Cracking the genetic code of human virus by using open source bioinformatics tools

Usman Sumo Friend Tambunan[*] and Arli Aditya Parikesit

Department of Chemistry, Faculty of Mathematics and Natural Sciences , University of Indonesia , Depok 16424. Indonesia

ABSTRACT

Viral infection is a very serious threat to humanity. It causes malicious diseases, such as HIV/AIDS, dengue, and Avian Influenza, therefore, novel method in virology to combat the viral infection is necessary. Bioinformatics provides outstanding tools for developing vaccines, PCR primers, mutation detection and drugs based design on genetic engineering principles. Those tools are mostly freeware. Algorithm from the computer science has made major contribution to them. Bioinformatics experiment greatly reduces the cost and time in wet laboratory experiment. Our labs has successfully designed PCR primers, vaccine, and mutations prediction. The vaccine design is elaborated for Dengue and HPV. The design has BLAST homology of more than 90%, and RSMD value of 0.1. Those data shown, that the design have identical structure with the native viral protein. However, their efficacy should be verified in the wet laboratory experiment. The future of medicine will greatly be shaped by advancement in bioinformatics.

| Bioinformatics | Mutation | Vaccines | Primer | Drug | Algorithm |

1. INTRODUCTION

Humanity is in the crisis of infectious diseases threat. Various microbial agents, for examples bacteria, plasmodium and virus, are causing uncureable diseases in the medical field. This article will discuss only the viral microbial agents. The emerging threats from biological virus are imminent. Some of the notorious virus are avian influenza, HIV/AIDS, and dengue. The Avian Influenza type A infection has caused more than 100 people dead in Indonesia. While HIV has infected more than 20 million people all over the world. The Dengue fever always made casualties every year in Indonesia and so far there is not any effective drug or vaccine to tackle those viral infections [1].

The Science of Virology is in crisis. Since Edward Jenner invented the small pox vaccine at 1700s, Virology was again in need for the effective medication against the viral infection. However, some clues are available. Just like living organism (although virus is not technically alive), virus has genetic codes. It comprises of four alphabets: A, G, T, and C. They stand for the nucleotides which constitutes the viral genetic codes. Certain combination of nucleotide will code for proteins. It is essential for every human physiological function. For example, certain nucleotide will code for hemoglobin, an essential protein for transporting oxygen to cells. If the coding is wrong, the hemoglobin could not be coded, and the cells will be in jeopardy [2].

This insight is the key element for understanding the biological virus. Virus will code for certain proteins in human cell. They are essential for the existence of the virus. Virus need those proteins for replicating, viral capsid and/or envelope, and weakening the human immune system. The primary questions in Virology are how could we detect the viral infection and how could we prevent and cure it. Virologist have made many classical tools for detecting, preventing, and curing the viral infections. For examples: ELISA method for detecting viral antibody, attenuated viral vaccine for preventing viral infection, and anti viral drugs for curing the viral infection. However, the viral infection prevalence and virulence are on rise, and the classical method may not be sufficient to handle the threats. Certain type of virus, the RNA based virus, are mutated very quickly, so the conventional methods would not sufficient to cope with them. The threat could be halted if the viral genetic code is cracked. RNA based virus is considered the most hazardous virus, because it is prone to mutation. A special security measure for conducting experimentation is necessary. A laboratory with BSL3+ (Biological Safety Level 3+) standard is obligatory for RNA based virology experimentation. The certain hazard of virology experimentation must be reduced by new bioinformatics method. Henceforth, the wet laboratory experiment could be done for other less hazardous experiment [2, 3].

Since the elucidation of DNA structure by Watson-Crick in 1953, the field of Molecular Biology was growing rapidly. Virology research was boosted as well by this

Corresponding author at: Department of Chemistry, Faculty of Mathematics and Natural Sciences , University of Indonesia , Depok 16424. Indonesia
E-mail addresses: usman@ui.ac.id (Usman Sumo Friend Tambunan)

achievement. Researcher gathered huge amount of DNA, RNA, and Protein sequences, without knowing how to solve problems with it. During the 70's, molecular biologist must stretch the sequences on the wall, in order to observe them. Data from wet experiment are in large amount, utilities to gathering important information from them are in needs. Stretching sequences on the wall is definitely impractical. Then, Information Technology (IT) was still in its infancy. Computer was utilized only in certain institution, and not as widespread as nowadays [3].

That was changed in the year 1981, when IBM PC (IBM Personal Computer) was invented. PC became available for non IT researcher. PC could be easily operated by non informaticians for solving various research related problems. Suddenly, the opportunity for virologist to understand the complexity of virus is wide open. The achievement in IT was made the data mining for viral genetic code possible. PC provides wide range of operating systems. The most well known are Windows and LINUX/UNIX. Recently, Apple ported its hardware architecture from Power PC to Intel Pentium. It made MacOSX one of the favorite operating system for researcher. LINUX/UNIX was the first well know platform to develop bioinformatics tools. Later on, they were ported to Windows and Macintosh platform. Important Bioinformatics tools are available in LINUX/UNIX, Windows, and MacOSX platforms [4].

The basic principle of applied bioinformatics research is the freely available of sequence data. During the Human Genome Project in 90's, there was bitter quarrel between National Institute of Health (NIH) and Celera corporation in USA, to decide weather the DNA/RNA/Protein sequence should be freely available or not. However, compromise has been made, and major research sequences are available for free. Only some sequences are proprietary, for patent purpose. The Open Source implementation was one of the key elements for made this development possible and therefore, boosting the bioinformatics research. Important sequences are available in NCBI websites at http://www.ncbi.nih.nlm.gov. They could be downloaded for free. The official name of the databank is Genbank. Major bioinformatics tools, such as Bioedit, ClustalX, Swiss Model, and CLC free workbench are freeware [4,5].

Our laboratory has worked on in silico HPV, influenza type A and dengue experiment. We have developed certain method to design PCR primer, vaccine and drug design for dengue [6,7,8].

2. ALGORITHM OF BIOINFORMATICS METHOD FOR PCR PRIMER DESIGN

2.1 Insight

Polymerase Chain Reaction (PCR) method made the fast and accurate detection of viral fingerprint possible. Basically, PCR method is a procedure to isolate the viral

genome from certain sample, e.g. saliva and blood. Then, the available viral genome is amplified in the thermocycler machine, in certain cycles and temperature. After the amplification process, the amplicons was electrophorated, and the gel was photographed in isolated UV chamber. When the photo shows certain band, then the virus is exist. If it is not, then the virus is not exist. PCR is a very convenient method for detecting virus in patient samples. However, if new strain of virus became available, the existing PCR detecting tools would not be sufficient. The essential part of PCR method is its primer. The oligonucleotides must bind with the certain part of viral genome. If there are new strain of virus exist, novel design of primer is necessary. Fortunately, Bioinformatics tools for PCR Primer Design is available. Previously, *in silico* PCR design was made for HPV (Human Pappilomma Virus). However, the crucial point is to compute which sequence data for *in silic*o experiment. Strong background in Virology and Molecular Biology is necessary for mining the correct data.

2.2 Our research methodology

Standard steps for making novel PCR primers are as follow: The viral genomes from the Genbank websites was downloaded. Utilization of other genomic database, such as EBI or DDBJ are possible. The available viral genomes sequences was stacked in the text editor application. The FASTA header has been included. The sequences was loaded in Clustalw application, for multiple sequences alignment process. This process will search for the homologies between different sequences. Clustalw could be downloaded from EBI websites. The sequences, which has been multiple aligned, was loaded in specific sequences editor, for example the Bioedit. It could be downloaded from http://www.mbio.ncsu.edu/BioEdit/bioedit.html. The editor could help us search for certain conserved region in the genome. The conserved region was tested in primer integrity application, such as netprimer. This step could determine the integrity and stability of the primer by using Netprimer. The link is http://www.premierbiosoft.com/netprimer/index.html .

Table 1: The Selected conserved region templates as PCR primer candidates for HPV

Nucleotide sequence	Alignment	Region	BLASTn result and position on genome according to HPV type
5'-ACAGGCTATGGTGCTATGGA-3'	Type 16; 18; 52	52	Type 16 (6248-6264)
			Type 61 (6373-6383)
			Type 68 (2223-2242)
			Type 70 (6190-6209)

Those steps will produce readily used PCR primer

design. There is another step, which is not the job for the bioinformaticians. The PCR primer design must be sent to the molecular biology company, for having its biochemicals. After they sent it, we could start the PCR method wet experimentation. The method could be applied in any biomedicals laboratory, which need reliable and robust primer design [9,10]. Our labs has found several primer candidates for HPV detection.

The most direct method for producing an MSA (Multiple Sequence Alignment) uses the dynamic programming technique to identify the globally optimal alignment solution. For proteins, this method usually involves two sets of parameters: a gap penalty and a substitution matrix assigning scores or probabilities to the alignment of each possible pair of amino acids based on the similarity of the amino acids' chemical properties and the evolutionary probability of the mutation. For nucleotide sequences a similar gap penalty is used, but a much simpler substitution matrix, wherein only identical matches and mismatches are considered, is typical. The scores in the substitution matrix may be either all positive or a mix of positive and negative in the case of a global alignment, but must be both positive and negative, in the case of a local alignment. In the latter case it is essential that the average score be less than 0.

Motif finding, also known as profile analysis, is a method of locating sequence motifs in global MSAs that is both a means of producing a better MSA and a means of producing a scoring matrix for use in searching other sequences for similar motifs.

A variety of methods for isolating the motifs have been developed, but all are based on identifying short highly conserved patterns within the larger alignment and constructing a matrix similar to a substitution matrix that reflects the amino acid or nucleotide composition of each position in the putative motif [6, 9].

2.3 The methodology of other laboratory

The most important part for PCR product prediction is finding appropriate primer annealing sites on the template. One important program is VPCR (Virtual PCR), which has designed by Cao et al. However, it does not deal with degenerated PCR primer.

New Algorithm based on information theory has been developed. One information source was obtained by converting the primer sequences to numeric vector of the potential full hydrogen bond numbers, and the second was created as a vector of the hydrogen bond numbers formed between the primer and its potential binding site on the template. An information coefficient was computed for determining the similarity between the two information sources as a criterion to locate primer annealing sites, and predict products. A computer program, SPCR (Simulated PCR), based on this algorithm was developed to predict PCR product, and its performance was evaluated by replicating 4 cases of laboratory PCR experiment *in silico*, and performing comparisons between the predicting results of the program and VPCR. It could be downloaded from http://moleco.sjtu.edu.cn/ SPCR [10].
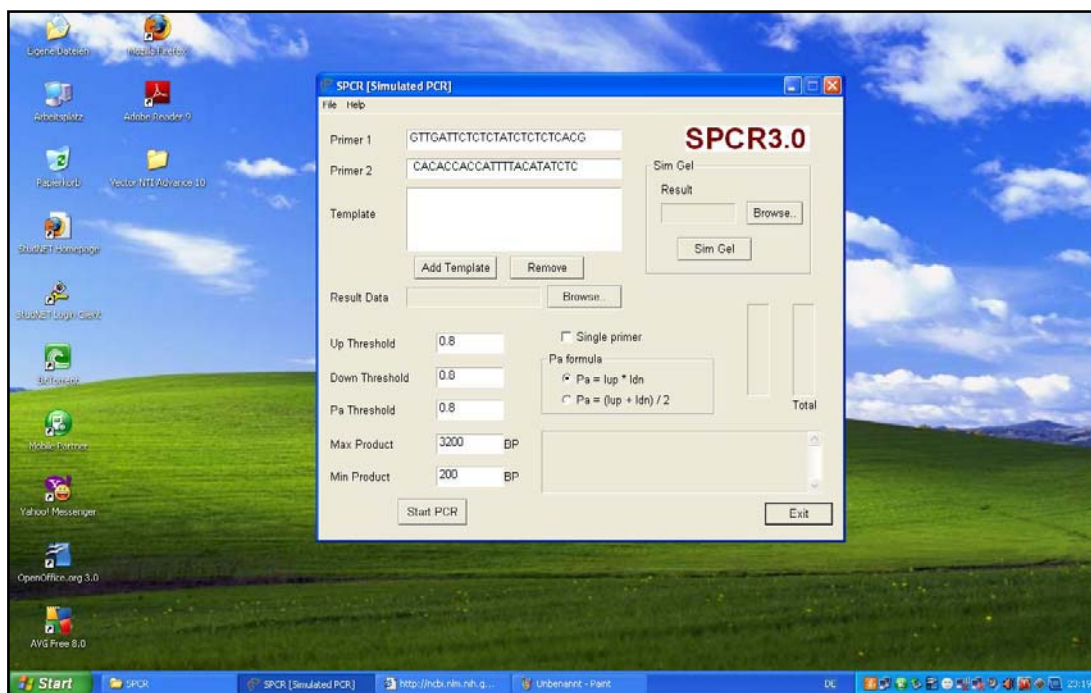

Figure 1 : The SPCR application. Shown in the figure, the primer entry box. After executing it, the amplicon result will be shown. It is useful as simulation, before running the real

3.1 ALGORITHM OF *IN SILICO* VACCINE DESIGN FOR INCREASING THE CAPABILITY OF OUR IMMUNE SYSTEM.

3.1. Insights

The well known method to prevent viral infection is vaccination. Vaccination is the method to increase the ability of human immune system, so it could handle the viral infection by itself. There are two important type of lymphocyte (white blood cell), which involved in this process. They are B-cells and T-cells. The B cells are responsible for producing antibodies, and T cells are responsible for antigenic recognition. T-cell epitopes of supertype HLA alleles will be crucial in the design. The principle of vaccine production is as follows: The original virus was isolated from the patient, and attenuated it in the bioreactor with certain chemicals and physical treatment. The classic attenuated virus vaccine has certain weakness, such as the availability of viral genome. If the viral genome still exist, the possibility of new infection in the patient is imminent. Nowadays, virologist has used the common method in molecular biology, such as genetic engineering, for designing vaccine. New genetically engineered vaccines do not have the viral genome, so it will be much safer. It only has the certain part of the virus, for examples its epitopes (protein as antigenic fingerprint). The vaccine was produced by coding the design to microbial plasmid, then it will transfected to recombinant microbe under certain conditions. However, when using ordinary wet experiment method, vaccine design could take a long time. The difficulty of genetically engineered microbe to grow in normal medium made this experiment a slow process. Aid from bioinformatics method to fasten the process is necessary. *In silico* Recombinant vaccine design was made previously for HPV. The important point is to compute the right epitopes for the experiment.

Khan et al has developed novel methodology for designing vaccine. It's a combined immunoinformatics and molecular strategy for vaccine development. Based upon the growing number of bioinformatics tools and antigen sequences available in public databases for identifying the pathogen peptides, the in silico prediction of T-cell epitopes can greatly reduce the list of candidate epitopes. Such a shortlist is then the strating point for molecular experiments that can validate the vaccine targets based on the biological function of the selected antigen sequences [11].

3.2 Our research methodology

The general steps for designing genetically engineered vaccine by bioinformatics method are as following: The protein sequence from Swiss Prot protein database was downloaded. We were downloading L1 HPV and E DENV protein sequences. Utilization of other protein database, such as NCBI or EBI is possible. The available viral proteomes sequences was stacked in the text editor application. The FASTA header must be included. The

sequences was loaded in special protein alignment toolbox, such as T-coffee, for multiple alignment procedure. Conventional off line application would not sufficient, because of the large number of sequences. The most conserved sequences region was utilized as vaccine backbones. The backbones could be beneficial to create multivalent vaccine design. This type of vaccine could prevent the infections of broad range of virus strain. Available epitopes in the vaccine backbones was found by using T-cell and B-cell epitope prediction tools. MULTIPRED is the T-cell epitope prediction tool, and CEP is the B-cell epitope prediction tool. The epitope was arranged in proper order, by substituting the low binder T-cell epitopes with the predicted high binder ones. The determination, whether the epitope has high or low binder were elucidated by the provided binding score. Else, the T-cell epitopes of HLA alleles should not overlap the B-cell epitopes. The vaccine protein sequence was verified by BLASTp tools for comparing it with the real virus protein sequence. If it is homologue, then the vaccine design is correct.

Vaccine protein sequence will be reversed translated to DNA sequences by using pBLASTn tools. The vaccine DNA sequence will be transformed into plasmid, by using plasmid editor Pdraw32. The next step is to produce the vaccine design with the wet laboratory method. The plasmid design will be sent to biotechnological company for producing its biochemical regent. After it, the plasmid biochemical will be transfected to certain recombinant microbe, for examples yeast. It will be produced in bioreactor with certain medium and treatment [7, 8, 12].

```
>ANN1_vaccine
MRCIGISNRDFVEGVSGGSWVDIVLEHGSCVTT
MAKNKPFVCRRTFVDEAKQPATLRKYCIEAKL
TNTTTESRCPTQGGEPTLNEEQDKRFVCKHSMV
DRGWGNGCGLFGKGGIVTCAMFTCKKNMEGK
IVYVCKHTYVDVVTPHSGEEHAVGNDTGKHG
KEVKITPQSSITEAELTGYGTVTMECSPRTGLDF
NEMVLLQMKDKAWLVHRQWFLDLPLPWLPG
ADTQGSNWIQKETLVTFKNPHAKKQDVVVLG
SQEGAMHTALTGATEIQMSSGNLLFTGHLKCR
LRMDKLQLKGMSYSMCTGKFKVVKEIAETQH
GTIVIRVQYEGDGSPCKIPFEIMDLEKRHVLGRL
ITVNPIVTEKDSYRGAKRMAIFGDSYIIIGVEPG
QLKLNWFKK
```

Figure 2 : The Peptide sequence of Dengue vaccine [7]. Shown here is the ANN1 vaccine. Red color sequence is E DENV-1 epitope. Blue color sequence is E DENV-3 epitope. Green color sequence is E DEV-4 epitope. The E DENV-2 protein is the backbone of the vaccine.

Our laboratory has found cVLP Vaccine design for HPV and peptide vaccine design for Dengue virus. We have succeeded in modeling their 3D structure by using homology modeling method. We search for their homologue templates in the PDB web sites, and align it with our cVLP sequences. The alignment was done by Deep View molecular modeling program. The program has option to submit the alignment to SWISS MODEL server. After the submission, downloaded model could be visualized by Deep View [7,8]. We verified the validity of our protein

model, by comparing their homology with the native protein. The BLAST search was conducted for verifying the sequence homology, while the VAST alignment was utilized for verifying the structural homology.

We have successfully determined the vaccine sequence of HPV and Dengue. Some of the sequences are in Figures 2 and 3.

KVVSTDEYVARTNIYYHAGTSRLLAVGHPYFPIKKPNNNKILVPKVSGLQYRVFRIHLPDPNKFGFPDTSFYNP
DTQRLVWACVGVEVGRGQPLGVGISGHPLLNKLDDTENASAYAANAGVDNRECISMDYKQTQLCLIGCKPPI
GEHWGKGSPCTQVAVQPGDCPPLELINTVIQDGDMVDTGFGAMDFTTLQANKILHYHLLPLSICKYPDYIKM
VSEPYGDSLFFYLRREQMFVRHLFNRAGTVGENVPDDLYIKGSGSTANLASSNYFPTPSGSMVTSDAQIFNKP
YWLQRAQGHNNGICWGNQLFVTVVDTTRSTNMSLCAAISTSETTYKNTNFKEYLRWGFGLTPPPIFQLCKITL
TADVMTYIHSMNSTILEDWNFGLQPPPGGTLEDTYRFVTSQAIACQKHTPPAPKEDPLKKYTFWEVNLKEKFS
ADLDQFPLGRKFLLQLGL

Figure 3 : The cVLP sequence of HPV vaccine [8]. Shown here is the HMM1 vaccine. Blue color sequence is L1 HPV-18 epitope. Red color sequence is L1 HPV-52 epitope. The cVLP of L1 HPV-16 os the backbone of the vaccine.

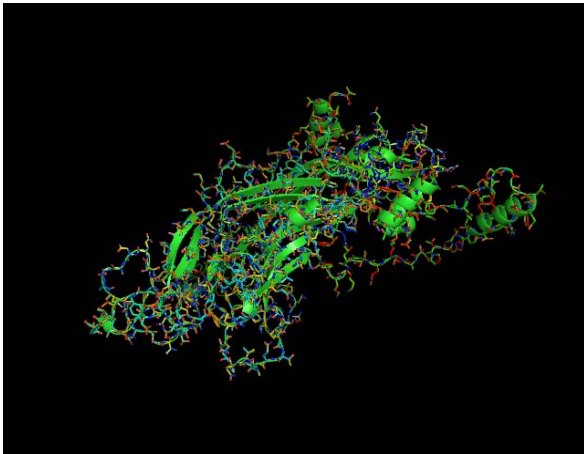Some of the cVLP structure we have is as follows:

Figure 4 : ANN1 cVLP L1 HPV Vaccine Design [8]. The vaccine was visualized by using MacPymol application. The ribbons inside the chains are the vaccine backbone.
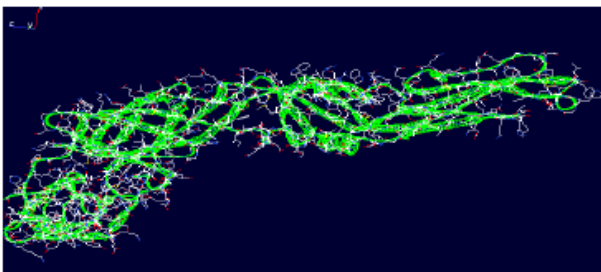

Figure 5 : 3D structure of HMM1 Peptide E Dengue Vaccine[7] The vaccine was visualized by using Deep-View application.

The validity of our design was verified by using BLAST and VAST. The ANN1 cVLP L1 HPV vaccine design has 96 % BLAST homology and 0.1 RSMD with native L1 HPV 16 protein, Meanwhile, the HMM1 Peptide Dengue Vaccine has percent 93.9 % BLAST homology and 0.1 RSMD with native E-DENV-2 protein. The 0.1 RSMD means that the artificial structure is identical with the native structure [8,9].

Hidden Markov models are probabilistic models that can assign likelihoods to all possible combinations of gaps, matches, and mismatches to determine the most likely MSA or set of possible MSAs. HMMs can produce a single highest-scoring output but can also generate a family of possible alignments that can then be evaluated for biological significance. HMMs can produce both global and local alignments. Although HMM-based methods have been developed relatively recently, they offer significant improvements in computational speed, especially for sequences that contain overlapping regions.

An artificial neural network (ANN), usually called "neural network" (NN), is a mathematical model or computational model based on biological neural networks. It consists of an interconnected group of artificial neurons and processes information using a connectionist approach to computation. In most cases an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase.

In more practical terms neural networks are non-linear statistical data modeling tools. They can be used to model complex relationships between inputs and outputs or to find patterns in data [4].

3.3   The methodology of other laboratory.

Singh et al have another application in the field of immunoinformatics. They have conducted research on HLA Class I and II binding of the T-cell epitopes of Hepatitis B Virus (HBV). HBV is highly pathogenic, and it can't be cultured easily as it always requires high level of biosafety containment. Synthetic peptides can be used as vaccines to induce either humoral or cell-mediated immunity. It requires an understanding of the nature of T-cell and B-cell epitopes. Bioinformatics tools were used for the analysis of genome of hepatitis B virus. The complete sequences of hepatitits B virus (NC_003977) were retrieved from www.ncbi.mlm.nih.gov. The open reading frames were identified from the whole genome using software viz, Generunner, ORF finder, and DNA star. The expected molecular weight, and isoelectric point (PI) value were verified using Generunner and ExPasy (http://www.expasy.org). Propred (http://www.imtech.res.in/raghava/propred) and Propred I (http://www.imtech.res.in/raghava/propred1/) immunoinformatics tools are available to predict the antigenic epitopes in the complete protein primary sequences. The prediction of epitopes in Polymerase, large-s-protein, middle-s-protein, s-protein, x-protein,

precore/core protein, core and e-antigen, protein of HBV were investigated by *in silico* method. Total 50 epitopes were predicted for class I MHC and 55 epitopes for class II MHC molecules for these proteins [13].

Table 2 : Epitopes of S-Protein. They are the possible candidates for HBV Vaccine

No	T-Cell Epitopes	Amino Acid Position	No of MHC Class II binding sites
1	LRRFJIFLF	76-84	17
2	FVQWFVGLS	178-186	8
3	FLLTRILTI	19-27	7
4	LGPLLVLQV	8-16	6
5	FLFILLLCL	83-91	16
6	SLDSWWTSL	31-39	6
7	SSWAFARFL	154-162	7
8	LYNILSPFL	205-213	6

Khan et al has utilized spesific method for designing vaccine. The data collection step was done by using ABK structural rule-based approach. The pathogens with multiple groups, pan-group consensus sequences are obtained by aligning consensus sequences derived from each of the different groups. The evolutionary stability of the peptides were determined by information entropy. The functions of conserved sequences can be elucidated that comprise data on protein families, domains, and functional sites. The distribution of conserved sequences in nature are investigated as well, by using BLAST. Algorithm for prediction of HLA binding peptides were utilized by using NetCTL, MULTIPRED, and TEPITOPE. The prediction is already proven in the wet labs, and it is helpful in immunological studies [11].
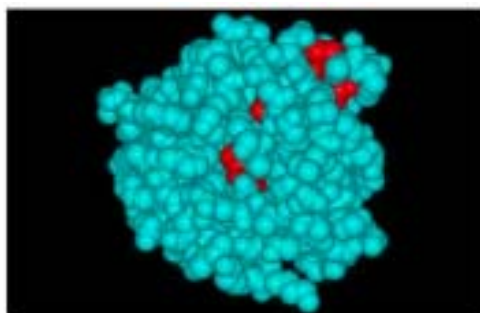

Figure 6 : Molecular location of NS3 Dengue pan-serotype conserved sequence on the protein's 3D structure [11]

4. ALGORITHM OF IN SILICO MUTATION STUDIES OF VIRAL PROTEIN

4.1 Our research methodology

Our laboratory has worked on in silico H5N1 virus

project. This virus was originally recognized only as the causative agent of fowl plague in 1955. However, in the last few years, the occurence of highly pathogenic avian influenza A (H5N1) virus also began to threaten human safety after fowl to human transfection case and was reported to increase. This problem arises due to the virus tendency to mutate and recombine with genetical material of other influenza virus. Very limited human to human transmission of the H5N1 strain was documented in healthcare workers and family members with contact. Our labs has worked with Banten province's haemagglutinin and neuraminidase amino acids from the virus. Both were examined for conserved region, mutation sites, secondary structural change, hydrophobicity, and post translational behavior. The methods are outlined as following.

The Haemagglutinin and Neuraminidase database of type A H5N1 Indonesian strain was downloaded. Then, the database similarity screening was done by BLAST. The conserve region was found by Bioedit. Secondary structures were predicted using NNPREDICT. Hydrophobicity was detected by Protscale. Post translational modification were mined using ScanProsite.

Our study has revealed that all mutation occur outside the conserved region, except at position 40 and 252 for neuraminidase. There is no post translational modification occur at any mutation sites. The *in silico* study cannot prove that the founded Indonesian strain is a new one. We conclude that the existing mutations might only case of antigenic drift. Based on phylogenetic tree analysis and 3D homology modeling, the mutation is insignificant and an H5N1 type A virus Banten strain cannot spread from human to human. [14]

The information from *in silico* study could be used for developing a vaccine for H5N1. The haemaglutinin and neuramidase protein could be used as peptide vaccine for H5N1.

Homology modeling, also known as comparative modeling of protein refers to constructing an atomic-resolution model of the "target" protein from its amino acid sequence and an experimental three-dimensional structure of a related homologous protein (the "template"). Homology modeling relies on the identification of one or more known protein structures likely to resemble the structure of the query sequence, and on the production of an alignment that maps residues in the query sequence to residues in the template sequence. The sequence alignment and template structure are then used to produce a structural model of the target. Because protein structures are more conserved than DNA sequences, detectable levels of sequence similarity usually imply significant structural similarity.

Phylogenetic trees among a nontrivial number of input sequences are constructed using computational phylogenetics methods. Distance-matrix methods such as neighbour-joining or UPGMA, which calculate genetic distance from multiple sequence alignments, are simplest to implement, but do not invoke an evolutionary model. Many sequence alignment methods such as ClustalW also create trees by using the simpler algorithms (i.e. those based on

distance) of tree construction. Maximum parsimony is another simple method of estimating phylogenetic trees, but implies an implicit model of evolution (i.e. parsimony). More advanced methods use the optimality criterion of maximum likelihood, often within a Bayesian Framework, and apply an explicit model of evolution to phylogenetic tree estimation. Identifying the optimal tree using many of these techniques is NP-hard [5], so heuristic search and optimization methods are used in combination with tree-scoring functions to identify a reasonably good tree that fits the data.

4.2 The methodology of other laboratory.

Flaviviruses (FV), (+) strand RNA viruses in the same genus as hepaci- and pestiviruses, are responsible for many emerging human encephalitic and hemorrhagic diseases. The ~10,5 kb genome encodes a single polyprotein that is cleaved into 10 viral proteins. The Flavitrack on line application was designed to ease the identification of conserved functional areas, and to group viruses according to their phenotypic characteristics. The application was designed by Misra et al. The link is http://carnot.utmb.edu/flavitrack .

The database contains all publicly available full-genome flavivirus sequences and provides access to sequences analysis tools. Flavitrack will eventually also contain structures or 3D models for all flavivirus proteins, allowing combined sequence/structure analysis to characterize common B- and T-cell epitopes, account for the functional effects of mutations, and determine highly conserved areas.

A list of mutants and variant sequences, tabulated according to derived strains, location of mutations, corresponding altered phenotypes, and the references for each mutation has been included in Flavitrack. It also provides access to PCPMer in-house program (http://born.utmb.edu/BinZhou/PCPMer), which can be used to automatically visualize areas that are highly conserved on structures of FV proteins. PCPMer results, coupled with the mutation data, aid in identifying structural motifs associated with viral function or lethality. This enables a user to correlate the genotype with viral characteristics.

The sequences of flaviviruses has been archived in a relational database. It is designed to aid in identifying surface exposed clusters of conserved amino acids and correlating these with data on mutational data for altered phenotypes, vector type and disease characteristics [15].

Table 3: The samples of various mutation data from flavitrack database

Accession	Protein Name	Mutation	Natural/Artificial	Phenotype
L04561	DEN2 Mexican Medium plaque 1	E Protein Asp390Asn	Natural Mutation – obtained from clone of DEN2 Mexican strain	Produced medium lytic plaque, Medium neurovirulence, var. mortality(<20%), tremor, seizure, paralysis
L04561	DEN2 Mexican Large plaque 2	E Protein Asp390His	Natural Mutation – obtained from clone of DEN2 Mexican strain	Produced large lytic plaque, high virulence, 70-90% mortality
L04561	DEN2 Mexican Small plaque 2	E Protein Asp390Asn	Natural Mutation – obtained from clone of DEN2 Mexican strain	Produced small lytic plaque, mild symptoms (inactivity, slow movement, shaking, tremor and inability to stand), low virulence
L04561	DEN2 Mexican Small plaque 1	E Protein Asp390Asn	Natural Mutation – obtained from clone of DEN2 Mexican strain	Produced small lytic plaque, mild symptoms (inactivity, slow movement, shaking, tremor and inability to stand), low virulence
L04561	DEN2 Mexican Small plaque 4	E Protein Asp390Asn	Natural Mutation – obtained from clone of DEN2 Mexican strain	Produced small lytic plaque, mild symptoms (inactivity, slow movement, shaking, tremor and inability to stand), low virulence
M29095	DEN2 Mexican Small plaque 3	E Protein Lys388Asn, Asp390Asn	Natural Mutation – obtained from clone of DEN2 Mexican strain	Produced small lytic plaque, Medium neurovirulence, var. mortality (<20%), tremor, seizure, paralysis

5. THE NECESSITY OF WET LABORATORY EXPERIMENTS

The Genomic and proteomic data in the Genbank are the result of Wet experiments laboratory. Wet experiment is still necessary for gathering primary data. The bioinformatics method could not be utilized without those primary data. However, the data must be converted into important information by bioinformatics tools. Special skills in bioinformatics, and knowledge in virology is necessary for computing the data into the necessary information. At least, the bioinformaticians must be familiar with the basic function of computer operating system, and must have a deep insight into the latest theory of molecular biology, for examples genetic or protein engineering, DNA hybridization, PCR method, Protein separation, and central dogma. The wet experiment is considered necessary, for giving the complete insight to the world of virology. The advantage of bioinformatics aid to wet experiment is to reduce the number of unnecessary experimentation, therefore saving the necessary biochemical regents for other important tasks. The upcoming financial crisis, which made the price of regents rising and the not readily available critical biochemicals regents will made bioinformatics an important supporting techniques for wet laboratory [16].

Bioinformatics tools made the ready-to-use information for tackling the viral infection threat. *in silico* PCR primer design and Recombinant vaccine design are the answers for it. The wet experiment laboratory is still necessary for producing the biochemicals of primer and vaccine. However, the aid of computational tools would greatly enhanced the effectivity and robustness of therapeutic and phrophylatic agents design [16].

6. FUTURE OF VIROLOGY AND BIOINFORMATICS

There are some breakthrough in the field of biotechnology, for example the widespread availability of HPV vaccine, Anti Retroviral (ARV) drugs, and the Polio Vaccine. However, there are some imminent threats. The threat of deadly contagious diseases, such as Avian Influenza, Dengue, and HIV/AIDS could jeopardize the world of medical science, if there is no other novel methods in progress. The implementation of Bioinformatics Open Source and the rapid gathering of wet experiment data will made the availability of important anti viral agent near in the future.

The high mutation rate of RNA based virus could eventually be cracked, and their protein coding in the host cell could be inhibited by the anti viral agent. The design of anti viral agent would made the utilization of 3D protein modeling necessary. This approach could create a complete modeling of inhibitor and viral antigen possible. The increasing computational power and easy of use of Bioinformatics tool will strengthen the bioinformatics research.

The IT industry has provided strong and robust computing power, with low cost expediture. Nowadays, a powerful low cost multiprocessor computers are available, which made the modeling of complicated proteins and sophisticated drug design possible. The major computer operating system, such as MacOSX, Linux, and Windows are already supporting open source bioinformatics software. They could do the functionalities of the commercial software, with the same robustness.

Nowadays, the field of bioinformatics is growing. The *In Silico* (Bioinformatics) experiment will be considered as important as wet experiment by biologist and/or biotechnologist. The *In silico* approach did not designed to replace wet experiment, but it's in order to supplement it. Open source implementation will help bioinformaticians to solve viral threat in efficient and effective manner. There will be more robust bioinformatics tools available in the future for solving crucial virology related problems. [17,18].

Our laboratory has successfully designing primer and vaccine for therapeutics. However, the efficacy of the design must be proven in the wet labs experiment. Synthesizing them by using latest molecular biology instrument is crucial for progressing towards clinical trial. Conducting it will require us to form strong cooperation with faculty of medicine in our university. We already have cooperation with them, and will verify our design in the future.

ACKNOWLEDGEMENT

REFERENCES

[1]   http://www.depkes.go.id , 27 Mei 2008.
[2]   Zuckerman, Arie., Banatvala, Jangu., Patison, John., Griffits, Paul., Schoub, Barry. 2004. Principles and Practice of Clinical Virology. John Willey and Sons. West Sussex.
[3]   Glick, B. R., Pasternak, J. J. 1990. Molecular Biotechnology; Principles and Applications of Recombinant DNA. ASM Press. Washington D.C.
[4]   Baldi, P., Brunak, S. 2001. Bioinformatic The Machine Learning Approach. MIT.
[5]   Mount, D.W. 2004. Bioinformatics: Sequence and Genome Analysis. Second edition. CSHL Press. New York.
[6]   Tambunan, Usman Sumo Friend., Butar-Butar, Herbert Wybert., Umbas, Radya., and Hidayah, Zulfa. 2007. Genome. Biotechnology 6 (1): 93-96.

[7]   Tambunan, Usman Sumo., Parikesit, Arli Aditya., Hendra., Ichsan Taufik, Rizky. Amelia, Fitri,. Syamsudin, 2009. Science Alert Online Journal of Biology.

[8]   Tambunan, Usman Sumo Friend., Parikesit, Arli Aditya. Tochary, Theo., Sugiono, Dedy. 2007. Makara journal of Science. University of Indonesia. pp. 56-62.

[9]   Claverie, J. M., Notredame, C. 2003. Bioinformatics for Dummies. Wiley Publishing Inc. New York.

[10]   Cao, Youfang., Wang, Lianjie,. Xu, Kexue, Kou, Chunhai., Zhang, Yulei., Wei, Guifang., He, Junjian., Wang, Yunfang,. Zhao, Liping. 2005. BMC Bioinformatics. 6:190

[11]   Khan, Asif M., Miotto, Olivo., Heiny, A.T., Salmon, Jerome., Srinivasan, K.N., Nascimento, Eduardo.,Marques, Ernesto T., Brusic, Vladimir., Tan, Tin Wee., August, J Thomas. 2006. Cell Immunol, 224(2):141-147

[12]   Kulkarni-Kale, U., Bhosie, S., Kolaskar, A.S. 2005. Nucleic Acid Research. vol 33, No. 3, pp. 4322-4330.

[13]   Singh, Vijay., Indramani, Chaudhary, Dharmendra Kumar, Somvanshi, Pallavi. 2009. JCSB Vol 2. 69-73. T

[14]   Tambunan, Usman Sumo., Himawan, Oksya,. Tockary, Theo. 2008. P 18-24. Asian Network for Scientific Information.

[15]   Misra, Millind., and Schein, Catherine H. 2007. Oxford Journal of Bioinformatics Volume 23, Number 19. 2645-2647.

[16]   Westhead, D.R., Parish, J.H. & Twyman, R.M. 2002. Instant Notes Bioinformatics. BIOS. Oxford.

[17]   Kolls, A., Sherris, J. 2000. HPV Vaccines: promises and challenges. PATH. Seatlle

[18]   Notredame, C. 1998. Utilization of Genetic Algorithm for analyzing biological Sequences. University of Paul Sabatier. Toulouse