2-1-2011

In silico design of selective high affinity ligands against HCV using novel computational diology tools

Reem Rafik Al Olabi

The American University in Cairo

School of Science

Biotechnology Program

In silico design of selective high affinity ligands against HCV using novel computational biology tools

A thesis submitted to

The Biotechnology Graduate Program

in partial fulfillment for the requirements of

The Master of Science

By Reem Rafik Al Olaby

Bachelor of Pharmaceutical Sciences - Ain Shams University

Under the supervision of:

Dr. Hassan Azzazy, Dr. Rod Balhorn, and Dr. Adam Zemla

June/2010

DEDICATION AND ACKNOWLEDGEMENTS

No words can express how much I am grateful to many great people in my life. My wonderful parents, without their encouragement and love I wouldn't have reached that far. My great loving husband who sacrificed and done all what he could do so that I could accomplish what I have always dreamt of. He is a great example of a wonderful, confident and caring husband. My lovely daughter who was such a sweet heart and although she is still an infant she was an angel and was never demanding although it is her right. My great advisor-Dr. Hassan Azzazy- who has really helped me a lot in learning so many things I didn't know before and benefiting me with his experienced advices. Although being busy he always provided me with all what I needed so as to accomplish my mission. Special thanks goes to Dr. Rania Siam for her continuous support and encouragement. I am really grateful to Mrs. Sawsan Mardini and Yasmine Ibrahim who were always there for me and supported me with all what support means from all aspects so as to achieve my goals and get my masters degree. Mr. Auchi who has thankfully funded me throughout my masters degree. I could never forget Dr. Rod Balhorn, who was a great co-advisor and who really helped me a lot to make it to the United States so as to continue my thesis which depends greatly on a field he is the pioneer in. He also made the thesis path feel so smooth and easy. I would also like to thank Dr. Adam Zemla, my co-advisor, for being a powerful collaborator, providing us with E2 homology model and trying to optimize it. I really appreciate the approval of Dr. Monique Balhorn to be a co-advisor and would like to thank her for her support. Deep appreciation must go to Dr. Arthur Olson, Rodney Harris and Stefano Forli and all Dr. Arthur's team for providing such an informative training which made me understand how to use autodock very well and was an eye opener to all the great features of autodock. Dr. Arthur also helped me get an NBCR account which I must also acknowledge. The NBCR account was a great award that my thesis depended on greatly. Special thanks go to Dr. Wilfred Li and Jane Ren from NBCR for facilitating my mission.

To you all, I dedicate this thesis, which I wish to help decrease the burden caused by HCV worldwide one day.

I would like to end by thanking Allah for all what he granted me and blessed me with.

ABSTRACT

The American University in Cairo

In silico design of selective high affinity ligands against HCV using novel computational biology tools

Reem Rafik Al Olabi

Dr. Hassan Azzazy
Dr. Rod Balhorn
Dr. Adam Zemla

Background: Hepatitis C virus (HCV) infects 170 million patients worldwide. The absence of an effective mean of treatment or prophylaxis makes HCV infection a serious public health problem. Generating selective high affinity ligands (SHALs) against HCV could provide a solution to this public health burden. HCV E2 glycoprotein is required for HCV entry into host cells. It binds to CD81, a host receptor protein that belongs to the tetraspanin family and plays a critical role in viral invasion. HCV protease is essential for the cleavage of non structural proteins, in addition to preventing the phosphorylation of human interferon regulatory factor 3 and thus prevents the anti-viral response.

Objective: To design a cocktail of SHAL-based inhibitors against several target proteins such as CD81, HCV E2, and HCV protease and optimize the currently available E2 homology models.

Methods and findings: Different homology modeling techniques such as AS2TS, Phyre, ROSETTA and TASSER, were used to obtain reliable models of HCV NS3 serine protease and HCV E2 glycoprotein. LGA was used to structurally analyze the models in addition of clustering the obtained models and finding the closest structural templates using StralCP. Auto Dock Tools 1.5.6 was used to prepare the crystal structures of the CD81-LEL protein (1G8Q and 1IV5), HCV protease, HCV polymerase and the homology model of HCV E2 by deleting water molecules, adding polar hydrogens, and assigning Gasteiger charges and to create a grid bounding box, which provided the desired grid parameter file using 0.375 A spacing. Autoligand, an AutoDock tool, was used to identify several binding sites on the protein targets. Fill points were created using a 1 A grid, and the calculations were performed using 10 to 210 fill points. AutoDock 4.2 was used to screen 30,000 ligands obtained from different libraries (NCI_DSII, Sigma and Asinex) and identify small

molecules that might bind to each site. The docking results were analyzed and the top 20 ligands for each binding site on the target proteins were ranked according to selection criteria required for the design of promising SHALs. Distances between pairs of bound ligands were estimated and used to design several SHALs that should bind selectively to the target proteins.

Conclusion: New computational tools have been used to design *in silico* several SHAL-based inhibitors that might have the potential to prevent both HCV entry into hepatocytes and the production of inflammatory cytokines that accelerate liver damage when targeting E2-CD81 interaction. It might alter protein processing and viral replication if HCV protease was targeted. By targeting HCV RNA dependent RNA polymerase, the HCV replication could be blocked. In addition, blocking the complexation of NS3 with the NS4A co-factor will render it non functional and thus block the replication process and disrupt the HCV life cycle. If a reliable E2 homology model was developed based on different studies and validations, it could help generate selective high affinity ligands that block the earliest phase of the HCV life cycle.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: LITERATURE REVIEW

Hepatitis C virus is considered a global public health burden, infecting approximately 170 million patients worldwide [1]. HCV was discovered 15 years ago by Houghton *et al.* paving the way for the development of different drugs and diagnostics against HCV [2]. Determining the HCV structure, translated proteins, their functions and the replication machinery made it easier for research groups worldwide interested in this field to develop novel HCV antiviral drugs. Interferon has been used to treat chronic HCV infection since 1986, which means before non A non B (NANB) infection was known as HCV. The therapy was optimized by developing peginterferon (PEG-IFN) so as to improve the pharmacokinetics of IFN and increase the response to treatment [3, 4]. Although effective, the monotherapy interferon (IFN) or combination therapy (IFN/Ribavirin) led to many side effects making this treatment patient non-compliant. The side effects are caused by the non selectivity of IFN which affects both the normal cells and the infected hepatocytes. This made different research groups think of alternatives that target either the host proteins involved in HCV infection [5 , 6, 7, 8] or HCV proteins [9, 10, 11]. Several approaches were tried in developing new therapeutics against HCV [12, 13, 14, 15, 16]. Recently, a novel approach was developed to generate selective high affinity ligands against a protein on lymphoma [17, 18, 19, 20]. This approach proved to be successful in developing highly selective and robust SHALs with high specificity towards the target, thus leading to the idea of applying such an approach to develop drugs against HCV targeting different levels of HCV life cycle such as the entry phase (E2-CD81-LEL interaction) and the polyprotein cleavage phase (HCV NS3-4A serine protease complex).

Targeting HCV proteins to develop anti-HCV drugs

HCV E2 glycoprotein

Several approaches are being used to develop anti-HCV drugs and vaccines against HCV E2 glycoprotein. Developing monoclonal and polyclonal antibodies is one of the approaches used. XTL Biopharmaceuticals reported that a combination of the monoclonal antibodies AB68 and AB65 (AB6865) was more effective than each one alone [21]. Both monoclonal antibodies were obtained by immortalizing peripheral blood mononuclear cells obtained from the blood of HCV infected patients. It was observed that AB6865 combination can identify different conformational epitopes on E2 and immune-precipitate different strains of HCV particles. The viral clearance takes place upon formation of immune complexes by endocytosis since that AB6865 induces phagocytosis of the immune complexes by neutrophils [21-24]. Bavituximab® is another monoclonal antibody developed by Peregrine pharmaceuticals, that acts against a unique target, phosphatidyl serine. Phosphatidylserine is a lipid molecule normally found in the interior of cellular membranes and becomes exposed on the cell surface of cell membranes of viruses and host infected cells during replication. It has been reported that the presence of the phosphatidyl serine on the

surface of the membranes of viruses creates a masking effect and thus helps the virus to evading the host immune response. This monoclonal antibody can be used against several viruses such as HIV, CMV, influenza and HCV. When comparing Bavituximab with other monoclonal antibodies, it was found to be more advantageous because it has high specificity towards cells infected with viruses only. In addition it targets a molecule in the host and thus won't be affected by any viral mutations [25, 26]. Human hepatitis C immune globulin or Civacir is a polyclonal antibody obtained from human plasma enriched with HCV polyclonal antibodies collected from screened donors. It is used to prevent the re-infection of HCV infected patients who had a liver transplant [27]. Another approach is to develop a vaccine that triggers passive immunity against HCV. It was found that immunoglobulin G obtained from HCV infected patients could be a good candidate for developing such a vaccine. Vanwolleghem *et al*. injected 8 chimeric mice with immunoglobulin G obtained from HCV infected patients, which were then challenged with 100% infectious dose of the acute phase HCV. Five out of eight mice developed immunity against HCV [27]. Law *et al*. [28] were able to isolate polyclonal antibodies that bind specifically to HCV E2 and neutralize several HCV strains. Full length immunoglobulins G were obtained from 3 antibody-antigen binding fragments that bound to 3 antigenic regions on E2. Monoclonal antibodies were obtained from these immunoglobulins and were injected intraperitoneally in mice with high levels of human liver chimerism. These mice were then challenged by HCV infected human serum, but the monoclonals didn't show the desired [28]. A third approach involves carbohydrate binding agents (CBA) against HCV E2. Since HCV envelope glycoproteins are highly glycosylated, carbohydrate binding agents (CBA) such as cyanovirin-N might prove to be promising in blocking HCV entry into the cells [29]. Several host receptors were found to be involved in HCV entry by either mediating viral attachment (such as LDL-R and Heparan Sulphate) [30] or by interacting with HCV E2 glycoprotein (such as CD81 and scavenger receptor BI (SR-BI)) [31-33]. Based on these findings new approaches were chosen such as developing small molecules that mimic the D-helix of CD81 and is known to interact with HCV E2 [12].

E2 Homology Models

There is no crystal structure available for HCV E2 glycoprotein, nor are there reliable homology models except for that developed by Yagnik *et al*. [34]. It is strongly believed that developing a reliable model of HCV E2 will be considered a break through because it will open another path for new category of anti-HCV drugs based on fragment based drug design and selective high affinity ligands approach.

Yagnik *et al*. [34] applied several fold recognition methods (such as TOPITS and THREADER2) to develop an HCV E2 model using the envelope protein E of Tick Borne Encephalitis virus as a template. In addition they constructed two truncated E2 proteins by PCR. These two proteins (N2 and H strains) were used in heparin and CD81 binding studies. The aim of that study was to map the results obtained from experimental studies they performed onto their E2 model with the goal of identifying important

residues involved inE2:CD81 interaction and a heparin binding domain. The results obtained in this study were also used to propose a rough model for the quaternary structure of the envelope glycoproteins E1 and E2 complex [34]. **Figure 1** shows the developed E2 model of E2 with the significant residues involved in binding to CD81 and the E1:E2 heterodimeric association [34].

The C-terminal ends of the E1 and E2 glycoproteins play important roles in E1-E2 heterodimerization and retention in (ER). Charloteaux *et al.* [35] examined the C-terminal sequences of E1 and E2 in 25 HCV strains and used these sequences to generate several 3-D models. They found out that both domains of E1 and E2 should have a configuration of one amphipathic α helix followed by a pair of transmembrane strands. After generating 3-D models and minimizing the energy, the molecular hydrophobicity potentials (MHP) calculation and the IMPALA procedure (a Monte Carlo minimization of energy based simulation) were used to evaluate the C-terminal domains' interaction with the membranes. They concluded that the created 3-D models were consistent with the experimental studies in that both C-terminal domains are involved in ER retention and in E1-E2 heterodimerization [35].

Yu *et al.* [36] created a structural model of the HCV E2 protein using the Yagnik *et al.* sequence alignment. The template used was the crystal structure of the tick-borne encephalitis virus (TBEV) soluble E protein. The ectodomain of HCV E2 in this model was similar to that of domains I and II of TBEV E protein and HCV E1 glycoprotein was similar to domain II of TBEV E protein. They proposed that E1 and E2 glycoproteins form a tetramer that forms the basic building block of the outer shell of HCV where as in case of flavivirus (TBEV) the outer shell is formed of E-protein dimers (**Figure 2**) [36].

Spiga *et al*. [37] reported a 3-D structure for the HCV E2 homodimer that was developed based on a secondary structure prediction using PsiPred. This was followed by fold recognition using GenThreader v.2.1. software. They used both the TBEV E-protein as a template and the alignment of the consensus sequence of six HCV E2 genotypes (1a, 1b, 2a, 2c, 3a, 4c) in building the E2 model using DeepView v.3.7 software. The model's sequence identity with the template was 14% and the structure exhibited little secondary structure (mainly B-sheets). In addition, molecular docking simulations were conducted to predict the HCV E2 dimeric assembly. The aim of this study was to find surface exposed protein sequences that could be immunoreactive in the HCV E2 glycoprotein (E2 mimotopes) and also common in most genotypes. Nine fragments were selected in the developed model (**Figure 3**), synthesized on nitrocellulose membranes and tested through binding assays using phosphate conjugated anti-HCV human antibodies. This approach was considered important because it might be used to develop diagnostic kits [37].

HCV NS3-4A serine protease

The HCV NS3 protein, a bifunctional HCV enzyme, consists of 2 domains, an N-terminal serine type proteinase domain and a C-terminal NTPase/helicase domain [8, 4, 38]. NS4A is an important co-factor required for the NS3 protease domain to function [4]. Targeting HCV enzymes such as HCV serine protease is considered a promising approach for developing new therapeutics for HCV. Several structures are present in the protein database for NS3 serine protease (ex. PDB ID: 3KN2) facilitating the development of drugs against it. Table 1 shows an example of HCV serine protease inhibitors under clinical trials [39-45].

Host receptors involved in HCV entry

SR-BI

Scavenger receptor class 1B (SR-BI) is a polyprotein composed of 509 amino acids which belongs to CD36 family. Scarselli *et al.* have shown that human SR-BI is one of the putative receptors involved in HCV entry into the hepatocytes [7]. Bartosch *et al*. conducted a study to observe the effect of pre-incubation of the Huh-7 cell line with anti-SR-BI polyclonal antibodies. It was shown that these polyclonal antibodies reduced the infectivity of HCV pseudoparticles (HCVpp) [31]. Another study conducted recently by Lavillette *et al*. has led to the finding that SR-BI expression silencing reduced HCVpp infectivity as well [46]. HDL was found to be a natural SR-BI ligand that is internalized into the cell via non-clathrin dependent endocytosis. This led to the proposing SR-BI as one of the receptors involved in triggering the internalization of HCV into hepatocytes [7]. Catanese *et al*. generated a panel of MAbs against human SR-B1, and 2 of the antibodies (3D5 and C167) were found to block the soluble E2-SR-BI interaction. They prevented the infection of Huh-7.5 with HCVcc in a dose dependent manner. Their study helped in proving that anti-SRBI MAbs might be promising HCV therapeutic agents [5].

LDL-R

Wunschmann *et al*. proposed LDL-R to be a putative receptor for HCV. They found that LDL-R might be involved in binding and entry of HCV into the host hepatocytes [47]. Mas Mraques *et al*.[48] conducted a study aiming at observing the effect of three single-nucleotide alterations within LDLR on the course of HCV infection and response to antiviral therapy. They concluded that these alterations affect the response to combination therapy and the self-limitation of HCV infection [48]. Owen et al.[49] proposed that anti-apolipoprotein E (apoE) antibody and beta-VLDL itself can block HCV entry by preventing HCV interaction with LDL-R thus blocking downstream interactions with other receptors. In

addition, siRNA and treatment with 25-hydroxycholesterol was found to decrease the HCV infection of cells [49].

Tight junction proteins

Two tight junction proteins, claudin-1 (CLDN-1) and occludin (OCLN) were found to be involved in HCV entry into the host cells [50]. Krieger et al. conducted a study to develop anti-CLDN-1 antibodies and observe their effects on HCV-host cell interaction. They found out that these antibodies blocked CLDN-1-E2 interaction in addition to blocking the CLDN-1-CD81 association which is essential for internalization of HCV into the host. They concluded that targeting CLDN-1 might be promising in decreasing the infectivity of HCV by preventing the interaction of HCV E2 glycoprotein with the cell surface in addition to altering the CD81-CLDN-1 interaction [14].

CD81-LEL

Several research groups have found that the CD81- large extracellular loop (CD81-LEL) plays a key role in HCV entry into cells by binding to the HCV E2 glycoprotein [8, 51, 52]. Higginbottom *et al*. and Drummer *et al.* used mutational studies to identify residues that contribute to the E2-CD81-LEL interaction [52]. This information is important because it can be used to direct the design of a series of selective high affinity ligand-based inhibitors that block the E2-CD81-LEL interaction. [52,53]. Zhang *et al.* discovered a separate, additional function for CD81 in the HCV life cycle. These studies showed that CD81-LEL is important for efficient HCV genome replication [54]. In addition, E2-CD81-LEL interaction induces several immuno-modulatory effects, among which include *in vitro* a co-stimulatory signal in naive and antigen-experienced T cells that leads to the production of pro-inflammatory cytokine gamma interferon. Thus the CD81-E2-LEL interaction may play a role in T-cell-mediated liver inflammation and contribute to liver damage. In addition, the interaction of these two proteins down regulates T cell receptors and suppresses the activity of natural killer (NK) cells [53]. Meuleman *et al*. showed that using anti-CD81 as a prophylactic treatment completely protected human liver-uPA-SCID infection with different genotypes of HCV [55].

A novel approach to develop anti-HCV drugs

A novel approach was proposed by Balhorn *et al*. against lymphoma [17-20]. It was based on the generation of selective high affinity ligands (SHALs) which are small synthetic molecules that mimic the

targeting properties of monoclonal antibodies. They are created by identifying two or three sets of molecules using computer docking methods, such as AutoDock, that are predicted to exhibit high binding affinity to the desired binding site in the target protein. One pair or more of these small molecules are linked together producing a series of bi- , tri- or bis-bidentate SHALs. When compared to monoclonal antibodies, they were found to be robust, easy to synthesize, and 1/50th the size of MAbs making blood clearance faster and giving SHALs the ability to reach smaller blood vessels and hence targets more readily. The binding affinity of SHALs are expected to be significantly higher (by a factor of 100 to 1,000,000) than that of each ligand alone. The SHALs developed by Balhorn *et al*. against HLA-DR10 showed a binding affinity of nanomolar to picomolar Kd's to cell lines expressing HLADR10.

In this project, this approach was implemented in the *in silico* design of SHALs against different targets involved in HCV infection. SHALs against CD81-LEL were designed using the data provided regarding the significant residues in CD81-LEL that are considered contact residues for HCV E2. Another set of SHALs were designed against HCV NS3 serine protease targeting the residues involved in the complexation of NS3 with its co-factor NS4A that is important for the function of NS3 serine protease. Our structural model of E2 was created using standard homology modeling methods. Based on an alignment provided by Yagnik *et al*. [34], the AS2TS system was used to create the initial model of E2. The detailed modeling of insertion, deletion and some loop regions was performed using the fragment-based method implemented in LGA_modeler program. Side chains were added by using SCWRL system. Then, in order to remove some steric clashes, and to refine structural model several iterations of MD simulation were performed. This last step in our modeling effort was assisted by using the Procheck system to verify the stereochemical quality of created final E2 model.

Summary and conclusion

HCV is a challenging virus due to its genomic diversity and not having the full picture of its mechanism of entry into the host cells. Several research groups identified important residues in HCV E2, CD81-LEL and HCV serine protease that are important in the completion of HCV life cycle. Using these information different approaches were used to develop drugs for treating HCV infection, either targeting HCV proteins or host receptors. Each of these targets has its advantages and disadvantages. Targeting HCV is better when the drugs target receptors in the liver specifically and prevent adverse events in other organs that might affect the normal homeostasis of the body is not a desired outcome, but targeting the host receptors is better when dealing with the continuous mutation of HCV and how it could develop resistance to drugs targeting it. For the picture to be complete, the mechanism of HCV entry must be completely unraveled clarifying the exact role of each of the already identified receptors involved in the entry and the ones that are not yet identified. SHALs approach was used in this project to target HCV

infection by designing SHALs against HCV and host proteins, such as HCV NS3 serine protease and CD81-LEL respectively. The anti-NS3 SHAL aims at altering its complexation with NS4A which is an important co-factor for the functioning of NS3 protease. The anti-CD81-LEL SHAL aims at blocking the E2-CD81-LEL interaction and thus preventing HCV entry, replication and propagation in addition to restoring the innate immunity of the host.

CHAPTER 2: INTRODUCTION

Hepatitis C virus

History

In early 1970s researchers discovered that there were more than the two known serotypes of hepatitis, hepatitis A and B. A third type was discovered and it was called non A non B (NANB) hepatitis. This discovery was a result of extensive work, which began by monitoring the participants in the study analyzing an abnormal pattern of aminotransferases bi-monthly for 6 to 12 months. This was followed by transfusion studies that were an eye opener on acute hepatitis. It was believed that the transfusion associated hepatitis would be hepatitis B but it wasn't. This was proven when the samples were found to be non reactive to hepatitis B tests. The stored sera from the transfusion studies were then used to inoculate chimpanzees, which later on developed biochemical abnormalities (e.g. Aminotransferases) indicating that NANB type was a transmissible agent. After 15 years of research, hepatitis C was identified by Houghton *et al.* [2]. Specific serological tests were developed and a comparison between NANB and hepatitis C confirmed that they are the same [56].

HCV genomic diversity

There are six major genotypes of HCV worldwide with RNA sequences that differ by >30 %. These genotypes are further classified into subtypes, isolates and quasispecies [57]. The continuous mutation of HCV is due to lack of efficient proof reading by the RNA dependent RNA polymerase in addition to the lack of the 5' to 3' exonuclease activity which leads to accumulation of mutations and viral escape [58]. The host immune's response is one of the major factors that affect the persistence of HCV infection leading either to viral clearance or to chronic infection. It is believed that there are positive and negative selection processes in which the beneficial mutations are selected and favoured over deleterious ones [59]. The HCV genotype is one of the most important predictive viral factors for the response to therapy. Other factors affecting response to therapy include sex and age of the patient, the initial viral load and the presence or absence of liver fibrosis [58].

Diagnosis

Biochemical markers such as alanine aminotransferase (ALT) and aspartate aminotransferase (AST) may be elevated in the blood of patients with HCV and thus indicate the need for further testing. However, 40% of infected patients can have normal levels of ALT and AST making these enzymes of low significance in the diagnosis of HCV [60]. Fibro Test-Acti Test, is a non-invasive blood test and an alternate to liver biopsy for determining the level of fibrosis and necroinflammation that may be caused by HCV [61].

Immunoassays are widely used for diagnosis of HCV infection. Four generations of enzyme immunoassays that detect antibodies against epitopes of different HCV proteins have been developed and used for serodiagnosis of HCV [61]. The core antigen has been proposed as a plasma marker for use in confirming an active infection [62]. Trak-C is an immunoassay that can detect free and antibody-bound core antigens and is used for screening blood donors and monitoring HCV treatment [62, 63].

Molecular assays are crucial for the management of patients infected with HCV. Qualitative RNA tests are mainly used to confirm the presence of active HCV infection and for screening of blood donors. Quantitative tests are used for monitoring the response to therapy and prognosis. Genotyping tests guide therapeutic strategies and aid the prediction of patient response to therapy [62].

Current treatment

HCV is currently treated using peginterferon (PEG-IFN) and ribavirin (RBV) [64]. Great efforts are being carried out to produce new drugs with higher efficacy, and better patient compliance by reducing their side effects, frequency of administration and easier routes of administration. Among the classes that are being developed are NS3 protease inhibitors, NS5B polymerase inhibitors, ribavirin analogs and cyclophilin inhibitors [65].

Epidemiology

Hepatitis C virus infects approximately 3% of the world population. In the developed countries, the chronic HCV infection accounts for 70% of the patients, and acute HCV for 20%. 40% of the infected patients' progresses to end stage cirrhosis, 60% to hepatocellular carcinoma and 30% require liver transplantation [66]. Chronic hepatitis C prevalence ranges from 0.1% to 14.5% in different countries [67, 1001,1002] and Egypt has the highest HCV prevalence worldwide (14.5%) [1002]. According to the WHO, about 20% of Egyptian blood donors are anti-HCV positive [1001].

HCV genome and proteins

HCV is a positive strand-RNA virus possessing a genome of approximately 9.5 kilobases which encodes a large polyprotein containing 3010 aminoacids [4]. The processing and cleavage of the polyprotein is performed using the machinery of the host and enzymes produced by the virus. Ten structural and non-structural proteins are produced upon cleavage of the polyprotein. Either pathway involves peptidase cleavage signals (**Figure 4**) [4]. There are two untranslated regions (UTRs) one at the 5' end and one at the 3' end of the RNA [4, 68]. The structural proteins are the envelope glycoproteins denoted E1 and E2 and the core protein. There are two other proteins of unknown significance known as "p7" and "F" proteins [4, 69]. As for the non-structural proteins; they are denoted NS2, NS3, NS4A, NS4B, NS5A, and NS5B [4, 69]. The structural proteins are released by the host machinery in the endoplasmic

reticulum, whereas the non structural proteins are released following cleavage by NS2-3 and NS3/4A protease. The C terminus of the capsid protein undergoes further processing triggered by a signal peptide peptidase [68]. Table 2 provides a brief description of the HCV proteins and their functions. Since the main HCV targets in this phase are HCV Envelope glycoproteins, and HCV NS3/4A complex, the next few paragraphs will provide more details on their nature and function [68].

a) Envelope glycoproteins

E1 and E2 glycoproteins are type I transmembrane proteins that possess a C-terminal transmembrane domain and an N-terminal ectodomain that is N-linked to glycans. The glycans were found to play a role in virus entry and glycoprotein folding [4]. Since HCV envelope glycoproteins are highly glycosylated, carbohydrate binding agents (CBA) such as cyanovirin-N might prove to be promising in blocking HCV entry into the cells [29]. Both proteins are cleaved from the polyprotein by host signal peptidase cleavage [3]. E1 and E2 form a natural non-covalent heterodimer that is retained in the endoplasmic reticulum (ER) [3, 4]. Their retention in ER contributes to HCV infection conversion into the chronic form because it cannot be detected by the host immune system and also proves that the virus is released from the host cells by budding and exocytosis [4]. Several host receptors were found to be involved in HCV entry by either mediating viral attachment such as LDL-R and Heparan Sulphate [30] or by interacting with HCV E2 glycoprotein such as CD81 and scavenger receptor BI (SR-BI) [31-33]. These interactions trigger HCV movement to the tight junctions and its uptake by Claudin-1 and its uptake via Claudin-1 and occludin [30]. Two hypervariable regions (HVR) have been identified in the HCV E2 glycoprotein sequence. HVR-1 is formed from the first 27 amino acids of the E2 N-terminus (ectodomain), and it largely contributes to the escape of the virus from the host immune response through what is called the viral-host evasion strategy by generating “quasispecies”. HVR-1 has also been shown to modulate HCV entry. As for HVR-2, it has been described as an HCV entry modulator [3, 4]. Viral glycoproteins such as HCV E2 are considered to be a promising target to generate drug against because it plays an important role in host cell interaction and thus if blocked will dissect the HCV life cycle in its earliest phase. To achieve the objective of generating SHALs against E2, several HCV E2 homology models were constructed based on the previously developed homology model by Yagnik *at al*. [34] and using different structure prediction systems [69-71].

b) NS3/4A protein

The HCV NS3 protein, which is bifunctional, consists of 2 domains, an N-terminal serine type proteinase domain and a C-terminal NTPase/helicase domain [3, 4, 15, 72]. NS4A is an important co-factor required for the NS3 protease domain to function. NS4A participates in the folding of the N-

terminal protease domain of NS3 by contributing one of its beta strands to the final structure. NS4A also leads to the change in the NS3 protease domain conformation, repositioning the catalytic triad. When NS3 is not associated with NS4A it is found to diffuse throughout the cytoplasm and nucleus, and this is attributed to not having a transmembrane domain. On the other hand NS4A is a membrane protein, thus by non-covalent association between NS3 serine protease and NS4A central domain, it will be associated with ER and ER-like membranes. In addition, NS4A stabilizes the NS3 protease domain against proteolytic cleavage [4]. The NS3-4A protease complex is essential for cleaving the remaining HCV NS proteins (NS4B, NS5A and NS5B "RNA dependent RNA polymerase") from the polyprotein. When HCV NS3-4A protease is inhibited, downstream NS proteins cleavage is halted preventing replication, and thus blocking further propagation of HCV [73]. Many companies and research groups have developed HCV protease inhibitors using structure-based drug design [74]. These inhibitors could not be developed until the HCV protease crystal structure became available in 1996 [75]. The NS3/4A complex is considered a significant target for development of anti-HCV drugs [15].

CD81 host receptor

Several research groups have found that the CD81- large extracellular loop (CD81-LEL) is the domain involved in HCV entry via interaction with HCV E2 glycoprotein [76, 38-40]. Higginbottom et al. identified 4 amino acid residues that are of large significance in the E2-CD81 interaction using mutational studies. They found that the D196E mutation reduced the binding to E2. In addition mutations F186L and E188K inhibited binding of CD81 to E2 where as T163A enhanced their fusion [8] (**Figure 6A**). Drummer et al. identified an E2 binding site in CD81-LEL to occupy around 806 $A^{\circ 2}$. The most important contact residues they found within this site were Ile182, Phe186, Asn184, and Leu162 [53] (**Figure 6B**) [77-81]. The identification of these residues provides new targets for the design of selective high affinity ligands that could inhibit E2-CD81 interaction. Zhang et al. also discovered a novel function for CD81 in HCV life cycle. They found that CD81 is important for efficient HCV genome replication [15].

This same group also discovered that the E2-CD81 interaction had several immuno-modulatory implications, among which include inducing in vitro a co-stimulatory signal in naive and antigen-experienced T cells leading to production of pro-inflammatory cytokine gamma interferon. Thus the CD81-E2 interaction might play a role in T-cell-mediated liver inflammation and lead to liver damage. In addition, the interaction of these two proteins down regulates T cell receptors and suppresses the activity of natural killer (NK) cells [53]. Brazzoli *et al.* found out that CD81 is a central regulator for several events within the HCV life cycle. CD81 was found to activate Rac, Rho, and Cdc42 (Rho GTPase family) mediating the contact between HCV E2/CD81 complex with CLDN-1 and

occludin tight junction proteins to internalize HCV into the cell. Thus inhibiting CD81-HCV E2 interaction will block these events as stated and will massively reduce the infectivity of HCV [82].

Proposed model of HCV entry into hepatocytes

HCV entry into the host cells involves several host receptors such as CD81 [33, 83, 84], the tight junction protein Claudin-1 (CLDN-1) [6, 14, 33, 83], scavenger receptor class BI [8, 33, 84]. LDL-R [33, 83], mannose binding lectins DC-SIGN and L-SIGN [33], heparan sulphate proteoglycans [30, 33] and the asialoglycoprotein receptor [33].

Moradpour *et al.* [85] proposed that HCV particles become associated with low density lipoproteins (LDL) and very low density lipoproteins (VLDL). Thus it was concluded that LDL receptors are involved in the entry. Both glycosaminoglycans (GAGs) and LDL-R function as primary receptors that mediate the attachment of HCV to CD81 and SR-BI [30, 85]. CD81 and SR-BI trigger HCV movement towards the tight junction protein; CLDN-1 which act as co-receptor [85]. CLDN-1 act at a late stage of entry which is followed by the internalization of HCV through Clathrin mediated endocytosis. It is suggested that HCV then moves inside the host cell in a low pH endosome where the acidification of the endosome helps mediate an HCV glycoprotein: endosomal membrane fusion (**Figure 5A**) [85]. It was found that the envelope proteins of other flaviviruses and alphaviruses (such as Semliki Forest virus and Dengue virus) possess an internal fusion peptide (class II fusion protein) that rearranges and trimerizes when exposed to low pH. This led to an assumption that HCV could involve class II fusion proteins in its internalization process [85, 86]. Budkowska [30] proposed a recent model of HCV entry based on the interaction of HCV: VLDL complex with LDL-R and/or heparan sulphate proteoglycan. It was also proposed that HCV might interact directly with GAGs and/or SR-BI. SR-BI then approaches and forms a complex with CD81 which acts as a post entry receptor for HCV triggering signaling cascades that will help internalize HCV into the host cell. This is followed by the movement of HCV particles to CLDN-1 and OCLDN. OCLDN is considered as “the final entry key” in the HCV internalization process (**Figure 5B**) [30].

Selective high affinity ligands

Selective high affinity ligands (SHALs) are small synthetic molecules that mimic the targeting properties of monoclonal antibodies mainly in their targeting properties [87]. They are created by identifying two or three sites and corresponding sets of molecules using computer docking methods such as AutoDock that are predicted to exhibit high binding affinity to the desired binding site in the target protein. The components of SHALs are the best 2, 3 or 4 ligands of choice, and the linkers. There are specific criteria of selection that the top ligands must satisfy (**Table 3**). They are as follows:

General criteria:

a. The ligands must have a free amino or carboxyl group
b. If the ligand has both one carboxyl and one amino group it would be better than just one carboxyl or amino because if both groups are present the charged group will increase the ligand's solubility after modification and attachment to the linker.
c. The ligands can't have 2 or more amino or carboxyl groups because of the difficulty of controlling which group will attached to the linker.

Specific criteria:

a. The free carboxyl or amino group of the ligand must be pointing away from the surface and is not predicted to interact with the protein
b. The free amine or carboxyl must point toward the second site.
c. Highly hydrophobic molecules are not favourable because they will reduce the solubility of the final product and might bind in more than one region.
d. Compounds that are known to be highly toxic must be avoided.
e. There must be a commercial source from which the ligand can be purchased because synthesizing the ligands may be difficult.
f. Detergents must be avoided because they tend to bind to many proteins.
g. Try to avoid compounds that might bind to more than one site on a protein (e.g. deoxycholate) must be avoided.
h. Compounds that are known substrates, inhibitors or activators of other proteins/enzymes mustn't be selected, because the ultimate goal is to design selective SHALs against the desired target and thus decrease the side effects, but many small molecules that are commercially available happen to be non specific.
i. Avoid peptides (natural sequences; FMOC, carbobenzoxy or other modified amino acids could be used because their side chains are modified so as that proteases can't detect it) because they can be cleaved by proteases and they will be unstable.
j. Molecules that seem to interact with the surface (e.g. charges on surface and uncharged regions) in such a way that they are oriented in a particular direction should be selected. If they could flip and still bind due to symmetry or extend up into the solvent, they are not suitable. Because we need good points of interaction that we must be sure that fit inside the binding site, not on the surface of the protein exposed to solvent. But they should not be excluded specially for the third site.
k. Compounds that might exist in different forms (e.g. enol vs ketol, etc) must be avoided. Because here it means that half of the population will be in the wrong tautomers.

l. Compounds that appear to interact through several points to the cavity could be selected because they will increase the binding affinity to the protein. The more context the higher the affinity.
m. Compounds with functional groups that are highly reactive to the protein (to amines, carboxyls, etc) should not be chosen unless their reactivity can be used to create a permanent bond once the ligand binds.
n. Ligands that are known to bind tightly to blood proteins (e.g. iodine containing compounds) must be avoided. The body has a system to recover iodide very efficiently and thus blood proteins tend to interact with iodinated compounds, where they become sequestered for long time which is totally unfavourable and non specific.
o. Compounds that have disulfides in them (-S-S-) should not be used because they will be split by reducing agents.

The linker design is determined based on several factors:

a) The distance between the ligands of choice
b) The location of the free amino or carboxyl group present in the ligand
c) If the 2 groups in the two ligands are different (one amino and one carboxyl) or the same (2 amino or 2 carboxyl)
d) The end result which is the desired dosage form and method of administration being oral or intravenous (IV).

The components of the linker could be a combination of any of the following according to the above factors (**Figure 7**):

- Polyethylene glycol (PEG):

  PEG has several advantages when being used as a linker component. A PEG unit short is (≈ 10 Å) and two or more PEG units can be linked together to increase the linker's length. PEG is inexpensive, and several clinical studies have shown that it is safe and causes no significant side effects [88]. Also the 2 oxygen atoms present in the PEG moiety, increase the solubility of the SHAL and may contribute to the SHALs binding affinity by being able to form H-bonds with the protein. PEG is used mainly to give length to the linker.

- Lysine

  Lysine is used as a branching point making SHALs mimic the structure of an antibody. Lysine could be replaced with glutamic acid in the following cases:

a. If the distance between the 2 ligands was shorter than the length of a lysine
b. If both free groups in the 2 ligands were amino groups since glutamic acid has 2 carboxyl groups.

- Carbohydrate (CHO) linker to increase drug bioavailability when administered orally [89]:

  Points to consider:

  a. It is absorbed better in the gastrointestinal tract
  b. It is more soluble due to having more hydroxyl groups
  c. Its disadvantage is that since there are many hydroxyl groups, it is hard to control and block them and this can complicate the synthesis. Another disadvantage is the instability of some of the rings that may open up during the synthesis of SHALs.

Protein structure modeling methods

SHALs can be developed using either computational approach entirely in order to speed up the process or a completely experimental approach (i.e. High throughput screening methods) to identify the target. In the computational approach a reliable 3D structure of the target protein is used. There are several computational protein modeling techniques that can be used to obtain a good model of a protein that doesn't have a crystal structure by X-ray, NMR or others. These programs are helpful when there are sets of data available regarding the target protein and other homologous proteins from other organisms to serve as templates. We applied some of these techniques to obtain reliable models of HCV NS3 serine protease for different HCV genotypes and compares them to crystal structures present. In addition, although HCV E2 is one of the most challenging proteins to be modeled, several models were created using different protein structure prediction techniques (**Table 4**).

Amino acid Sequence to Tertiary Structure (AS2TS) system is one of the tools used in protein structure modeling [90]. It translates the amino acid sequences into standard protein databank (PDB) atomic coordinates. This system uses the closely related proteins from PDB to produce a set of 3D models for the desired target protein like HCV E2 or HCV NS3 serine protease in our case.

Protein Structure and Function prediction Resource (PSiFR) is another system used in predicting the structure and function of proteins. Some proteins are of unknown functions, and thus special tools are needed to predict their function. The protein functional annotation depended mainly on the comparison of different protein sequences with known and unknown functions so as to try to find proteins with nearly similar functions. The need for new computational functional annotation tools based on structure and not sequence became needed recently due to the appearance of new proteins with sequences that had no similarity with other proteins, in addition, the protein tertiary structure is more conserved than the protein

sequence. PSiFR is a system based on TASSER threading technique for protein structure and function prediction [70, 71].

ROSETTA is one of the software packages that is also used for protein structure and function prediction. It is used to predict protein structures in the presence or absence of experimental data of the target protein. In addition, it is used for protein-protein and protein-fragment docking experiments together with designing new proteins and redesigning previously existing ones. ROSETTA works by identifying the relevant conformation space when predicting the structure of a protein, and sequence space in case of protein design. This is followed by evaluating the energy of the obtained structural models [91].

Protein Homology/analogY Recognition Engine (Phyre) is a fold recognition-based protein structure prediction system [72]. Structural Classification of Proteins (SCOP) database is the source of known protein structures used by Phyre server. These known structures are updated using newer depositions in the Protein Data Bank (PDB). Each of these structures' sequence is scanned against a non-redundant sequence database obtaining a profile which is then saved in the 'fold library', where both the predicted structure and experimental structure of this protein are stored [72].

Besides modeling, it is important to compare the different protein structures, models and fragments so as to obtain a reliable protein classification. Local-global alignment (LGA) is a method used to compare different protein structures and fragments. An LGA scoring function could help classify groups of proteins in addition to ranking the level of similarity between two structures [92].

STRucture ALignment-based Clustering of Proteins (STRALCP), is an algorithm that is used to determine regions in a given set of proteins that possess structural similarities. Those regions are then used for clustering, yielding protein classification schemes based on these structural similarities. STRALCP algorithm has several advantages such as its ability to cluster the same SCOP family separately by detecting the structural differences between their domains. Another advantage is being able to group proteins in different clusters based on the difference in length or if multi-domain proteins are found in different conformations [93] such as CD81-LEL which was found to be present in 2 different conformations [94].

Critical assessment of methods of protein structure prediction (CASP) is an experiment performed every year by the protein structure prediction groups throughout the world to assess different protein structure prediction methods. To date eight CASP workshops were conducted to determine the capabilities and limitations of current tools used in modeling protein structure from sequence. In addition, they are used to track the different improvements done in these tools. Assessments of methods are based on analyzing large number of blind predictions of protein structure [95].

Most drugs developed or are being developed to disrupt ligand-receptor or protein-protein interactions. Thus, the presence of the experimental structure or molecular model is critical for the development of these kinds of drugs. If a structure is not available, computational molecular docking techniques are applied to protein models so as to predict potential bin ding sites, determine the different modes of interaction between the ligand and the target protein, and determine the binding affinity and energy of different ligands.

AutoDock

AutoDock is a group of tools that have been used for many years to perform virtual screens of small molecules binding to proteins molecular modeling [1003], and predict ligand-receptor interactions. It consists of two main programs. AutoDock is used to dock the ligand to the region of choice pre-defined by grid maps. AutoGrid, the second program, pre-calculates these grids [1003]. AutoDock Tools (ADT) is the graphical user interface [95, 96, 1003] that helps visualize the grid box and determine the desired site, in addition to analyzing the dockings [1003]. The scoring function of AutoDock uses a subset of the AMBER force field, implementing the united-atom model [95, 96]. AutoDock has many applications such as fragment based drug design, lead optimization, virtual screening, protein-protein docking, and combinatorial library design [1003]. In this study AutoDock was used to prepare the CD81-LEL receptor, calculate the grid map files and perform virtual ligand screening using the National Biomedical Computation Resources (NBCR) cluster [1004]. AutoLigand, a recently developed AutoDock tool, is used to predict potential binding sites in known protein structures. It works by finding the best fill points which have the largest possible interaction energy with the protein and accordingly help find the best virtual screening hits [97]. The top hits in this study were 80% superimposed with the results of AutoLigand. This could further help in determining the best ligands that could design promising SHALs.

CHAPTER 3: MATERIALS AND METHODS

I. Analysis and homology modeling of HCV proteins

The sequence of HCV genotype 1b (isolate H) was used to develop homology models for some of the HCV proteins and categorize the models that could be created using this sequence according to the availability of reliable crystal structures which will affect the accuracy of the obtained model (**Table 5**). The sequence was obtained from **UniProtKB/Swiss-Prot P27958; POLG_HCVH (Figure 8)**. The AS2TS system [90] was used to obtain models for HCV core protein (**Figure 8A**), E1 glycoprotein (F**igure 8B**), NS2-3 protease (F**igure 8C**), NS3 serine protease (Fi**gure 8D**), NS4A (F**igure 8E**), NS4B (**Figure 8F**), NS5A (F**igure 8G**) and HCV RNA polymerase (NS5B) (F**igure 8H**). The AS2TS server is available online at: http://proteinmodel.org/AS2TS/as2ts.html.

II. Homology modeling of HCV E2 glycoprotein

Theoretical models were created using several modelling systems [70, 72, 90] and also using the HCV E2 alignment published by Yagnik et al. [34] (**Figure 9**). The AS2TS system [90] was used to obtain an HCV E2 3D model (tertiary structure) based on the alignment model of Yagnik et al. [34] AS2TS was also used to create a set of additional draft 3D models based on sequence alignments calculated by programs such as Smith–Waterman [99], FASTA [100], BLAST and PSI-BLAST [101]. HCV E2 sequence was first threaded through a representative PDB structure library (with a pair-wise sequence identity cut-off of 70%) using TASSER system [70] to search for the possible folds. The continuous fragments are obtained by excision of the threading aligned regions and reassembling them into full-length models. Ab initio modeling (Rosetta method [91]) was used to build the missing regions [70]. The full models are then created by docking all the obtained domains together. Metropols Monte Carlo simulation was used to do the docking of the domains. The aim of the overall process is to obtain the best domain orientation that is similar to the full length I-TASSER model but has minimum steric hindrance [70]. The server used is: http://zhang.bioinformatics.ku.edu/I-TASSER. The same sequence was submitted to the Phyre server http://www.sbg.bio.ic.ac.uk/phyre/ [72].This was followed by viewing the PSI-Blast pseudo-multiple sequence alignment to see the results of scanning the sequence against an up-to-date non redundant protein sequence library. The number of homologs with low E-values was identified to verify that the used alignment is highly reliable. This is followed by alignment interpretation, structure homology detection fold recognition, and finally by viewing the “Alignment Accuracy” to determine the predicted accuracy at each position of the alignment [72]. After obtaining the five models (F**igures 10A-E**), one of the CD81-LEL binding regions identified by Yagnik *et al.* [34] was selected and highlighted using AutoDock Tools 1.5.6 in each model (F**igures 11A-E)**. The aim of this step was to determine the percentage of similarity in this binding site in the five models.

III. Homology modeling of HCV NS3 serine protease

The AS2TS system [89] was used to obtain HCV NS3 serine protease models (tertiary structure) for the 6 HCV genotypes. The sequences were obtained from pubmed (http://www.ncbi.nlm.nih.gov/pubmed/) where the accession numbers of the polyproteins of genotypes 1 to 6 are AAD56196, YP_001469630, YP_001469631, YP_001469632, YP_001469633, and YP_001469634 respectively. The regions corresponding to NS3 serine protease in genotypes 1 to 6 were used to create the models included the amino acid residues : 1056-1204, 1060- 1206, 1062-1210, 1057-1204, 1057-1205, and 1061 to 1209 (**Figures 12A-F**). After obtaining the 6 models of NS3 serine protease, they were analyzed using the LGA system [92]. LGA calculates the best superposition between proteins in addition to identifying the regions of local similarity between the compared models. The system is available at: http://proteinmodel.org/AS2TS/LGA/lga.html [92].

Another step was done simultaneously with modeling, which is sequence alignment of the 6 protein sequences used in modeling of NS3 serine protease. ClustalW [102] was used to perform multiple sequence alignment of the used protein sequences from the six genotypes. The best match was calculated for the selected sequences and they were lined up both as text (F**igure 13A**) and in graphical representation (F**igure 13B**) to make it easier to see the similarities between the sequences, the conserved regions and the differences.

IV. *In silico* design of SHALs against CD81-LEL opened (PDB ID: 1G8Q) and closed (PDB ID: 1IV5) conformations

Preparation of CD81-LEL structure and prediction of binding sites

AutoDock Tools (ADT) 1.5.6 [96, 97,1003] was used to prepare the CD81-LEL crystal structures obtained from the Protein Data Bank (PDB) by deleting water molecules, adding polar hydrogens, assigning Gasteiger charges, and creating grid bounding boxes with a 1 A spacing for use with AutoLigand and a 0.375A spacing for use with AutoDock 4.2. ADT was also used to prepare the positive control ligand for the docking experiment by detecting its root and selecting the number of torsions. AutoGrid 4.2 [1003] was used to pre-calculate the grids and create the map files that were used by AutoLigand to predict the potential binding sites in CD81-LEL. AutoLigand employs a grid-based representation of carbon, oxygen, and hydrogen atom centers to identify atomic centers of ligand atoms that could bind to atoms on the protein surface and maximize the ligand's affinity within the given site. AutoLigand can also be used to identify the best binding site on a protein by filling in a contiguous set of affinity points using three steps: a) flood fill, b) local migration, and c) ray casting [98]. By varying the number of fill points, AutoLigand determines the best total affinity per volume of the fill as the best binding site. To identify potential ligand sites, AutoLigand was used to scan the entire CD81-LEL protein using fill

sizes from 10 to 210 fill points. The number of grid points in X,Y,Z (grid box) that were used to guide AutoLigand was 40, 18 and 38 respectively. The grid centre points of the X,Y, and Z coordinates were 3.144, 34.966, and 15.812 respectively.

Virtual screening

AutoDock 4.2 [96, 97, 1003] was used to perform virtual screening runs using a subset of the ZINC small molecule database containing 10,000 molecules taken from the National Cancer Institute-Diversity Set II (NCI_DSII), Sigma, and Asinex libraries. The parameters were set at 100 for the number of genetic algorithm (GA) runs, 150 as the population size, and a maximum number of generations of 25000. The Lamarckian genetic algorithm in AutoDock was used to perform the docking experiments. Docking results were sorted by the lowest binding energy in addition to specific ligand selection criteria that would facilitate the design and synthesis of the best SHALs. The virtual screening runs were performed using the National Biomedical Computation Resources (NBCR) computer cluster [1004]. Vision [85] was used to construct the computational workflows that were used for virtual screening using the NBCR cluster.

Ligand evaluation and in silico design of SHALs

Several criteria were considered in the selection of the ligands used to design SHALs. During the initial examination of the list of ligands predicted to bind to each site by AutoDock, only ligands containing one free carboxyl group or one amino group (or one of each) were selected. In the most highly ranked cases, these amino or carboxyl groups were not buried in a cavity nor did they interact with the protein surface. They were exposed to solvent and were predicted by AutoDock to bind to the protein with the functional group pointed in the general direction of the second ligand binding site. Preference was given to ligands that were predicted to form multiple contacts with atoms or amino acid residues in or around the perimeter of the targeted cavities. Ligands containing two or more free amino or carboxyl groups were only considered for use in creating SHALs that might need additional charge to increase their solubility. Molecules that were highly hydrophobic, highly charged, known to be toxic, exist in more than one form (such as enol-keto forms), or contained disulfide bonds were avoided. After filtering the ligand sets to remove the molecules that did not meet these criteria, the predicted binding energy was used to further rank the top hits. A set of ligands predicted to have the lowest binding energy were then used to design two prototype SHALs. The distance between the functional groups on pairs of ligands that would be linked together was measured and a linker of suitable length was designed using a combination of lysine and miniPEG molecules. The linker could be varied by incorporating PEG moieties between the ligands to adjust the length of the linker and by inserting lysine residues at key points to enable linker branching [17].

V. *In silico* design of SHALs against HCV NS3 serine protease genotype 4

This part started by preparing the HCV NS3 serine protease genotype 4 model and prediction of binding sites using AutoDock Tools (ADT) 1.5.6 [96, 97,1003]. The number of grid points in X, Y, Z (grid box) that used to guide AutoLigand were 78, 44 and 46 respectively. The grid center points of the XYZ coordinates were 24.293, 67.762 and 35.552 respectively. AutoDock 4.2 [96, 97,1003] was then used to perform virtual screening runs against around 10,000 molecules taken from the National Cancer Institute-Diversity Set II (NCI_DSII), Sigma, and Asinex libraries. The parameters used were 100 as the number of genetic algorithm (GA) runs, 150 as the population size, and a maximum number of generations of 25000. The Lamarckian genetic algorithm 4.2 was used to do the docking experiments. Docking results were sorted by the lowest binding energy in addition to specific ligand selection criteria that would facilitate the design and synthesis of the best SHALs. The virtual screening runs were performed using the National Biomedical Computation Resources (NBCR) clusters [1004]. Vision [81] was used to construct the computational workflows that were used for virtual screening using the NBCR cluster. The virtual screening was done against the region involved in NS3-4A complexation (**Figures 14).** This was followed by analyzing the obtained results from virtual screening runs and identifying the virtual screening hits that fit into the criteria of selection of ligands suitable for generating SHALs as stated in (**III.c**). Where molecules that were highly hydrophobic, highly charged, known to be toxic, exist in more than one form (such as enol-keto forms), or contained disulfide bonds were avoided. After filtering the ligands and obtaining the top hits according to the criteria of selection and lowest energy, the distance between the functional groups that would be linked together on the two bound ligands was measured and a linker of suitable length was designed using a combination of lysine and miniPEG molecules. Each linker's length could be varied slightly using the PEG moieties to adjust the length of the linker and using the lysines to enable linker branching [17].

CHAPTER 4: RESULTS AND DISCUSSION

I. General overview of structural modeling and analysis of HCV proteins (genotype 1a isolate H)

Some of the HCV proteins represent a big challenge to model and start targeting to generate anti-HCV drugs. It is worth trying and exerting effort to obtain reliable models because it is believed that if these proteins were targeted they will help generate highly specific anti-HCV drugs. As seen in **table 5**, HCV proteins were categorized by the AS2TS system [90] according to the availability of their crystal structures and enough data to generate reliable homology models of these proteins. In general models that lie in category A are of good accuracy due to sequence identity with other available structures > 45% and coverage of > 75% of the overall sequence. Only two HCV proteins lie in this category, they are highlighted in green in table 5, HCV NS3 serine protease (F**igure 8D**) and HCV RNA polymerase (NS5B) (**Figure 8H**). As for category B models, they are of medium accuracy having sequence identity > 20% and coverage of > 75% of the overall sequence, and currently there is no HCV protein that lies in this category. Proteins in category C1 are considered of low accuracy, with sequence identity of > 15% and coverage of > 50% of the overall sequence. Only HCV NS2-3 protease could help create a C1 model (F**igure 8C**). Category C2 models are usually very low in quality or only small fragments could be created. Five C2 models could be obtained based on five HCV proteins which are HCV core protein (F**igure 8A**), E1 glycoprotein (**Figure 8B**), NS4A (**Figure 6E**), NS4B (F**igure 6F**), and NS5A (F**igure 8G**). As for C3 models, they are not reliable models due to lack of enough data, in addition to lack of at least one crystal or NMR structure. HCV E2 and P7 proteins lie in this category which makes it the hardest challenge of all to create models for them.

II. Homology modeling of HCV E2 glycoprotein

Although creating models of HCV E2 glycoprotein is challenging, several models were created (**Figures 10A-10E**) using different modeling techniques. The sequence alignment used was that of Yagnik *et al.* [34]. This group created a truncated HCV E2 homology model and identified significant residues involved in CD81-LEL binding to E2. In addition they identified residues involved in dimerization of E1-E2 [34]. AS2TS was one of the modeling methods we used to create the HCV E2 model where the structural template was the envelope glycoprotein of the tick-borne encephalitis virus (PDB entry: 1SVB). The sequence identity between the E2 and the template was 7.2% which is extremely low (F**igure 10A**). I-TASSER/threading system was another method used to obtain an E2 model which had 6% of sequence identity with the crystal structure of protoporphyrinogen IX oxidase (PDB entry: 1SEZ) (**Figure 10B**). Another model created by TASSER had 8% sequence identity with the polyamine oxidase extracted from yeast (PDB entry: 1XPQ) (F**igure 10C**). PSiFR system was used to create a model that had a sequence

identity of 15% with E.coli enzyme, quinol-fumarate reductase (PDB entry: 1KF6) which is still considered a low percentage (**Figure 10D**). The last method used was Phyre method which created a small fragment that had a sequence identity of 7% with the structural template PDB 1NB4 (F**igure 10E**).

After creating the HCV E2 models, one of the regions identified by Yagnik *et al*. [34] to be significant in CD81-LEL binding to E2 was selected in each model so as to determine the percentage of similarity in all five models (**Figures 11A-E**). It was noticed that the selected region (green region) in the five models had some similarity which means it could be the most reliable part to be targeted and to run virtual screening against. This step will be carried out later on when obtaining more accurate E2 model.

III. Homology modeling of NS3 serine protease glycoprotein

AS2TS was used to create six models of HCV NS3 serine protease corresponding to the 6 HCV genotypes. **Figures 16A** to **16F** represent the HCV NS3 serine protease models for genotypes 1 to 6, respectively. Several types of comparisons were done for these created models. LGA was used for analyzing the models and comparing them to each other on the local basis (specific sequences' similarities) and globally (throughout the whole alignment.) A dendrogram obtained at the end (F**igure 17**) showed that genotypes 1 (**Figure 16A**) and 4 (F**igure 16D**) were the most similar models. It was also concluded from observing the dendrogram that genotype 6 is more similar to genotypes 1, 4, followed by 5 and 3. Finally, genotype 2 is the least similar to all other genotypes. The same order of similarities was confirmed by a purely sequence-based analysis performed using ClustalW (see Figure 13A). Another type of comparison was conducted using ClustalW to identify the conserved regions, and how they differ among the 6 genotypes. **Figure 13B** shows that starting from approximately amino acid 80 in the 6 genotypes the percentage of conserved regions increases leading to the conclusion that this is a good region to target so as to obtain a drug that could act on all genotypes. This could be achieved by either targeting each NS3 serine protease genotype alone and selecting a region taking amino acid 80 as a marker and do virtual screening runs against it or make a model for the whole conserved region as a separate pocket, do virtual screening runs against it as a whole and generate SHALs against it which might then act as therapeutics against all genotypes.

IV. In silico design of SHALs against CD81-LEL crystal structures

Identification of Target Sites on CD81-LEL

Kitadokoro *et al.* determined the 3D structure of CD81-LEL using X-ray crystallography [103,104]. Two different crystal forms of CD81-LEL (PDB Ids: 1G8Q and 1IV5) were reported [103, 104]. In the 1G8Q structure the C and D helices form a cleft-like motif within the site of E2 binding which could be targeted for inhibitor development. The 1IV5 conformation is considered a closed form of the E2 structure

in which this cleft is absent. Any ligands binding to the closed conformation would involve interactions with 1IV5 in more shallow and surface exposed sites than those observed on 1G8Q [94]. Molecular dynamics studies performed by Neugebauer *et al*. have suggested that the 1IV5 structure may be the physiologically relevant conformation. This was attributed to the closure of the cleft in 1G8Q that occurred after around 50 ps of molecular dynamic simulations. The 1G8Q conformation was also considered to be less stable because more amino acid residues were found to be outside the favoured energy region of Ramachandran plot. In addition, 1IV5 structure had more crystallographic contacts, which would suggest that the structure is more stable of the two conformations [94].

In this study we used both conformations as targets for virtual screening runs to identify ligands predicted to bind to the E2 contact residues that were reported previously. Based on mutation studies, Higginbottom *et al*. [52] identified four residues that were considered to be essential for the HCV E2 protein to bind to CD81-LEL. The D196E mutation in CD81 was observed to reduce binding to E2. In addition mutations F186L and E188K inhibited binding of CD81 to E2 where as T163A enhanced their fusion (**Figure 6A**) [52]. Drummer *et al.* identified an E2 binding site in CD81-LEL of around 806 $A^{\circ 2}$. The most important contact residues they found within this site were Ile182, Phe186, Asn184, and Leu162 (**figure 6B**) [53]. The identification of these residues provided the information needed to design a series of selective high affinity ligands to block the E2-CD81 interaction.

The AutoLigand fill points and plot analysis

AutoDock is one of the most widely used docking programs. Recent program modifications and the addition of new computational tools that support the program have made the AutoDock even more powerful. The AutoDock suite contains two main programs, AutoDock and AutoGrid. AutoDock is used to dock the ligand to the region of choice pre-defined by grid maps. The scoring function of AutoDock employs a subset of the AMBER force field, implementing the united-atom model [96,97]. AutoGrid, the second program, pre-calculates these grids [1003]. AutoDock Tools (ADT) is the graphical user interface [96, 97,1003] that helps visualize the grid box and determine the desired site, in addition to analyzing the dockings [1003]. AutoLigand, a recently developed AutoDock tool, is used to predict potential binding sites in known protein structures. It works by finding the best fill points which have the largest possible interaction energy with the protein and using these fill points to identify the best virtual screening hits [98].

AutoLigand [98] was used to select the best binding sites in CD81-LEL. The protein database (pdb) files that are generated represent the fill points that correspond to the different binding sites. AutoLigand identified 5 binding sites in 1G8Q and 4 binding sites in 1IV5. The best binding sites were determined by plotting the total energy per volume (Kcal/mol A^3) vs. the volume of the cavity (**Figure 18**). As shown in Figure 18, the plots contain a different number of points having different shapes and colors.

In the plot, each group of points having the same shape and colour represent one of the predicted binding sites.

One of the binding sites predicted for 1G8Q (F**igure 19A**) corresponded to the region that had most of the E2 contact residues (Ile182, Phe186, Asn184, Glu188, Asp196) [52]. Another group of fill points were found in a neighboring cavity located on the opposite site of the protein (**Figure 19B**). Both regions were predicted to be good binding sites by AutoLigand because they had the largest possible interaction energy with CD81-LEL. Consequently, these two sites were selected as the primary sites for use in small molecule docking and SHAL development. **Figure 20** shows two other regions that were predicted by AutoLigand to be good ligand binding sites. These sites were used to identify additional small molecules that could be linked to ligands predicted to bind to the first two sites to enable the creation of tridentate SHALs (SHALs with three ligands linked together). Tridentate SHALs have been shown to exhibit higher affinities for their protein targets than bidentate SHALs containing two ligands [100]. The higher the affinity of the inhibitor, the more effective the molecule should be in competing for E2 binding.

AutoLigand identified 4 binding sites on the surface of the closed 1IV5 conformation. One of these sites was found to overlap with the region containing the residues that participate in the binding of E2 to CD81-LEL. This site and two others (**Figure 21**) were selected and used in virtual screening runs performed with the 1IV5 conformation of CD81. One unexpected observation made during our analysis of the fill point energy plot for several of these sites was that sites on the closed conformation might prove to be as good as those identified on the open conformation. The predicted binding energies were as low, and in some cases lower, than those predicted for the open conformation binding sites.

Docking and analysis of ligands predicted to bind to the selected sites

More than twenty new molecular docking programs have been developed since DOCK was first introduced by Kuntz in 1982 [34,35]. While these programs use different algorithms to dock two molecules together, a variety of methods for handling the flexibility of molecules and the influence of water, and one of several scoring functions that evaluate and rank the predicted interactions and the final conformations of the molecular components, many of these programs have demonstrated they can predict how molecules interact [106, 107]. A number of the more popular programs, such as AutoDock [1003], GOLD [108], Flex [109], and DOCK [110] are particularly useful when it comes to predicting how small molecules bind to the surfaces of proteins. Run on supercomputers or computer clusters, hundreds of thousands of small molecules can be docked to a site on the protein surface using these programs and evaluated for their potential to bind to the site in as little as 24 hrs. The AutoDock program suite has many

applications, and it is frequently used in fragment based drug design, lead optimization, virtual screening, protein-protein docking, and combinatorial library design [1003].

Docking runs were performed for three sites located on both the open and closed CD81-LEL conformations (1G8Q and 1IV5). The list of ligands predicted to bind to each site were ranked according to binding energy and how well the ligand's atoms mapped onto the fill points for the site. In addition to the fill points defining the rough shape of ligands that would fit best into the cavity, specific fill points are also color coded to identify particular atoms (carbon, hydrogen, nitrogen or oxygen) in the ligand that would interact optimally with the surface of the protein in the regions surrounding the ligand. An example is shown in **Figure 22**. The fill points predicted for this site by AutoLigand are colored red (for hydrogen acceptors such as oxygen or nitrogen), blue for hydrogen atoms, or gray for carbons (**Figure 22A**). One of the better ligands predicted to bind to this site (**Figure 22B**) have atoms that superimpose well with the fill point map (**Figure 22C**) While the superimposition does not need to match perfectly, the points of contact on the protein are considered to be good if the majority of the different atom types in the molecule (75-80%) approximate the same location as the fill points.

The top ranking ligands were further evaluated based on several additional criteria. The most important is the requirement that selected ligands contain a single amino or carboxyl group (or one of each). Ligands that contain either group can easily be conjugated to a linker (miniPEG or a lysine) using the simple peptide-based chemistry used to rapidly synthesize SHALs [18]. The second requirement involved the orientation of the bound ligand. Ligands were only selected if the free carboxyl or amino group were predicted by AutoDock to point out toward solvent and in the general direction of the second binding site. This would ensure the ligand could still bind to the protein when conjugated to the linker, and it would enable both ligands to bind to their respective cavities after being linked together without having to adopt an unfavourable conformation. In addition, those ligands that formed multiple contacts/interactions with the protein (such as hydrogen bonds, salt bridges, Van Der Waals interactions) were considered to be better than those predicted to make only one or two contacts.

Applying these criteria, a set of ligands was selected for two adjacent target sites on both the open and closed conformations of CD81-LEL. Table 6 shows an example of 10 virtual screening hits predicted by AutoDock to bind to Site 1 on the open 1G8Q conformation. These molecules are arranged in a descending order from good to poor. Ligand 1 was ranked as one of the best candidate ligands because it has several atoms that interact with the protein's side chains surrounding the cavity through a series of hydrogen bonds and Van Der Waals interactions. In addition this ligand has one free amino group and one free carboxyl group which enable the ligand to be used to create both bidentate and tridentate SHALs. Ligand 3 also appeared to make good contacts with the protein surface, but it was not ranked as highly due to the expectation that the ligand, which is more hydrophobic, might contribute to

the insolubility of any SHALs that contain the molecule. Ligand 7 made good contacts with the protein surface, but it didn't exhibit the desired selectivity. This ligand was predicted to bind to more than one site on the protein. Ligand 10 was considered poor because it is highly hydrophobic. This would not only limit its solubility, but it might also direct the molecule to concentrate in adipose tissue. The highly conjugated ring structure might also lead to the generation of a metabolic product that could be mutagenic or carcinogenic.

Eight examples of ligands predicted to bind to site 2 and 3 on the open conformation of CD81 LEL are shown in Table 7. Of those predicted to bind to site 2, ligand 1 is considered to be better candidate because the ligand, which is not overly large, makes several contacts inside the binding site in addition to having a free carboxyl group. Ligand 2 was ranked as intermediate because it contained two amino groups. This would help with SHAL solubility, but it would limit how the ligand could be incorporated into a SHAL. Conjugation to the linker could only be performed using the carboxyl group (otherwise multiple products would be produced) and the presence of the two free amines would require the ligand be incorporated into the SHAL as the last synthesis step. Ligand 3 was ranked as fair because it was predicted to have fewer contact points with the binding site. Ligand 4 was ranked as poor, primarily because of its small size, minimal contact with the surface, and its lack of specificity. The same ligand was also predicted to bind to other sites. The ligands shown for site 3 were ranked in the same manner.

Similar results were observed for the ligands predicted to bind to the three sites identified on the closed conformation of CD81-LEL (PDB ID: 1IV5). Even though this conformation appears to have fewer deep cavities that might serve as suitable ligand binding sites, a number of good binding sites were identified by AutoLigand. Tables 8 and 9 show twelve examples of ligands that were predicted to bind to the three best sites shown in Figure 23. We observed for the ligands predicted to bind to sites on the open conformation of CD81-LEL, a similar proportion of ligands were identified that fit well within the binding site and made multiple contacts with the surface. Ligand 1 (Table 9, site 3) is an example of one of the better ligands. This ligand was predicted to form three hydrogen bonds with atoms on the surface of the protein (**Figure 24**). Two or three (depending on which ligand functional group, carboxyl or amino, is used to attach the ligand to the linker) of these hydrogen bonds would be expected to remain even after the ligand is incorporated into a SHAL.

In addition to identifying ligands that can be used to create a series of small molecule inhibitors that block HCV invasion, the molecules predicted to bind to the sites on the two different conformations of CD81 LEL can also be used to help determine which CD81 LEL conformation should be targeted for therapeutic development. The eighteen molecules shown in Tables 6 and 7 and the twelve molecules shown in Tables 8 and 9 will be used in a follow-on experimental study to determine which ligands bind to a recombinant form of CD81 LEL. These experiments should not only confirm the presence (or absence)

of the open or closed conformations, but it may be possible to use the two ligand sets to obtain information on the dynamics of exchange between the two conformational states.

In silico design of SHALs and controlling the length of linkers for optimum *in silico* binding

Both bi- and tri-dentate SHALs have been designed using the ligands predicted to bind to different sites on the open conformation of CD81 LEL. Functionalized mini-PEGs and protected (FMOC) lysine residues were used to link the ligands using a form of chemistry that would generate amide linkages. The distance between two ligands was defined by the number of mini-PEG molecules used to separate them. Lysines were introduced at specific spots to create branch points. Figure 25 shows an example of a bidentate SHAL constructed using two ligands, one predicted to bind to site 1 (ligand 4)(**Figure 25A**) and the other predicted to bind to site 2 (ligand 2) (**Figure 25B**) on the open form of CD81 LEL. To create this SHAL, the carboxyl group of ligand 2 was conjugated to the alpha amino group of a lysine. The amino group of a functionalized mini-PEG was linked to the carboxyl group of the same lysine and the amino group of ligand 4 was conjugated to the carboxyl group at the other end of the mini-PEG (**Figure 25C**). Ligands 2 and 4 were selected because they had the desired free carboxyl and amino groups respectively.

Carbodiimide conjugation chemistry could be used to link free amino and carboxyl groups together because the chemistry is well established (it has been used for many years for peptide synthesis), efficient, inexpensive and fast. A tridentate SHAL was also designed using a similar approach (**Figure 26**). Both SHALs were designed to demonstrate how the ligands might be combined to create candidate inhibitors that bind selectively to CD81. However, before additional effort is put into designing other SHALs for synthesis and experimental testing, the sets of ligands predicted to bind to the various sites must first be tested experimentally to determine which ligands actually bind to a recombinant form of the CD81 LEL protein. The results of the experiments will then be used to identify which combinations of ligands should be used to create the SHALs. Confirming ligand binding by experiment is an essential step in the process because the docking results are only predictions. Based on our previous docking studies, typically 25-65% of the ligands predicted to bind will actually end up binding to the protein.

V. *In silico* design of SHALs against HCV NS3 serine protease genotype 4 model common conserved region

HCV genotype 4 is the most prevalent in Egypt [111], thus HCV NS3 serine protease genotype 4 is the target of choice as a start for the *in silico* design of SHALs. Several important residues in HCV NS3 serine protease have been identified to form the zinc binding site (Cys 97, 99, 145 and His 149) [112].

Also there is the active site (trypsin like motif) which constitutes His57, Asp81 and Ser139 [113]. The region targeted in this study is the NS3-4A complexation region which mainly includes valine, leucine, isoleucine and glycine residues [112]. This region was selected in NS3 serine protease genotype 4 (**Figure 15**). Since this region hasn't been targeted by others it was important to confirm that this region woulad have a good binding site. This was achieved by using AutoLigand to predict all possible binding sites on the protein. The results showed that among the potential binding sites predicted there was one good site in this region (F**igure 27D**- F**igure 28**). This gave us more confidence that this site could be targeted and might help generate promising SHALs specifically against genotype 4. After doing virtual screening runs against several sites within this region and analyzing the ligands according to the criteria of selection, an *in silico* designed bi-dentate SHAL was created as an example where it included PEG and lysine moieties to give length and branching property to the SHAL (F**igure 29**). After doing the experimental studies on the obtained virtual screening hits, synthesizing the SHALs and testing them, the success or failure of such an approach could be confirmed.

CHAPTER 5: CONCLUSION

The newly developed AutoDock tool, AutoLigand, enhanced the functionality of AutoDock and proved to be helpful in identifying potential ligand binding sites on the surface of CD81-LEL. Two sites on the open conformation of CD81 LEL that were selected by AutoLigand were located in regions identified by mutational studies to be within the site of E2 binding. These and other sites on both the closed and open CD81-LEL conformations were used to perform virtual ligand screening runs using AutoDock. Sets of ligands were identified for multiple sites on both conformations of the protein. Ligands predicted to bind to the best sites on the open conformation of CD81 LEL were then used to design a bidentate and tridentate SHAL as a demonstration of the process.

Thirty ligands were also identified for use in a future experimental binding studies aiming at determining which conformation of CD81-LEL, open or closed, should be targeted for inhibitor development. Eighteen of these ligands were molecules predicted to bind to sites located on the open conformation of CD81 LEL. Twelve were molecules predicted to bind to sites located on the closed conformation of CD81 LEL. Our next step will be to test both sets of ligands experimentally and use the results to 1) provide information about whether or not sites in the open or closed (or both) conformations exist long enough for the ligands to bind, 2) provide experimental confirmation that the new AutoDock tools enhanced our ability to select the best ligands, and 3) determine which ligand sets should be used for SHAL design and synthesis. Different combinations of the molecules that bind to CD81-LEL will then be linked together using solid phase chemistry to produce a series of SHALs, and these moleules will be tested to confirm that they bind to the target protein CD81 and to determine their affinities for the protein. The best SHALs will then be tested in a series of HCV cell culture assays using HCV pseudoparticles (HCVpp) to test their ability to inhibit HCV infectivity. Based on the previous work of others, we expect that some of these SHALs will not only block the binding of the virus and virus invasion by inhibiting HCV E2-CD81 interaction [51-54], but may also block the production of cytokines that leads to the induction of liver damage [53] and alter the replication of HCV [54].

The CASP workshops have shown repeatedly that the most accurate modeling approach is the homology modeling method. Thus HCV NS3 serine protease and HCV E2 models were created using several modeling systems. The models of NS3 serine protease are considered the most reliable, whereas HCV E2 was considered a challenging target because there were no homologs available. Thus only models of lower quality could be created. These models need to be optimized when more studies on HCV E2 are conducted to identify its complete 3D structure. Modeling is important when there is no specific experimentally solved protein structure available but through extensive computational searches possible homologs can be identified and used to predict its structure and function.

This thesis had several positive outcomes; one is doing an analytical study on the HCV genome in relation to the ability to obtain accurate models. Six models of HCV NS3 serine protease of the six HCV genotypes were created and multiple sequence alignment was done using ClustalW so as to see the similarities and differences between the enzyme in the 6 genotypes. Another promising point is the creation of HCV E2 model which when optimized and become reliable enough will be a promising target to develop vaccines and powerful anti-HCV drugs.

Working with different protein targets and predicting sites that might be promising using new computational biology tools such as AutoLigand to generate SHALs; might be an effective approach targeting different levels of HCV life cycle such as the entry (CD81-LEL), translation (NS3 serine protease) and replication (CD81-LEL/NS3 serine protease). This approach could be applied to generate diagnostic tests for HCV by acting as carriers and being specific to a certain target, they could also act as vaccines by carrying an immunogenic part of HCV E2, but this approach needs more research so as to see how to make such a vaccine trigger sustained immunity against HCV.

This phase focused on doing an analytical study on the HCV genome in relation to the ability to obtain accurate models. In addition six models of HCV NS3 serine protease of the six HCV genotypes were created and multiple sequence alignment was done using ClustalW so as to see the similarities and differences between the enzyme in the 6 genotypes. Another promising point is the creation of HCV E2 model which when optimized and become reliable enough will be a promising target to develop vaccines and powerful anti-HCV drugs.

Future plans

The next phase aims at conducting the experimental studies that will include the following (**Figure 30**):

1- Binding assays to identify the best ligands that could be used to synthesize SHALs and validate computational hits.
2- Synthesizing SHALs using the solid phase synthesis approach.
3- Binding assays of SHALs to be sure that after the synthesis the binding affinity increased and was specific (negative and positive controls).
4- Cell culture assays to see if the synthesized SHALs inhibited the infectivity of HCV using HCV pp (pseudoparticles) and HCV cc (cell culture).
5- *In vivo* studies using the HCV mouse models [115,116].

By doing the experimental studies, all the computational biology work implemented will be verified and will show to be suitable approach in targeting HCV or not. The results of our modeling and docking studies suggest that this approach will be effective.

CHAPTER 6: LIST OF TABLES AND FIGURES

Table 1. Selective NS3/4A serine protease inhibitors [39-45].

Drug	Clinical trial phase
Telaprevir	**Phase III**
Boceprevir	**Phase III**
TMC435350	**Phase II**
ITMN-191	**Phase I**
MK-7009	**Phase I**
BI201335	**Phase II**

Table 2. HCV proteins reproduced from [4] with permission.

Protein	Molecular mass KDa	Function
Core	21	RNA binding, nucleocapsid
E1	31-35	Envelope glycoprotein, associate with E2
E2	70	Envelope glycoprotein, associate with E2, receptor binding
P7	7	Ion channel
NS2	21	Component of NS2-3 proteinase
NS3	69	N-terminal proteinase domain/ C-terminal NTPase/Helicase domain
NS4A	6	NS3/4A proteinase co-factor
NS4B	27	Induces membrane alterations
NS5A	56-58	Phosphoprotein
NS5B	68	RNA dependent RNA polymerase

Table 3. Comparison between suitable and non suitable ligands for the generation of SHALs.

Suitable ligands	Non suitable ligands
Have free amino and/ or carboxyl group	**Lack an amino or a carboxyl group** **Have 2 amino or 2 carboxyl groups**
The free amino/carboxyl group points towards the second site of choice and away from the surface	**The free amino/carboxyl group points towards the surface and opposite to the desired second site of choice**
Hydrophilic molecule	**Hydrophobic molecule**
Non toxic compounds	**Highly toxic compounds**
Commercially available	**Not available commercially**
Non detergents	**Detergents**
Compounds with high specificity	**Compounds that bind to more than one site**
Modified aminoacids	**Peptides**
Compounds having functional groups with moderate to low reactivity	**Compounds with highly reactive functional groups**
Compounds that lack disulfide bonds	**Compounds that possess disulfide bonds**

Table 4. Different protein modeling and structure analysis techniques and their characteristics [70,71,90-93].

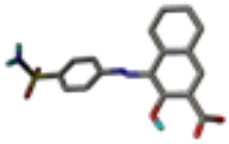
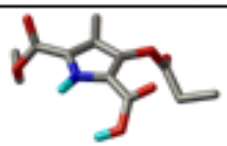
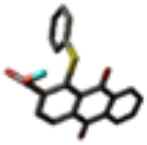
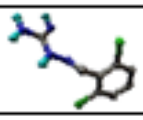
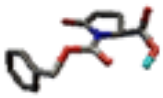
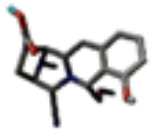
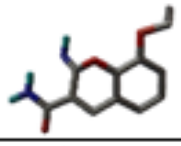
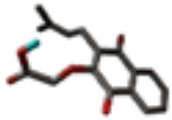
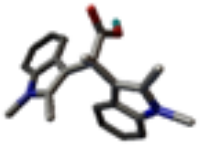
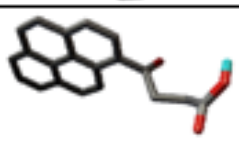
Protein modeling technique	Characteristics
Amino acid Sequence to Tertiary Structure (AS2TS) [90]	Translates the amino acid sequences into standard protein databank (PDB) atomic coordinates using closely related proteins from PDB to produce a set of 3D models for the desired target protein.
Protein Structure and Function prediction Resource (PSiFR) [70,71]	Predicts the structure and function of proteins by comparing different protein sequences with known and unknown functions so as to try to find proteins with nearly similar functions
ROSETTA [91]	Predicts protein structures in the presence or absence of experimental data of the target protein via identifying the relevant conformation space when predicting the structure of a protein, and sequence space in case of protein design.
Protein Homology/analogY Recognition Engine (Phyre) [72]	Folds recognition-based protein structure prediction system by scanning known protein structures against a non-redundant sequence database obtaining a profile which is then saved in the 'fold library', where both the predicted structure and experimental structure of this protein are stored.
Local-global alignment (LGA) [92]	Compares different protein structures and fragments and help classify groups of proteins in addition to ranking the level of similarity between two structures.
STRucture ALignment-based Clustering of Proteins (STRALCP) [93]	Determines regions in a given set of proteins that possess structural similarities. Those regions are then used for clustering, yielding protein classification schemes based on these structural similarities.

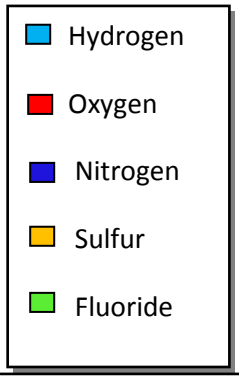
Table 5. The categorization of HCV proteins according to the accuracy of the structural models that could be created using AS2TS system.

Model	Length	Accuracy
Core protein_P21	191	Category_C2
Envelope gp_E1	192	Category_C2
Envelope gp_E2	363	Category_C3 NO
P7	63	Category_C3 NO
Protease NS2-3	217	Category_C1
Serine protease NS3	147	Category_A
NS4A	54	Category_C2
NS4B	261	Category_C2
NS5A	448	Category_C2
NS5B –RNA polymerase	591	Category_A

Grouping criteria (E-value <0.02)			
Categories	Sequence Id.	Coverage	Quality
Category_A	>45	>75	Good
Category_B	>20	>75	Medium
Category_C1	>15	>50	Low
Category_C2	Very low homology or small fragments		
Category_C3	No reliable models were created		

Table 6. Small molecules predicted by AutoDock to bind to Site 1 (main site) on the open conformation of CD81-LEL (figure 3). The molecules are listed according to the assessed quality of the ligand and its interaction with CD81-LEL in descending order. Atom color codes: hydrogen (light blue), nitrogen (dark blue), oxygen (red), sulfur (yellow), and fluorine (green). Criteria used to define the quality of the ligands are: Good - makes more than 5 contacts with protein, predicted to be selective and not predicted to bind to multiple sites, not too hydrophobic; Intermediate - makes 4-5 contacts with protein, hydrophobic interactions contribute to binding; Fair – makes 3-4 contacts with protein; Poor – makes less than 3 contacts with protein, large conjugated ring structures, molecule is hydrophobic and does not fit well inside the cavity.

Ligand	Structure	Potential quality
1		Good
2		Good
3		Intermediate
4		Intermediate
5		Intermediate
6		Fair
7		Fair
8		Poor
9		Poor
10		Poor



Good:

- >5 interacting contact points
- Predicted to be selective based on interaction, hydrophobicity and size.

Intermediate:

- 4-5 interacting contact points
- Predicted to be of less selectivity due to higher hydrphobicity

Fair:

- 3-4 interacting contact points

Poor:

- <3 interacting contact points
- Has large number of rings that doesn't fit well into the binding site, thus not selective.

Table 7. Small molecules predicted by AutoDock to bind to Site 2 and Site 3 on the open conformation of CD81-LEL (figure 3). The molecules are listed according to the assessed quality of the ligand and its interaction with CD81-LEL in descending order. Atom color codes: hydrogen (light blue), nitrogen (dark blue), oxygen (red), sulfur (yellow), and fluorine (green). Criteria used to define the quality of the ligands are: Good - makes more than 5 contacts with protein, predicted to be selective and not predicted to bind to multiple sites, not too hydrophobic; Intermediate - makes 4-5 contacts with protein, hydrophobic interactions contribute to binding; Fair – makes 3-4 contacts with protein; Poor – makes less than 3 contacts with protein, large conjugated ring structures, molecule is hydrophobic and does not fit well inside the cavity.

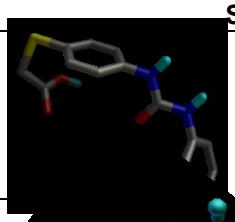
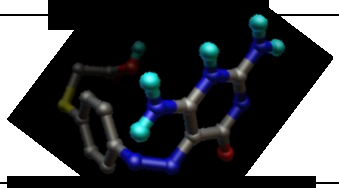
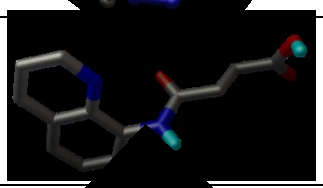
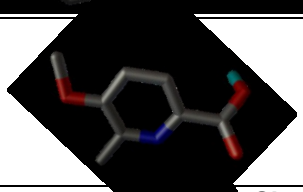
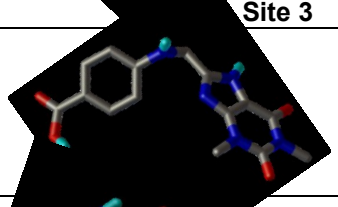
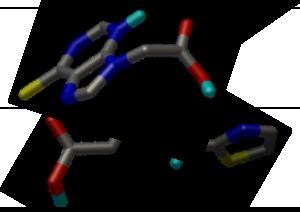
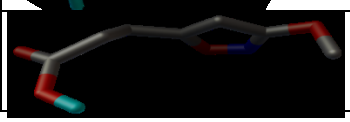
#	Ligand	Potential quality
	Site 2	
1		Good
2		Intermediate
3		Fair
4		Poor
	Site 3	
1		Good
2		Intermediate
3		Fair
4		Poor

Table 8: Small molecules predicted by AutoDock to bind Site 1 on the closed conformation of CD81-LEL. The molecules are listed according to the assessed quality of the ligand and its interaction with CD81-LEL in descending order. Atom color codes: hydrogen (light blue), nitrogen (dark blue), oxygen (red), sulfur (yellow), and fluorine (green). Criteria used to define the quality of the ligands are: Good - makes more than 5 contacts with protein, predicted to be selective and not predicted to bind to multiple sites, not too hydrophobic; Intermediate - makes 4-5 contacts with protein, hydrophobic interactions contribute to binding; Fair – makes 3-4 contacts with protein; Poor – makes less than 3 contacts with protein, large conjugated ring structures, molecule is hydrophobic and does not fit well inside the cavity.

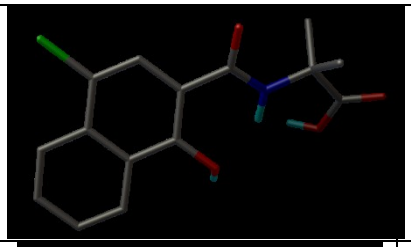
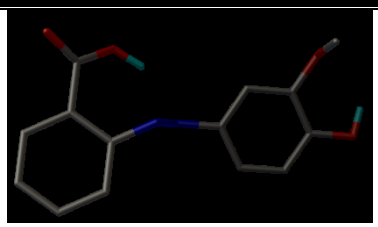
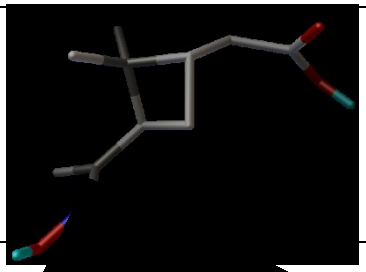
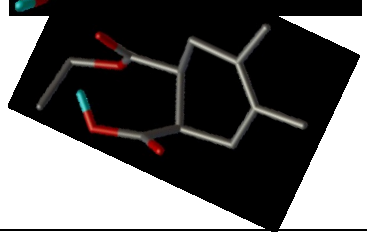
Ligand	Structure	Potential quality
	Site 1	
		Good
2		Intermediate
3		Fair
4		Poor

Table 9. Small molecules predicted by AutoDock to bind Site 2 and Site 3 on the closed conformation of CD81-LEL. The molecules are listed according to the assessed quality of the ligand and its interaction with CD81-LEL in descending order. Atom color codes: hydrogen (light blue), nitrogen (dark blue), oxygen (red), sulfur (yellow), and fluorine (green). Criteria used to define the quality of the ligands are: Good - makes more than 5 contacts with protein, predicted to be selective and not predicted to bind to multiple sites, not too hydrophobic; Intermediate - makes 4-5 contacts with protein, hydrophobic interactions contribute to binding; Fair – makes 3-4 contacts with protein; Poor – makes less than 3 contacts with protein, large conjugated ring structures, molecule is hydrophobic and does not fit well inside the cavity.

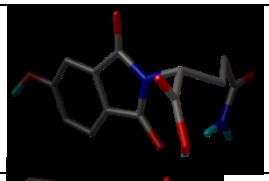
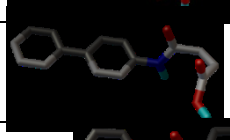
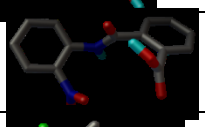
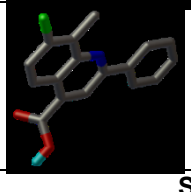
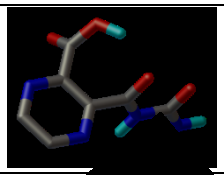
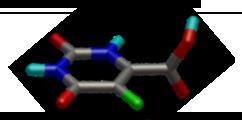
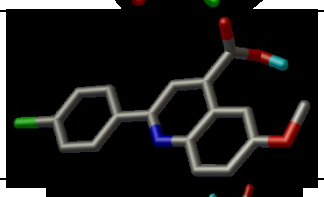
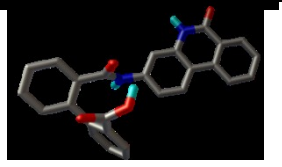
Ligand	Structure	Potential quality
Site 2		
1		Good
2		Intermediate
3		Fair
4		Poor
Site 3		
1		Good
2		Intermediate
3		Fair
4		Poor

Figure 1. E2 homology model . Reproduced from [34].

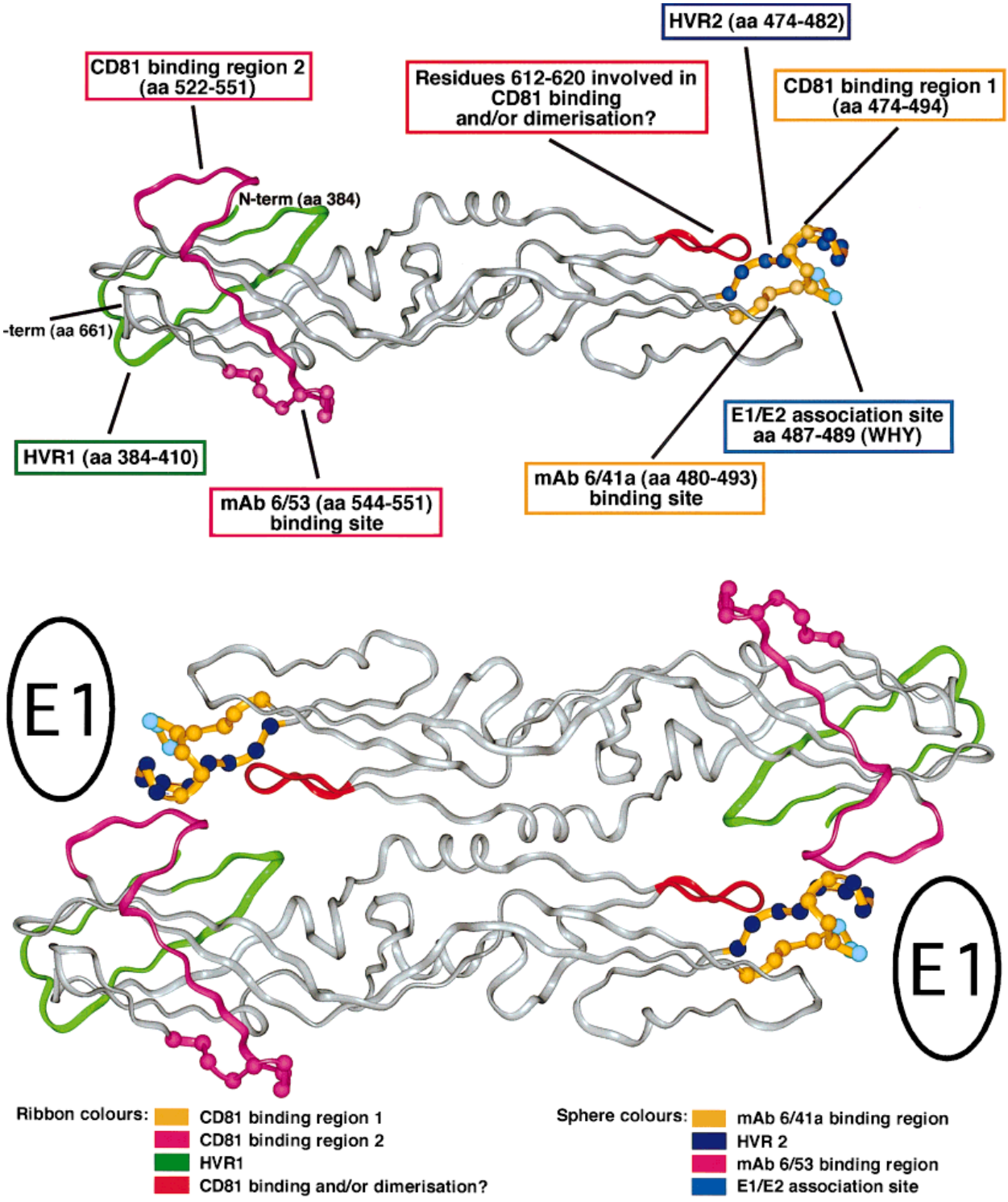
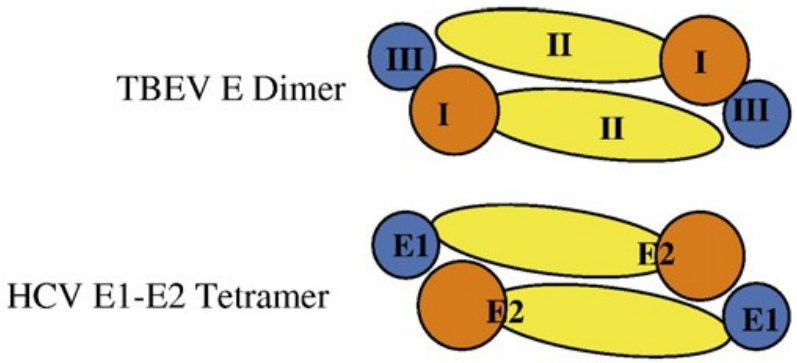
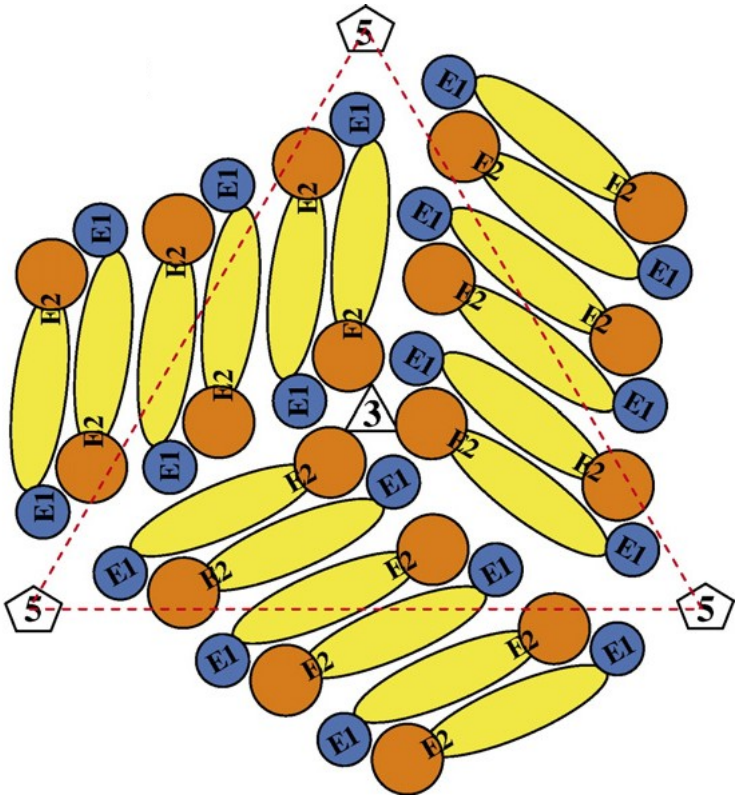


Figure 2. The difference between the envelope glycoproteins of HCV and TBEV. (A) The components of the outer shell of both TBEV and HCV where E1-E2 tetramer forms the outer shell incase of HCV and E-protein dimer forms the outer shell incase of TBEV. (B) The proposed outer shell structure of HCV formed of E1-E2 tetramers. Reproduced from [35].



(A)



(B)

Figure 3. The predicted (A) monomeric and(B) dimeric models of HCV E2. (C & D) The nine mimotopic fragments that were chosen by Spiga *et al*. Reproduced from [36].

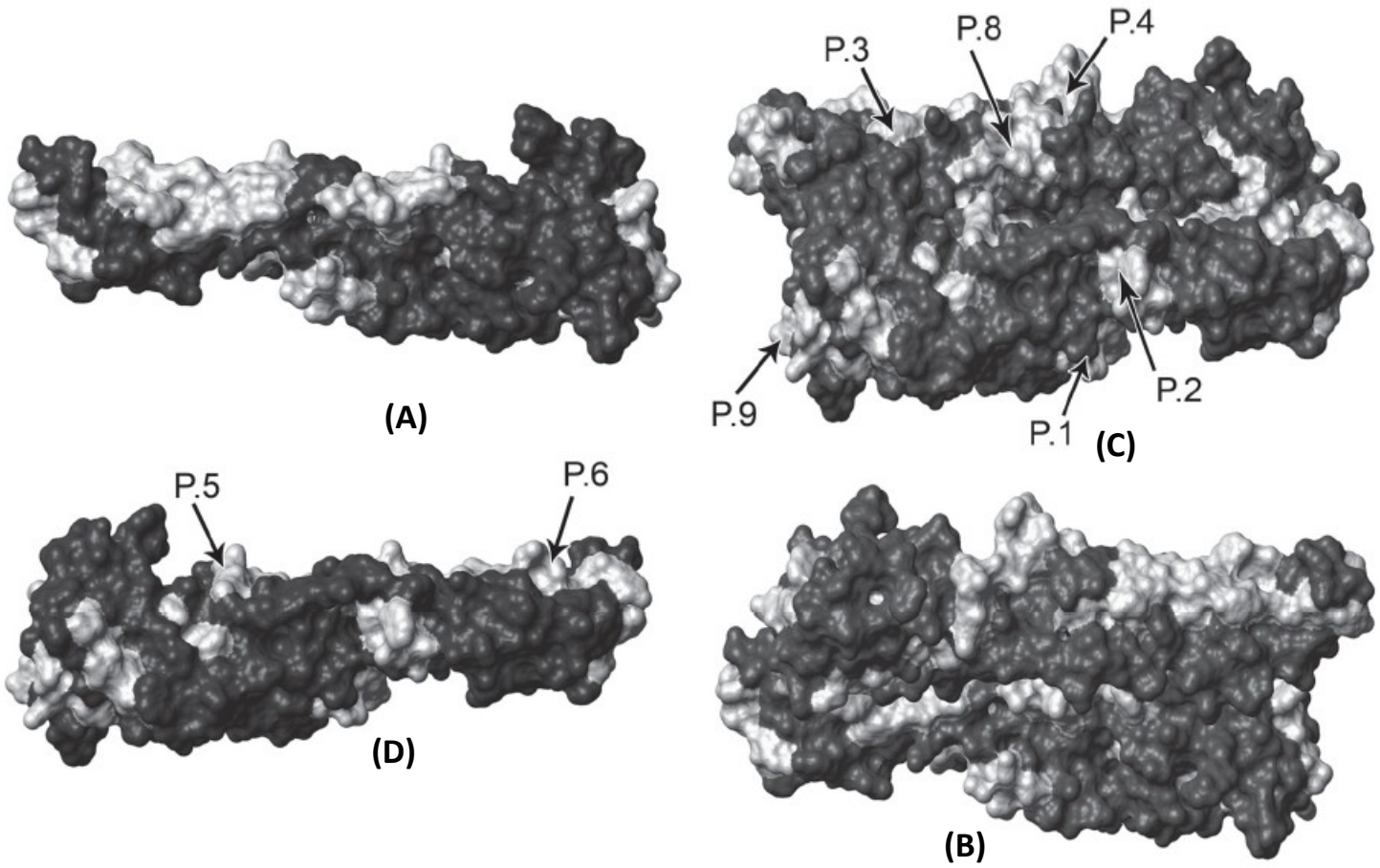


Figure 4. HCV genome . Reproduced from[116].

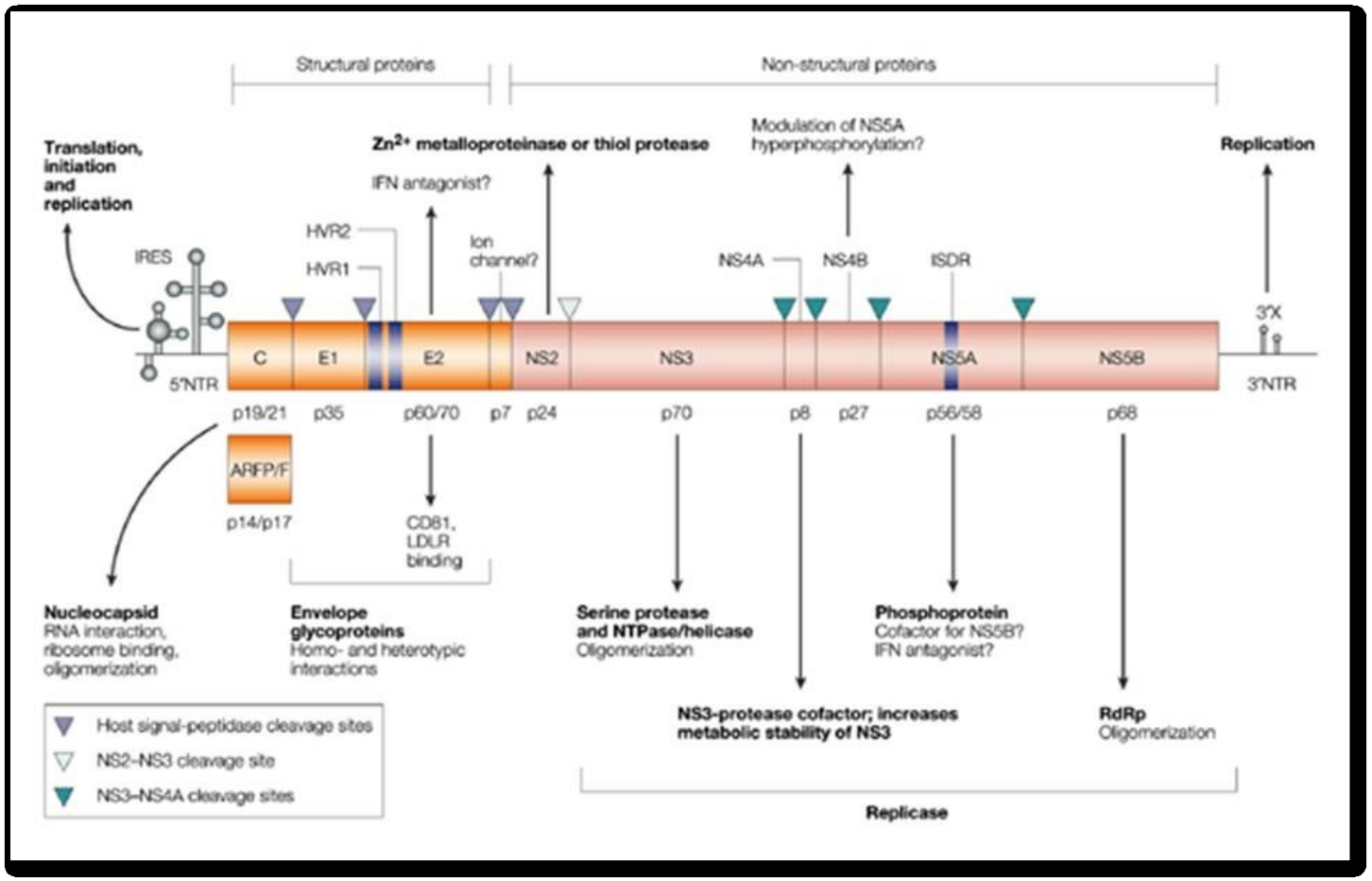
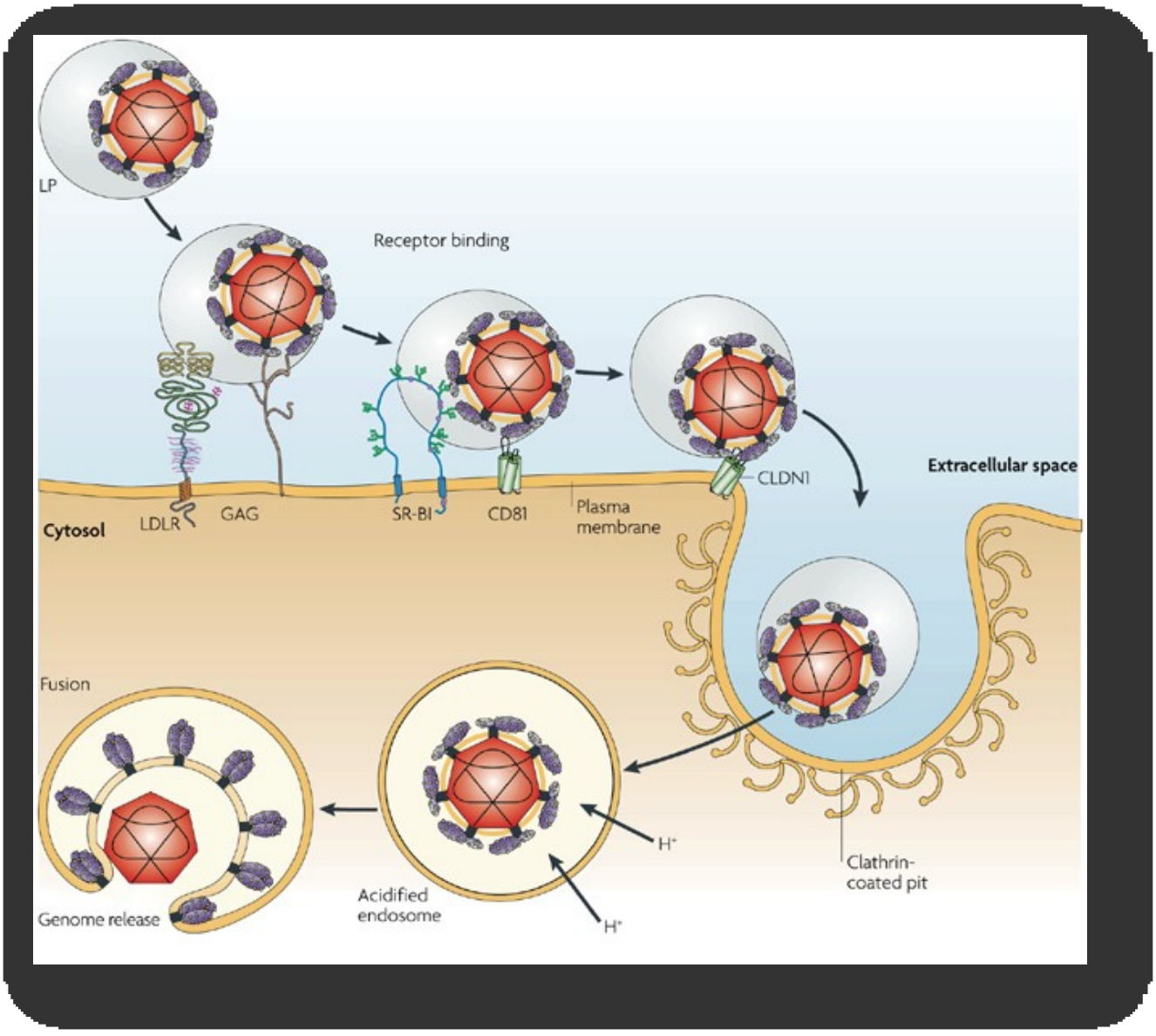
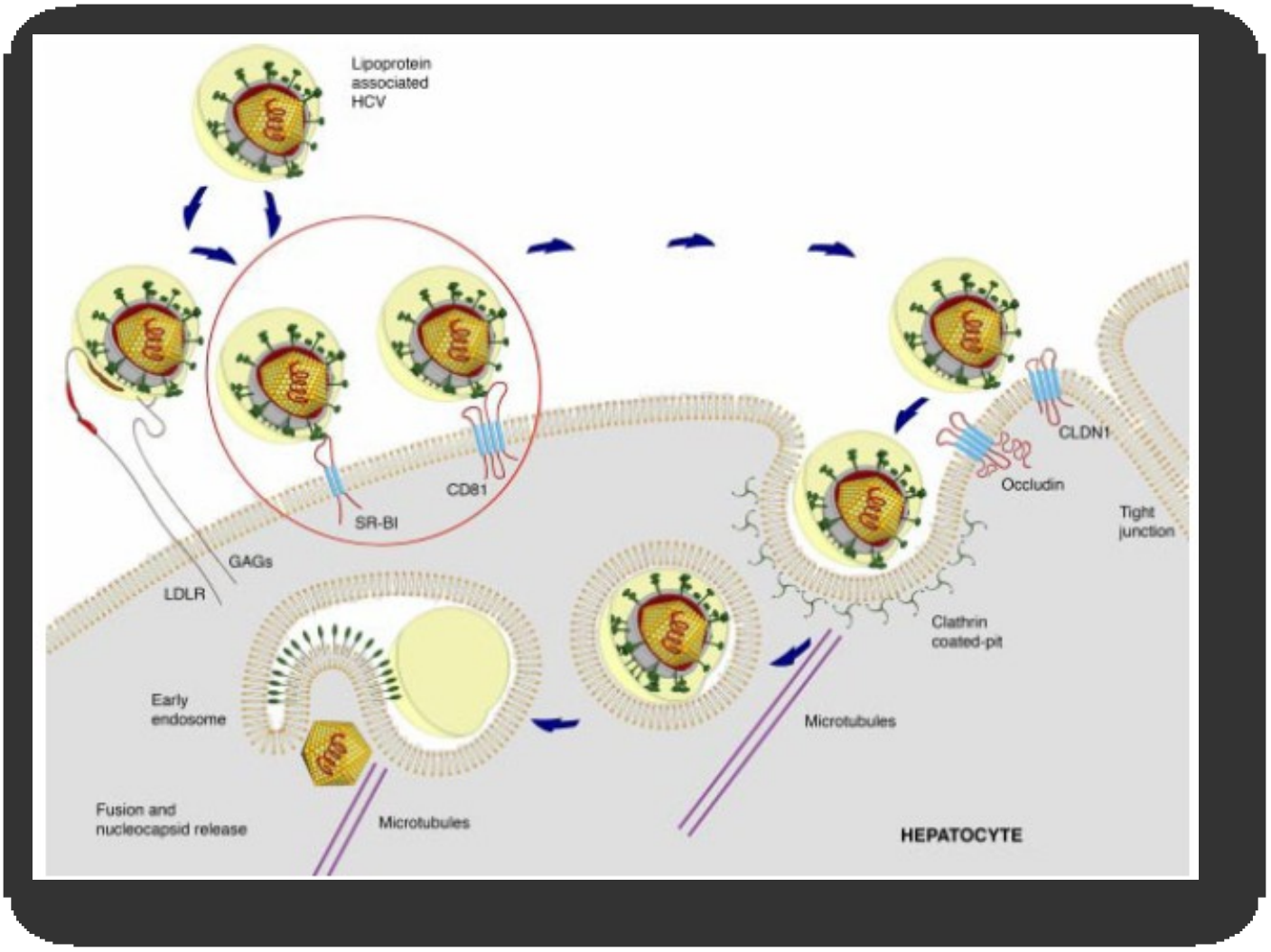


Figure 5. Proposed models for HCV entry into the host cells. (A) LDL-R and Glycosaminoglycan plays the initial role by mediating the attachment of HCV associated with liporptoeins to CD81 and SRBI which upon interacting will trigger HCV movement to the tight junction proteins (Such as claudin-1 (CLDN1)) followed by clathrin mediated endocytosis. (B) A proposed model that differsfrom model (A) in having an extra tight junction protein (OCLDN) involved in the HCV entry by acting as the final key in HCV entry process. (A) and (B) are reproduced from [85] and [30] respectively.



(A)


Lipoprotein
associated
HCV
CD81
SR-BI
GAGs
LDLR
CLDN1
Occludin
Tight
junction
Clathrin
coated-pit
Early
endosome
Microtubules
Fusion and
nucleocapsid release
Microtubules
HEPATOCYTE

(b)

Figure 6: CD81-LEL open conformation (PDB ID: 1G8Q). The colored residues are amino acids that have been identified to contribute to the binding of the HCV protein E2 to CD81-LEL. (A) Front view of the protein showing the four contact residues Leu162, Ile182, Asn184, and Phe186. (B) Back side of the CD81-LEL protein showing the other three contact residues Thr163, Glu188 and Asp196 [52, 53]. Figure prepared using AutoDock Tools version 1.5.6.

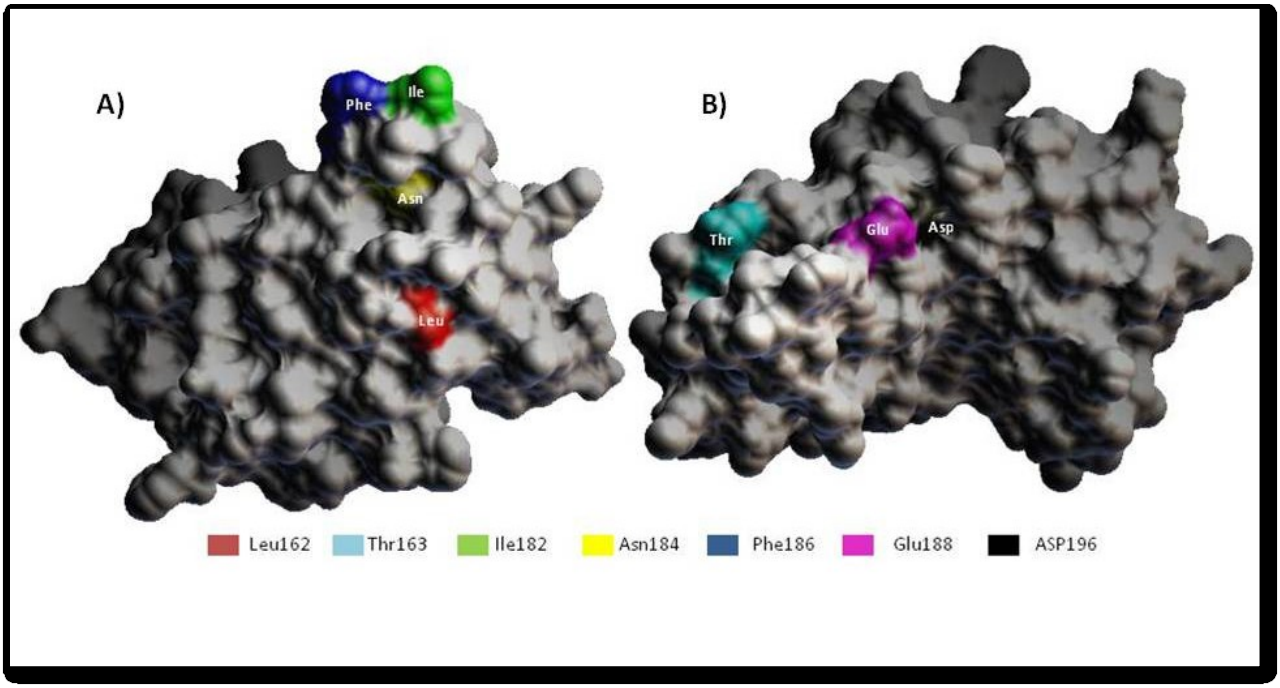


Figure 7: Components of selective high affinity ligands (SHALs)

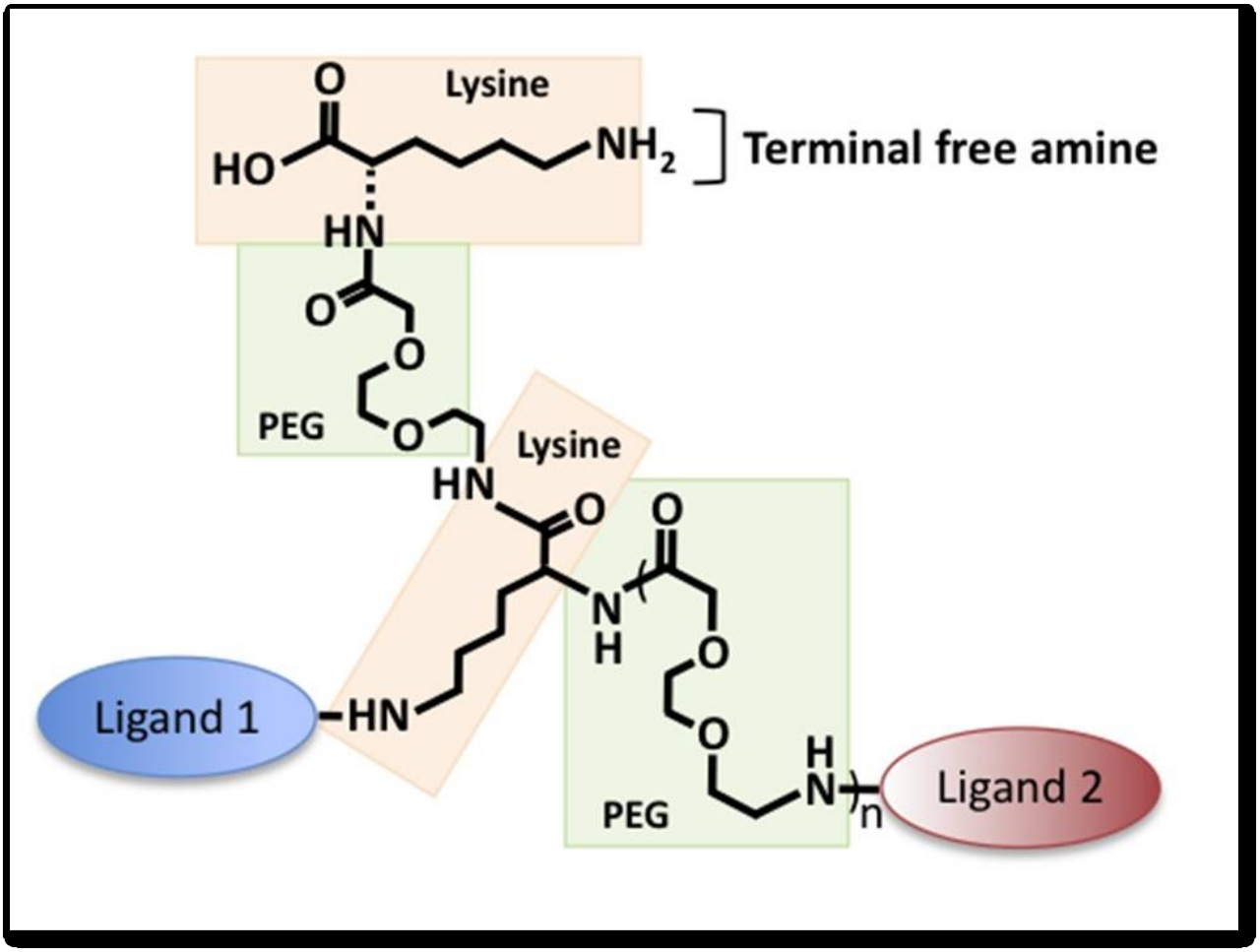


Figure 8: The complete genome of HCV genotype 1a isolate H (complete genome (UniProtKB/Swiss-Prot P27958; POLG_HCVH): 3011 aa).

Figure 9: FASTA format of the protein sequence of some HCV proteins, their modeling trials, and their sequence identity percentage with known PDB structures.

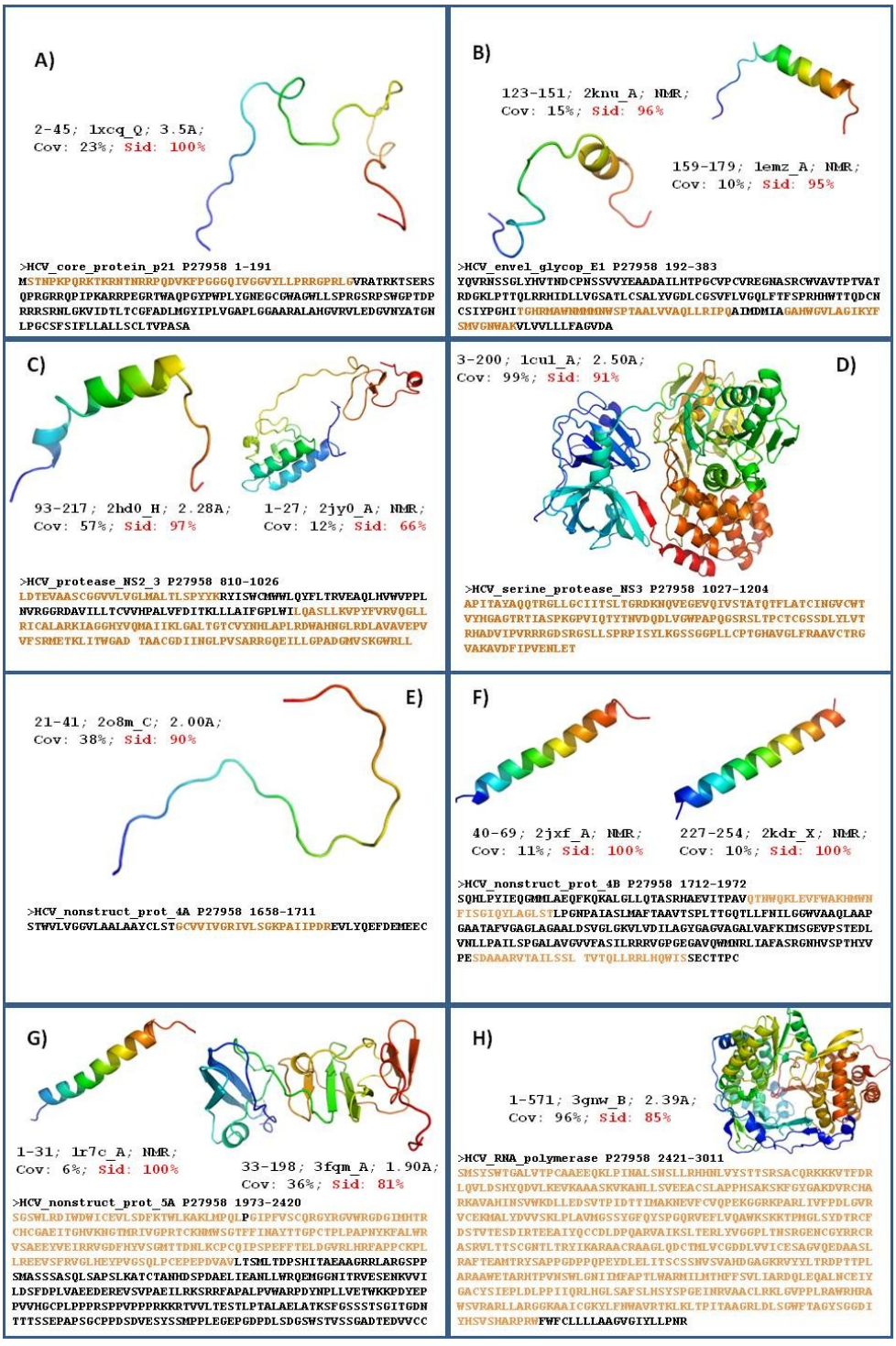


Figure 10: The HCV E2 glycoprotein sequence alignment used in obtaining the E2 homology models and obtained from [34].

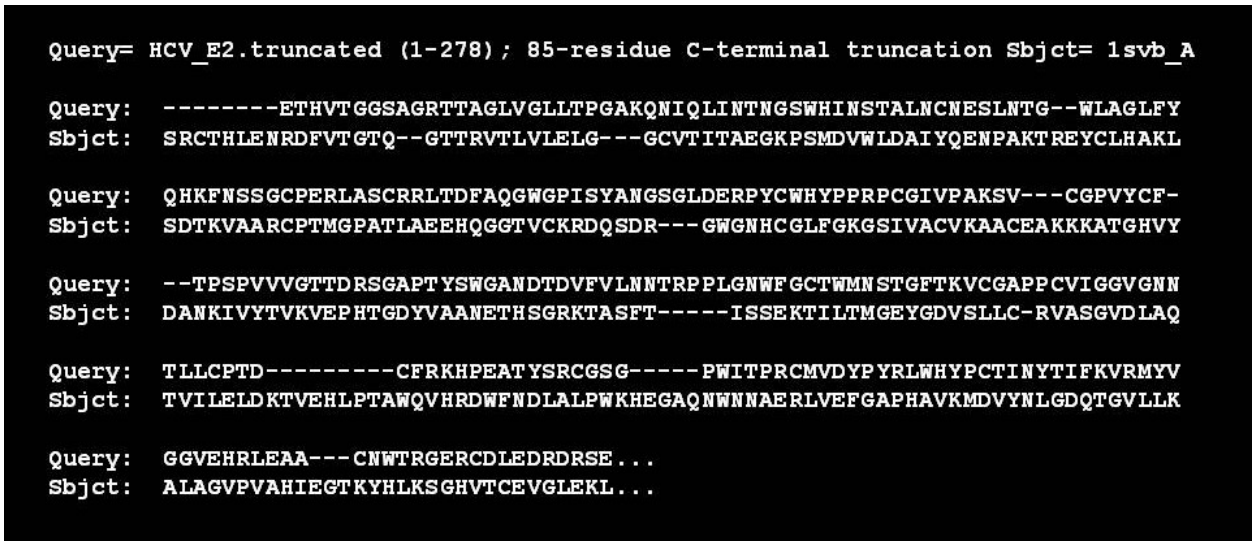
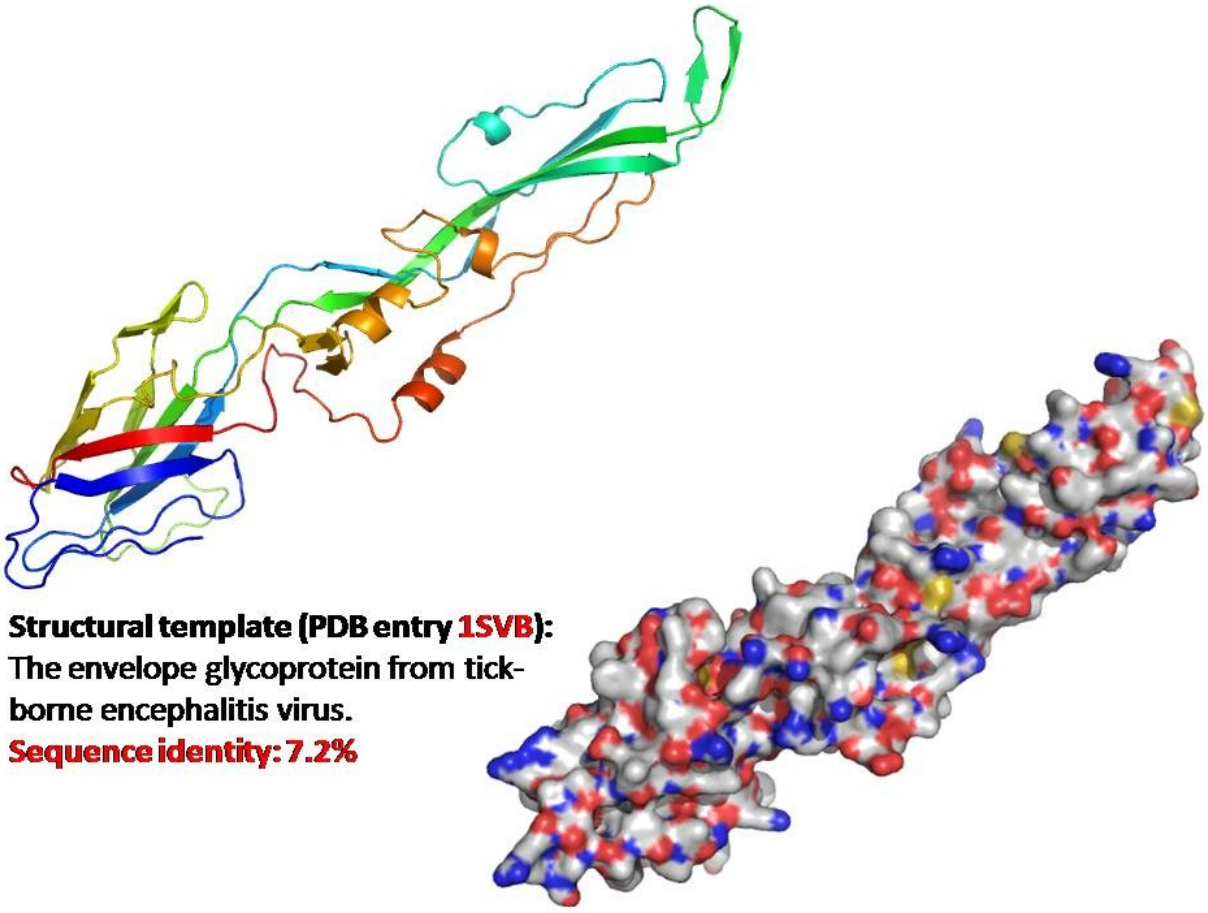
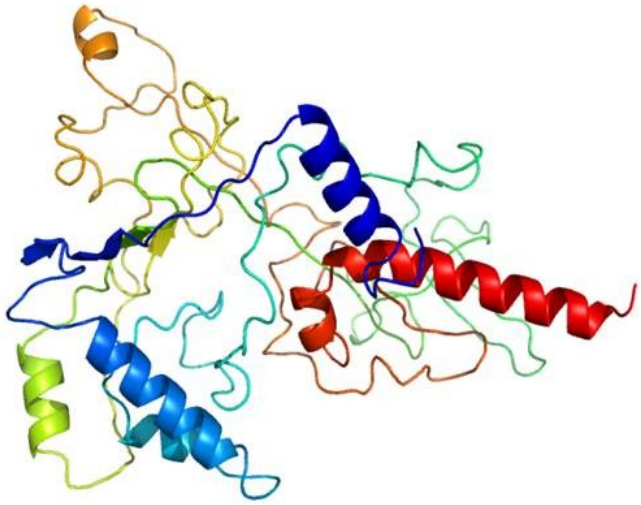


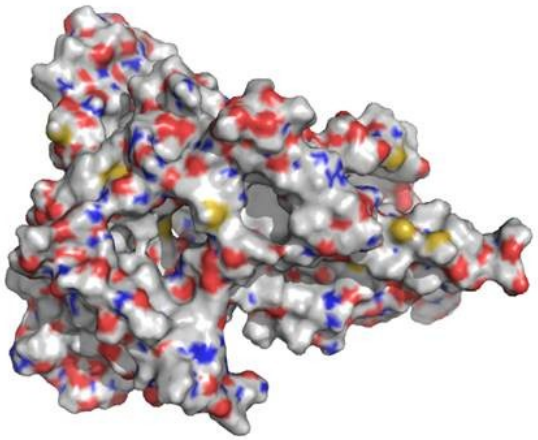
Figure 11: Theoretical model of the Envelope Glycoprotein E2 created by AS2TS system (A), I-TASSER threading/ab-initio system (B, C), PSiFR threading system (D), and PHYRE fold recognition system (D).



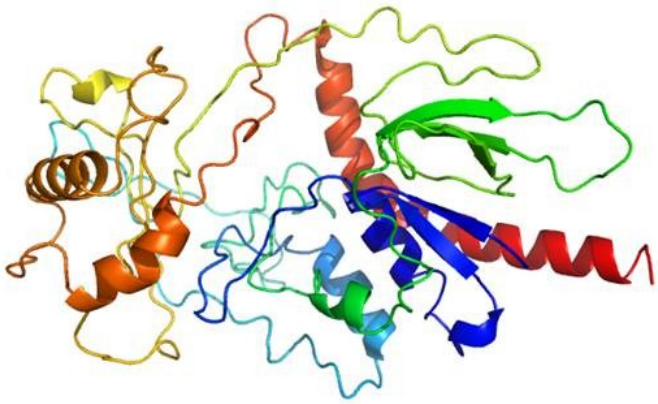
(A)



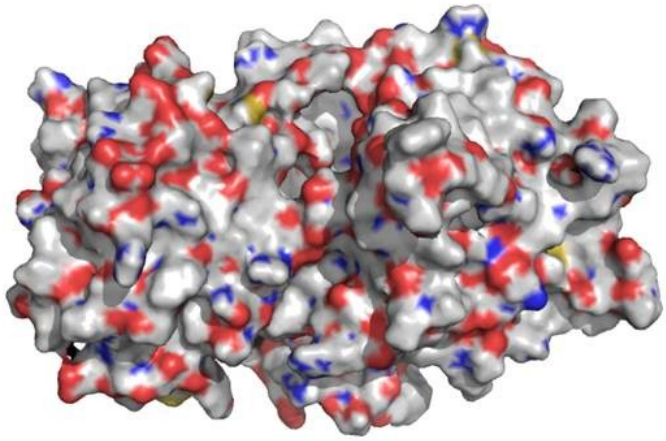
Structural template PDB 1SEZ:
Crystal structure of protoporphyrinogen IX oxidase: a key enzyme in haem and chlorophyll biosynthesis.
Sequence identity: 6.0%



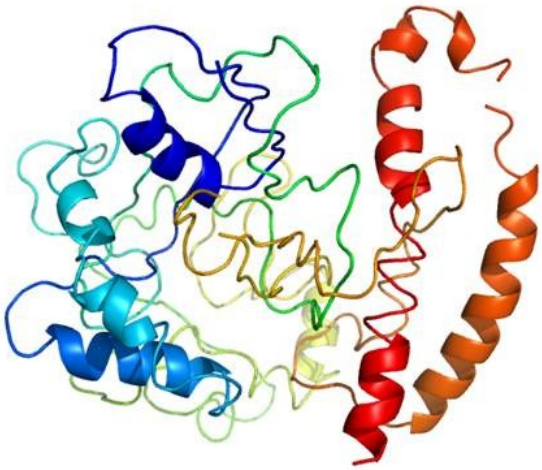
(B)



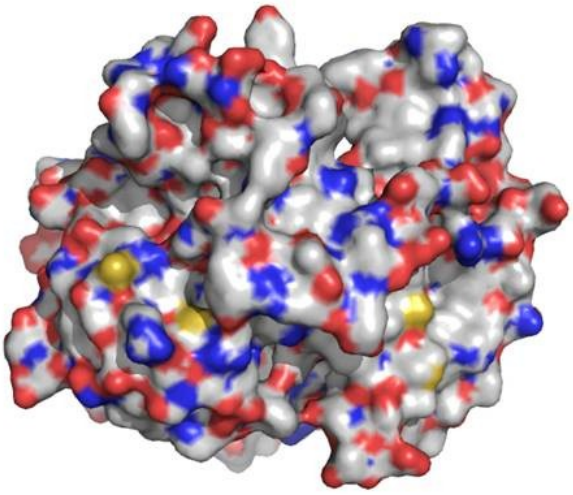
Structural template PDB 1XPQ:
Crystal structure of fms1, a polyamine oxidase from yeast.
Sequence identity: 8.0%



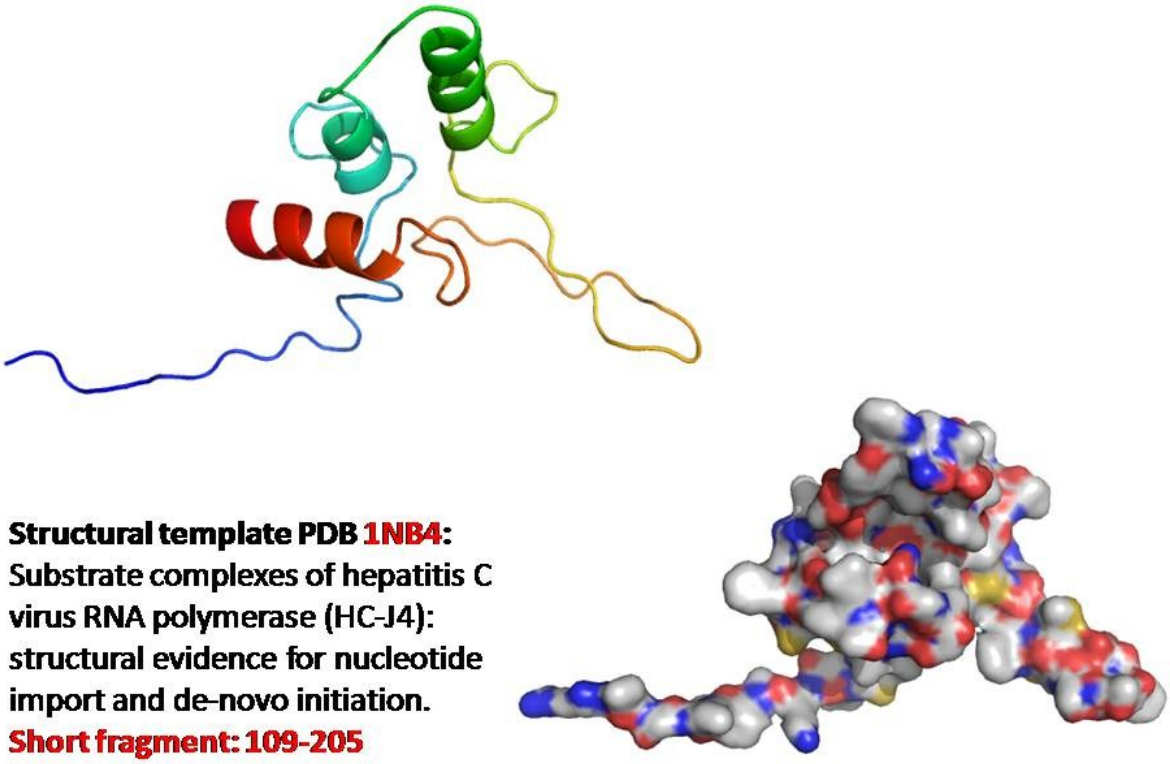
(C)



Structural template PDB 1KF6: Crystallographic studies of the Escherichia coli quinol-fumarate reductase with inhibitors bound to the quinol-binding site.
Sequence identity: 15.0%



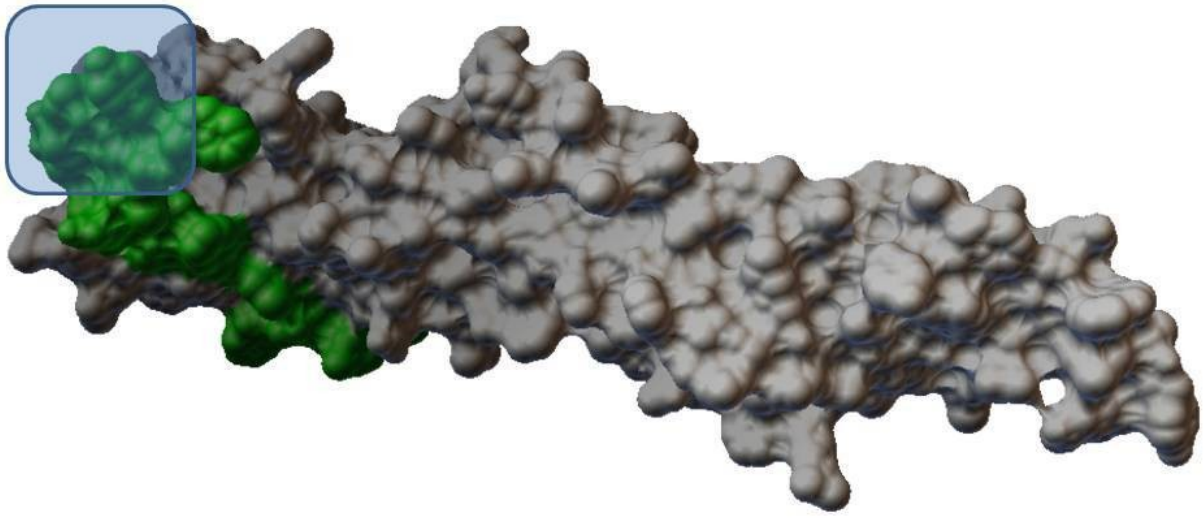
(D)


Structural template PDB 1NB4:
Substrate complexes of hepatitis C virus RNA polymerase (HC-J4): structural evidence for nucleotide import and de-novo initiation.
Short fragment: 109-205
Sequence identity: 7.0%

(E)

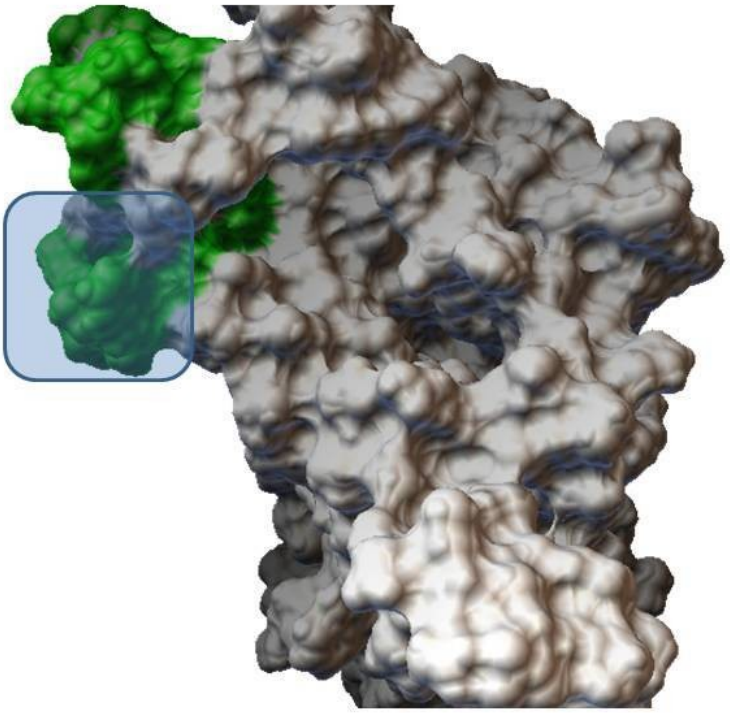
Figure 12: Comparing one of the identified CD81-LEL binding sites on HCV E2 by Yagnik et al. in the 4 models; AS2TS model (A), TASSER models (B, C), PSiFR model (D), and Phyre model (E) to see if the binding pocket have certain percentage of similarity.

AS2TS



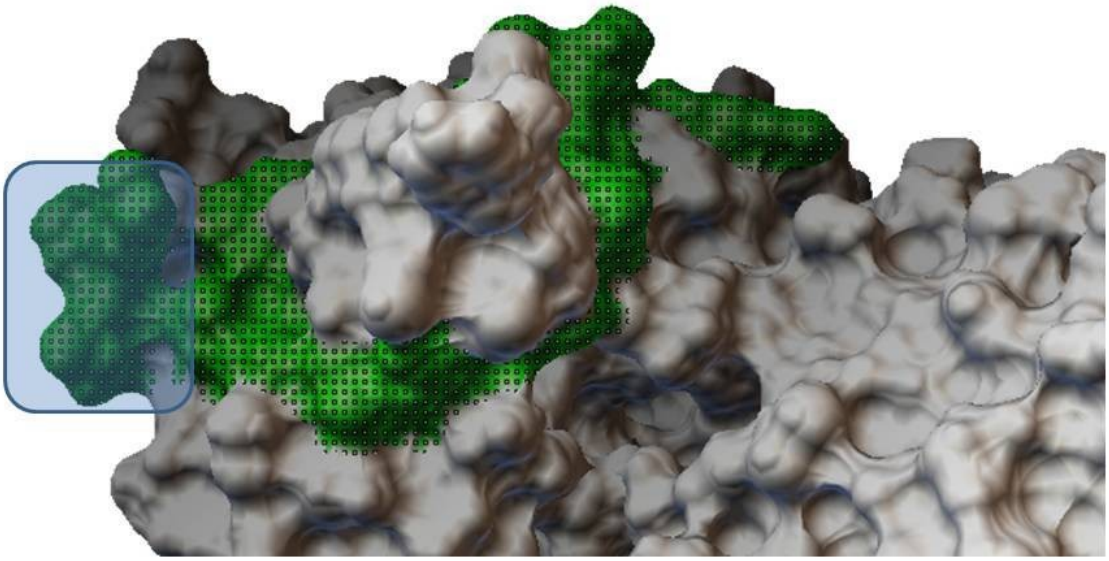
(A)

TASSER I



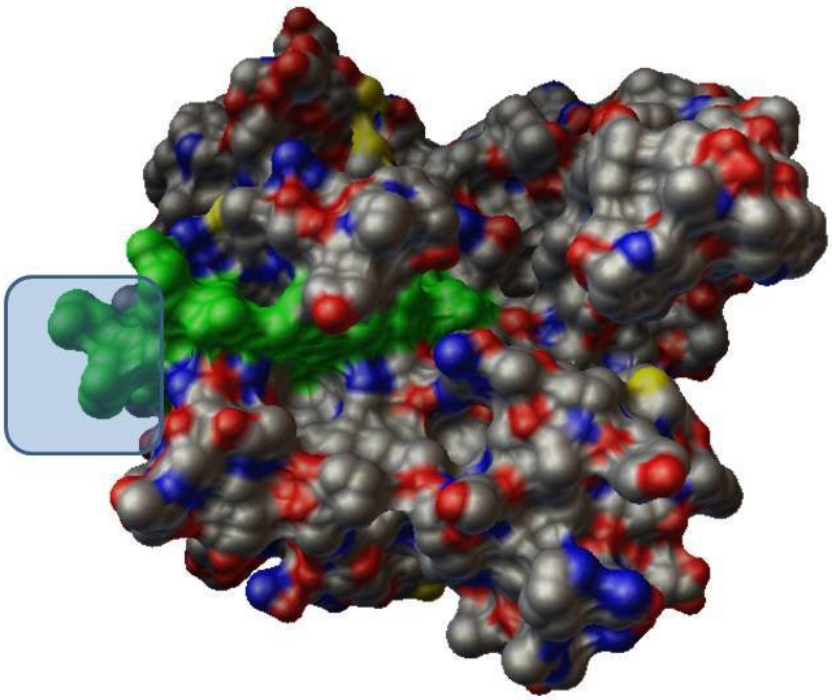
(B)

TASSER II



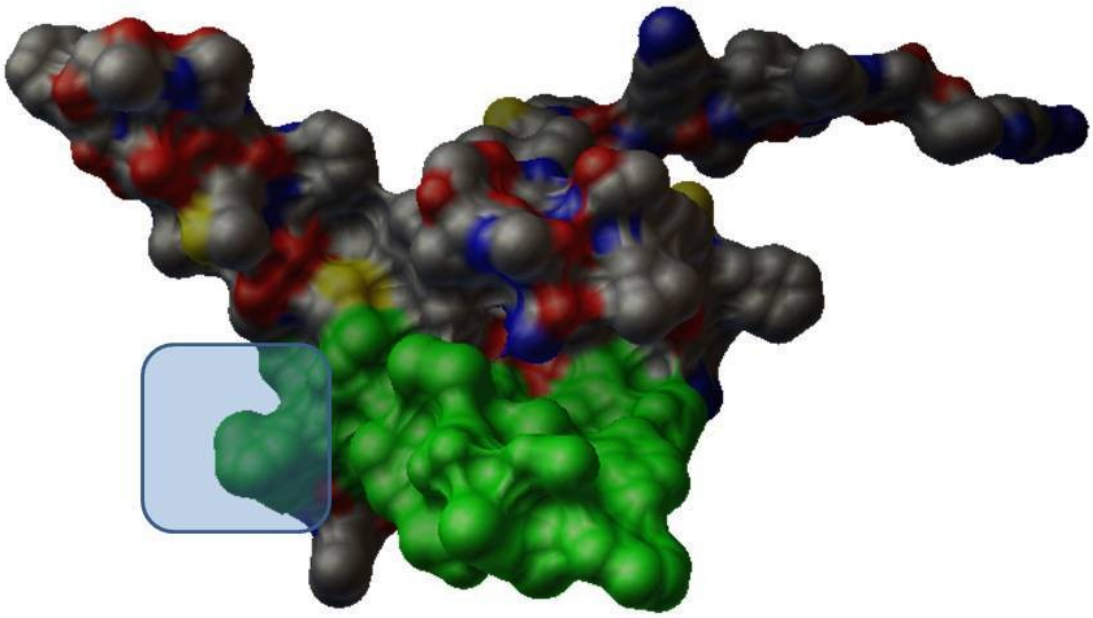
(C)

PSIFR



(D)

PHYRE



(E)

Figure 13: HCV NS3 serine protease sequences for the six HCV genotypes. NS3 serine protease sequences of genotypes 1 (A), 2 (B), 3 (C), 4 (D), 5 (E) and 6 (F).

A) >gi|5918961|gb|AAD56196.1|AF165

EGEVQVVSTATQSFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVDLDLVGWQAPPGARSL
TPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVA
KAVDFIPVESMETT

B) >gi|157781213|ref|YP_001469630.

AGEIQVLSTVTQSFLGTSISGVLWTVYHGAGNKTLAGSRGPVTQMYSSAEGDLVGWPSPPGTKSLEP
CTCGAVDLYLVTRNADVIPARRRGDKRGALLSPRPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVA
KSIDFIPVETLD

C) >gi|157781217|ref|YP_001469631.

TGEVQVLSTATQTFLGTTVGGVIWTVYHGAGSRTLAGAKHPALQMYTNVDQDLVGWPAPPGAKSLE
PCACGSSDLYLVTRDADVIPARRRGDSTASLLSPRPLACLKGSSGGPVMCPSGHVAGIFRAAVCTRGVA
KSLQFIPVETLSTQ

GEVQVLSTATQSFLGTAVNGVMWTVYHGAGAKTISGPKGPVNQMYTNVDQDLVGWPAPPGVRSL
APCTCGSADLYLVTRHADVIPVRRRGDTRGALLSPRPISILKGSSGGPLLCPMGHRAGIFRAAVCTRGV
AKAVDFVPVESLETT

E) > gi|157781211|ref|YP_001469633.

EGEVQFLSTATQTFLGICINGVMWTLFHGAGSKTLAGPKGPVVQMYTNVDKDLVGWPSPPGKGSLT
RCTCGSADLYLVTRHADVIPARRRGDTRASLLSPRPISYLKGSSGGPIMCPSGHVVGVFRAAVCTRGVA
KALEFVPVENLETT

F) >gi|157781215|ref|YP_001469634.

EGEVQVVSTATQSFLATTINGVLWTVYHGAGSKNLAGPKGPVCQMYTNVDQDLVGWPAPLGARSLA
PCTCGSSDLYLVTRGADVIPARRRGDTRAALLSPRPISTLKGSSGGPLMCPSGHVVGLFRAAVCTRGVA
KALDFIPVENMDTT

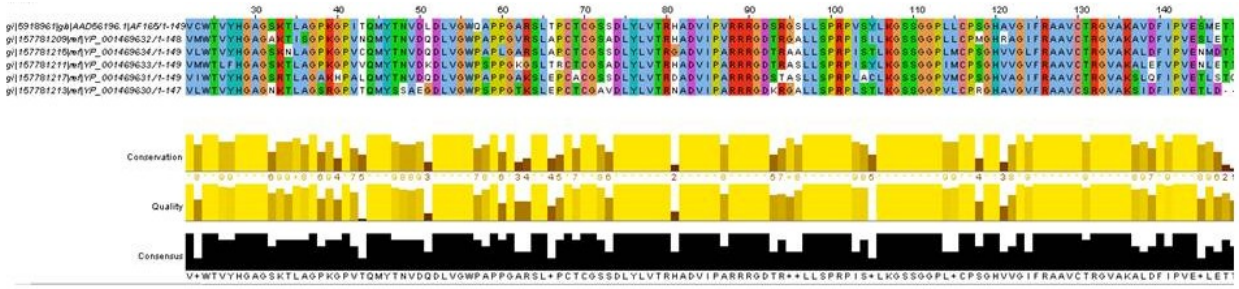
Figure 14: HCV NS3 serine protease sequence alignment of the 6 genotypes so as to determine the conserved regions (A), and a graphical representation (B) [102].

```
gi|5918961|gb|AAD56196.1|AF165    EGEVQVVSTATQSFLATCVNGVCWTVYHGAGSKTLAGPKGPITQMYTNVD 50
gi|157781209|ref|YP_001469632.    -GEVQVLSTATQSFLGTAVNGVMWTVYHGAGAKTISGPKGPVNQMYTNVD 49
gi|157781215|ref|YP_001469634.    EGEVQVVSTATQSFLATTINGVLWTVYHGAGSKNLAGPKGPVCQMYTNVD 50
gi|157781211|ref|YP_001469633.    EGEVQFLSTATQTFLGICINGVMWTLFHGAGSKTLAGPKGPVVQMYTNVD 50
gi|157781217|ref|YP_001469631.    TGEVQVLSTATQTFLGTTVGGVIWTVYHGAGSRTLAGAKHPALQMYTNVD 50
gi|157781213|ref|YP_001469630.    AGEIQVLSTVTQSFLGTSISGVLWTVYHGAGNKTLAGSRGPVTQMYSSAE 50

gi|5918961|gb|AAD56196.1|AF165    LDLVGWQAPPGARSLTPCTCGSSDLYLVTRHADVIPVRRRGDSRGSLLSP 100
gi|157781209|ref|YP_001469632.    QDLVGWPAPPGVRSLAPCTCGSADLYLVTRHADVIPVRRRGDTRGALLSP 99
gi|157781215|ref|YP_001469634.    QDLVGWPAPLGARSLAPCTCGSSDLYLVTRGADVIPARRRGDTRAALLSP 100
gi|157781211|ref|YP_001469633.    KDLVGWPSPPGKGSLTRCTCGSADLYLVTRHADVIPARRRGDTRASLLSP 100
gi|157781217|ref|YP_001469631.    QDLVGWPAPPGAKSLEPCACGSSDLYLVTRDADVIPARRRGDSTASLLSP 100
gi|157781213|ref|YP_001469630.    GDLVGWPSPPGTKSLEPCTCGAVDLYLVTRNADVIPARRRGDKRGALLSP 100

gi|5918961|gb|AAD56196.1|AF165    RPVSYLKGSSGGPLLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMETT 149
gi|157781209|ref|YP_001469632.    RPISILKGSSGGPLLCPMGHRAGIFRAAVCTRGVAKAVDFVPVESLETT 148
gi|157781215|ref|YP_001469634.    RPISTLKGSSGGPLMCPSGHVVGLFRAAVCTRGVAKALDFIPVENMDTT 149
gi|157781211|ref|YP_001469633.    RPISYLKGSSGGPIMCPSGHVVGVFRAAVCTRGVAKALEFVPVENLETT 149
gi|157781217|ref|YP_001469631.    RPLACLKGSSGGPVMCPSGHVAGIFRAAVCTRGVAKSLQFIPVETLSTQ 149
gi|157781213|ref|YP_001469630.    RPLSTLKGSSGGPVLCPRGHAVGVFRAAVCSRGVAKSIDFIPVETLD-- 147
```

(A)



(B)

Figure 15: Vision interface showing an example of the virtual screening workflow [85].

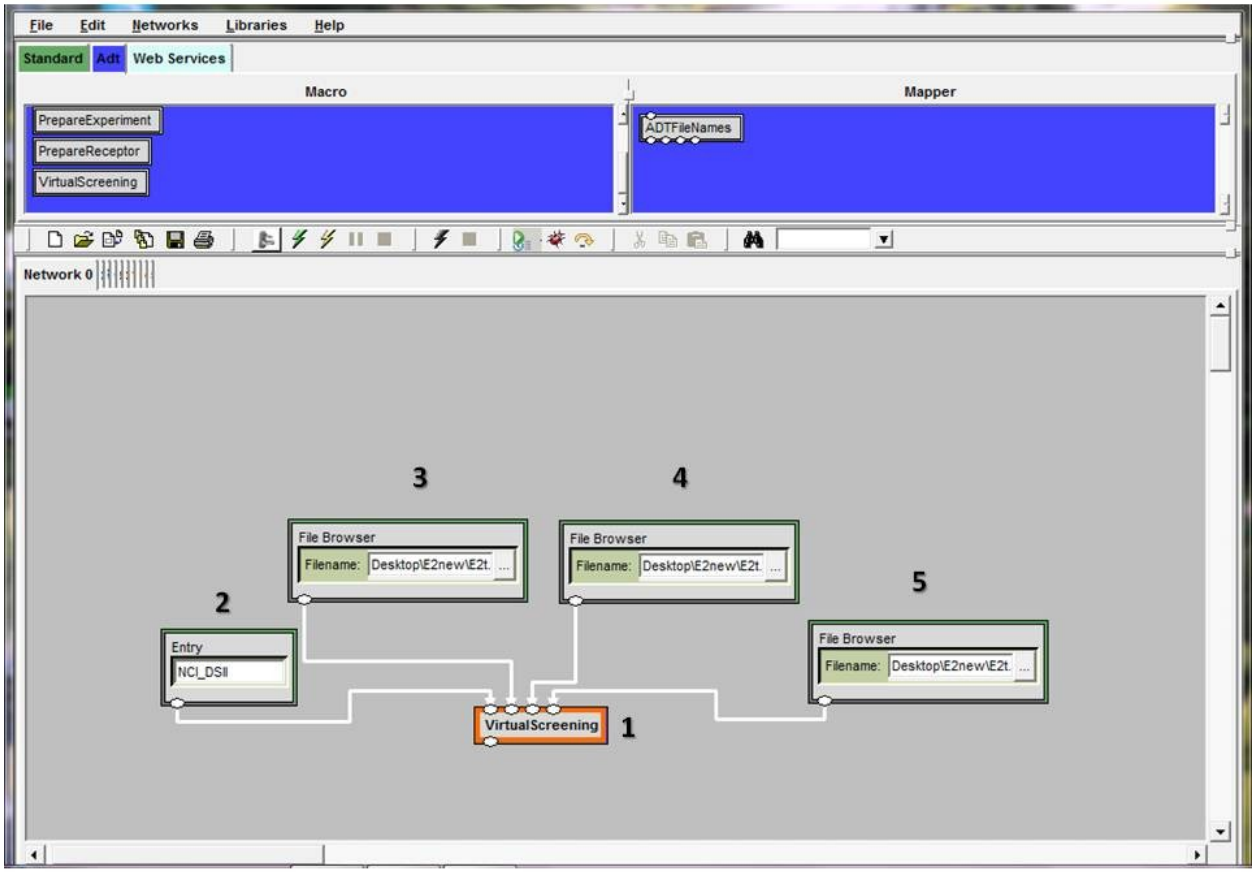


Figure 16: Significant residues in NS3-4A genotype 4 complexation selected using AutoDock Tools 1.5.6

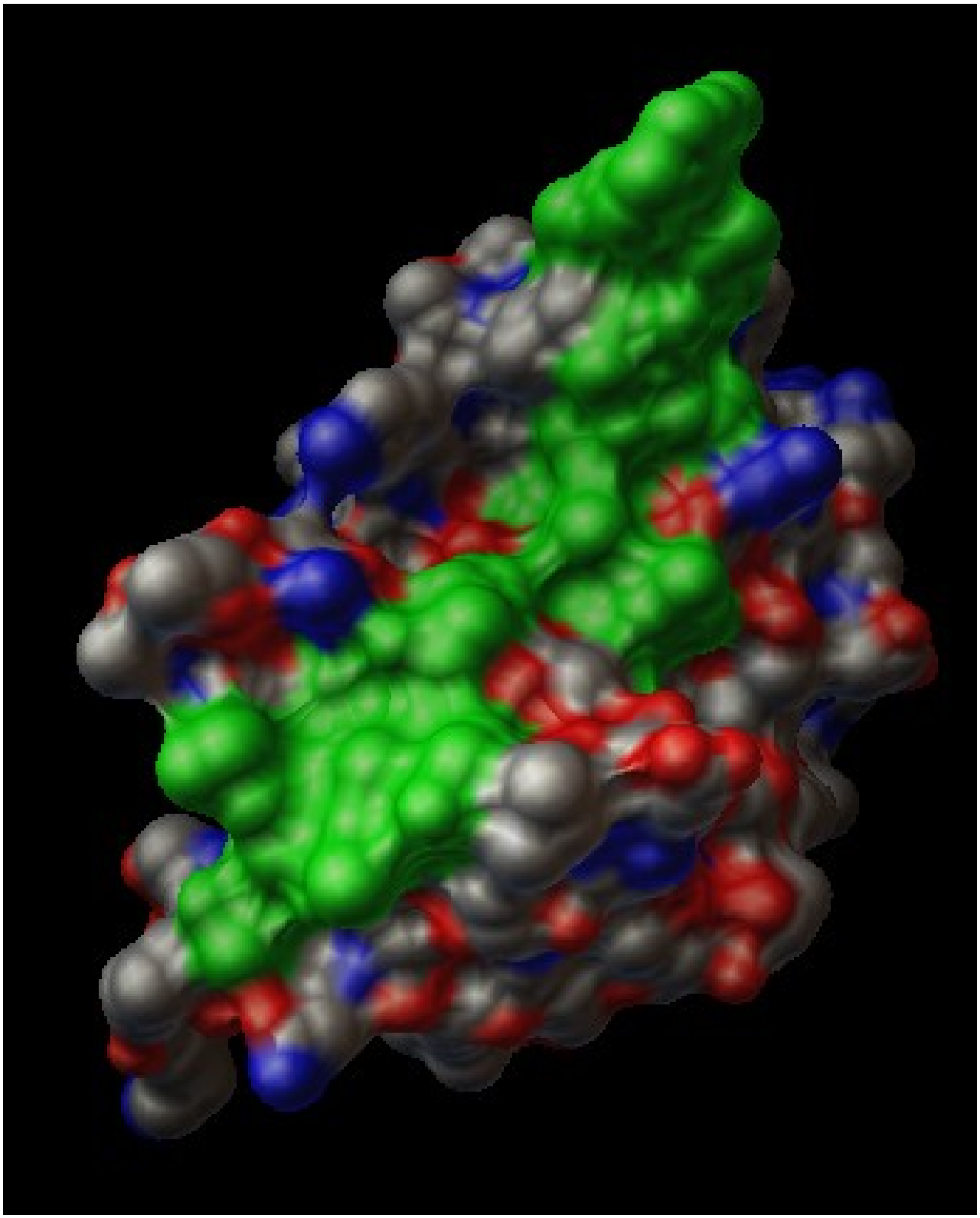
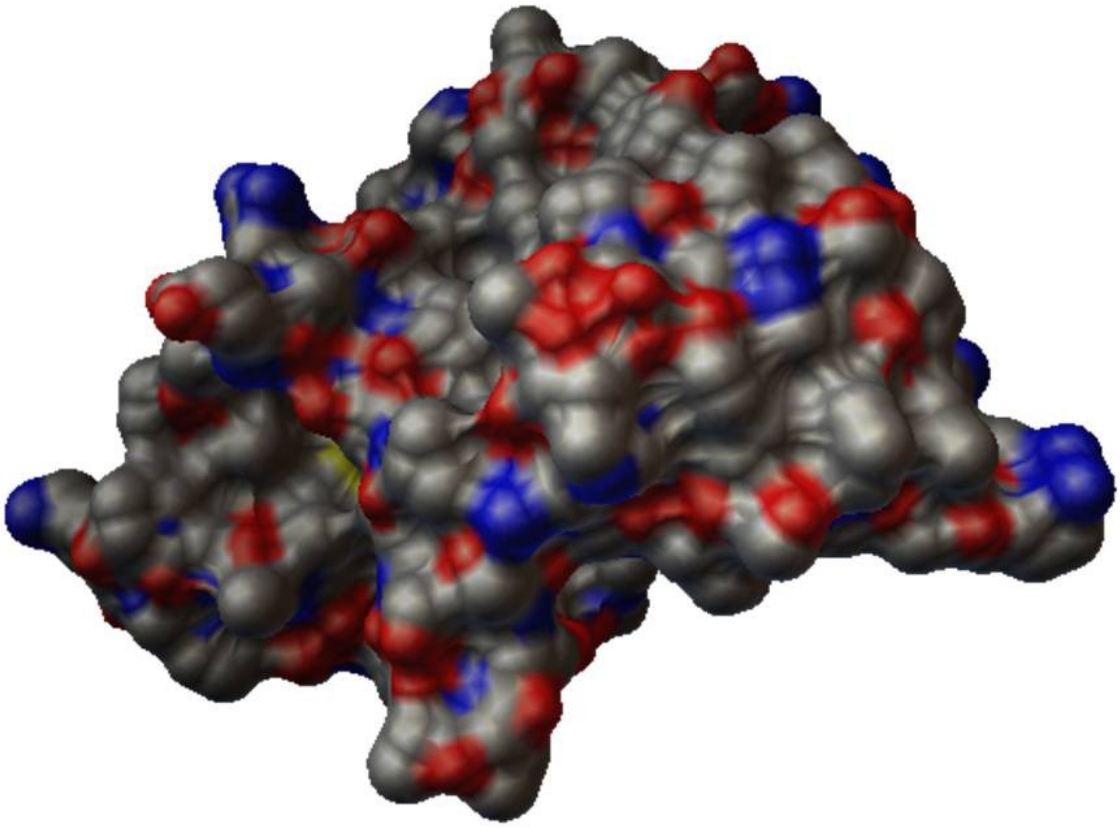
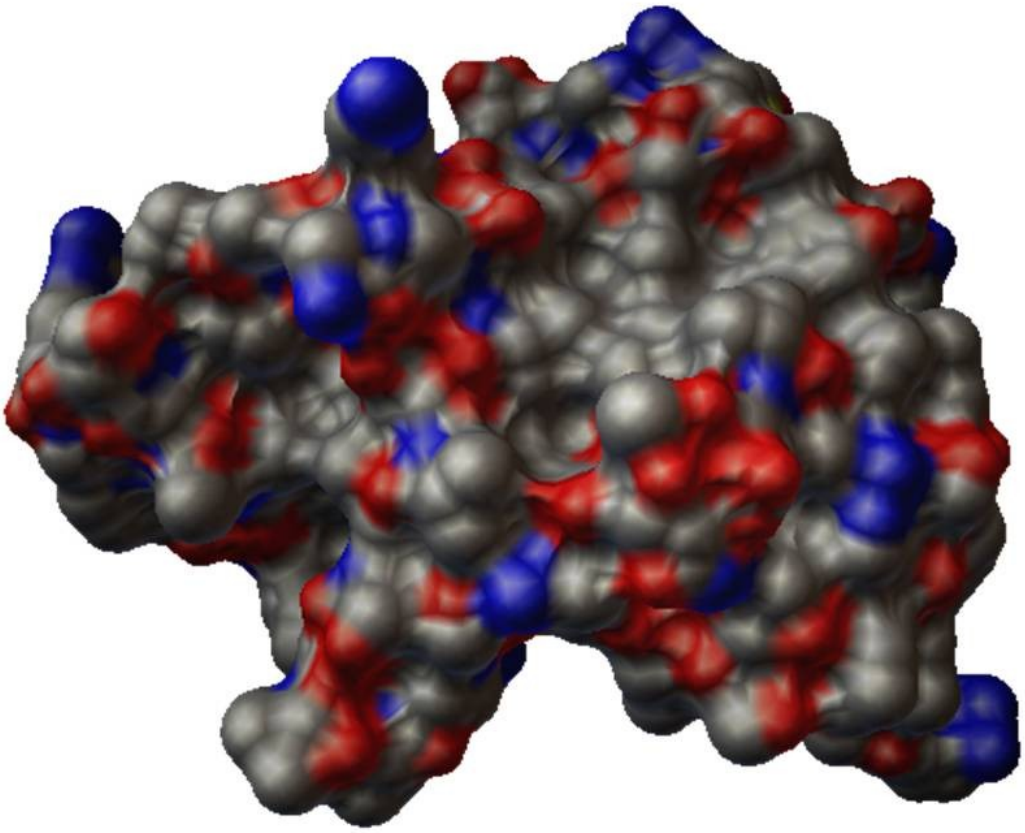


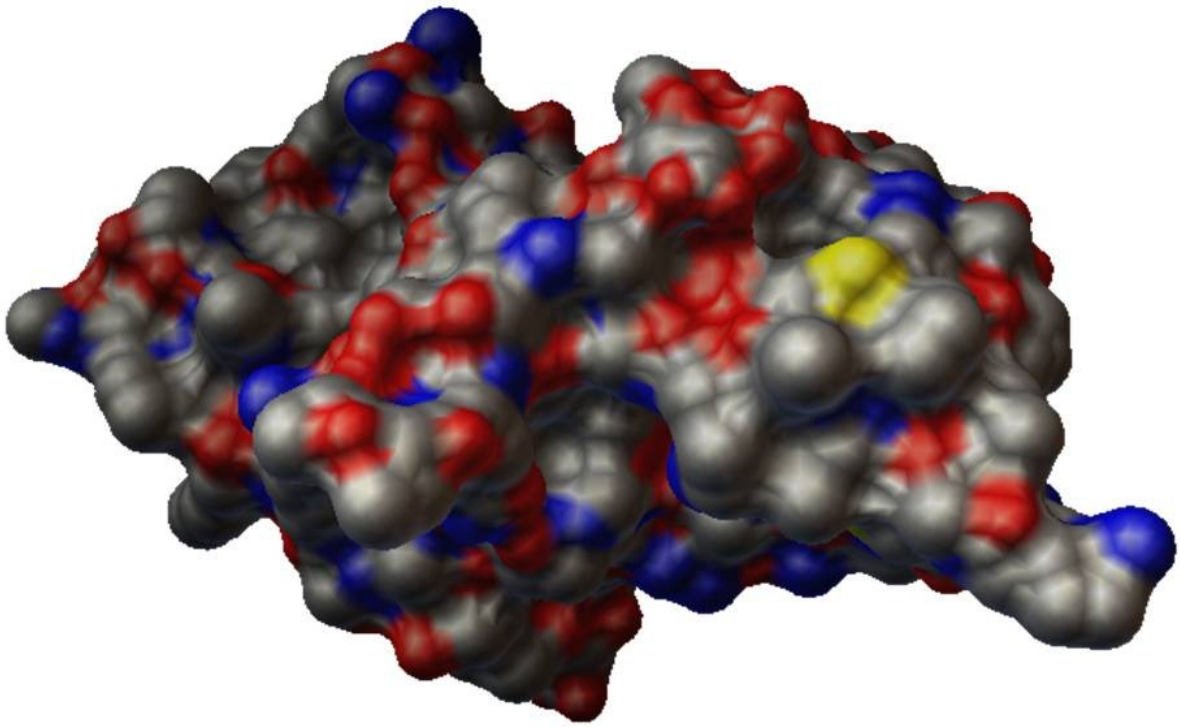
Figure 17: The created NS3 serine protease models of the six HCV genotypes; GT 1 (A), GT 2 (B), GT 3 (C), GT 4 (D), GT 5 (E), and GT 6 (F).



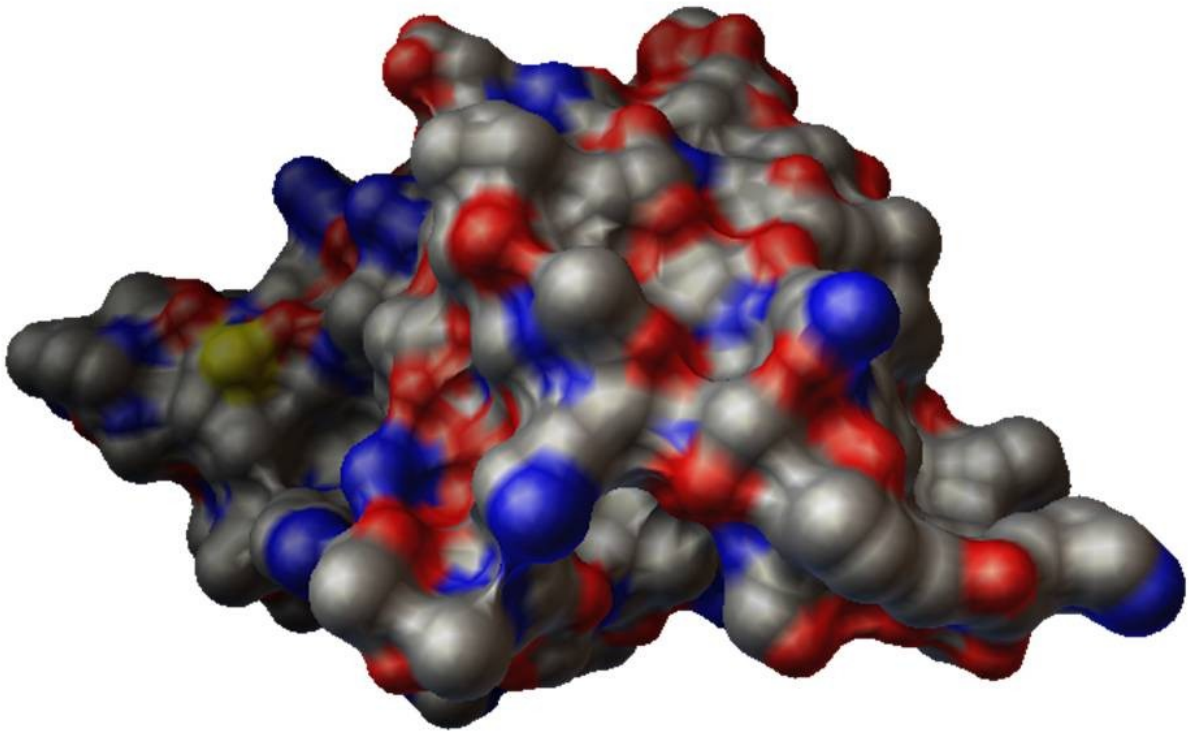
(A)



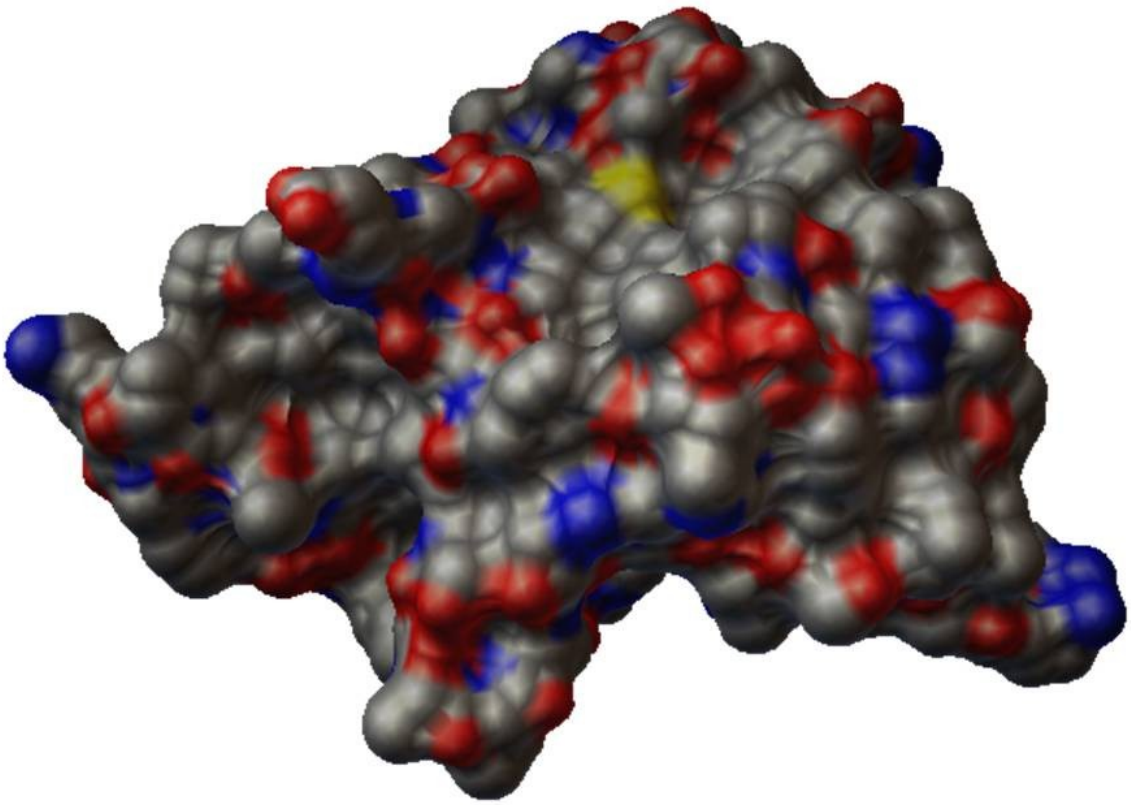
(B)



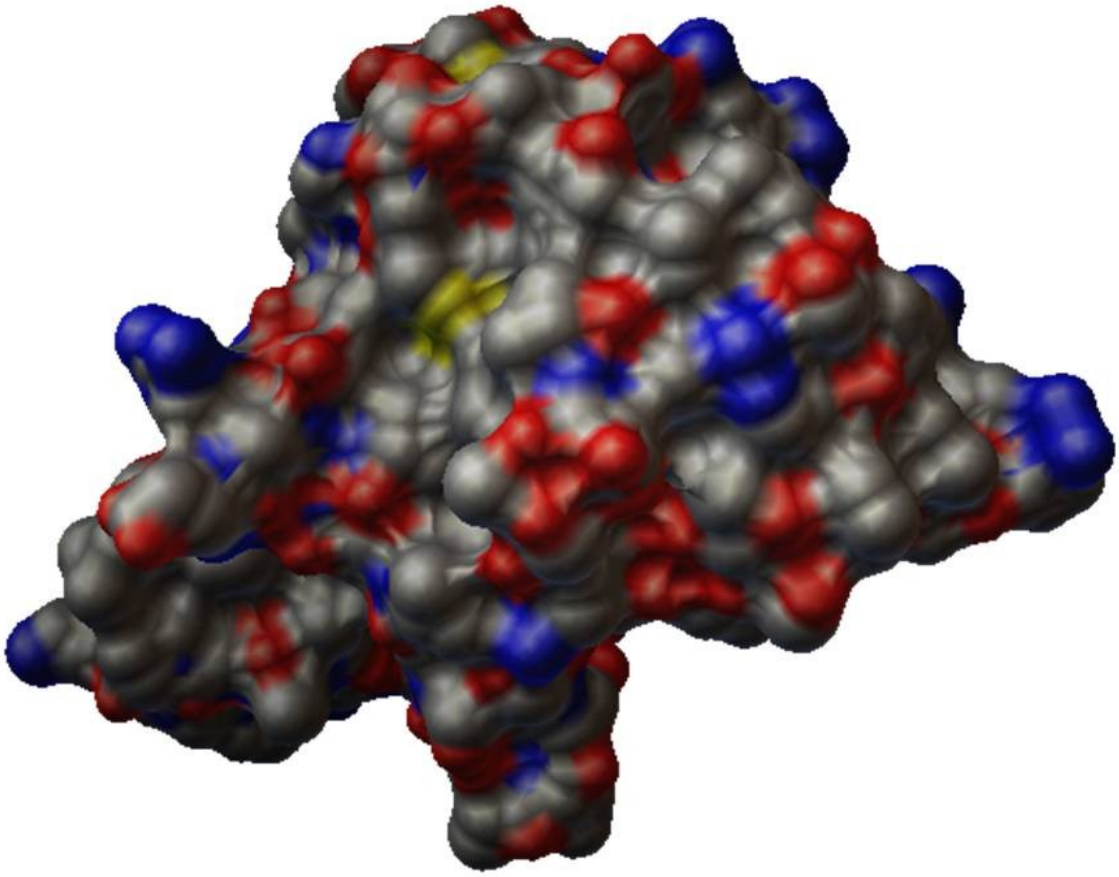
(C)



(D)



(E)



(F)

Figure 18: The dendrogram representing sequence similarity-based clustering of 6 genotypes of HCV serine proteases NS3. For each genotype of NS3 serine protease a structural model was created by AS2TS system, and sequence similarities were calculated from structural alignments using Local – Global Alignment (LGA) analysis.

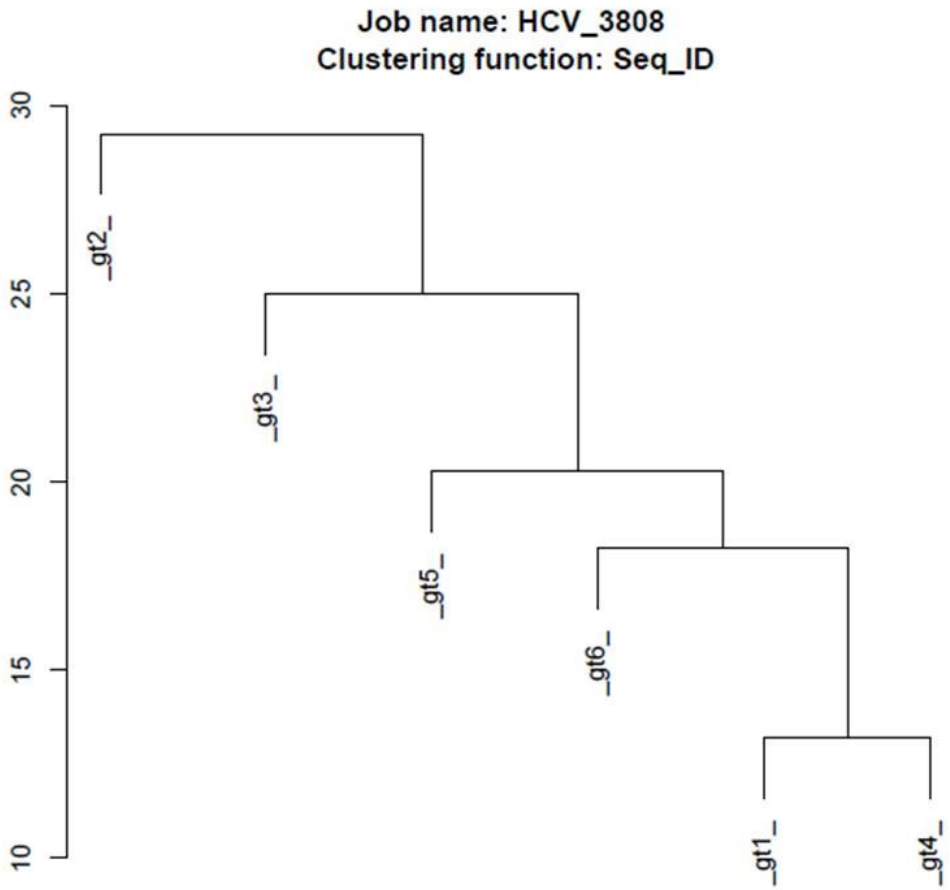
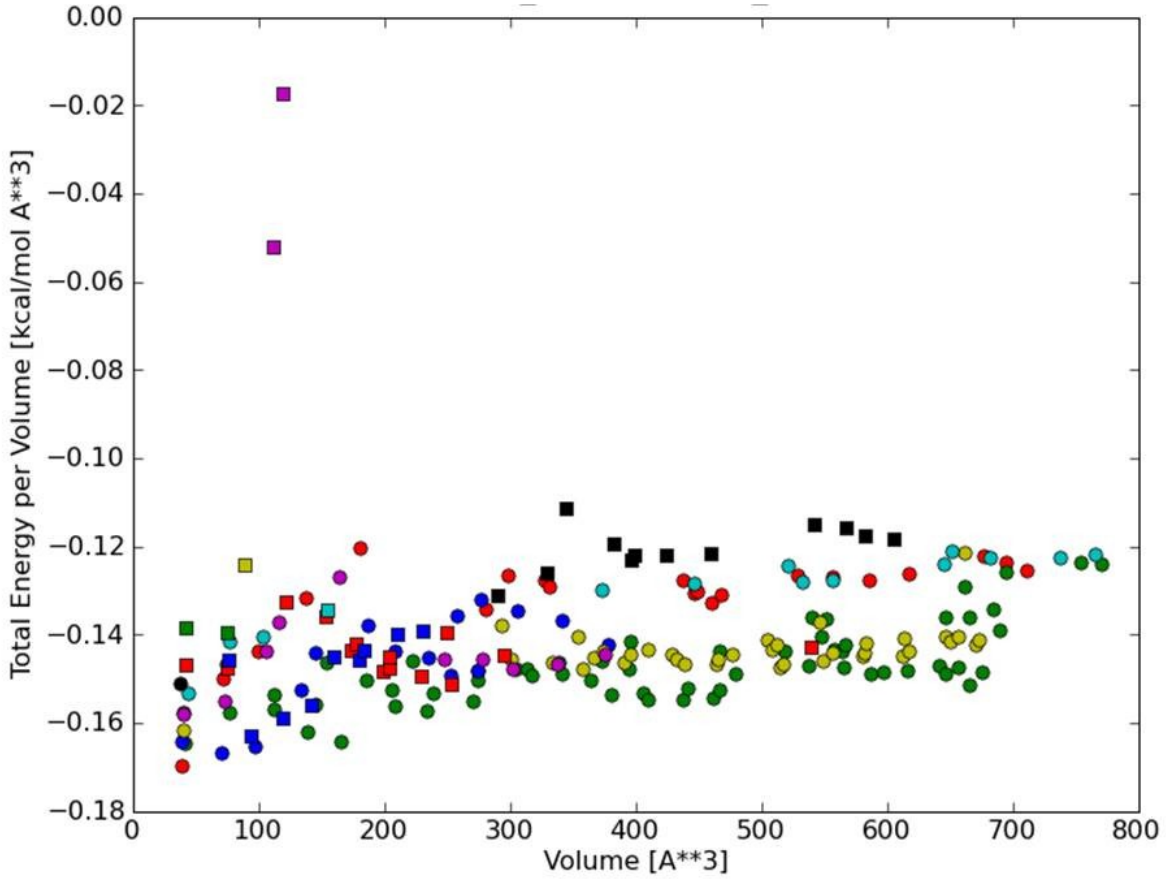
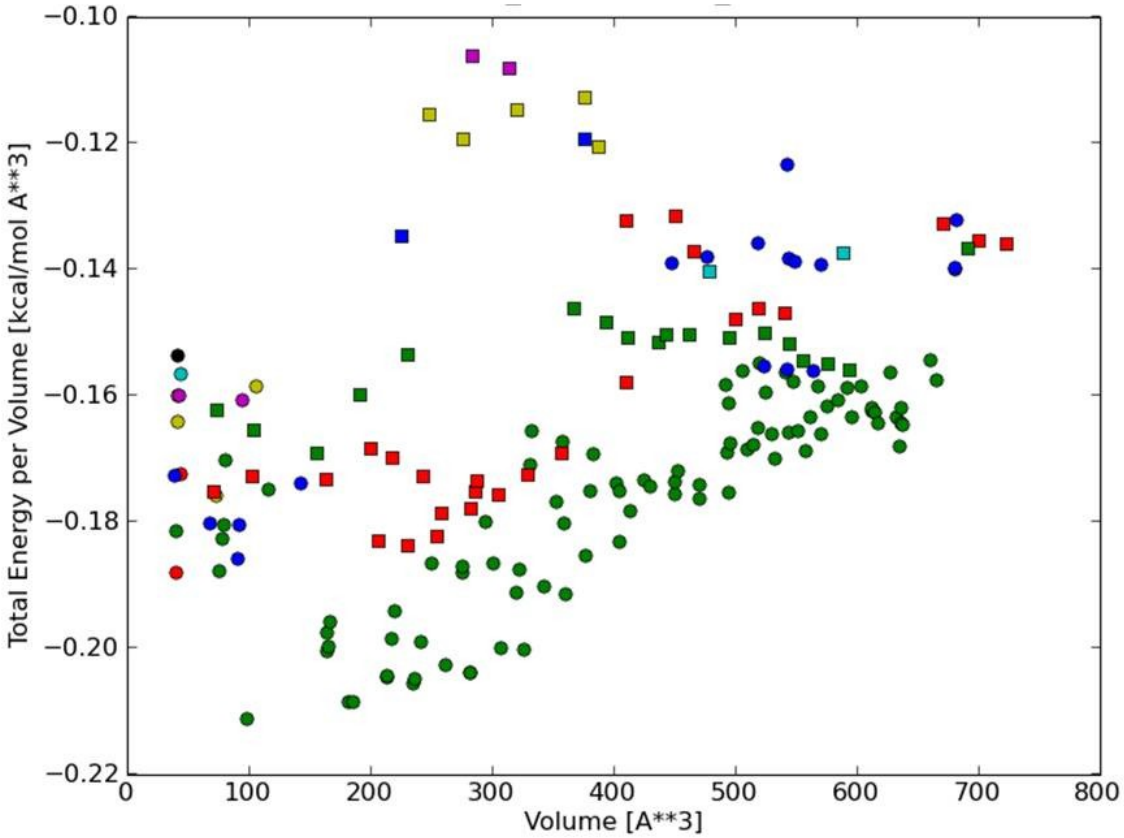


Figure 19: Predicted free energy data for ligand binding sites identified on the surface of CD81-LEL by AutoLigand. Data for fill points within a given site are shown as a set of colored squares or circles. (A) CD81-LEL open conformation (PDB ID: 1G8Q). (B) CD81-LEL closed conformation (PDB ID: 1IV5). Figure prepared using AutoDock Tools version 1.5.6.

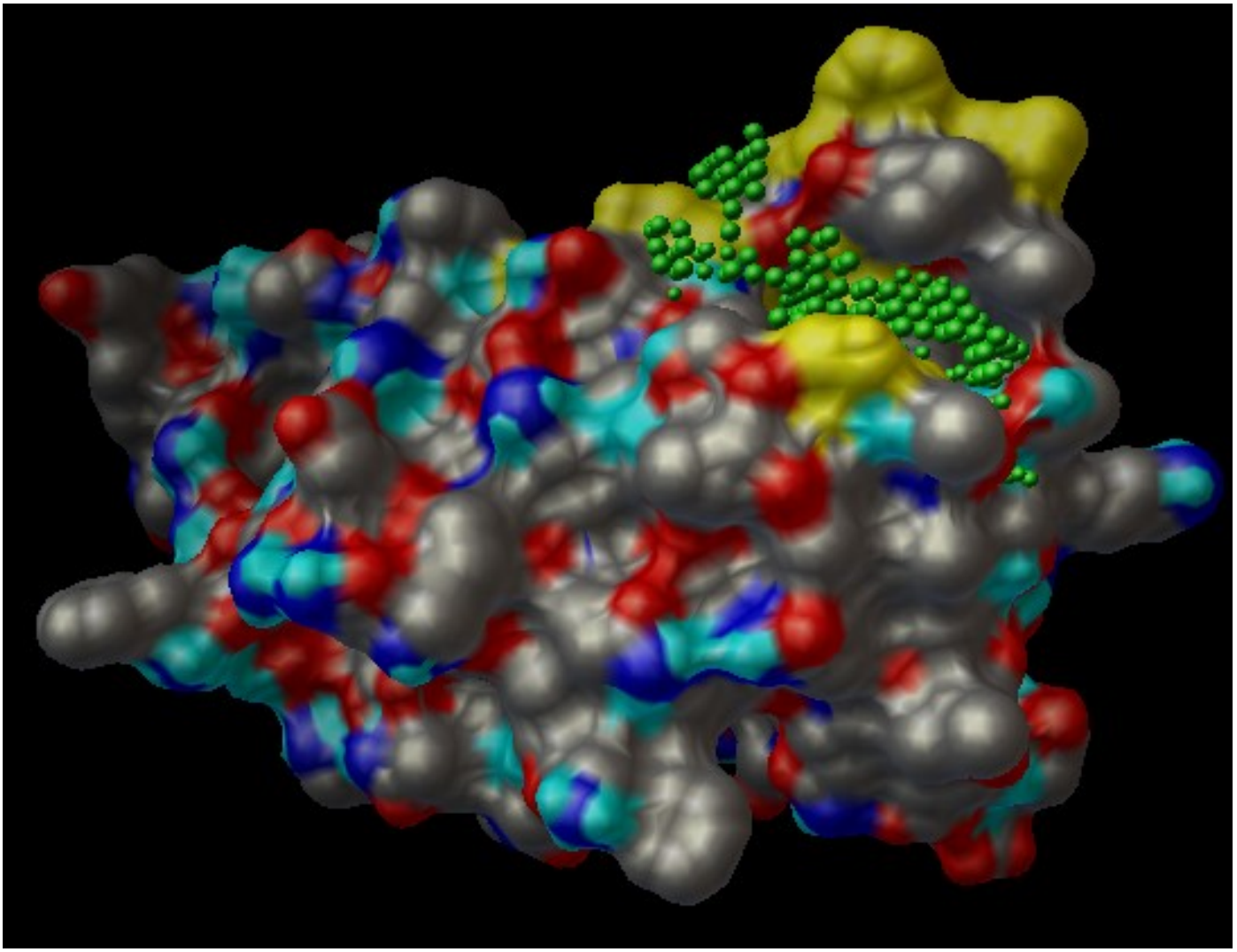


(A)

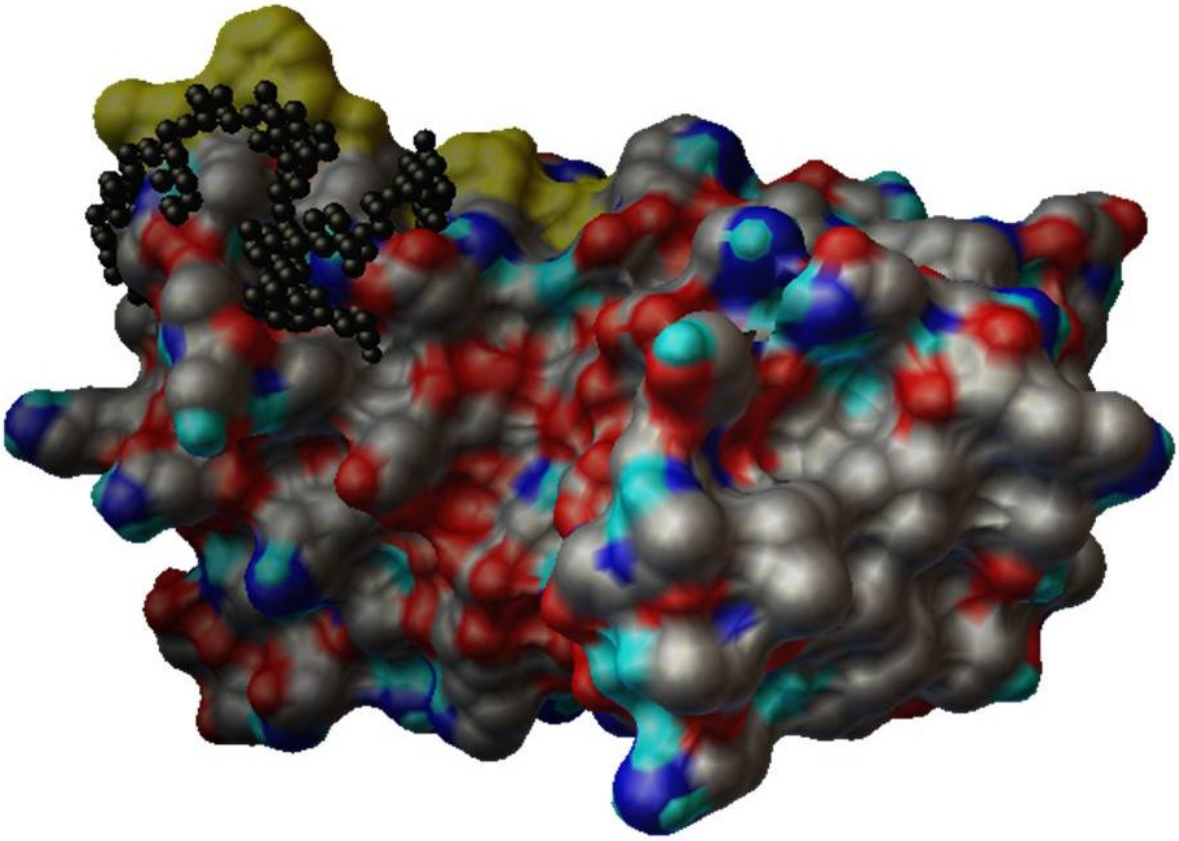

−0.10
−0.12
−0.14
−0.16
−0.18
−0.20
−0.22
Total Energy per Volume [kcal/mol A**3]
0
100
200
300
400
500
600
700
800
Volume [A**3]

(B)

Figure 20: Two ligand binding sites identified by AutoLigand on the open conformation of CD81-LEL (PDB ID: 1G8Q). These two sites were selected as docking targets based on their proximity to the amino acid residues that contact E2 (colored in yellow) and the low free energy (high affinity) predicted for ligands that would bind in this site. (A) The green spheres define the ligand binding site predicted by Autoligand to be the best binding site. Ligands binding to this site would interact directly with several key amino acids in the E2 binding site. The green spheres correspond to the green circle fill points in figure 16A located between 500 and 600 A^3. (B) The black spheres identify a second binding site predicted by AutoLigand on the opposite side of the protein. Ligands binding to this site should also contribute to the disruption of E2 binding. The black spheres correspond to the black square fill points shown in figure 16A located between 550 and 650 A^3. Figure prepared using AutoDock Tools version 1.5.6



(A)



(B)

Figure 21: Two additional binding sites (shown filled with green spheres) predicted by AutoLigand on the open conformation of CD81-LEL (PDB ID: 1G8Q). The amino acids that contact E2 are colored yellow. These binding sites correspond to the fill points shown as green circles in Figure 16A that are located between 200 and 300 A^3. (A) The green fill points identify Site 4. (B) The green fill points identify Site 3. Figure prepared using AutoDock Tools version 1.5.6. Figure prepared using AutoDock Tools version 1.5.6.

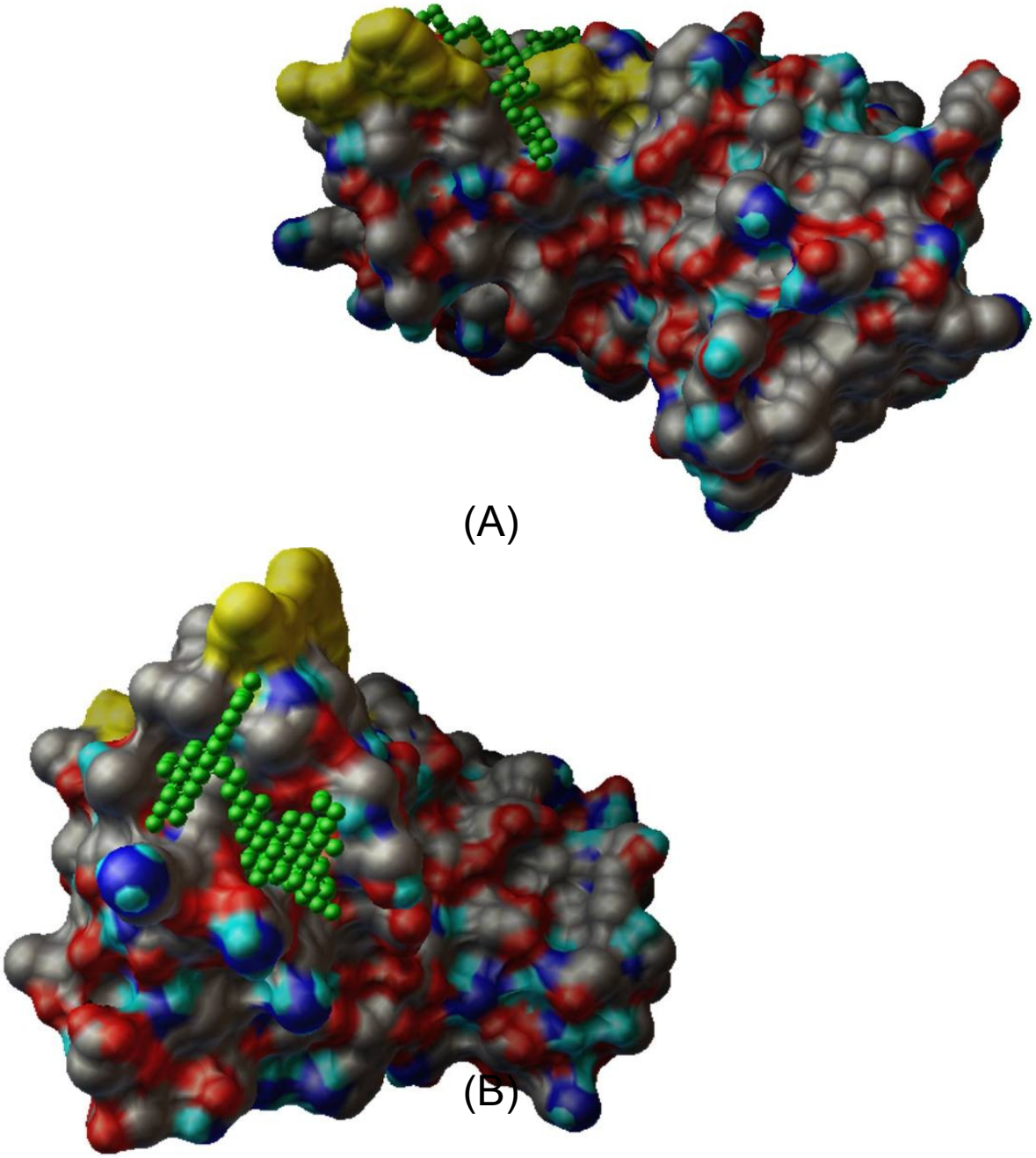


Figure 22: Ligand binding sites predicted by AutoLigand on the surface of the closed conformation of CD81-LEL (PDB ID: 1IV5). Site 1 (the cluster of blue fill points) was located between 500 and 700 A^3 in Figure 3B, and Site 2 (the smaller cluster of green fill points located to the left of the large blue cluster) was located between 500 and 600 A^3 in Figure 3B. Site 3 is marked by a small group of green fill points just barely visible on the far left side of the protein. The yellow residues are the amino acids that have been shown to contribute to E2 binding. Figure prepared using AutoDock Tools version 1.5.6.

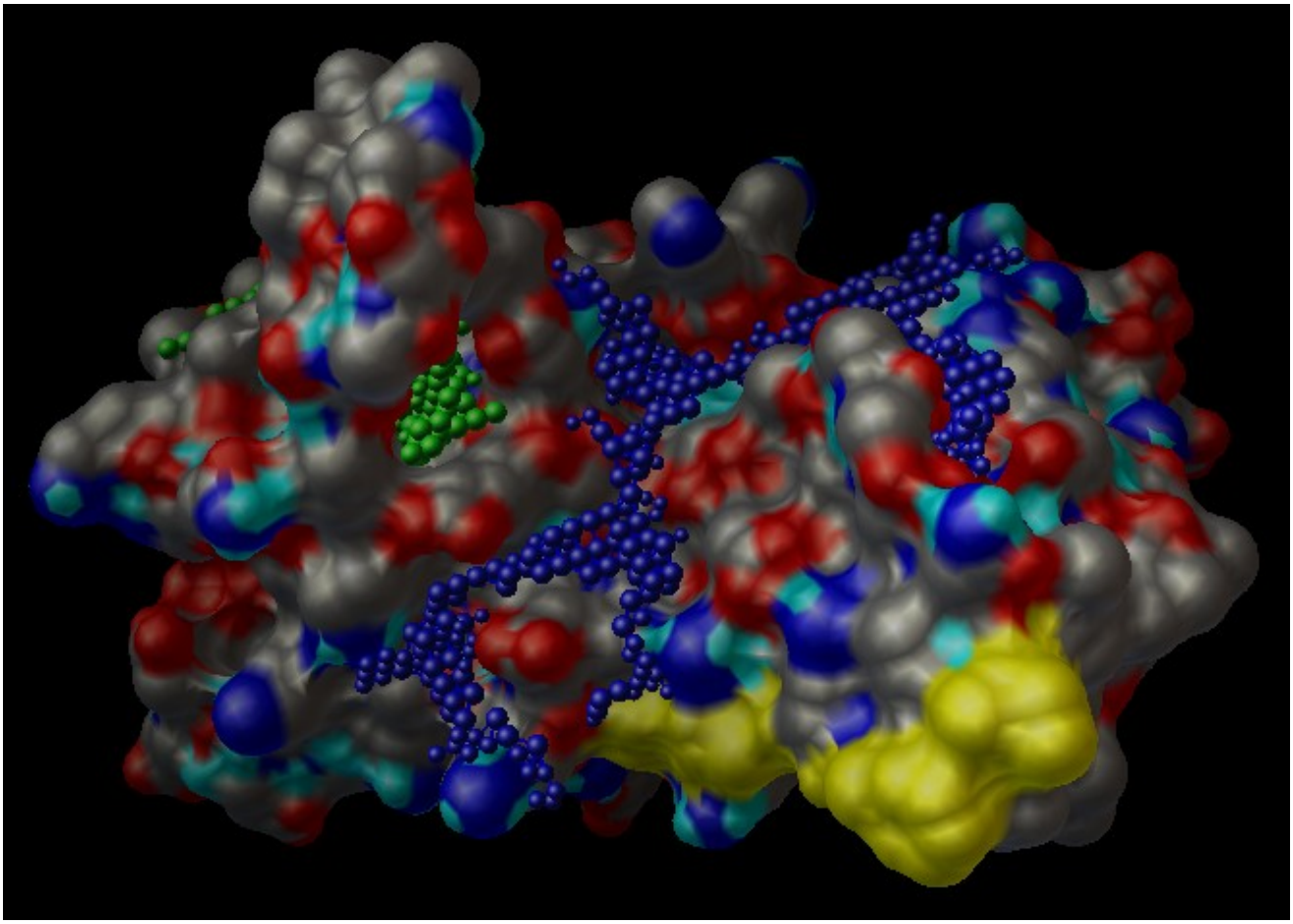
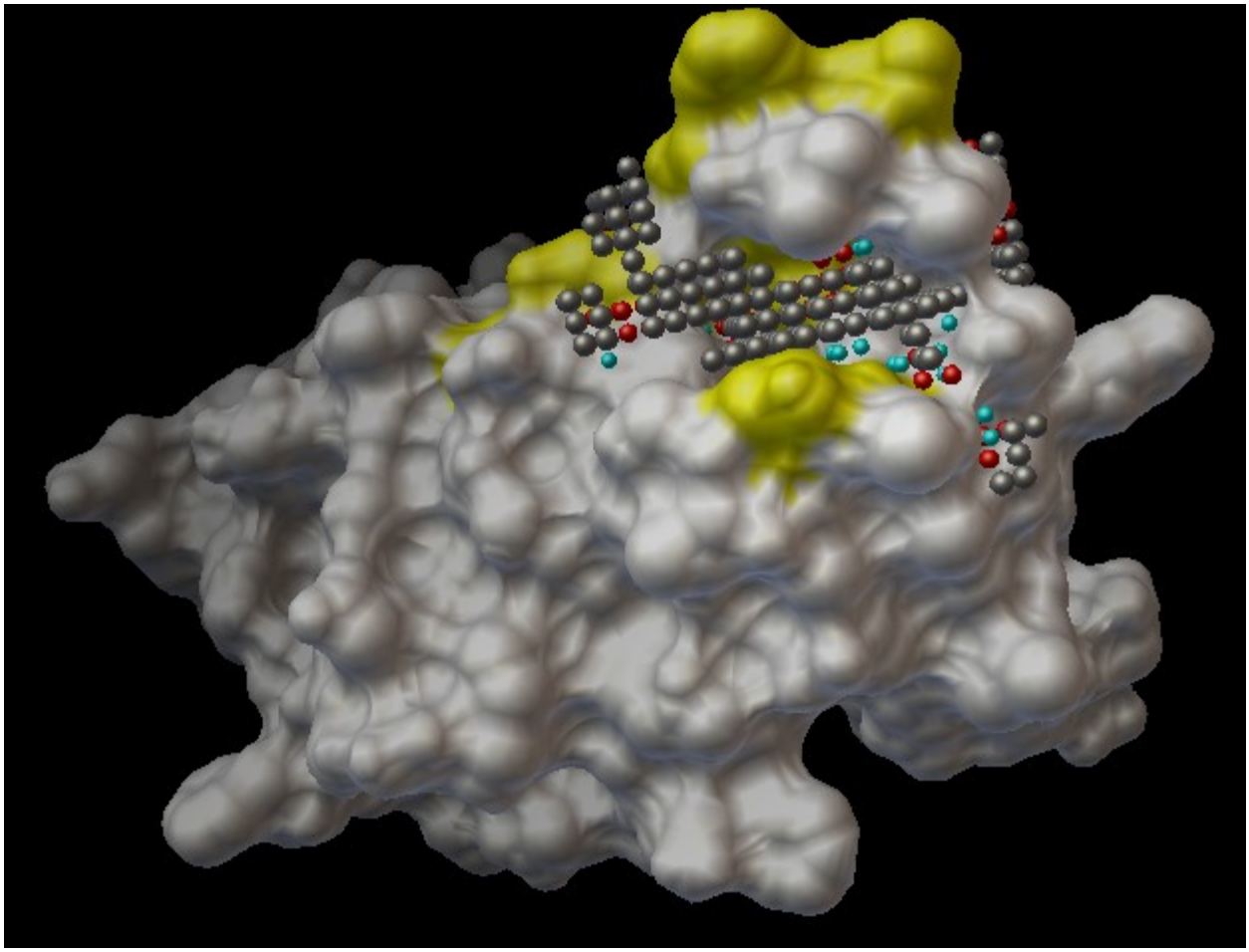
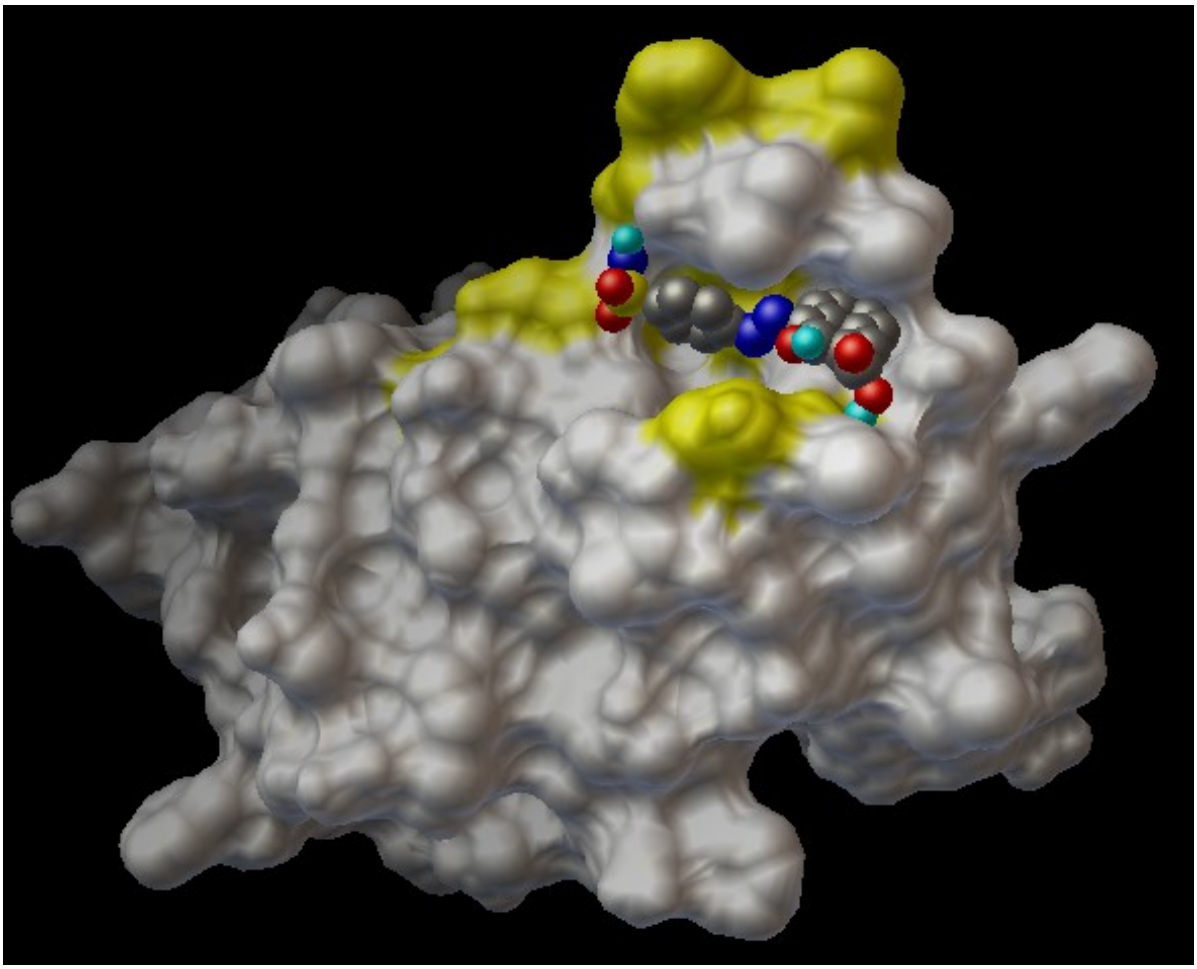


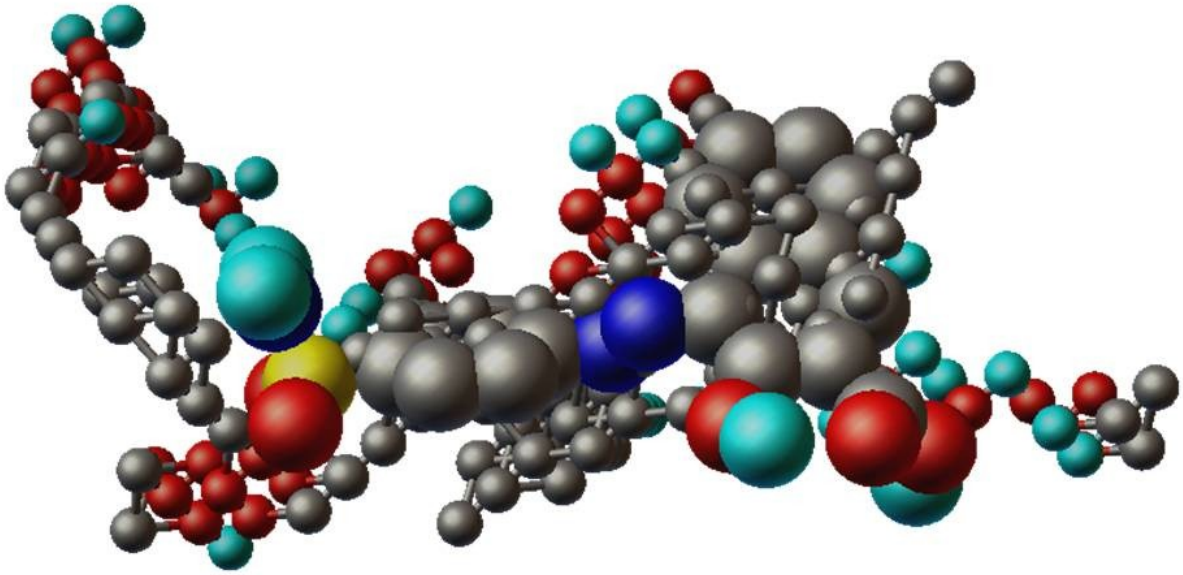
Figure 23: AutoLigand fill points not only identify cavities on the surfaces of proteins, but they also predict the structural features of ligands that would bind with the best affinity and selectivity to the protein at these sites. (A) The fill points provided by AutoLigand define the rough shape of ligands that would fit best into the cavity. Specific fill points are also color coded (gray, carbon; light blue, hydrogen; and red, hydrogen acceptors oxygen and nitrogen) to identify particular atoms in the ligand that would interact optimally with the protein's atoms or functional groups in the regions surrounding the ligand. (B) Ligand 1 (table 6) is shown bound to CD81-LEL in the location and orientation predicted by AutoDock. (C) The superposition of fill points and atom types in Ligand 1 is high (75-80%) indicating that this ligand should bind well in this particular site. Figure prepared using AutoDock Tools version 1.5.6.



(A)



(B)



(C)

Figure 24. Ligand 1 (table 9) is shown binding to Site 2 in the closed conformation of CD81-LEL (PDB ID: 1IV5). The three hydrogen bonds that contribute to the stability of the predicted interaction between the ligand and protein are shown as gray spheres. Figure prepared using AutoDock Tools version 1.5.6

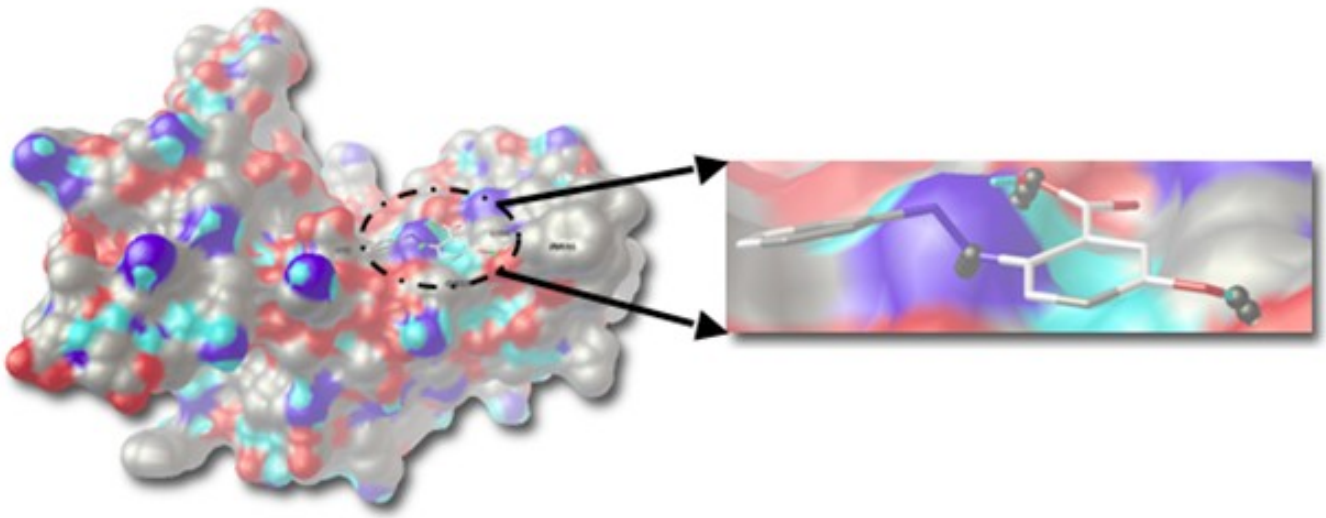
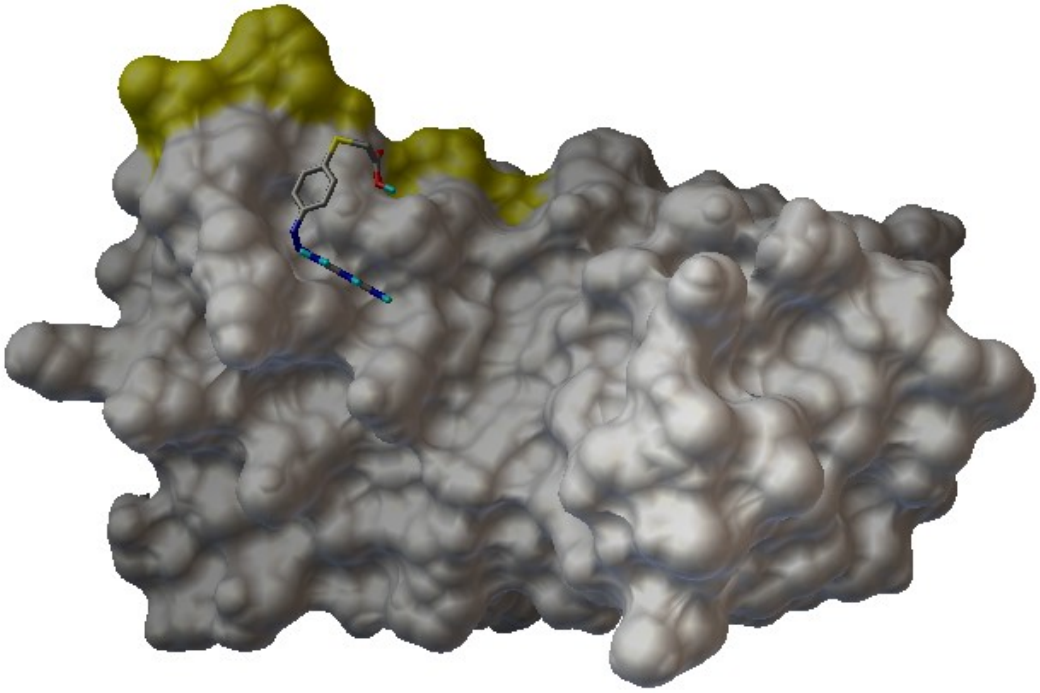
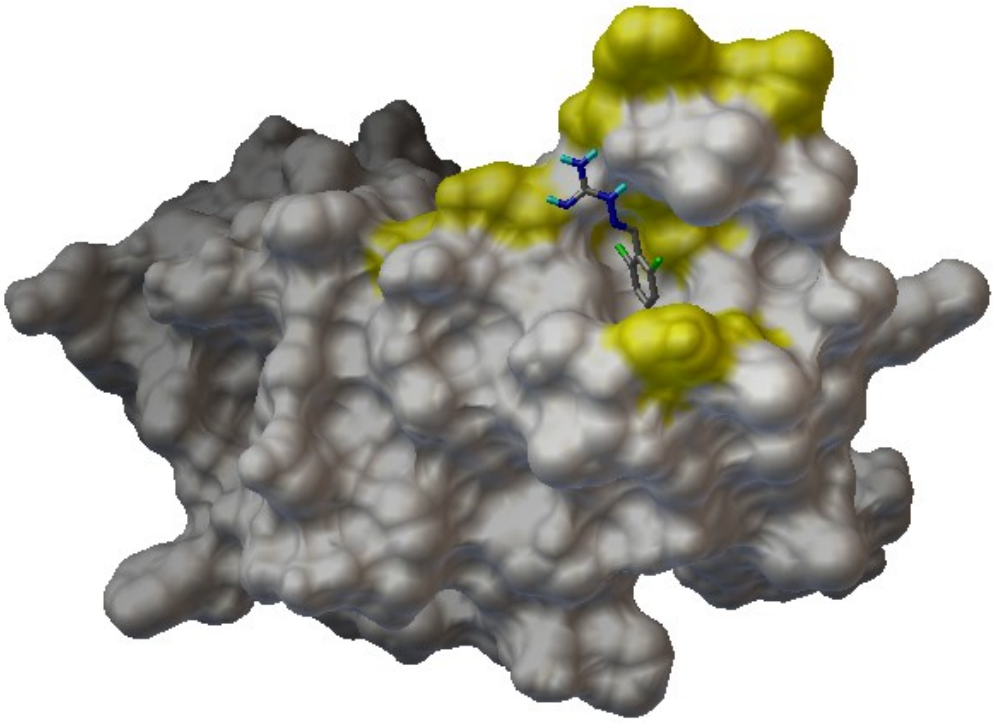


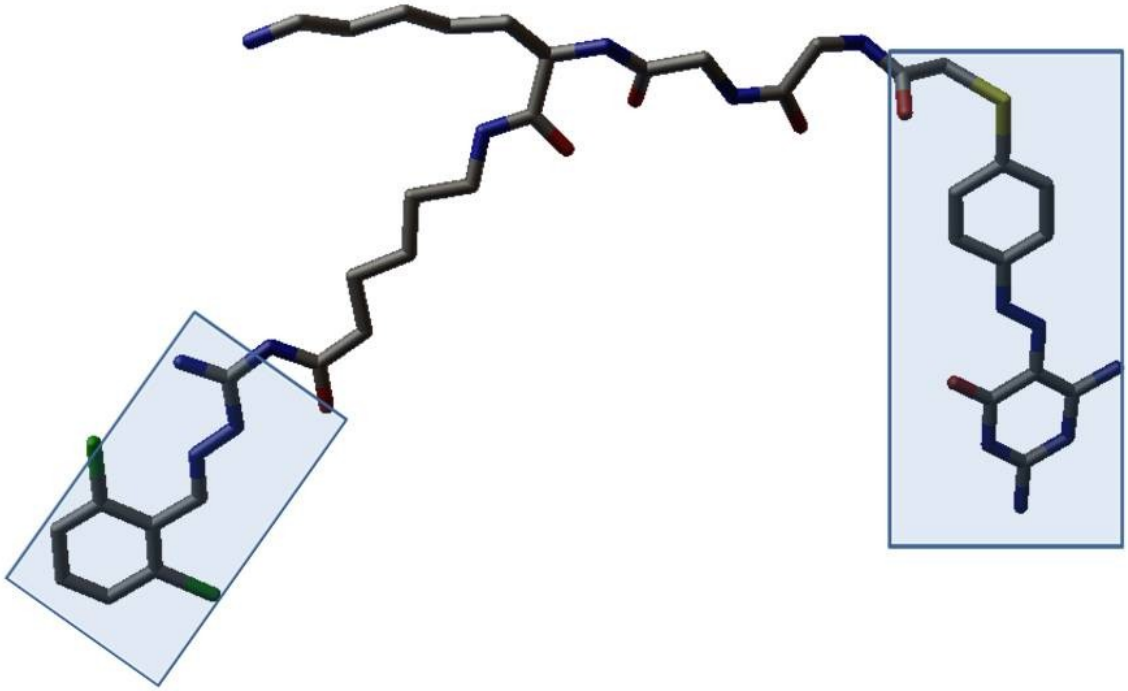
Figure 25: An example of a bidentate SHAL designed using two ligands predicted to bind to Site 1 and Site 2 on the open conformation of CD81-LEL. (A) Location and orientation of ligand 2 (table 7) binding predicted by AutoDock. (B) Location and orientation of ligand 4 (table 6) binding predicted by AutoDock. (C) Structure of the bidentate SHAL. The linker was created using a single miniPEG and lysine and the resulting spacing between ligands was sufficient to allow the individual ligands to bind to the predicted sites.



(A)



(B)



(C)

Figure 26. An example of a tridentate SHAL designed to bind to three sites on the open conformation of CD81-LEL. This SHAL was designed using three ligands in which one ligand (Ligand 2) has both a free amino and a free carboxyl group.

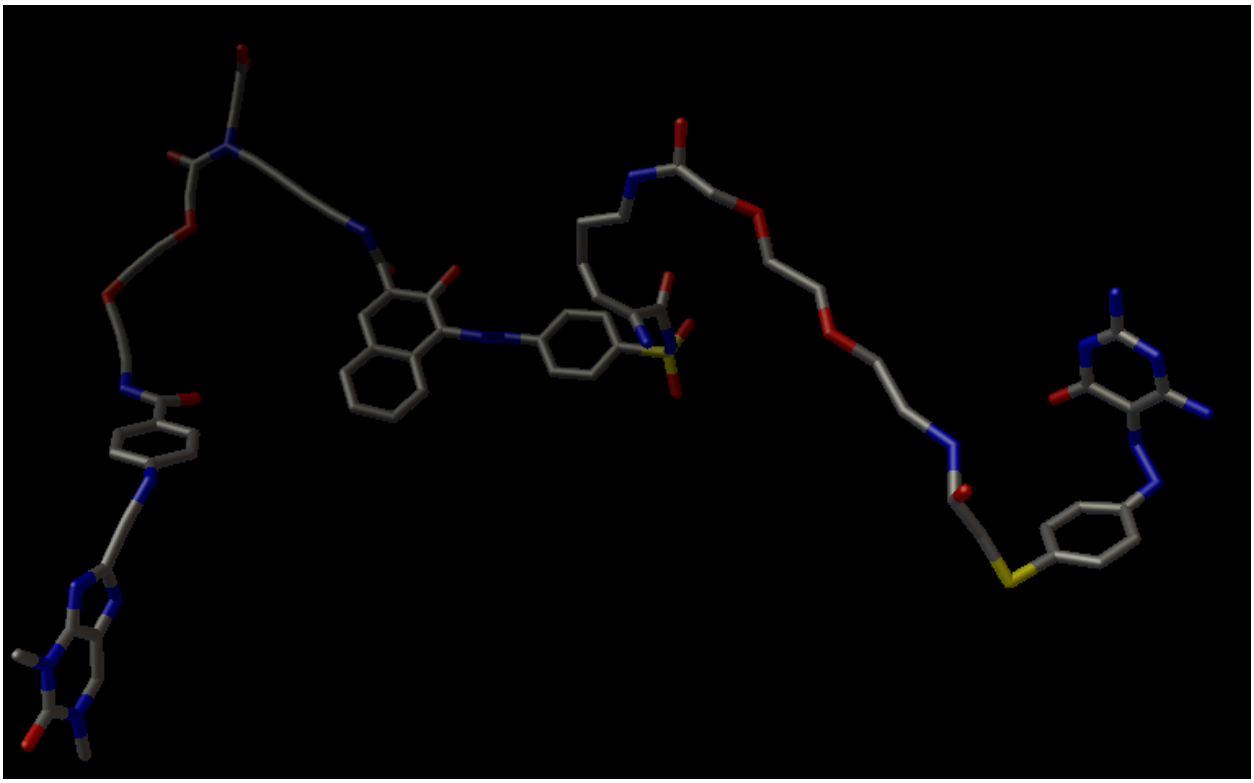
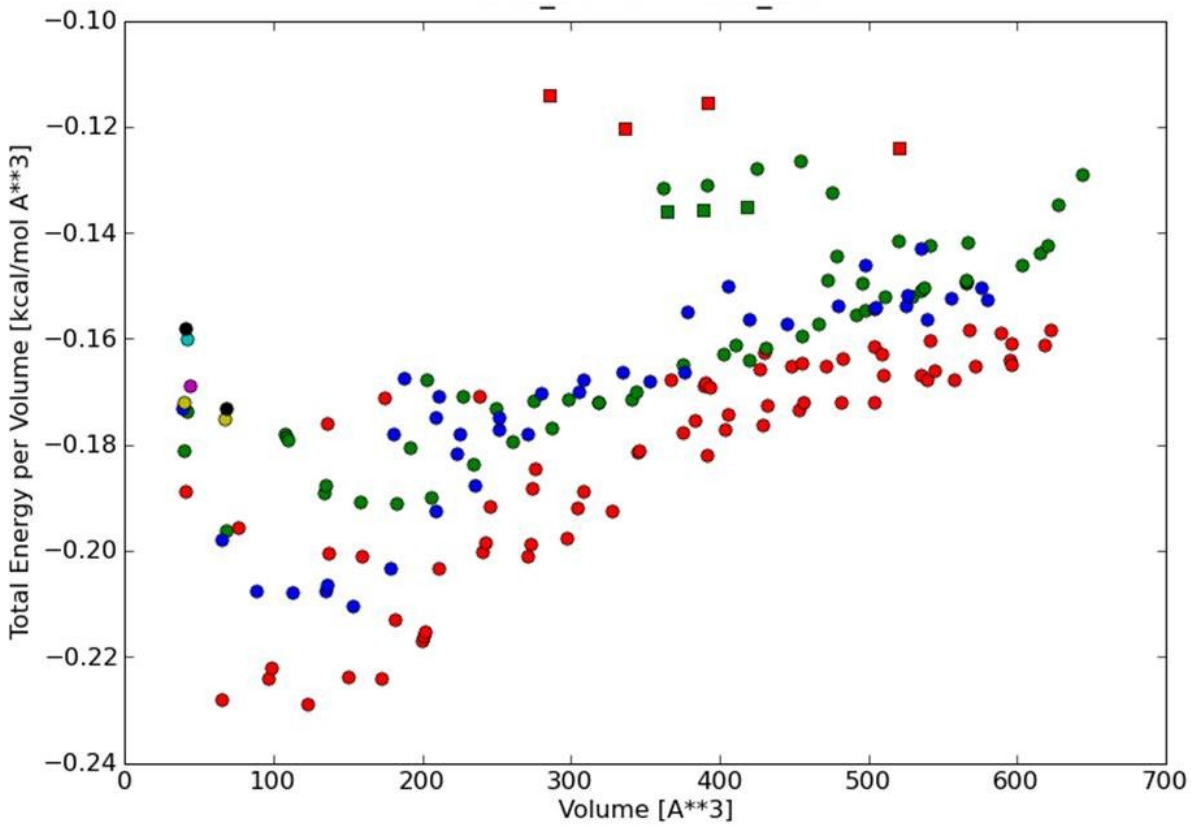
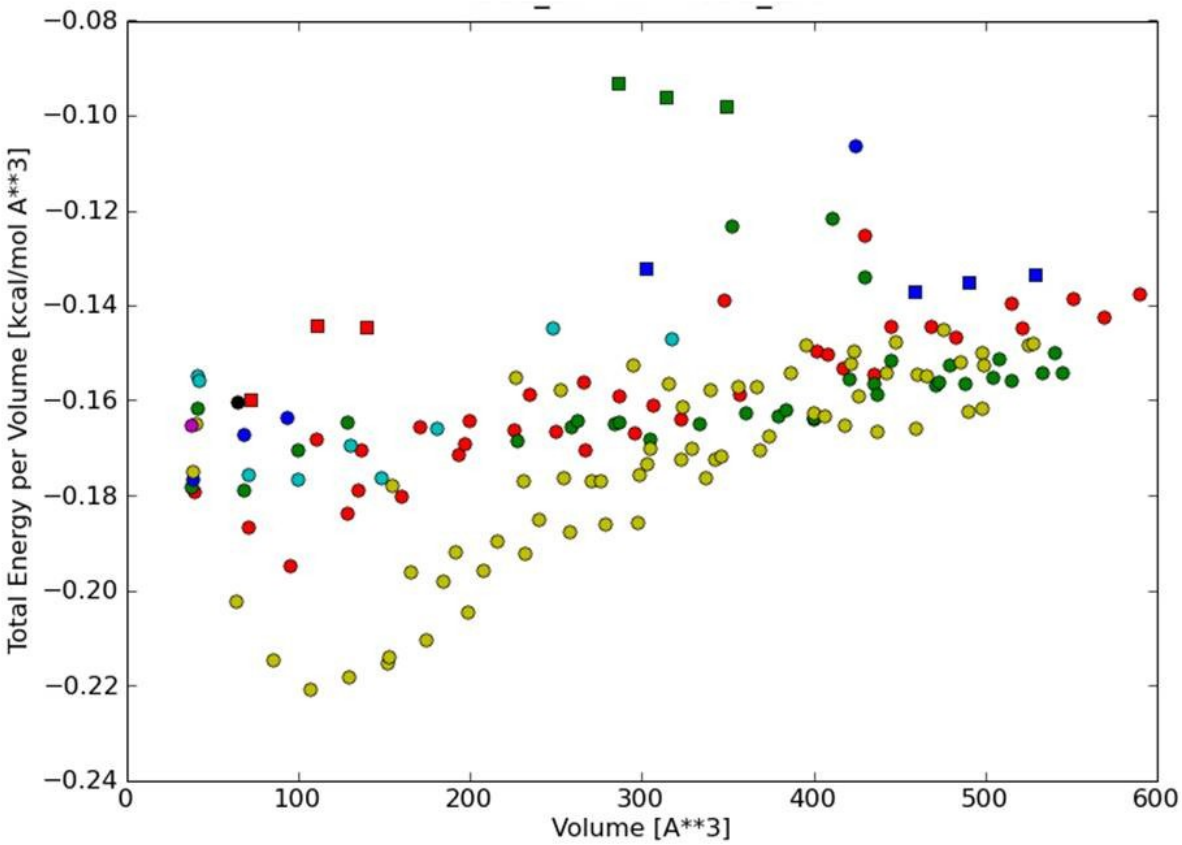


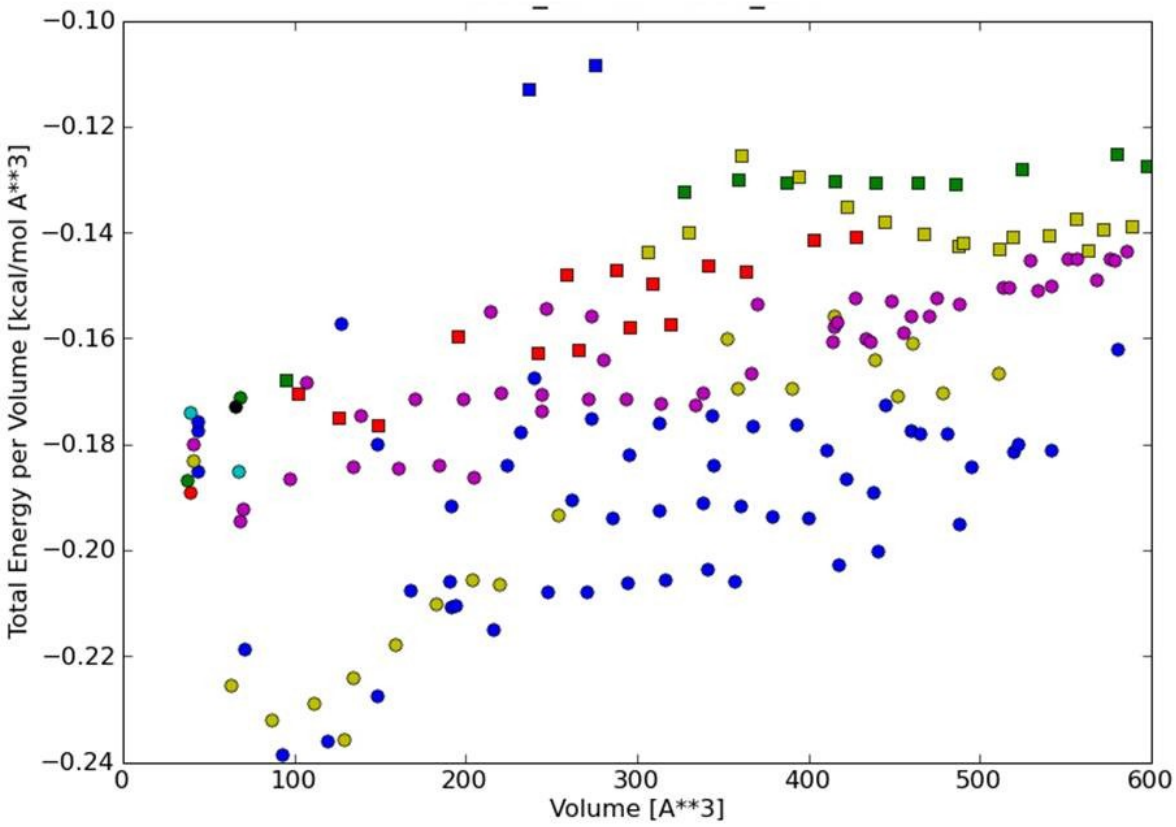
Figure 27: Predicted binding sites using plot_set_results for HCV NS3 protease genotype 1 (A), 2 (B), 3 (C), 4 (D), 5 (E), and 6 (F).



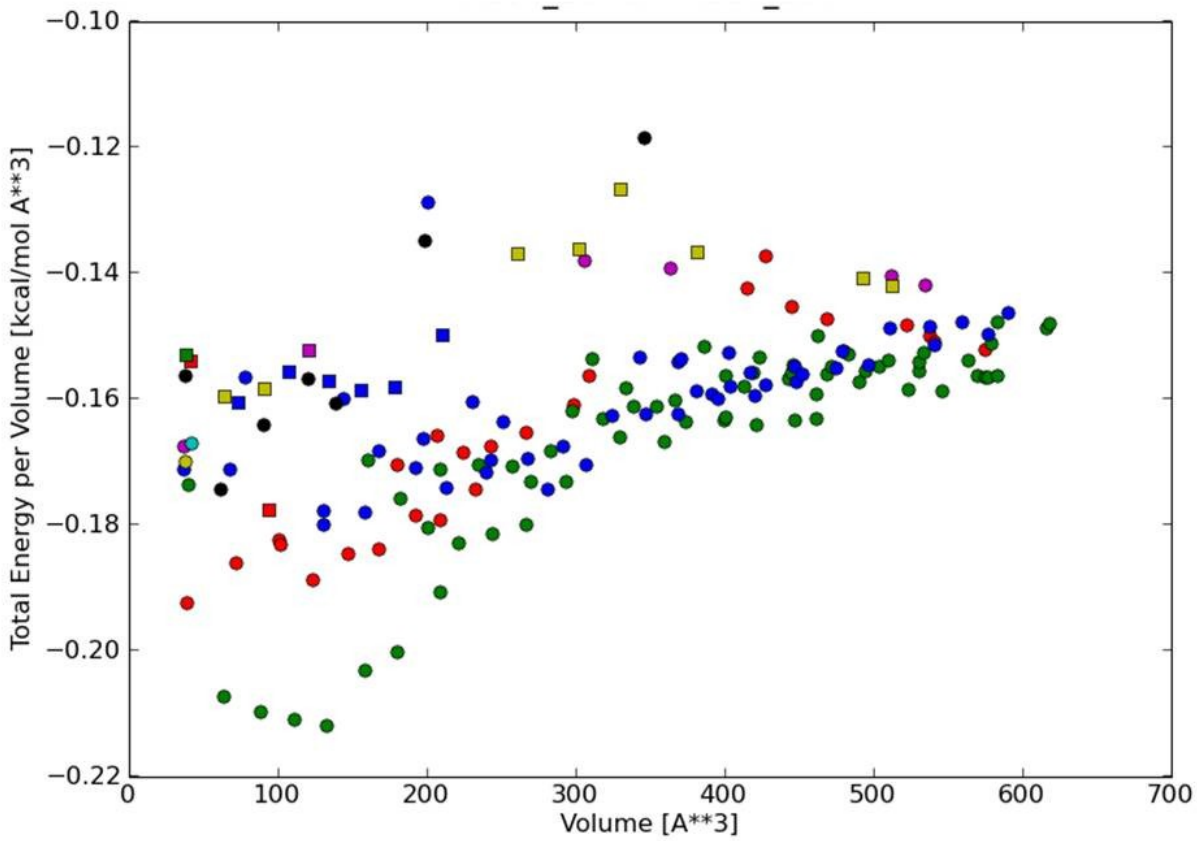
(A)


−0.08
−0.10
−0.12
−0.14
−0.16
−0.18
−0.20
−0.22
−0.24
Total Energy per Volume [kcal/mol A**3]
0
100
200
300
400
500
600
Volume [A**3]

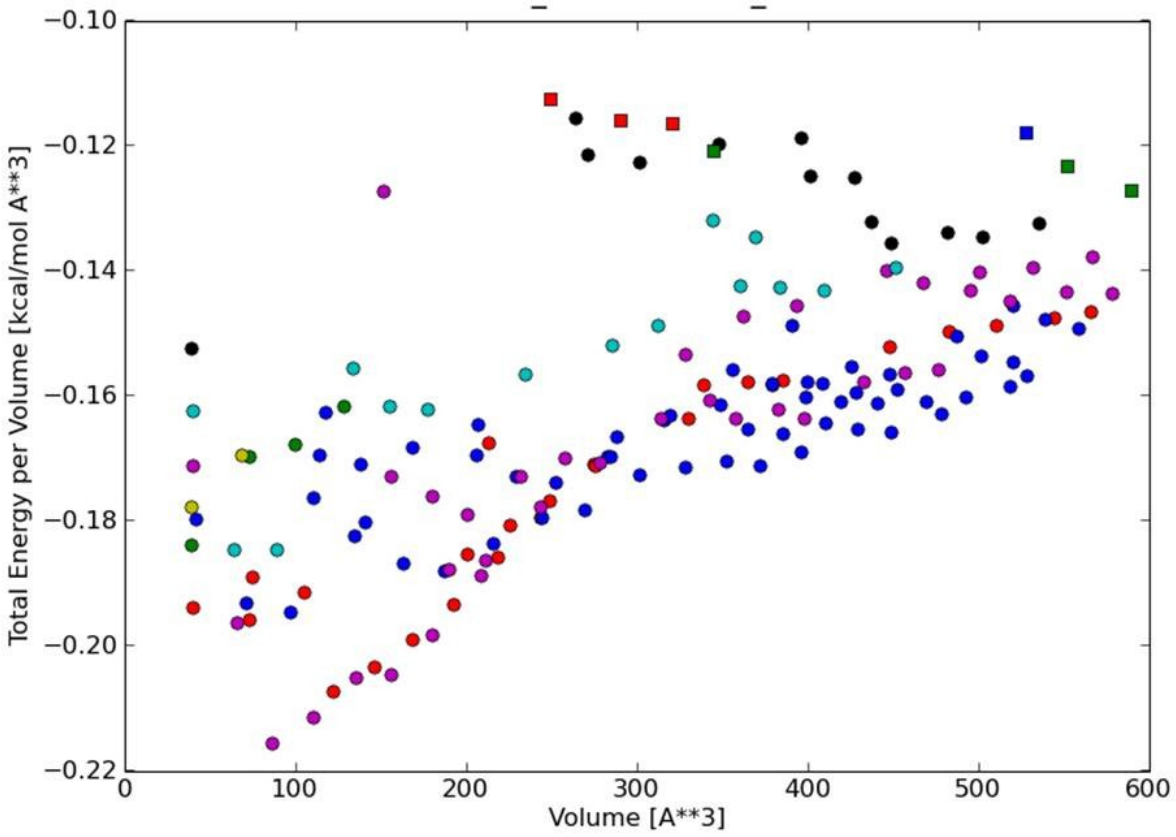
(B)


−0.10
−0.12
−0.14
−0.16
−0.18
−0.20
−0.22
−0.24
Total Energy per Volume [kcal/mol A**3]
0
100
200
300
400
500
600
Volume [A**3]

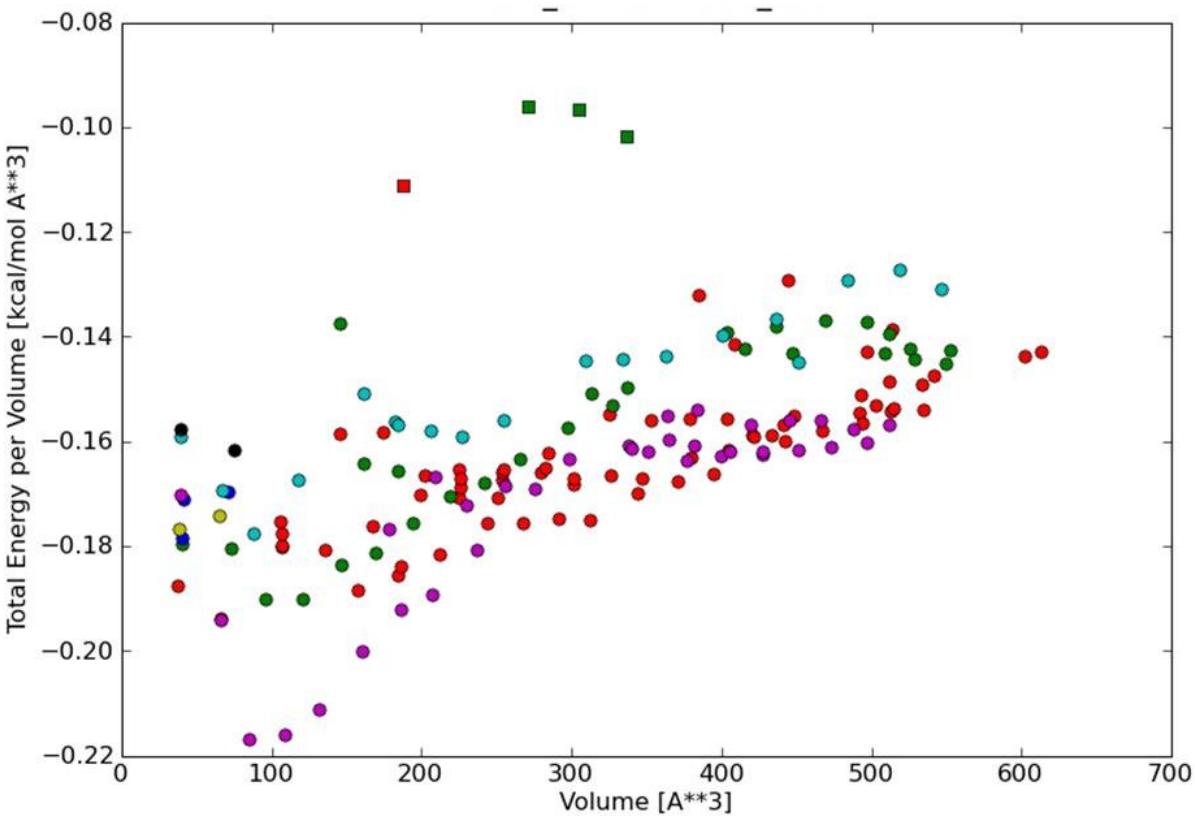
(C)


Total Energy per Volume [kcal/mol A**3]
Volume [A**3]
−0.10
−0.12
−0.14
−0.16
−0.18
−0.20
−0.22
0
100
200
300
400
500
600
700

(D)


Total Energy per Volume [kcal/mol A**3]
Volume [A**3]
−0.10
−0.12
−0.14
−0.16
−0.18
−0.20
−0.22
0
100
200
300
400
500
600

(E)


−0.08
−0.10
−0.12
−0.14
−0.16
−0.18
−0.20
−0.22
Total Energy per Volume [kcal/mol A**3]
0
100
200
300
400
500
600
700
Volume [A**3]

(F)

Figure 28: NS3 residues involved in complexation of NS3-4A together with one of the autoligand predicted sites in genotype 4 represented by the blue balls that correspond to the blue circles in the plot figure 23-D that are in the volume range of 400 to 600 A^3.

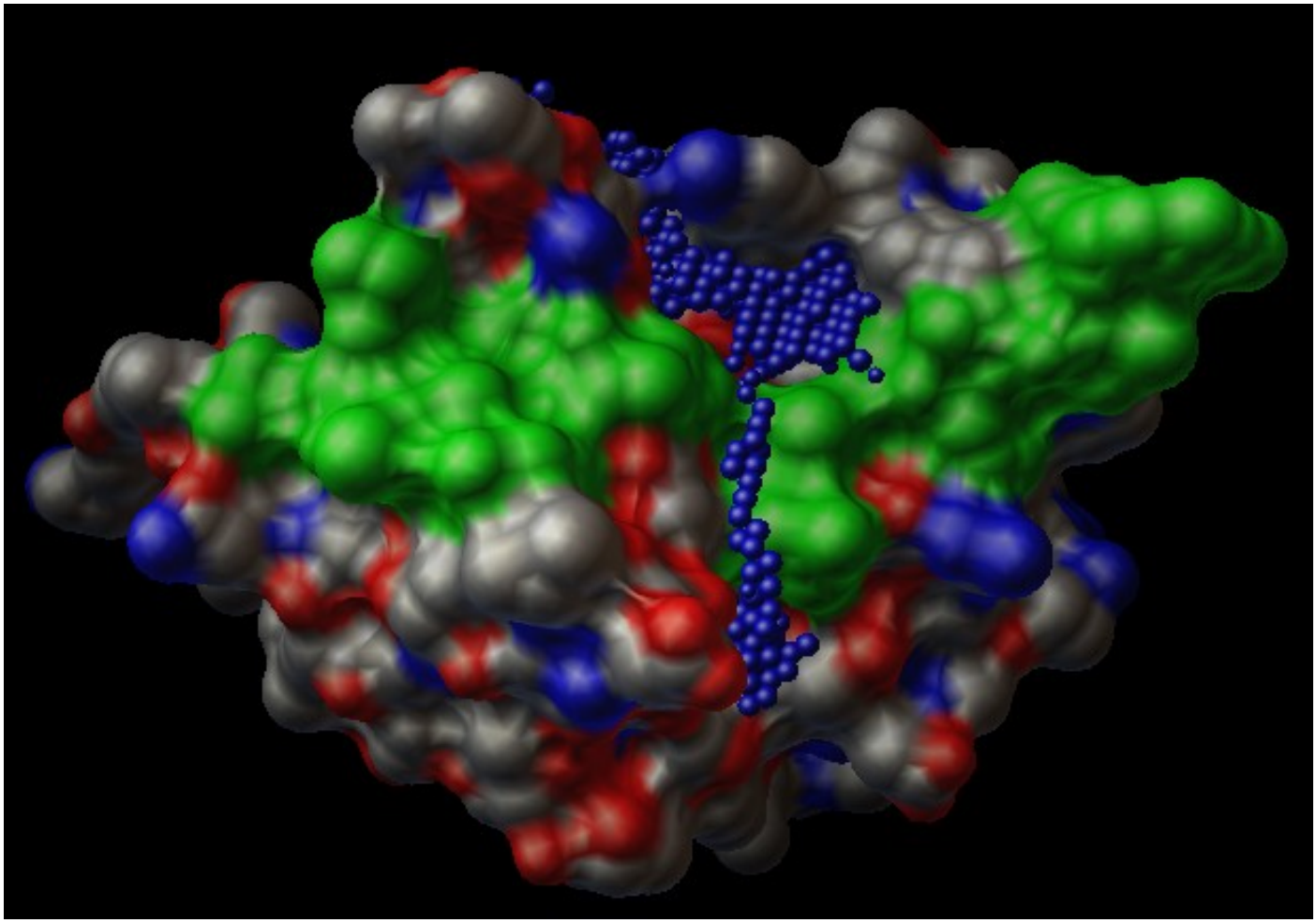


Figure 29: An example of an *in silico* designed SHAL against NS3 serine protease genotype 4 model. The distance between the amino group of one ligand and the carboxyl group of the other was around 20 A (A). The SHAL was designed according to the distance between the 2 ligands and their orientation (B).

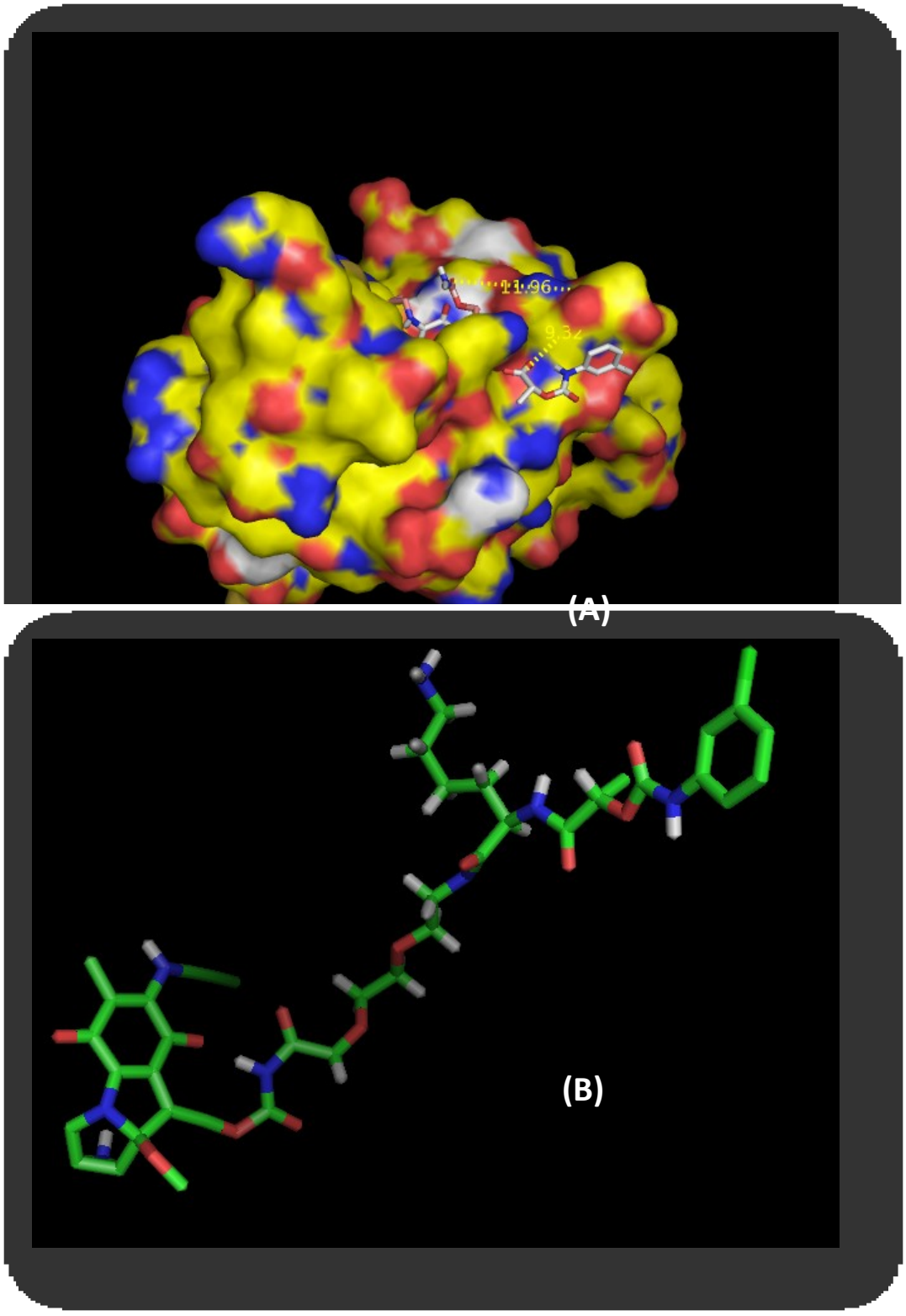
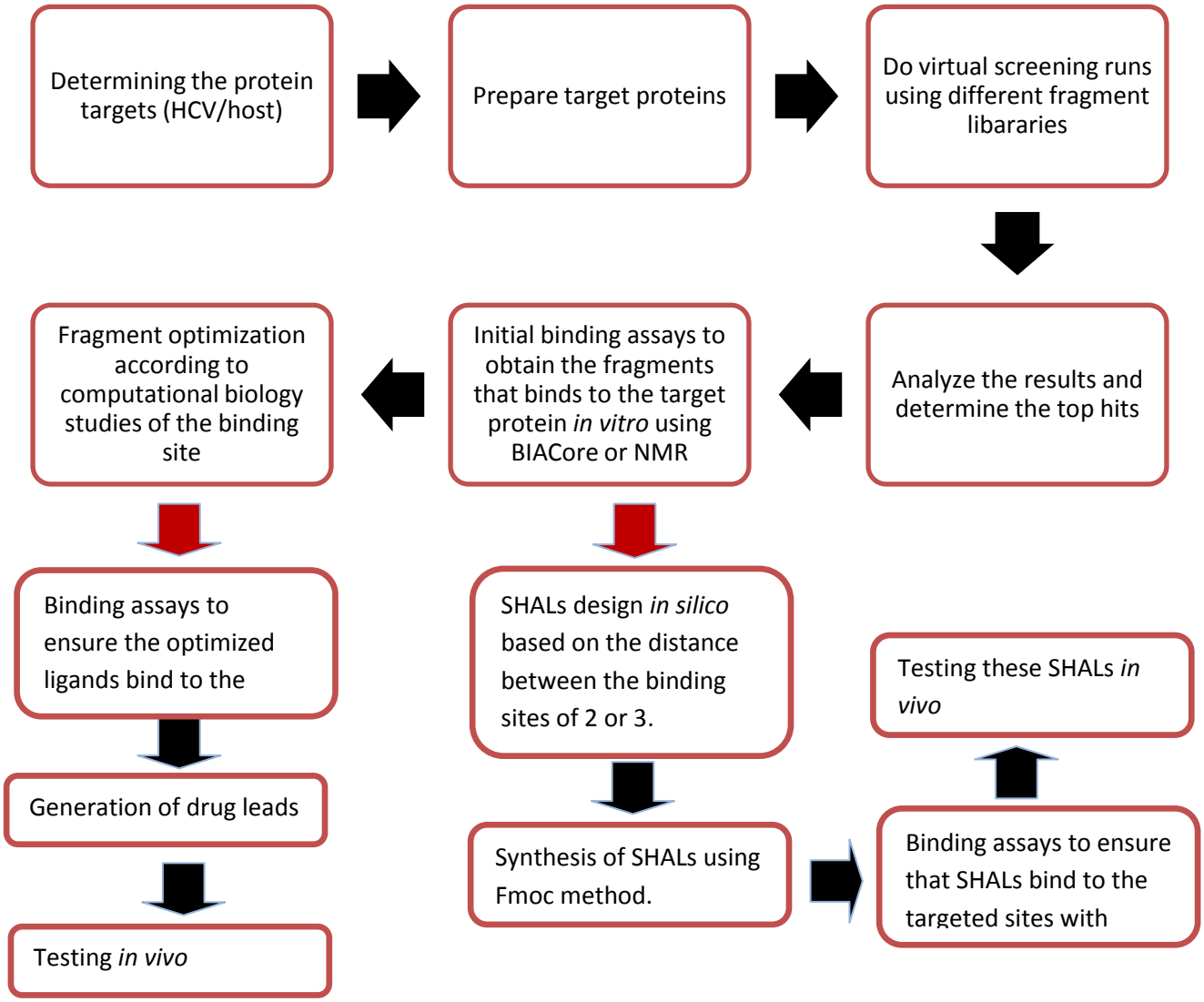


Figure 30. The flowchart of the overall process of generating SHALs against different HCV targets.



REFERENCES

1 Lavanchy, D.(2009).The global burden of hepatitis C. Liver Int. **29 Suppl 1,** 74-81

2 Houghton, M.(2009).Discovery of the hepatitis C virus. Liver Int. **29 Suppl 1,** 82-88.

3 Krekulova, L., Rehak, V. and Riley, L. W. (2006).Structure and functions of hepatitis C virus proteins: 15 years after. Folia Microbiol. (Praha)**. 51,** 665-680.

4 Dubuisson, J.(2007).Hepatitis C virus proteins. World J. Gastroenterol. **13,** 2406-2415.

5 Catanese, M. T., Graziani, R., von Hahn, T., Moreau, M., Huby, T., Paonessa, G., Santini, C., Luzzago, A., Rice, C. M., Cortese, R., Vitelli, A. and Nicosia, A. (2007).High-avidity monoclonal antibodies against the human scavenger class B type I receptor efficiently block hepatitis C virus infection in the presence of high-density lipoprotein. J. Virol. **81,** 8063-8071.

6 Evans, M. J., von Hahn, T., Tscherne, D. M., Syder, A. J., Panis, M., Wolk, B., Hatziioannou, T., McKeating, J. A., Bieniasz, P. D. and Rice, C. M. (2007).Claudin-1 is a hepatitis C virus co-receptor required for a late step in entry. Nature**. 446,** 801-805.

7 Scarselli, E., Ansuini, H., Cerino, R., Roccasecca, R. M., Acali, S., Filocamo, G., Traboni, C., Nicosia, A., Cortese, R. and Vitelli, A. (2002).The human scavenger receptor class B type I is a novel candidate receptor for the hepatitis C virus. EMBO J. **21,** 5017-5025.

8 Pileri, P., Uematsu, Y., Campagnoli, S., Galli, G., Falugi, F., Petracca, R., Weiner, A. J., Houghton, M., Rosa, D., Grandi, G. and Abrignani, S. (1998).Binding of hepatitis C virus to CD81. Science**.** 282, 938-941.

9 Melagraki, G., Afantitis, A., Sarimveis, H., Koutentis, P. A., Markopoulos, J. and Igglessi-Markopoulou, O. (2007).Identification of a series of novel derivatives as potent HCV inhibitors by a ligand-based virtual screening optimized procedure. Bioorg. Med. Chem. 15, 7237-7247.

10 Owsianka, A. M., Tarr, A. W., Keck, Z. Y., Li, T. K., Witteveldt, J., Adair, R., Foung, S. K., Ball, J. K. and Patel, A. H. (2008).Broadly neutralizing human monoclonal antibodies to the hepatitis C virus E2 glycoprotein. J. Gen. Virol. 89, 653-659.

11 Perera, R., Khaliq, M. and Kuhn, R. J. (2008).Closing the door on flaviviruses: entry as a target for antiviral drug design. Antiviral Res. 80, 11-22.

12 VanCompernolle, S. E., Wiznycia, A. V., Rush, J. R., Dhanasekaran, M., Baures, P. W. and Todd, S. C. (2003).Small molecule inhibition of hepatitis C virus E2 binding to CD81. Virology. 314, 371-380.

13 Zeisel, M. B., Koutsoudakis, G., Schnober, E. K., Haberstroh, A., Blum, H. E., Cosset, F. L., Wakita, T., Jaeck, D., Doffoel, M., Royer, C., Soulier, E., Schvoerer, E., Schuster, C., Stoll-Keller, F., Bartenschlager, R., Pietschmann, T., Barth, H. and Baumert, T. F. (2007).Scavenger receptor class B type I is a key host factor for hepatitis C virus infection required for an entry step closely linked to CD81. Hepatology. **46**, 1722-1731.

14 Krieger, S. E., Zeisel, M. B., Davis, C., Thumann, C., Harris, H. J., Schnober, E. K., Mee, C., Soulier, E., Royer, C., Lambotin, M., Grunert, F., Dao Thi, V. L., Dreux, M., Cosset, F. L., McKeating, J. A., Schuster, C. and Baumert, T. F. (2010).Inhibition of hepatitis C virus infection by anti-claudin-1 antibodies is mediated by neutralization of E2-CD81-claudin-1 associations. Hepatology. **51**, 1144-1157.

15 Kwong, A. D., McNair, L., Jacobson, I. and George, S. (2008).Recent progress in the development of selected hepatitis C virus NS3.4A protease and NS5B polymerase inhibitors. Curr. Opin. Pharmacol. 8, 522-531.

16 Chockalingam, K., Simeon, R. L., Rice, C. M. and Chen, Z. (2010).A cell protection screen reveals potent inhibitors of multiple stages of the hepatitis C virus life cycle. Proc. Natl. Acad. Sci. U. S. A. **107**, 3764-3769.

17 Balhorn, R., Hok, S., Burke, P. A., Lightstone, F. C., Cosman, M., Zemla, A., Mirick, G., Perkins, J., Natarajan, A., Corzett, M., DeNardo, S. J., Albrecht, H., Gregg, J. P. and DeNardo, G. L. (2007).Selective high-affinity ligand antibody mimics for cancer diagnosis and therapy: initial application to lymphoma/leukemia. Clin. Cancer Res. **13**, 5621s-5628s.

18 Balhorn, R., Hok, S., DeNardo, S., Natarajan, A., Mirick, G., Corzett, M. and Denardo, G. (2009).Hexa-arginine enhanced uptake and residualization of selective high affinity ligands by Raji lymphoma cells. Mol. Cancer**. 8**, 25.

19 DeNardo, G. L., Natarajan, A., Hok, S., Mirick, G., DeNardo, S. J., Corzett, M., Sysko, V., Lehmann, J., Beckett, L. and Balhorn, R. (2008).Nanomolecular HLA-DR10 antibody mimics: A potent system for molecular targeted therapy and imaging. Cancer Biother. Radiopharm. **23**, 783-796.

20 Hok, S., Natarajan, A., Balhorn, R., DeNardo, S. J., DeNardo, G. L. and Perkins, J. (2007).Synthesis and radiolabeling of selective high-affinity ligands designed to target non-Hodgkin's lymphoma and leukemia. Bioconjug. Chem. **18**, 912-921.

21 Eren, R., Landstein, D., Terkieltaub, D., Nussbaum, O., Zauberman, A., Ben-Porath, J., Gopher, J., Buchnick, R., Kovjazin, R., Rosenthal-Galili, Z., Aviel, S., Ilan, E., Shoshany, Y., Neville, L., Waisman, T., Ben-Moshe, O., Kischitsky, A., Foung, S. K., Keck, Z. Y., Pappo, O., Eid, A., Jurim, O., Zamir, G., Galun, E. and Dagan, S. (2006).Preclinical evaluation of two neutralizing human monoclonal antibodies against hepatitis C virus (HCV): a potential treatment to prevent HCV reinfection in liver transplant patients. J. Virol. **80,** 2654-2664.

22 Galun, E., Terrault, N. A., Eren, R., Zauberman, A., Nussbaum, O., Terkieltaub, D., Zohar, M., Buchnik, R., Ackerman, Z., Safadi, R., Ashur, Y., Misrachi, S., Liberman, Y., Rivkin, L. and Dagan, S. (2007).Clinical evaluation (Phase I) of a human monoclonal antibody against hepatitis C virus: safety and antiviral activity. J. Hepatol. **46,** 37-44.

23 Schiano, T. D., Charlton, M., Younossi, Z., Galun, E., Pruett, T., Tur-Kaspa, R., Eren, R., Dagan, S., Graham, N., Williams, P. V. and Andrews, J. (2006).Monoclonal antibody HCV-AbXTL68 in patients undergoing liver transplantation for HCV: results of a phase 2 randomized study. Liver Transpl. **12,** 1381-1389.

24 Borgia, G.(2004).HepeX-C (XTL Biopharmaceuticals). Curr. Opin. Investig Drugs**. 5,** 892-897.

25 Bavituximab. Peregrine Pharmaceuticals Inc. Fall Winter 2008. Available at: http://www.peregrineinc.com/images/stories/media/siteFiles/20081121_BavituximabAV.pdf. Accessed February 7, 2009.

26 Soares, M. M., King, S. W. and Thorpe, P. E. (2008).Targeting inside-out phosphatidylserine as a therapeutic strategy for viral diseases. Nat. Med. **14,** 1357-1362.

27 Davis, G. L., Nelson, D. R., Terrault, N., Pruett, T. L., Schiano, T. D., Fletcher, C. V., Sapan, C. V., Riser, L. N., Li, Y., Whitley, R. J., Gnann, J. W.,Jr and Collaborative Antiviral Study Group (2005).A randomized, open-label study to evaluate the safety and pharmacokinetics of human hepatitis C immune globulin (Civacir) in liver transplant recipients. Liver Transpl. **11,** 941-949.

28 Law, M., Maruyama, T., Lewis, J., Giang, E., Tarr, A. W., Stamataki, Z., Gastaminza, P., Chisari, F. V., Jones, I. M., Fox, R. I., Ball, J. K., McKeating, J. A., Kneteman, N. M. and Burton, D. R. (2008).Broadly neutralizing antibodies protect against hepatitis C virus quasispecies challenge. Nat. Med. **14,** 25-27.

29 Helle, F., Wychowski, C., Vu-Dac, N., Gustafson, K. R., Voisset, C. and Dubuisson, J. (2006).Cyanovirin-N inhibits hepatitis C virus entry by binding to envelope protein glycans. J. Biol. Chem. **281**, 25177-25183.

30 Budkowska, A.(2009).Mechanism of cell infection with hepatitis C virus (HCV)--a new paradigm in virus-cell interaction. Pol. J. Microbiol. **58**, 93-98.

31 Bartosch, B., Vitelli, A., Granier, C., Goujon, C., Dubuisson, J., Pascale, S., Scarselli, E., Cortese, R., Nicosia, A. and Cosset, F. L. (2003).Cell entry of hepatitis C virus requires a set of co-receptors that include the CD81 tetraspanin and the SR-B1 scavenger receptor. J. Biol. Chem. **278**, 41624-41630.

32 Bartosch, B. and Cosset, F. L. (2006).Cell entry of hepatitis C virus. Virology. **348**, 1-12.

33 uerel, L., Voisset, C. and Dubuisson, J. (2006).Hepatitis C virus entry: potential receptors and their biological functions. J. Gen. Virol. **87,** 1075-1084.

34 Yagnik, A. T., Lahm, A., Meola, A., Roccasecca, R. M., Ercole, B. B., Nicosia, A. and Tramontano, A. (2000).A model for the hepatitis C virus envelope glycoprotein E2. Proteins. **40**, 355-366.

35 Charloteaux, B., Lins, L., Moereels, H. and Brasseur, R. (2002).Analysis of the C-terminal membrane anchor domains of hepatitis C virus glycoproteins E1 and E2: toward a topological model. J. Virol. 76, 1944-1958.

36 Yu, X., Qiao, M., Atanasov, I., Hu, Z., Kato, T., Liang, T. J. and Zhou, Z. H. (2007).Cryo-electron microscopy and three-dimensional reconstructions of hepatitis C virus particles. Virology. 367, 126-134.

37 Spiga, O., Padula, M. G., Scarselli, M., Ciutti, A., Bernini, A., Venditti, V., Prischi, F., Falciani, C., Lozzi, L., Bracci, L., Valensin, P. E., Caudai, C. and Niccolai, N. (2006).Structurally driven selection of human hepatitis C virus mimotopes. Antivir Ther. 11, 917-922.

38 Massariol, M. J., Zhao, S., Marquis, M., Thibeault, D. and White, P. W. (2010).Protease and helicase activities of hepatitis C virus genotype 4, 5, and 6 NS3-NS4A proteins. Biochem. Biophys. Res. Commun. 391, 692-697.

39 Reesink, H. W., Zeuzem, S., Weegink, C. J., Forestier, N., van Vliet, A., van de Wetering de Rooij,J., McNair, L., Purdy, S., Kauffman, R., Alam, J. and Jansen, P. L. (2006).Rapid decline of viral RNA in hepatitis C patients treated with VX-950: a phase Ib, placebo-controlled, randomized study. Gastroenterology**. 131,** 997-1002.

40 Sarrazin, C., Rouzier, R., Wagner, F., Forestier, N., Larrey, D., Gupta, S. K., Hussain, M., Shah, A., Cutler, D., Zhang, J. and Zeuzem, S. (2007).SCH 503034, a novel hepatitis C virus protease inhibitor, plus pegylated interferon alpha-2b for genotype 1 nonresponders. Gastroenterology**. 132,** 1270-1278.

41 Reesink, H. W., Fanning, G. C., Farha, K. A., Weegink, C., Van Vliet, A., Van 't Klooster, G., Lenz, O., Aharchi, F., Marien, K., Van Remoortere, P., de Kock, H., Broeckaert, F., Meyvisch, P., Van Beirendonck, E., Simmen, K. and Verloes, R. (2010).Rapid HCV-RNA decline with once daily TMC435: a phase I study in healthy volunteers and hepatitis C patients. Gastroenterology**. 138,** 913-921.

42 Seiwert, S. D., Andrews, S. W., Jiang, Y., Serebryany, V., Tan, H., Kossen, K., Rajagopalan, P. T., Misialek, S., Stevens, S. K., Stoycheva, A., Hong, J., Lim, S. R., Qin, X., Rieger, R., Condroski, K. R., Zhang, H., Do, M. G., Lemieux, C., Hingorani, G. P., Hartley, D. P., Josey, J. A., Pan, L., Beigelman, L. and Blatt, L. M. (2008).Preclinical characteristics of the hepatitis C virus NS3/4A protease inhibitor ITMN-191 (R7227). Antimicrob. Agents Chemother. **52**, 4432-4441.

43 McCauley, J. A., McIntyre, C. J., Rudd, M. T., Nguyen, K. T., Romano, J. J., Butcher, J. W., Gilbert, K. F., Bush, K. J., Holloway, M. K., Swestock, J., Wan, B. L., Carroll, S. S., DiMuzio, J. M., Graham, D. J., Ludmerer, S. W., Mao, S. S., Stahlhut, M. W., Fandozzi, C. M., Trainor, N., Olsen, D. B., Vacca, J. P. and Liverton, N. J. (2010).Discovery of vaniprevir (MK-7009), a macrocyclic hepatitis C virus NS3/4a protease inhibitor. J. Med. Chem. **53,** 2443-2463.

44 Wright D, Miller JL, Valentine J, et al. (2008). Safety, tolerability, and pharmacokinetic data following single- and multiple-dose administration of MK-7009, a hepatitis C virus non-structural 3/4a protease inhibitor, to healthy male subjects. Hepatology. **48,** A1910.

45 Manns MP, Bourlie`re M, Benhamou Y, et al. (2008). Safety and antiviral activity of BI201335, a new HCV NS3 protease inhibitor, in treatment-naive patients with chronic hepatitis C genotype-1 infection given as monotherapy and in combination with Peginterferon alfa 2a and Ribavirin. Hepatology. **48,** A1849.

46 Lavillette, D., Tarr, A. W., Voisset, C. *et al*. (2005). Characterization of host-range and cell entry properties of the major genotypes and subtypes of hepatitis C virus. Hepatology **41**, 265–274.

47 Wunschmann, S., Medh, J. D., Klinzmann, D., Schmidt, W. N. & Stapleton, J. T. (2000). Characterization of hepatitis C virus (HCV) and HCV E2 interactions with CD81 and the low-density lipoprotein receptor. J Virol 74, 10055–10062

48 Mas Marques, A., Mueller, T., Welke, J., Taube, S., Sarrazin, C., Wiese, M., Halangk, J., Witt, H., Ahlenstiel, G., Spengler, U., Goebel, U., Schott, E., Weich, V., Schlosser, B., Wasmuth, H. E.,

Lammert, F., Berg, T., Schreier, E. (2009). Low-density lipoprotein receptor variants are associated with spontaneous and treatment-induced recovery from hepatitis C virus infection. Infect Genet Evol. 9(5),847-52.

49 Owen, D. M., Huang, H., Ye, J., Gale, M. Jr. (2009). Apolipoprotein E on hepatitis C virion facilitates infection through interaction with low-density lipoprotein receptor. Virology. **394**(1),99-108.

50 Shufeng, L. (2009). Tight Junction Proteins Claudin-1 and Occludin Control Hepatitis C Virus Entry and Are Downregulated during Infection To Prevent Superinfection. JOURNAL OF VIROLOGY, p. 2011–2014.

51 Petracca, R., Falugi, F., Galli, G., Norais, N., Rosa, D., Campagnoli, S., Burgio, V., Di Stasio, E., Giardina, B., Houghton, M., Abrignani, S. and Grandi, G. (2000).Structure-function analysis of hepatitis C virus envelope-CD81 binding. J. Virol. 74, 4824-4830.

52 Higginbottom, A., Quinn, E. R., Kuo, C. C., Flint, M., Wilson, L. H., Bianchi, E., Nicosia, A., Monk, P. N., McKeating, J. A. and Levy, S. (2000).Identification of amino acid residues in CD81 critical for interaction with hepatitis C virus envelope glycoprotein E2. J. Virol. 74, 3642-3649.

53 Drummer, H. E., Wilson, K. A. and Poumbourios, P. (2002).Identification of the hepatitis C virus E2 glycoprotein binding site on the large extracellular loop of CD81. J. Virol. 76, 11143-11147.

54 Zhang, Y. Y., Zhang, B. H., Ishii, K. and Liang, T. J. (2010).Novel function of CD81 in controlling hepatitis C virus replication. J. Virol. 84, 3396-3407.

55 Meuleman, P., Hesselgesser, J., Paulson, M., Vanwolleghem, T., Desombere, I., Reiser, H. and Leroux-Roels, G. (2008).Anti-CD81 antibodies can prevent a hepatitis C virus infection in vivo. Hepatology. 48, 1761-1768.

56 Seeff LB. The history of the "natural history" of hepatitis C (1968-2009). *Liver Int.* 29 Suppl 1 89-99 (2009).

57 Le Guillou-Guillemette, H., Vallet, S., Gaudy-Graffin, C., Payan, C., Pivert, A., Goudeau, A. and Lunel-Fabiani, F. (2007).Genetic diversity of the hepatitis C virus: impact and issues in the antiviral therapy. World J. Gastroenterol. **13,** 2416-2426.

58 Timm, J. and Roggendorf, M. (2007).Sequence diversity of hepatitis C virus: implications for immune control and therapy. World J. Gastroenterol. **13,** 4808-4817.

59 Scott, J. D. and Gretch, D. R. (2007).Molecular diagnostics of hepatitis C virus infection: a systematic review. JAMA**. 297,** 724-732.

60 Poynard, T., Imbert-Bismut, F., Munteanu, M., Messous, D., Myers, R. P., Thabut, D., Ratziu, V., Mercadier, A., Benhamou, Y. and Hainque, B. (2004).Overview of the diagnostic value of biochemical markers of liver fibrosis (FibroTest, HCV FibroSure) and necrosis (ActiTest) in patients with chronic hepatitis C. Comp. Hepatol. **3,** 8.

61 Richter SS. Laboratory assays for diagnosis and management of hepatitis C virus infection. J.Clin.Microbiol. 40(12), 4407-4412 (2002).

62 Mederacke, I., Wedemeyer, H., Ciesek, S., Steinmann, E., Raupach, R., Wursthorn, K., Manns, M. P. and Tillmann, H. L. (2009).Performance and clinical utility of a novel fully automated quantitative HCV-core antigen assay. J. Clin. Virol. **46,** 210-215.

63 Medhi, S., Potukuchi, S. K., Polipalli, S. K., Swargiary, S. S., Deka, P., Chaudhary, A., Begum, N., Hussain, Z., Ahlawat, R. S. and Kar, P. (2008).Diagnostic utility of hepatitis C virus core antigen in hemodialysis patients. Clin. Biochem. **41,** 447-452.

64 Nelson, D. R.(2009).Hepatitis C drug development at a crossroads. Hepatology**. 50,** 997-999.

65 Shields WW, Pockros PJ. (2009). Novel drugs for hepatitis C virus. Curr Hepat Rep. **8**, 52-58.

66 Marcellin P. Hepatitis B and hepatitis C in 2009. *Liver Int.* 29 Suppl 1 1-8 (2009).

67 Eroglu C, Pinarbasi E. Hepatitis C Virus: Genome Organization, Viral Proteins and Implications in Disease Pathogenesis. Turk J Biol. 24, 253–269 (2000).

68 Chevaliez S, Pawlotsky JM. Hepatitis C viruses .Genome and molecular biology. In: HCV Genome and Life Cycle. Tan SL (Ed.), Horizon Scientific Press, Nowrich, UK, (2006).

69 Helle, F., Wychowski, C., Vu-Dac, N., Gustafson, K. R., Voisset, C. and Dubuisson, J. (2006).Cyanovirin-N inhibits hepatitis C virus entry by binding to envelope protein glycans. J. Biol. Chem. 281, 25177-25183.

70 Zhang, Y.(2008).I-TASSER server for protein 3D structure prediction. BMC Bioinformatics**. 9,** 40.

71 Pandit, S. B., Brylinski, M., Zhou, H., Gao, M., Arakaki, A. K. and Skolnick, J. (2010).PSiFR: an integrated resource for prediction of protein structure and function. Bioinformatics**. 26,** 687-688.

72 Kelley, L. A. and Sternberg, M. J. (2009).Protein structure prediction on the Web: a case study using the Phyre server. Nat. Protoc. **4,** 363-371.

73 Massariol, M. J., Zhao, S., Marquis, M., Thibeault, D. and White, P. W. (2010).Protease and helicase activities of hepatitis C virus genotype 4, 5, and 6 NS3-NS4A proteins. Biochem. Biophys. Res. Commun. 391, 692-697.

74 Kim, J. L., Morgenstern, K. A., Lin, C., Fox, T., Dwyer, M. D., Landro, J. A., Chambers, S. P., Markland, W., Lepre, C. A., O'Malley, E. T., Harbeson, S. L., Rice, C. M., Murcko, M. A., Caron, P. R. and Thomson, J. A. (1996).Crystal structure of the hepatitis C virus NS3 protease domain complexed with a synthetic NS4A cofactor peptide. Cell**. 87,** 343-355.

75 Zein, N. N.(2000).Clinical significance of hepatitis C virus genotypes. Clin. Microbiol. Rev. **13,** 223-235.

76 Levy, S., Todd, S. C., and Maecker, H. T. (1998). CD81 (TAPA-1): A molecule involved in signal transduction and cell adhesion in the immune system. Annu. Rev. Immunol. **16**, 89-109

77 Sanner, M. F.(1999).Python: a programming language for software integration and development. J. Mol. Graph. Model. 17, 57-61.

78 Sanner, M. F., Olson, A. J. and Spehner, J. C. (1996).Reduced surface: an efficient way to compute molecular surfaces. Biopolymers. 38, 305-320.

79 Bajaj, C, Park, S., Thane, A., (2002), A Parallel Multi-PC Volume Rendering System, ICES and CS Technical Report, University of Texas, 2002.

80 Bajaj, C, Pascucci, V., Schikore, D., (1996), Fast IsoContouring for Improved Interactivity, Proceedings of ACM Siggraph/IEEE Symposium on Volume Visualization, ACM Press, 39 - 46, San Francisco, CA

81 Michel F. Sanner, Daniel Stoffler and Arthur J. Olson. ViPEr a Visual Programming Environment for Python. 10th International Python Conference, February 2002.

82 Brazzoli, M., Bianchi, A., Filippini, S., Weiner, A., Zhu, Q., Pizza, M. and Crotta, S. (2008).CD81 is a central regulator of cellular events required for hepatitis C virus infection of human hepatocytes. J. Virol. **82,** 8316-8329.

83 Favre, D. and Muellhaupt, B. (2005).Potential cellular receptors involved in hepatitis C virus entry into cells. Lipids Health. Dis. **4,** 9.

84 Regeard, M., Trotard, M., Lepere, C., Gripon, P. and Le Seyec, J. (2008).Entry of pseudotyped hepatitis C virus into primary human hepatocytes depends on the scavenger class B type I receptor. J. Viral Hepat. **15,** 865-870.

85 Moradpour, D., Penin, F. and Rice, C. M. (2007).Replication of hepatitis C virus. Nat. Rev. Microbiol. **5,** 453-463.

86 Tscherne, D. M., Jones, C. T., Evans, M. J., Lindenbach, B. D., McKeating, J. A. and Rice, C. M. (2006).Time- and temperature-dependent activation of hepatitis C virus for low-pH-triggered entry. J. Virol. **80,** 1734-1741.

87 Albrecht, H., Cosman, M., Ngu-Schwemlein, M., Corzett, M., Curran, K. W., Dolan, C., Fang, X., DeNardo, S. J., DeNardo, G. L. and Balhorn, R. (2007).Recombinant expression of the beta-subunit of HLA-DR10 for the selection of novel lymphoma targeting molecules. Cancer Biother. Radiopharm. **22,** 531-542.

88 Working, P. K., Newman, S. M., Johnson, J., Cornacoff, J. B. (1997).Safety of Poly(ethylene glycol) and Poly(ethylene glycol) Derivatives. **. 680,** 45-57.

89 Francis, M. F., Piredda, M. and Winnik, F. M. (2003).Solubilization of poorly water soluble drugs in micelles of hydrophobically modified hydroxypropylcellulose copolymers. J. Control. Release**. 93,** 59-68.

90 Zemla, A., Zhou, C. E., Slezak, T., Kuczmarski, T., Rama, D., Torres, C., Sawicka, D. and Barsky, D. (2005).AS2TS system for protein structure modeling and analysis. Nucleic Acids Res. **33,** W111-5.

91 Kaufmann, K. W., Lemmon, G. H., Deluca, S. L., Sheehan, J. H. and Meiler, J. (2010).Practically useful: what the Rosetta protein modeling suite can do for you. Biochemistry**. 49,** 2987-2998.

92 Zemla, A.(2003).LGA: A method for finding 3D similarities in protein structures. Nucleic Acids Res. 31, 3370-3374.

93 Zemla, A., Geisbrecht, B., Smith, J., Lam, M., Kirkpatrick, B., Wagner, M., Slezak, T. and Zhou, C. E. (2007).STRALCP--structure alignment-based clustering of proteins. Nucleic Acids Res. 35, e150.

94 Neugebauer, A., Klein, C. D. and Hartmann, R. W. (2004).Protein-dynamics of the putative HCV receptor CD81 large extracellular loop. Bioorg. Med. Chem. Lett. 14, 1765-1769.

95 Moult, J., Fidelis, K., Kryshtafovych, A., Rost, B. and Tramontano, A. (2009).Critical assessment of methods of protein structure prediction - Round VIII. Proteins. 77 Suppl 9, 1-4.

96 Morris, G. M., Huey, R. and Olson, A. J. (2008).Using AutoDock for ligand-receptor docking. Curr. Protoc. Bioinformatics. Chapter 8, Unit 8.14.

97 Morris, G. M. , Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E. , Belew, R. K., Olson, A. J. (1998). Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function, J. Comput. Chem. **19**, 1639–1662.

98 Harris, R., Olson, A. J. and Goodsell, D. S. (2008).Automated prediction of ligand-binding sites in proteins. Proteins. **70**, 1506-1517.

99 Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res., 25, 3389–3402.

100 Smith,T.F. and Waterman,M.S. (1981) Identification of common molecular subsequences. J. Mol. Biol., **147**, 195–197.

101 Pearson,W.R. (1991) Searching protein sequence libraries: comparison of the sensitivity and selectivity of the Smith–Waterman and FASTA algorithms. Genomics**, 11**, 635–650.

102 Larkin, M. A., Blackshields, G., Brown, N. P. *et al.* (2007) ClustalW and ClustalX version 2. Bioinformatics. **23**, 2947-2948.

103 Kitadokoro, K., Galli, G., Petracca, R., Falugi, F., Grandi, G. and Bolognesi, M. (2001).Crystallization and preliminary crystallographic studies on the large extracellular domain of human CD81, a tetraspanin receptor for hepatitis C virus. Acta Crystallogr. D Biol. Crystallogr. 57, 156-158.

104 Kitadokoro, K., Bordo, D., Galli, G., Petracca, R., Falugi, F., Abrignani, S., Grandi, G. and Bolognesi, M. (2001).CD81 extracellular domain 3D structure: insight into the tetraspanin superfamily structural motifs. EMBO J. 20, 12-18.

105 Balhorn, R. and Balhorn Cosman, M., (2010) Selective High Affinity Ligands (SHALs): A New Class of Targeting Agents for Cancer Imaging and Therapy. In: Medical Radiology,

Radiation Oncology, Volume Therapeutic Nuclear Medicine, R.P. Baum, ed., Springer-Verlag, Germany, in press.

106 Sousa, S. F., Fernandes, P. A. and Ramos, M. J. (2006).Protein-ligand docking: current status and future challenges. Proteins. 65, 15-26.

107 Warren, G. L., Andrews, C. W., Capelli, A. M., Clarke, B., LaLonde, J., Lambert, M. H., Lindvall, M., Nevins, N., Semus, S. F., Senger, S., Tedesco, G., Wall, I. D., Woolven, J. M., Peishoff, C. E. and Head, M. S. (2006).A critical assessment of docking programs and scoring functions. J. Med. Chem. 49, 5912-5931.

108 Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W. and Taylor, R. D. (2003).Improved protein-ligand docking using GOLD. Proteins. 52, 609-623.

109 Soriano-Ursua, M. A., Trujillo-Ferrara, J. G. and Correa-Basurto, J. (2009).Homology modeling and flex-ligand docking studies on the guinea pig beta(2) adrenoceptor: structural and experimental similarities/ differences with the human beta(2). J. Mol. Model. 15, 1203-1211.

110 Ewing, T. J., Makino, S., Skillman, A. G. and Kuntz, I. D. (2001).DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. J. Comput. Aided Mol. Des. 15, 411-428.

111 Zekri, A. R., Bahnassy, A. A., Alam El-Din, H. M., Salama, H. M. (2009). Consensus siRNA for inhibition of HCV genotype-4 replication. Virology j. **6**: 13.

112 De Francesco, Raffaele et al (1998). The hepatitis C virus NS3 proteinase: structure and function of a zinc-containing serine proteinase. Therapies for Viral Hepatitis, **238**.

113 Love, Robert A. (1996). The crystal structure of hepatitis C virus NS3 proteinase reveals a trypsin-like fold and a structural zinc binding site. Cell87, 331-342.

114 Tan, S. L., Pause, A., Shi, Y. and Sonenberg, N. (2002).Hepatitis C therapeutics: current status and emerging strategies. Nat. Rev. Drug Discov. **1,** 867-881.

115 Fausto, N.(2001).A mouse model for hepatitis C virus infection? Nat. Med. **7,** 890-891.

116 Kamiya, N., Iwao, E., Hiraga, N., Tsuge, M., Imamura, M., Takahashi, S., Miyoshi, S., Tateno, C., Yoshizato, K. and Chayama, K. (2010).Practical evaluation of a mouse with chimeric human liver model for hepatitis C virus infection using an NS3-4A protease inhibitor. J. Gen. Virol. **91,** 1668-1677.

1001. www.who.int

1002. http://www.epidemic.org/

1003. AutoDock website: http://autodock.scripps.edu/

1004. NBCR website: https://www.nbcr.net/pub/wiki/index.php?title=CADD Pipeline