cn/metapocket/	Structure based
18	LIGSITE ^{CS}	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

konde root verb kungunyala, meaning “that which bends up” [112, 113] which is in reference to drying up or becoming contorted and signifies the cause of stooped posture developed due to the excruciating joint and muscle pain and other rheumatologic manifestations [114, 115]. The disease etiology consists of sudden onset of fever with arthralgia, which generally resolves within a few days [116, 117].

Female mosquitoes acquire the virus by taking blood from viremic vertebrate hosts (Fig. 2). The virus elicits a persistent infection and replicates at a high pace, especially in the salivary glands of the insects [118, 119]. In addition to salivary glands, it replicates in various other organs inside body cavity including gut, ovary, neural tissue, body fat etc. [120]. When this CHIKV loaded mosquito infects a healthy human, it transfers the virus into its blood stream. These virions through interaction with the receptors reach the target cells by endocytosis. The acidic environment of the endosome triggers conformational changes in the viral envelope that expose the E1 peptide [121-125], which mediates virus-host cell membrane fusion. This allows cytoplasmic delivery of the core and release of the viral genome in cytoplasm. The site of mRNA transcription is in the cell cytoplasm.

CHIKV is an enveloped, spherical bodied virus of about 70nm in diameter. The virion genome consists of a linear single-stranded (ss), positive-sense RNA molecule of approximately 11.8 kb length,

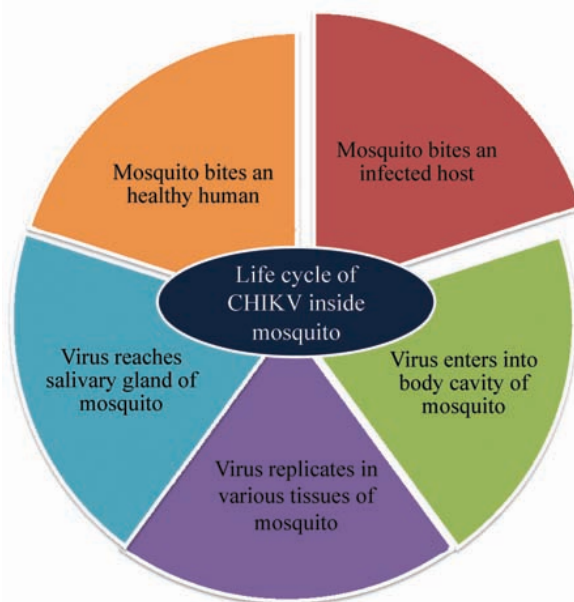


Fig. (2). Flow diagram depicting the movement of CHIKV from viremic host to mosquito and from mosquito to healthy host.

Table 5. A list of Softwares for Drug Design

Sl. No.	Softwares	URL	Description
1	Discovery studio	http://accelrys.com/products/discovery-studio/structure-based-design.html	Molecular modeling and <i>de novo</i> 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	http://autodock.scripps.edu/	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

where the 5' end is capped with a 7-methylguanosine while the 3' end is polyadenylated. The CHIKV genome is comprised of 30% A, 25% C, 25% G and 20% T (U) base pairs with two long open reading frames (ORF) that encode the non-structural (2474 amino acids) and structural polyproteins (1244 amino acids) [126-131].

The genomic organization of CHIKV is considered to be 5' cap-nsP1-nsP2-nsP3-nsP4-(junction)-C-E3-E2-6K-E1-Poly (A)3' (Fig. 3). The non-structural polyproteins (nsP1-4) located in an ORF of 7425 nucleotides get initiated by a start codon at position 77-79 and terminated by a stop codon at position 7499-7501. This polyprotein is autocatalytically cleaved to produce nonstructural proteins nsP1, nsP2, nsP3 and nsP4. In contrast, the structural polyproteins are located on an ORF of 3735 nucleotides with a start codon at position 7567-7569 and a stop codon at position 11299-11313. Likewise, this polyprotein is cleaved to produce the structural proteins namely the capsid protein (C), the glycoproteins E1, E2 and E3 and 6K [126, 132-136]. The polypeptides are cleaved into active proteins by viral and cellular proteases [137-148]. The functional properties of the active cleaved proteins are summarized in (Table 6).

Although no specific drugs are available, CHIK is usually treated with non-steroidal anti-inflammatory drugs (NSAIDs), with inconsistent success [149-172] (Table 7). Owing to the non-availability of a potential drug to cure the disease, there is an urgent need to adopt a skilled strategy to develop new therapeutics. We describe in the following section how computational approaches can help in reducing the time in arriving at potential lead molecules.

3. CALCULATIONS & RESULTS: APPLICATION OF THE G2H ASSEMBLY LINE TO CHIKV

The genome sequence of Chikungunya virus was retrieved from NCBI (http://www.ncbi.nlm.nih.gov/nucleotide/NC_004162). For gene prediction, the sequence was processed using *ChemGenome 3.0* (<http://www.scfbio-iitd.res.in/chemgenome/chemgenome3.jsp>) software [5, 6]. The results displayed the existence of two genes which were similar to the already published ones, essentially implying that in this case, 100% accuracy is achieved with *ChemGenome 3.0*. These nucleotide sequences were translated to protein sequences by *ChemGenome 3.0*. The proteins in CHIKV are polyproteins i.e. the sequence displayed in results contains sequences for all proteins coded by the gene. The individual proteins from polyprotein are cleaved during post translational processing. Till date no reliable computational approach is available to cleave the polyproteins, therefore the sequences were dissected manually for each protein, based on literature and experimental evidence to identify cleavage site. The *ChemGenome 3.0* results are archived at <http://www.scfbio-iitd.res.in/software/chemgenomeresult.jsp>.

The sequences extracted from *ChemGenome 3.0* served as inputs to *Bhageerath-H* (http://www.scfbio-iitd.res.in/bhageerath/bhageerath_h.jsp), a tertiary structure prediction server [28]. For each submitted sequence, five structures were returned by the server. The results received from *Bhageerath-H* are shown in (Fig. 4). As no homolog information is available to give strength to these structural models, all the five structures are considered as plausible candidates for the native, and considered for

Table 6. Functional Properties of Structural and Non-structural Proteins Found in *Chikungunya*

Protein Type	Proteins	Functions
NonStructural Proteins NP_690588.1	nsP1	❖ Viral methyl transferase domain (acts as cytoplasmic capping enzyme and transfers 7-methyl-GMP complex to mRNA, thus forming the cap structure)
	nsP2	❖ Viral RNA helicase domain and RNA trisphosphatase (part of the RNA polymerase complex) ❖ Peptidase C9 domain (cleaves four mature proteins from non structural polyprotein)
	nsP3	❖ Processing domain (crucial for minus strand and subgenomic 26S mRNA synthesis)
	nsP4	❖ Viral RNA dependent RNA polymerase domain (replicates genomic and antigenomic RNA and also transcribes 26S subgenomic mRNA which encodes for structural proteins)
Structural proteins NP_690589.2	C	❖ Peptidase_S3 domain (autocatalytic cleavage) ❖ Trypsin like serine protease domain
	E3	❖ Alphavirus E3 spike glycoprotein domain (tentative)
	E2	❖ Alphavirus E2 glycoprotein domain (virus attachment to host) ❖ Transmembrane domain
	6K	❖ Alphavirus E1 glycoprotein domain (virus glycoprotein processing and membrane permeabilization) ❖ Signal peptide domain ❖ Transmembrane domain
	E1	❖ Alphavirus E1 glycoprotein domain (class II viral fusion protein) ❖ Glycoprotein E dimerization domain (forms E1-E2 heterodimer in inactive state and E1 trimerizes in active state) ❖ Immunoglobulin E set domain ❖ Transmembrane domain

Table 7. A list of Drugs Available for Treating *Chikungunya* Fever

Drug	Category	Description
Chloroquine	Antirheumatic Agents / Antimalarials / Amebicides	It is believed to inhibit the heme polymerase activity
Aspirin	Anticoagulants / cyclooxygenase(COX) Inhibitors / PlateletAggregation Inhibitors	Irreversibly inhibits the activity of both types of cyclooxygenase (COX-1 and COX-2)
Ibuprofen	Anti-inflammatory Agents / COX Inhibitors / Analgesics / Nonsteroidal Anti-inflammatory Agents (NSAIAs)	A non-selective inhibitor of cyclooxygenase, an enzyme involved in prostaglandin synthesis via the arachidonic acid pathway
Naproxen	COX Inhibitors / Gout Suppressants	It is believed to be associated with the inhibition of cyclooxygenase activity
Ribavirin	Antiviral Agents / Antimetabolites	A potent competitive inhibitor of inosine monophosphate (IMP) dehydrogenase, viral RNA polymerase and messenger RNA (mRNA) guanylyl transferase (viral); may get incorporated into RNA in RNA viral species.
Prednisolone	Hormonal Glucocorticoids	The antiinflammatory actions of glucocorticoids are thought to involve phospholipase A2 inhibitory proteins, lipocortins
Acetaminophen	Analgesics, Non-Narcotic / Antipyretics	Inhibits various forms of cyclooxygenase, COX-1, COX-2, and COX-3 enzymes

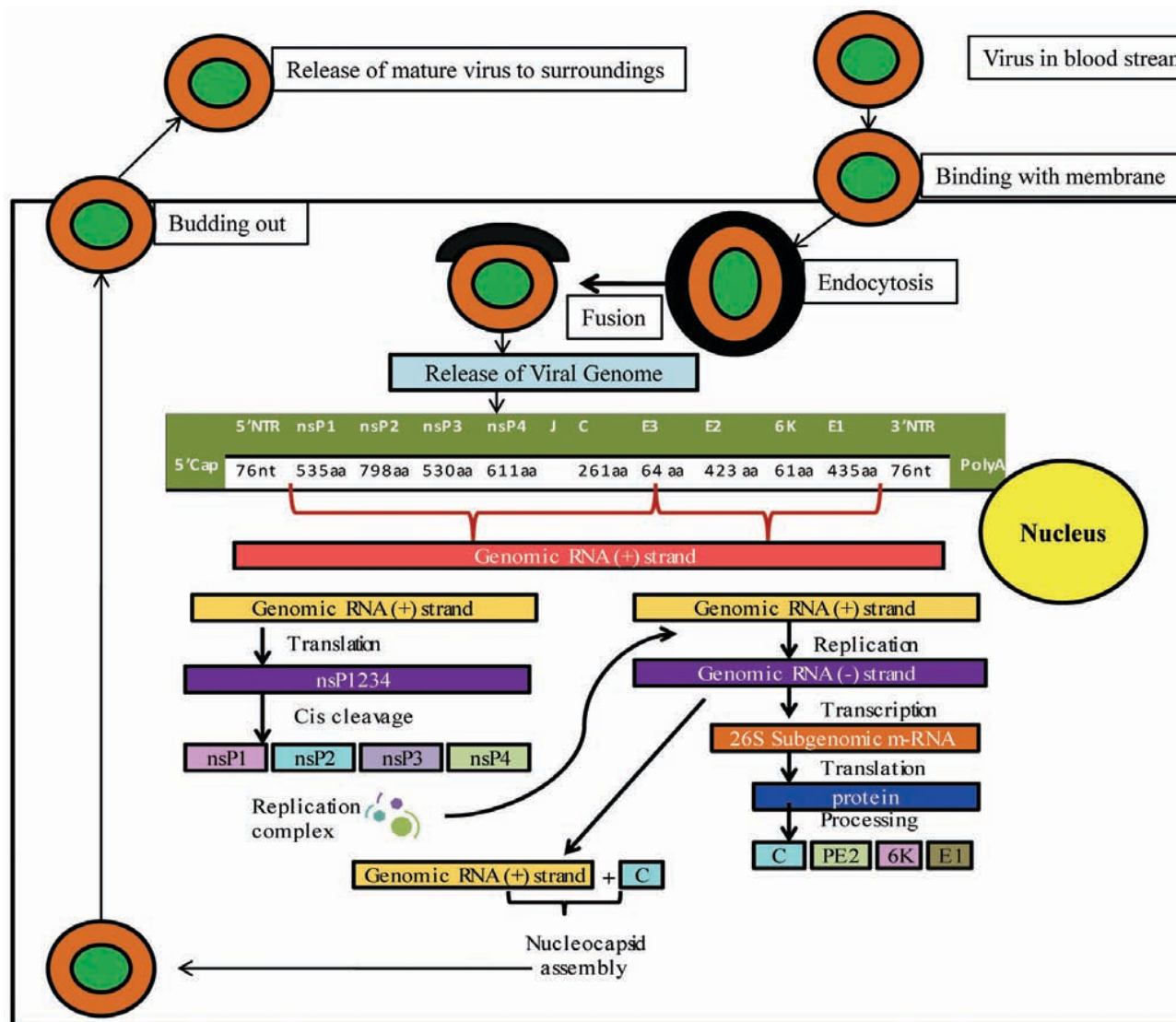


Fig. (3). The lifecycle of CHIKV in the host cell.

further studies. It may be noted that tertiary structure prediction of structural proteins associated with membranes is a nascent area with low success rate at this stage and hence the focus here has been on nonstructural proteins which can fold autonomously.

Most of the experimentally determined structures have some information of ligand binding domain/site but in the present scenario, CHIKV proteins lack the structural information, thus necessitating detection of ligand binding sites (active sites). In order to facilitate active site detection, an automated version of active site finder i.e. AADS (Automated active site docking and scoring) (http://www.scfbio-iitd.res.in/dock/ActiveSite_new.jsp) is utilized which predicts the potential binding site(s) and further performs the docking of the selected molecule to the top ten cavities in an automated mode [40]. Binding sites on each of the five structural models of each nonstructural protein are identified. Not all cavities determined by the active site identifier may be true binding sites with functional implication but one among them is very likely to be such a site. The additional cavities can be checked for their ability to act

as allosteric sites. The predicted top 10 binding sites are shown as black dots in the protein structures (Fig. 4).

In search of probable hits, the 10 cavities per structure identified by AADS are further subjected to RASPD (Rapid screening of preliminary drugs) (<http://www.scfbio-iitd.res.in/software/drugdesign/raspd.jsp>) software [41]. The RASPD returned more than 500 molecules against the predicted cavities of CHIKV proteins with -8.00 kcal/mol as the binding energy cutoff.

The *in silico* drug design beyond this stage involves rigorous docking and scoring [173, 174]. The hits identified from screening via RASPD above are further docked with their respective target site using *Sanjeevini* software (<http://www.scfbio-iitd.res.in/sanjeevini/sanjeevini.jsp>) which utilizes ParDOCK as a docking tool. For all the modeled structures, one molecule for each cavity has been proposed on SCFBio's CHIKV webpage which is accessible at <http://www.scfbio-iitd.res.in/software/chikv.jsp>. This webpage contains information on the genome annotation, protein

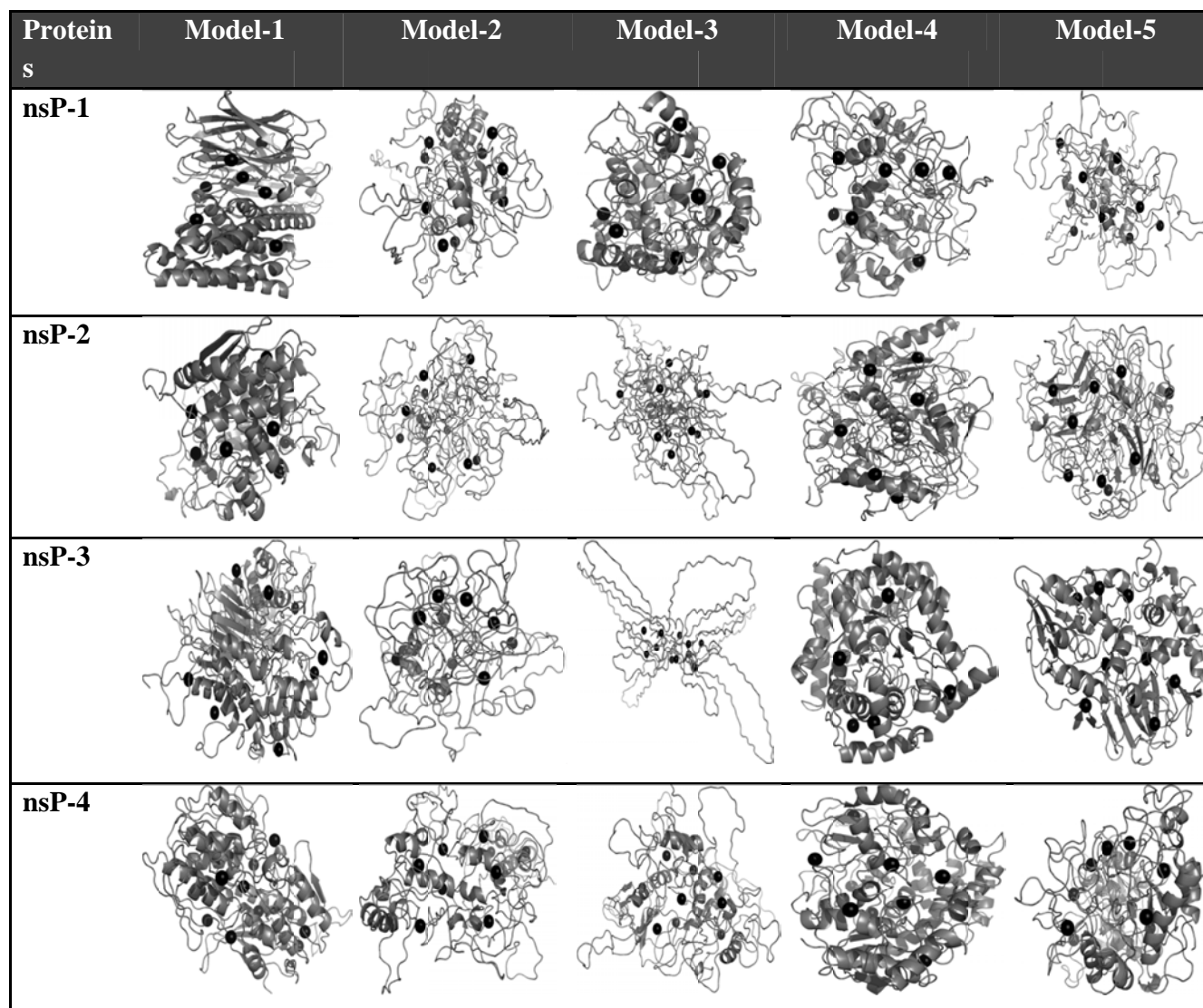


Fig. (4). An illustration of the protein structures of CHIKV predicted by *Bhageerath-H* shown along with their binding pockets.

tertiary structure prediction, and hit molecule identification and docking and scoring results of the complete genome to hit protocol.

The best 20 molecules selected against the nonstructural proteins of CHIKV are displayed in (Table 8). From here on, the *in silico* strategies go hand-in-hand with experimentation. In an iterative process of synthesis, testing, modification, docking and scoring, these molecules can be further improved to yield candidate drugs while taking care of the ADMET profiles [175-180].

4. DISCUSSION ON THE G2H ASSEMBLY LINE

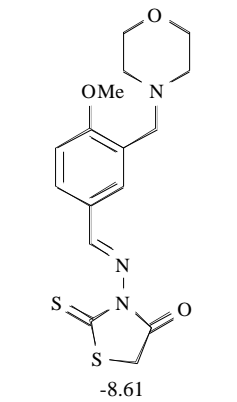
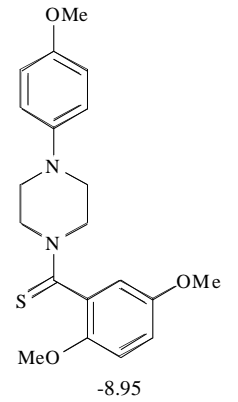
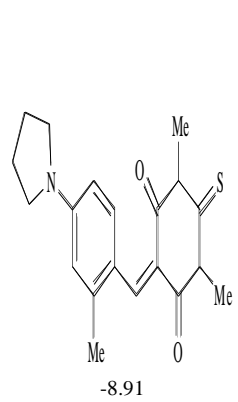
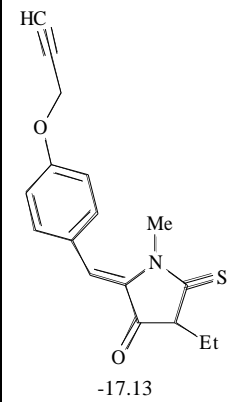
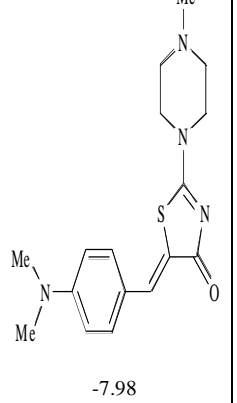
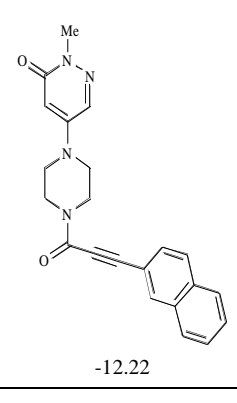
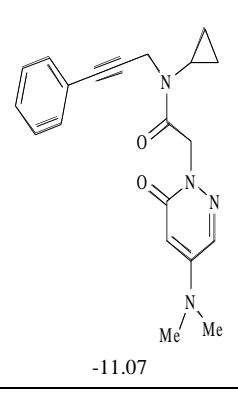
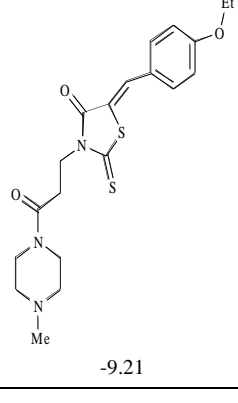
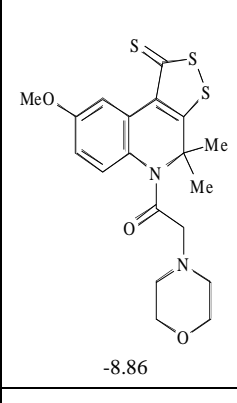
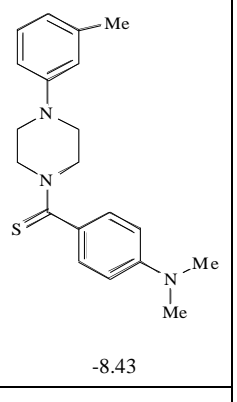
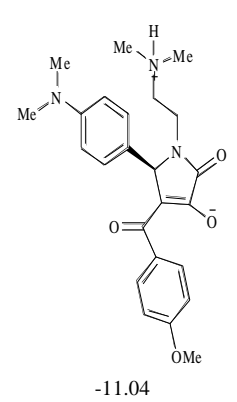
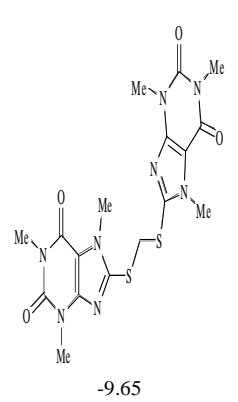
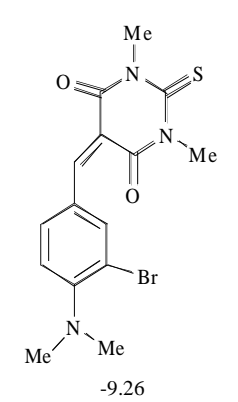
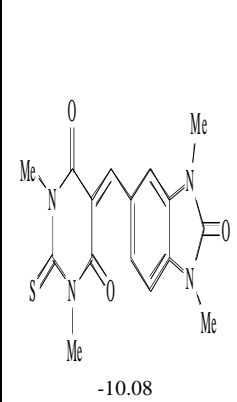
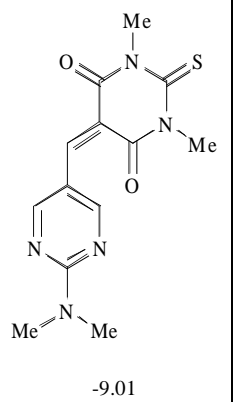
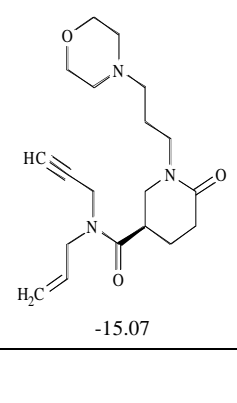
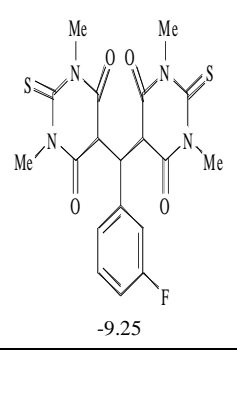
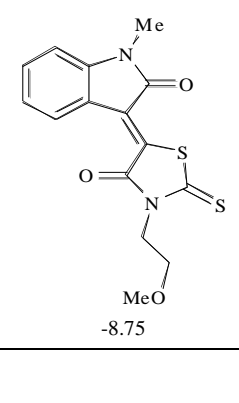
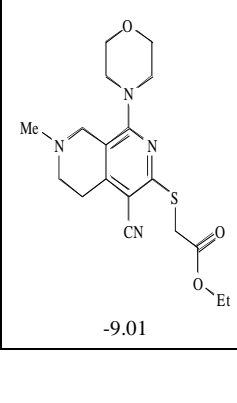
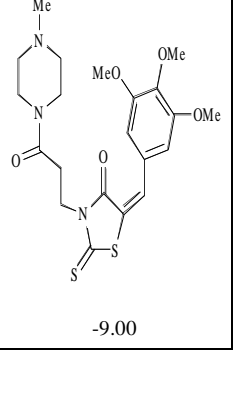
The wealth of information available from experimental host-pathogen interaction studies invites computational biologists to develop databases and newer computational methods to advance further focused experimentation. Consequently, bioinformatics is rapidly evolving into independent fields addressing 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.

A basic overview of the G2H technology is given in this review with an application to Chikungunya virus. G2H 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.

5. CONCLUSION & PERSPECTIVES

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.

Table 8. Structural Representations of 20 Molecules Showing high Affinity to the Nonstructural Proteins of CHIKV. (Computed Binding Energies are also Shown in kcal/mol Underneath Each Molecule)

Protein	Model-1	Model-2	Model-3	Model-4	Model-5
nsP1	 -8.61	 -8.95	 -8.91	 -17.13	 -7.98
nsP2	 -12.22	 -11.07	 -9.21	 -8.86	 -8.43
nsP3	 -11.04	 -9.65	 -9.26	 -10.08	 -9.01
nsP4	 -15.07	 -9.25	 -8.75	 -9.01	 -9.00

CONFLICT OF INTEREST

The authors confirm that this article content has no conflicts of interest.

ACKNOWLEDGEMENTS

This project is funded by the programme support to SCFBio from the Department of Biotechnology, Govt. of India. The authors gratefully acknowledge the help received from Ms. Garima Khandelwal, Ms. Priyanka Dhingra, Ms. Tanya Singh, Mr. Goutam Mukherjee, Mr. Avinash Mishra, Mr. Shashank Shekhar and Ms. Vandana Shekhar. The authors thank Dr. Aditya Mittal for useful discussions and a critical reading of the manuscript.

SUPPLEMENTARY INFORMATION ON CHIKUNGUNYA AT SCFBIO WEBSITE

Details of the results on genes and protein tertiary structures predicted, binding pockets, hit molecules identified and lead molecules proposed for synthesis are available for free download from the SCFBio website (<http://www.scfbio-iitd.res.in/software/chikv.jsp>). These results will be updated periodically with improvements in protocols for protein structure prediction and ADMET evaluations.

REFERENCES

- [1] Altschul SF, Madden TL, Schäffer AA, *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997; 25: 3389-402.
- [2] Li Y, Chia N, Lauria M, Bundschuh R. A performance enhanced PSI-BLAST based on hybrid alignment. *Bioinformatics* 2011; 27: 31-7.
- [3] Standley DM, Toh H, Nakamura H. Functional annotation by sequence-weighted structure alignments: statistical analysis and case studies from the Protein 3000 structural genomics project in Japan. *Proteins* 2008; 72: 1333-51.
- [4] Khandelwal G, Jayaram B. A phenomenological model for predicting melting temperatures of DNA sequences. *PLoS ONE* 2010; 5: e12433.
- [5] Dutta S, Singhal P, Agrawal P, *et al.* A physicochemical model for analyzing DNA sequences. *J Chem Inf Model* 2006; 46: 78-85.
- [6] Singhal P, Jayaram B, Dixit SB, Beveridge DL. Prokaryotic gene finding based on physicochemical characteristics of codons calculated from molecular dynamics simulations. *Biophys J* 2008; 94: 4173-83.
- [7] Khandelwal G, Jayaram B. DNA-water interactions distinguish messenger RNA genes from transfer RNA genes. *J. Am. Chem. Soc* 2012; 134: 8814-16.
- [8] Khandelwal G, Gupta J, Jayaram B. DNA energetics based analyses suggest additional genes in prokaryotes. *J Bio Sc* 2012; 37: 433-44.
- [9] Shenoy SR, Jayaram B. Proteins: Sequence to Structure and Function-Current Status. *Curr Protein Pept Sci* 2010; 11: 498-514.
- [10] Jayaram B, Dhingra P. Towards creating complete proteomic structural databases of whole organisms. *Curr Bioinform* 2012; 7: 424-35.
- [11] Leopold PE, Montal M, Onuchic JN. Protein Folding Funnels - a Kinetic Approach to the Sequence Structure Relationship. *Proc Natl Acad Sci USA* 1992; 89: 8721.25.
- [12] Jayaram B, Dhingra P, Lakhani B, Shekhar S. Bhageerath-targeting the near impossible: pushing the frontiers of atomic models for protein tertiary structure prediction. *J Chem Sci* 2012; 124: 83-91. doi:10.1007/s12039-011-0189-x
- [13] Narang P, Bhushan K, Bose S, Jayaram B. A computational pathway for bracketing native-like structures for small alpha helical globular proteins. *Phys Chem Chem Phys* 2005; 7: 2364-75.
- [14] Narang P, Bhushan K, Bose S, Jayaram B. Protein structure evaluation using an all-atom energy based empirical scoring function. *J Biomol Struct Dyn* 2006; 23: 385-406.
- [15] Arora N, Jayaram B. The Strength of Hydrogen Bonds in Alpha Helices. *J Comput Chem* 1997; 18: 1245-52.
- [16] Surjit B. Dixit, R. Bhasin, E. Rajasekaran and B. Jayaram. Solvation Thermodynamics of Amino Acids: Assessment of the Electrostatic Contribution and Force Field Dependence. *J. Chem. Soc. Farad. Trans* 1997; 93: 1105-13.
- [17] Mittal A, Jayaram B, Shenoy S, Bawa TS. A Stoichiometry driven universal spatial organization of backbones of folded proteins: Are there Chargaff's rules for protein folding?. *J Biomol Struct Dyn* 2010; 28: 133-42.
- [18] Mittal A, Jayaram B. Backbones of folded proteins reveal novel invariant amino-acid neighborhoods. *J Biomol Struct Dyn* 2011; 28: 443-54.
- [19] Mittal A, Jayaram B. The Newest View on Protein Folding: Stoichiometric and Spatial Unity in Structural and Functional Diversity. *J Biomol Struct Dyn* 2011; 28: 669-74.
- [20] Dill KA, Ozkan SB, Weikl TR, Chodera JD, Voelz VA. The protein folding problem: when will it be solved? *Curr Opin Struct Biol* 2007; 17: 342.6.
- [21] Kim DE, Blum B, Bradley P, Baker D. Sampling Bottlenecks in De novo Protein Structure Prediction. *J Mol Bio* 2009; 393: 249.60.
- [22] Dill KA, Chan HS. From Levinthal to pathways to funnels. *Nat Struct Biol* 1997; 4: 10.19.
- [23] Lindorff-Larsen K, Piana S, Palmo K, *et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* 2010; 78: 1950-8.
- [24] Berka K, Laskowski RA, Hobza P, Vondrasek J. Energy Matrix of Structurally Important Side-Chain/Side-Chain Interactions in Proteins. *J. Chem. Theory Comput* 2010; 6: 2191.203.
- [25] Cooper S, Khatib F, Treuille A, *et al.* Predicting protein structures with a multiplayer online game. *Nature* 2010; 466: 756.60.
- [26] Faver JC, Benson ML, He X, *et al.* The energy computation paradox and ab initio protein folding. *PLoS One* 2011; 6: e18868.
- [27] Jayaram B, Bhushan K, Shenoy SR, *et al.* Bhageerath: an energy based web enabled computer software suite for limiting the search space of tertiary structures of small globular proteins. *Nucleic Acids Res* 2006; 34: 6195-204.
- [28] Dhingra P, Mishra A, Rao S, Jayaram B. Bhageerath-H: A Homology ab initio Hybrid Webserver for Protein Tertiary Structure Prediction (Manuscript in preparation).
- [29] Mishra A, Rao S, Mittal A, Jayaram B. Capturing native protein structures with a physico-chemical metric (pcSM), *BBA-Proteins & proteomics*, 2013, under revision.
- [30] Dhingra P, Lakhani B, Jayaram B. Generating native-like structures of soluble monomeric proteins via Bhageerath-H - a homology/ab initio hybrid method. *J Comput Chem* 2013 Under revision.
- [31] Lipinski CA. Lead- and drug-like compounds: the rule-of-five revolution. *Drug Discov Today: Tech* 2004; 1: 337-41.
- [32] Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* 1997; 46: 3-26.
- [33] Lipinski CA. Drug-like properties and the causes of poor solubility and poor permeability. *J Pharmacol Toxicol Methods* 2000; 44: 235-49.
- [34] (a) Shaikh SA, Jain T, Sandhu G, Latha N, Jayaram B. From drug target to leads-sketching a physicochemical pathway for lead molecule design in Silico. *Curr Pharm Des* 2007; 13: 3454-70. (b) Shaikh SA, Jain T, Sandhu G, Soni A, Jayaram B. From drug target to leads-sketching a physicochemical pathway for lead molecule design in Silico. *Frontiers in Medicinal Chemistry*, 2012, 6, 324-360. Eds: Atta ur Rahman, Allen B. Reitz and M. Iqbal Choudhary, Bentham Publishers. doi:10.2174/97816080546401120601
- [35] Jayaram B, Dhingra P, Mukherjee G, Perumal V. "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", 2012, Chapter 3, pages 13-35.
- [36] Das A, Kalra P, Latha N, Jayaram B. In Silico Trends in Thermodynamics and Kinetics of Binding - New Tools for De Novo Drug Design, in *Recent Trends in Chemistry*, Editors, Dr. M. M. Srivastava, Shalini Srivastava, 2002 Chapter-15, pp 218-234.
- [37] Gupta A, Gandhimathi A, Sharma P, Jayaram B. ParDOCK: An All Atom Energy Based Monte Carlo Docking Protocol for Protein-Ligand Complexes. *Protein Pept Lett* 2007; 14: 632-46.
- [38] Jain T, Jayaram B. An all atom energy based computational protocol for predicting binding affinities of protein.ligand complexes. *FEBS Lett* 2005; 579: 6659-66.

- [39] Jain T, Jayaram B. Computational Protocol for Predicting the Binding Affinities of Zinc Containing Metalloprotein-Ligand Complexes. *Proteins* 2007; 67: 1167-78.
- [40] Singh T, Biswas D, Jayaram B. AADS-an automated active site identification, docking and scoring protocol for protein targets based on physico-chemical descriptors. *J Chem Inf Model* 2011; 51: 2515-27.
- [41] Mukherjee G, Jayaram B. A rapid scoring methodology based on physico-chemical descriptors of small molecules (RASPD) for identifying hits against a protein target. *Phys Chem Chem Phys* 2013; DOI:10.1039/C3CP44697B.
- [42] Shaikh SA, Ahmed SR, Jayaram B. A molecular thermodynamic view of DNA-drug interaction: A case study of 25 minor groove binders. *Arch Biochem Biophys* 2004; 429: 81-99.
- [43] Shaikh SA, Jayaram B. A Swift All-Atom Energy-Based Computational Protocol to Predict DNA-Ligand Binding Affinity and ΔT_m . *J Med Chem* 2007; 50: 2240-4.
- [44] Jayaram B, Latha N, Jain T, Sharma P, Gandhimathi A, Pandey VS. Sanjeevini: A comprehensive active site directed lead design software. *Indian J Chem A* 2006; 45: 1834-37.
- [45] Jayaram B, Singh T, Mukherjee G, Mathur A, Shekhar S, and Shekhar V. Sanjeevini: A freely accessible web-server for target directed lead molecule discovery. *BMC Bioinformatics* 2012; 13: S7. doi:10.1186/1471-2105-13-S17-S7
- [46] Kalra P, Reddy TV, Jayaram B. Free Energy Component Analysis for Drug Design: A Case Study of HIV-1 Protease Inhibitor Binding. *J Med Chem* 2001; 44: 4325-38.
- [47] Mukherjee G, Patra N, Barua P, Jayaram B. A Fast Empirical GAFF Compatible Partial Atomic Charge Assignment Scheme for Modeling Interactions of Small Molecules with Biomolecular Targets. *J Comput Chem* 2011; 32: 893-907.
- [48] Powers AM, Logue CH. Changing patterns of chikungunya virus: re-emergence of a zoonotic arbovirus. *J Gen Virol* 2007; 88: 2363-77.
- [49] Gubler DJ. Human arbovirus infections worldwide. *Ann N Y Acad Sci* 2001; 951: 13-24.
- [50] Thiboutot MM, Kannan S, Kawalekar OU, *et al.* Chikungunya: A Potentially Emerging Epidemic? *PLoS Negl Trop Dis* 2010; 4: e623.
- [51] Robinson MC. An epidemic of virus disease in Southern Province, Tanganyika Territory, in 1952-53. I. Clinical features. *Trans R Soc Trop Med Hyg* 1955; 49: 28-32.
- [52] Sumathy K, Ella KM. Genetic diversity of chikungunya virus, India 2006-2010: Evolutionary dynamics and serotype analyses. *J Med Virol* 2012; 84: 462-70.
- [53] Anyamba A, Linthicum KJ, Small JL *et al.* Climate teleconnections and recent patterns of human and animal disease outbreaks. *PLoS Negl Trop Dis* 2012; 6: e1465.
- [54] Ross RW. The Newala epidemic. III. The virus: isolation, pathogenic properties and relationship to the epidemic. *J Hyg* 1956; 54: 177-91.
- [55] Lumsden WH. An epidemic of virus disease in Southern Province, Tanganyika Territory, in 1952-53. II. General description and epidemiology. *Trans R Soc Trop Med Hyg* 1955; 49: 33-57.
- [56] Powers AM, Brault AC, Tesh RB, Weaver SC. Re-emergence of chikungunya and o'nyong-nyong viruses: evidence for distinct geographical lineages and distant evolutionary relationships. *J Gen Virol* 2000; 81: 471-9.
- [57] Arankalle VA, Shrivastava S, Cherian S, *et al.* Genetic divergence of Chikungunya viruses in India (1963-2006) with special reference to the 2005-2006 explosive epidemic. *J Gen Virol* 2007; 88: 1967-76.
- [58] Yergolkar PN, Tandale BV, Arankalle VA, *et al.* Chikungunya outbreaks caused by African genotype, India. *Emerg Infect Dis* 2006; 12:1580-3.
- [59] Ravi V. Re-emergence of chikungunya virus in India. *Indian J Med Microbiol* 2006; 24: 83-4.
- [60] Lahariya C, Pradhan SK. Emergence of chikungunya virus in Indian subcontinent after 32 years: A review. *J Vector Borne Dis* 2006; 43:151-60.
- [61] Tssetsarkin KA, McGee CE, Volk SM, Vanlandingham DL, Weaver SC, Higgs S. Epistatic roles of E2 glycoprotein mutations in adaptation of chikungunya virus to *Aedes albopictus* and *Ae. aegypti* mosquitoes. *PLoS One* 2009; 4: e6835.
- [62] Sam IC, Chan YF, Chan SY, *et al.* Chikungunya virus of Asian and Central/East African genotypes in Malaysia. *J Clin Virol* 2009; 46:180-3.
- [63] Staples JE, Breiman RF, Powers AM. Chikungunya fever: an epidemiological review of a re-emerging infectious disease. *Clin Infect Dis* 2009; 49: 942-8.
- [64] Vassil St. Georgiev. *Emerging and Re-emerging Infectious Diseases*. National Institute of Allergy and Infectious Diseases, NIH Infectious Disease, 2009, Part I, 23-28, DOI: 10.1007/978-1-60327-297-1_4.
- [65] Department of Health and Human Services; Centers for Disease Control and Prevention; National Institutes of Health. *Biosafety in Microbiological and Biomedical Laboratories (BMBL) 5th Edition*. Washington DC: US Government Printing Office; 2007. *Arboviruses and Related Zoonotic Viruses*.
- [66] World Health Organization. Communicable Diseases Branch. *Chikungunya fever, laboratory diagnosis of Chikungunya fevers*. Geneva: World Health Organization 2007.
- [67] Rezza G. Re-emergence of Chikungunya and other scourges: the role of globalization and climate change. *Ann Ist Super Sanita* 2008; 44: 315-8.
- [68] Epstein PR. Chikungunya fever resurgence and global warming. *Am J Trop Med Hyg* 2007; 76: 403-4.
- [69] Tabachnick WJ. Challenges in predicting climate and environmental effects on vector-borne disease epistemics in a changing world. *J Exp Biol* 2010; 213: 946-54.
- [70] Stock I. Chikungunya fever--expanded distribution of a re-emerging tropical infectious disease. *Med Monatsschr Pharm* 2009; 32: 17-26.
- [71] Gould EA, Gallian P, De Lamballerie X, Charrel RN. First cases of autochthonous dengue fever and chikungunya fever in France: from bad dream to reality! *Clin Microbiol Infect* 2010; 16: 1702-4.
- [72] Weaver SC, Reisen WK. Present and future arboviral threats. *Antiviral Res* 2010; 85: 328-45.
- [73] Moulay D, Aziz-Alaoui MA, Cadivel M. The chikungunya disease: modeling, vector and transmission global dynamics. *Math Biosci* 2011; 229: 50-63.
- [74] Lam SK, Chua KB, Hooi PS *et al.* Chikungunya infection - emerging disease in Malaysia. *Southeast Asian J Trop Med Public Health* 2001; 32: 447-51.
- [75] Laras K, Sukri NC, Larasati RP, *et al.* Tracking the re-emergence of epidemic chikungunya virus in Indonesia. *Trans R Soc Trop Med Hyg* 2005; 99: 128-41.
- [76] Mathew T, Tiruvengadam KV. Further studies on the isolate of Chikungunya from the Indian repatriates of Burma. *Indian J Med Res* 1973; 61: 517-20.
- [77] Pastorino B, Muyembe-Tamfum JJ, Bessaud M, *et al.* Epidemic resurgence of Chikungunya virus in Democratic Republic of the Congo: identification of a new central African strain. *J Med Virol* 2004; 74: 277-82.
- [78] Rezza G, Nicoletti L, Angelini K, *et al.* Infection with chikungunya virus in Italy: an outbreak in a temperate region. *Lancet* 2007; 370: 1840-6.
- [79] Borgherini G, Poubreau P, Staikowsky F, *et al.* Outbreak of chikungunya on Reunion Island: early clinical and laboratory features in 157 adult patients. *Clin Infect Dis* 2007; 44: 1401-7.
- [80] Lemant J, Boisson V, Winer A, *et al.* Serious acute chikungunya virus infection requiring intensive care during the Reunion Island outbreak in 2005-2006. *Crit Care Med* 2008; 36: 2536-41.
- [81] Selvavinayagam TS. Chikungunya fever outbreak in Vellore, South India. *Indian J Community Med* 2007; 32: 286-7.
- [82] Dwibedi B, Sabat J, Mahapatra N, Kar SK, Kerketta AS, *et al.* Rapid spread of chikungunya virus infection in Orissa: India. *Indian J Med Res* 2011; 133: 316-21.
- [83] Dutta P, Khan SA, Khan AM, Borah J, Chowdhury P, Mahanta J. First evidence of chikungunya virus infection in Assam, Northeast India. *Trans R Soc Trop Med Hyg* 2011; 105: 355-7.
- [84] Reiter P, Fontenille D, Paupy C. *Aedes albopictus* as an epidemic vector of chikungunya virus: another emerging problem? *Lancet Infect Dis* 2006; 6: 463-4.
- [85] H. Townson, M.B. Nathan. Resurgence of Chikungunya. *Trans R Soc Trop Med Hyg* 2008; 102: 308-9.
- [86] Srikanth P, Sarangan G, Mallilankaraman K, *et al.* Molecular characterization of Chikungunya virus during an outbreak in South India. *Indian J Med Microbiol* 2010; 28: 299-302.

- [87] Santhosh SR, Dash PK, Parida M, Khan M, Rao PV. Appearance of E1: A226V mutant Chikungunya virus in coastal Karnataka, India during 2008 outbreak. *Virol J* 2009; 6: 172.
- [88] Kannan M, Rajendran R, Sunish IP, *et al.* A study on Chikungunya outbreak during 2007 in Kerala, South India. *Indian J Med Res* 2009; 129: 311-5.
- [89] Schuffenecker I, Iteanu I, Michault A, *et al.* Genome microevolution of chikungunya viruses causing the Indian Ocean outbreak. *PLoS Med* 2006; 3: e263.
- [90] Manimunda SP, Sugunan AP, Rai SK, *et al.* Outbreak of Chikungunya Fever, Dakshina Kannada District, South India. *Am J Trop Med Hyg* 2010; 83: 751-54.
- [91] Ray P, Ratagiri VH, Kabra SK, *et al.* Chikungunya Infection in India: Results of a prospective hospital based multi-centric study. *PLoS ONE* 2012; 7(2): e30025. doi:10.1371/journal.pone.0030025.
- [92] Chem YK, Zainah S, Berendam SJ, Rogayah TA, Khairul AH, Chua KB. Molecular epidemiology of chikungunya virus in Malaysia since its first emergence in 1998. *Med J Malaysia* 2010; 65(1): 31-5.
- [93] M Naresh Kumar CV, Anthony Johnson AM, R Sai Gopal DV. Molecular characterization of chikungunya virus from Andhra Pradesh, India & phylogenetic relationship with Central African isolates. *Indian J Med Res* 2007; 126: 534-40.
- [94] Dash PK, Parida MM, Santhosh SR, *et al.* East Central South African genotype as the causative agent in reemergence of Chikungunya outbreak in India. *Vector Borne Zoonotic Dis* 2007; 7: 519-27.
- [95] Diallo M, Thonnon J, Traore-Lamizana M, Fontenille D. Vectors of Chikungunya virus in Senegal: current data and transmission cycles. *Am J Trop Med Hyg* 1999; 60: 281-6.
- [96] Tssetsarkin KA, Vanlandingham DL, McGee CE, Higgs S. A single mutation in chikungunya virus affects vector specificity and epidemic potential. *PLoS Pathog* 2007; 3: e201.
- [97] Vazeille M, Moutailler S, Coudrier D, *et al.* Two chikungunya isolates from the outbreak of La Reunion (Indian Ocean) exhibit different patterns of infection in the mosquito, *Aedes albopictus*. *PLoS One* 2007; 2: e1168.
- [98] Mourya DT, Yadav P. Vector biology of dengue & chikungunya viruses. *Indian J Med Res* 2006; 124: 475-80.
- [99] Pages F, Peyrefitte CN, Mve MT, *et al.* *Aedes albopictus* mosquito: the main vector of the 2007 Chikungunya outbreak in Gabon. *PLoS One* 2009; 4:e4691.
- [100] Vazeille M, Martin E, Mousson L, Failloux AB. Chikungunya, a new threat propagated by the cosmopolite *Aedes albopictus*. *BMC Proceedings* 2011; 5: O8.
- [101] N. Pardigon. The biology of Chikungunya: a brief review of what we still do not know. *Pathol Biol* 2009; 57: 127-32.
- [102] Martin E, Moutailler S, Madec Y, Failloux A-B. Differential responses of the mosquito *Aedes albopictus* from the Indian Ocean region to two chikungunya isolates. *BMC Ecology* 2010; 10: 8.
- [103] Ng LC, Tan LK, Tan CH, *et al.* Entomologic and virologic investigation of Chikungunya, Singapore. *Emerg Infect Dis* 2009; 15:1243-9.
- [104] Tssetsarkin KA, McGee CE, Higgs S. Chikungunya virus adaptation to *Aedes albopictus* mosquitoes does not correlate with acquisition of cholesterol dependence or decreased pH threshold for fusion reaction. *Virology* 2011; 8: 376.
- [105] Tssetsarkin KA, Chen R, Leal G, *et al.* Chikungunya virus emergence is constrained in Asia by lineage-specific adaptive landscapes. *Proc Natl Acad Sci USA* 2011; 108: 7872-7.
- [106] Haparachchi HC, Bandara KB, Sumanadasa SD, Hapugoda MD, Lai YL, Lee KS. Re-emergence of chikungunya virus in Southeast Asia: virological evidence from Sri Lanka and Singapore. *J Gen Virol* 2010; 91: 1067-76.
- [107] Volk SM, Chen R, Tssetsarkin KA, Adams AP, Garcia TI, Sall AA. Genome-scale phylogenetic analyses of chikungunya virus reveal independent emergences of recent epidemics and various evolutionary rates. *J Virol* 2010; 84: 6497-504.
- [108] Chretien JP, Anyamba A, Bedno SA, *et al.* Drought-associated Chikungunya emergence along coastal East Africa. *Am J Trop Med Hyg* 2007; 76: 405-7.
- [109] Anish TS, Vijayakumar K, Leela Itty Amma KR. Domestic and environmental factors of chikungunya-affected families in Thiruvananthapuram (Rural) district of Kerala, India. *J Global Infect Dis* 2011; 3: 32-6.
- [110] Chakkaravarthy VM, Vincent S, Ambrose T. Novel Approach of Geographic Information Systems on Recent Out-Breaks of Chikungunya in Tamil Nadu, India. *J Environ Sci Technol* 2011; 4: 387-94.
- [111] Lakshmi V, Neeraja M, Subbalaxmi MVS, *et al.* Clinical Features and Molecular Diagnosis of Chikungunya Fever from South India. *Clin Infect Dis* 2008; 46: 1436-42.
- [112] Her Z, Kam YW, Lin RT, Ng LF. Chikungunya: a bending reality. *Microbes Infect* 2009; 11: 1165-76.
- [113] Kondekar S, Gogtay NJ. Why Chikungunya is called chikungunya. *J Postgrad Med* 2006; 52: 307.
- [114] Sudeep AB and Parashar D. Chikungunya: an overview. *J Biosci* 2008; 33: 443-9.
- [115] Yazdani R, Kaushik VV. Chikungunya fever. *Rheumatology* 2007; 46: 1214-15.
- [116] De Andrade DC, Jean S, Clavelou P, Dallel R, Bouhassira D. Chronic pain associated with the Chikungunya Fever: long lasting burden of an acute illness. *BMC Infect Dis* 2010; 10: 31-7.
- [117] Suhrbier A, La Linn M. Clinical and pathologic aspects of arthritis due to Ross River virus and other alphaviruses. *Curr Opin Rheumatol* 2004; 16: 374-9.
- [118] Dubrulle M, Mousson L, Moutailler S, Vazeille M, Failloux A-B. Chikungunya Virus and *Aedes* Mosquitoes: Saliva Is Infectious as soon as Two Days after Oral Infection. *PLoS ONE* 2009; 4: e5895.
- [119] Chamberlain RW. Epidemiology of arthropod-borne togaviruses: The role of arthropods as hosts and vectors and of vertebrate hosts in natural transmission cycles. In *The Togaviruses: Biology, structure, replication*. R.W. Schlesinger, ed. (Orlando: Academic), 1980; pp. 175-227.
- [120] Mourya DT, Ranadive SN, Gokhale MD, Barde PV, Padbidri VS, Banerjee K. Putative Chikungunya virus-specific receptor proteins on the midgut brush border membrane of *Aedes aegypti* mosquito. *Indian J Med Res* 1998; 107: 10-4.
- [121] Kielian M, Rey FA. Virus membrane-fusion proteins: more than one way to make a hairpin. *Nat Rev Microbiol* 2006; 4: 67-76.
- [122] Marsh M, Helenius A. Virus entry: open sesame. *Cell* 2006; 124: 729-40.
- [123] Lu YE, Cassese T, Kielian M. The cholesterol requirement for Sindbis virus entry and exit and characterization of a spike protein region involved in cholesterol dependence. *J Virol* 1999; 73: 4272-8.
- [124] Ahn A, Schoepp RJ, Sternberg D, Kielian M. Growth and stability of a cholesterol-independent Semliki Forest virus mutant in mosquitoes. *Virology* 1999; 262: 452-6.
- [125] Solignat M, Gay B, Higgs S, Briant L, Devaux C. Replication cycle of chikungunya: a re-emerging arbovirus. *Virology* ; 393: 183-97.
- [126] Khan AH, Morita K, Parquet Md Mdel C, Hasebe F, Mathenge EG, Igarashi A. Complete nucleotide sequence of chikungunya virus and evidence for an internal polyadenylation site. *J Gen Virol* 2002; 83: 3075-84.
- [127] Ou JH, Strauss EG, Strauss JH. Comparative studies of the 3'-terminal sequences of several alphavirus RNAs. *Virology* 1981; 109: 281-9.
- [128] Ou JH, Trent DW, Strauss JH. The 3'-noncoding regions of alphavirus RNAs contain repeating sequences. *J Mol Biol* 1982; 156: 719-30.
- [129] Ou JH, Strauss EG, Strauss JH. The 5'-terminal sequences of the genomic RNAs of several alphaviruses. *J Mol Biol* 1983; 168: 1-15.
- [130] Pfeffer M, Kinney RM, Kaaden OR. The alphavirus 3'-nontranslated region: size heterogeneity and arrangement of repeated sequence elements. *Virology* 1998; 240: 100-8.
- [131] Schwartz O, Albert ML. Biology and pathogenesis of chikungunya virus. *Nat Rev Microbiol* 2010; 8: 491-500.
- [132] Simizu B, Yamamoto K, Hashimoto K, Ogata T. Structural proteins of Chikungunya virus. *J Virol* 1984; 51: 254-58.
- [133] Weaver, S. C., Frey, T. K., Huang, H. V., Kinney, R. M., Rice, C. M., Roehrig, J. T., Shope, R. E. & Strauss, E. G. (2005). *Togaviridae*. In *Virus Taxonomy: Eighth Report of the International Committee on Taxonomy of Viruses*, pp. 999-1008. Edited by C. M. Fauquet, M. A. Mayo, J. Maniloff, U. Desselberger & L. A. Ball. Amsterdam: Elsevier Academic Press.
- [134] Schlesinger, M. & Schlesinger, S. (1986). Formation and assembly of alphavirus glycoproteins. In *The Togaviridae and Flaviviridae*, pp. 121-148. Edited by S. Schelsinger & M. J. Schlesinger. New York: Plenum Publishing Corp.

- [135] Strauss EG, Strauss JH. (1986). Structure and replication of the alphavirus genome. In *The Togaviridae and Flaviviridae*, pp. 35-90. Edited by S. Schlesinger & M. J. Schlesinger. New York: Plenum Press.
- [136] Kääriäinen L, Takkinen K, Keränen S, Söderlund H. Replication of the genome of alphaviruses. *J Cell Sci Suppl* 1987; 7: 231-50.
- [137] Strauss JH, Strauss EG. The alphaviruses: gene expression, replication, and evolution. *Microbiol Rev* 1994; 58: 491-562.
- [138] Ten Dam E, Flint M, Ryan MD. Virus-coded proteinases of the Togaviridae. *J Gen Virol* 1999; 80: 1879-88.
- [139] Lemm JA, Rümenapf T, Strauss EG, Strauss JH, Rice CM. Polypeptide requirements for assembly of functional Sindbis virus replication complexes: a model for the temporal regulation of minus- and plus-strand RNA synthesis. *EMBO J* 1994; 13: 2925-34.
- [140] Voss JE, Vaney MC, Duquerois S, *et al.* Glycoprotein organization of Chikungunya virus particles revealed by X-ray crystallography. *Nature* 2010; 468:709-12.
- [141] Söderlund H, Ulmanen I. Transient association of Semliki Forest virus capsid protein with ribosomes. *J Virol* 1977; 24: 907-9.
- [142] Choi H-K, Tong L, Minor W, *et al.* Structure of Sindbis virus core protein reveals a chymotrypsin-like serine proteinase and the organization of the virion. *Nature* 1991; 354: 37-43.
- [143] Ulmanen, I., Söderlund, H., and Kääriäinen, L. Semliki Forest virus capsid protein associates with the 60 S ribosomal subunit in infected cells. *J Virol* 1976; 20: 203-10.
- [144] Garoff H, Simons K, Dobberstein B. Assembly of the Semliki Forest virus membrane glycoproteins in the membrane of the endoplasmic reticulum *in vitro*. *J Mol Biol* 1978; 124: 587-600.
- [145] Sariola M, Saraste J, Kuismanen E. Communication of post-Golgi elements with early endocytic pathway: regulation of endoproteolytic cleavage of Semliki Forest virus p62 precursor. *J Cell Sci* 1995; 108: 2465-75.
- [146] Wahlberg JM, Boere WAM, Garoff H. The heterodimeric association between the membrane proteins of Semliki Forest virus changes its sensitivity to low pH during virus maturation. *J Virol* 1989; 63: 4991-7.
- [147] Zhao H, Garoff H. Role of cell surface spikes in alphavirus budding. *J Virol* 1992; 66: 7089-95.
- [148] Suomalainen M, Liljestrom P, Garoff H. Spike protein-nucleocapsid interactions drive the budding of alphaviruses. *J Virol* 1992; 66: 4737-47.
- [149] Queyriaux B, Simon F, Grandadam M, Michel R, Tolou H, Boutin JP. Clinical burden of chikungunya virus infection. *Lancet Infect Dis* 2008; 8: 2-3.
- [150] Mourya DT, Mishra AC. Chikungunya fever. *Lancet* 2006; 368: 186-7.
- [151] Mohan A, Kiran DHN, Manohar IC, Kumar DP. Epidemiology, clinical manifestations and diagnosis of chikungunya fever: lessons learned from the re-emerging epidemic. *Indian J Dermatol* 2010; 55: 54-63.
- [152] Ali Ou Alla S, Combe B. Arthritis after infection with Chikungunya virus. *Best Pract Res Clin Rheumatol* 2011; 25: 337-46.
- [153] Brighton SW, Prozesky OW, De La Harpe AL. Chikungunya virus infection: a retrospective study of 107 cases. *S Afr Med J* 1982; 63: 313-15.
- [154] Vijayakumar KP, Anish TS, George B, Lawrence T, Muthukkutty SC, Ramachandran R. Clinical profile of chikungunya patients during the epidemic of 2007 in Kerala, India. *J Global Infect Dis* 2011; 3: 221-6.
- [155] Couderc T, Lecuit M. Focus on Chikungunya pathophysiology in human and animal models. *Microbes Infect* 2009; 11: 1197-1205.
- [156] Robin S, Ramful D, Le Seach' F, Jaffar-Bandjee MC, Rigou G, Alessandri JL. Neurologic manifestations of pediatric chikungunya infection. *J Child Neurol* 2008; 23: 1028-35.
- [157] Pakran J, George M, Riyaz N, *et al.* Purpuric macules with vesiculobullous lesions: a novel manifestation of chikungunya. *Int J Dermatol* 2011; 50: 61-9.
- [158] Parida MM, Santhosh SR, Dash PK, *et al.* Rapid and real-time detection of Chikungunya virus by reverse transcription loop-mediated isothermal amplification assay. *J Clin Microbiol* 2007; 45: 351-7.
- [159] Yap G, Pok KY, Lai YL, *et al.* Evaluation of chikungunya diagnostic assays: differences in sensitivity of serology assays in two independent outbreaks. *PLoS Negl Trop Dis* 2010; 4: e753.
- [160] Edwards CJ, Welch SR, Chamberlain J, *et al.* Molecular diagnosis and analysis of Chikungunya virus. *J Clin Virol* 2007; 39: 271-5.
- [161] Mishra B, Sharma M, Pujhari SK, *et al.* Utility of multiplex reverse transcriptase-polymerase chain reaction for diagnosis and serotypic characterization of dengue and chikungunya viruses in clinical samples. *Diagn Microbiol Infect Dis* 2011; 71: 118-25.
- [162] Alphey L, Benedict M, Bellini R, *et al.* Sterile-Insect Methods for Control of Mosquito-Borne Diseases: An Analysis. *Vector Borne Zoonotic Dis* 2010; 10: 295-311.
- [163] Levitt NH, Ramsburg HH, Hasty SE, Repik PM, Cole FE Jr, Lupton HW. Development of an attenuated strain of Chikungunya virus for use in vaccine production. *Vaccine* 1986; 4: 157-62.
- [164] Harrison VR, Eckels KH, Bartelloni PJ, Hampton C. Production and valuation of a formalin killed chikungunya vaccine. *J Immunol* 1971; 107: 643-7.
- [165] Edelman R, Tacket CO, Wasserman SS, Bodison SA, Perry JG and Mangiafico JA. Phase II safety and immunogenicity study of live chikungunya virus Vaccine TSI-GSD-218. *Am J Trop Med Hyg* 2000; 62: 681-5.
- [166] Akahata W, Yang ZY, Andersen H, *et al.* A VLP vaccine for epidemic Chikungunya virus protects nonhuman primates against infection. *Nat Med* 2010; 16: 334-8.
- [167] Anbarasu K, Manisenthil KK, Ramachandran S. Antipyretic, anti-inflammatory and analgesic properties of nilavembu kudineer choornam: a classical preparation used in the treatment of chikungunya fever. *Asian Pac J Trop Med* 2011; 4: 819-23.
- [168] Delogu I, Pastorino B, Baronti C, Nougairède A, Bonnet E, de Lamballerie X. *In vitro* antiviral activity of arbidol against Chikungunya virus and characteristics of a selected resistant mutant. *Antiviral Res* 2011; 90: 99-107.
- [169] Partidos CD, Weger J, Brewoo J, *et al.* Probing the attenuation and protective efficacy of a candidate chikungunya virus vaccine in mice with compromised interferon (IFN) signaling. *Vaccine* 2011; 29: 3067-73.
- [170] Wang D, Suhrbier A, Penn-Nicholson A, *et al.* A complex adenovirus vaccine against chikungunya virus provides complete protection against viraemia and arthritis. *Vaccine* 2011; 29: 2803-9.
- [171] De Lamballerie X, Ninove L, Charrel RN. Antiviral treatment of chikungunya virus infection. *Infect Disord Drug Targets* 2009; 9: 101-4.
- [172] Khan M, Santhosh SR, Tiwari M, Lakshmana Rao PV, Parida M. Assessment of *in vitro* prophylactic and therapeutic efficacy of chloroquine against Chikungunya virus in vero cells. *J Med Virol* 2010; 82: 817-24.
- [173] Shoichet BK, McGovern SL, Wei B, Irwin JJ. Lead Discovery Using Molecular Docking. *Curr Opin Chem Biol* 2002; 6: 439-46.
- [174] Cross JB, Thompson DC, Rai BK, *et al.* Comparison of several molecular docking programs: Pose prediction and virtual screening accuracy. *J Chem Inf Model* 2009; 49: 1455-74.
- [175] Di L, Kerns EH, Carter GT. Drug-like property concepts in pharmaceutical design. *Curr Pharm Des* 2009; 15: 2184-94.
- [176] Wang J. Comprehensive assessment of ADMET risks in drug discovery. *Curr Pharm Des* 2009; 15: 2195-219.
- [177] Raunio H. *In silico* toxicology - non-testing methods. *Front Pharmacol* 2011; 2: 33.
- [178] Zhu F, Qin C, Tao L, *et al.* Clustered patterns of species origins of nature-derived drugs and clues for future bioprospecting. *Proc Natl Acad Sci USA* 2011; 108:12943-8.
- [179] Tang W, Lu AY. Drug metabolism and pharmacokinetics in support of drug design. *Curr Pharm Des* 2009; 15: 2170-83.
- [180] Ferriz JM, Vinsová J. Prodrug design of phenolic drugs. *Curr Pharm Des* 2010; 16: 2033-52.