Welsh School of Pharmacy – Cardiff University
Ysgol Fferylliaeth – Prifysgol Caerdydd

A thesis submitted to the University of Cardiff, Wales for the Degree of

PHILOSOPHIAE DOCTOR

by

DIMITRIOS P.VLACHAKIS

"Computer-Aided Drug Design and Biological Evaluation of Novel Anti-Viral Agents"

2006

UMI Number: U584138

Dissertation Publishing

take the DIMHOLT road...

I ABSTRACT

In this thesis is presented a description of studies concerning the molecular modelling and biological evaluation of a set of novel antiviral agents for the helicase and polymerase proteins of *Flaviviridae*.

Viruses in this family are enveloped, have positive-sense RNA and are responsible for a variety of life threatening diseases. To date neither specific antiviral treatments exist nor are there any vaccines available for *Flaviviridae* infection. Thus there is an urgent need for new therapies.

The ultimate aim of this project was to design a coordinated *in silico* & *in vitro* protocol for the design and evaluation of novel *Flaviviridae* inhibitors.

That was achieved initially by establishing the three-dimensional structures of various *Flaviviridae* members by homology-based molecular modelling. In continuation, a set of small compound libraries was designed using a *de novo* structure-based drug design approach. Those compounds were screened *in silico* with the aid of molecular docking and a set of scoring algorithms. The best candidates were chosen to be chemically synthesised (not part of this thesis).

The genes of Hepatitis C and Dengue helicases as well as the Dengue NS3 domain (helicase and protease) were cloned in expression vectors and the proteins were produced and purified.

A novel biological assay was then established for the Hepatitis C helicase in order to evaluate the potency of the designed inhibitors *in vitro*. An attempt was finally made to feedback the computer model using the biological activity data of those compounds, in order to improve the cooperation levels between the *in silico* and the *in vitro* parts of this research.

II ACKNOWLEDGMENTS

I would like to thank my supervisor, Dr. Andrea Brancale for his constant guidance, support, patience, suggestions and comments on every aspect of my PhD, but also for being a great sport and a friend.

Moreover, I would like to thank Prof. Chris McGuigan for allowing me to participate in most of the activities of his group, helping me to develop a variety of skills.

I would like to thank Dr. Colin Berry for his support and supervision for the molecular biology part of my PhD.

Thanks also to Dr Claire Simons, for her support and extremely useful suggestions when it came to write my thesis.

Special thanks to Gareth, Tim, Dave, Jerome, Olivier, Mary-Rose, Federica, Felice, Antonella, Giovanna, Annette, Costantino, Plinio, Marco, Maria Chiara, Alessandro, Rita, Rina and everybody else from the WSP for making my PhD life in Cardiff fun!

Thanks to each and every one of the Hellenic Society people, who helped me organise some of the coolest parties to take place in Cardiff ever.

Last but not least, special thanks to my parents and to my sister for their love, support and care throughout my PhD.

III. ABBREVIATIONS

ALT	Alanine Aminotransferase
AS	Active Site
ATP	Adenosine triphosphate
GTP	Guanosine triphosphate
UTP	Uracile triphosphate
BDV	Border Diarrheoa Virus
BSA	Bovine Serum Albumin
BVDV	Bovine Viral Diarrheoa Virus
CoMFA	Comparative Multifragment Analysis
CSFV	Classical Swife Fever Virus
DD	Drug Design
dnDD	*De novo* Drug Design
DEN	Dengue
DENV	Dengue Virus
DHF	Dengue Haemoragic Fever
DNA	Deoxyribonucleic acid
ds	double stranded
DTT	Dithiothreitol
e.g.	*exempli gratia*
EC50	Effective Concentration
EDTA	Ethylenediaminetetraacetic acid
eq	equivalent
EU	European Union
GBV-A	Hepatitis GB Virus strain A
GBV-B	Hepatitis GB Virus strain B
GBV-C	Hepatitis GB Virus strain C
HAV	Hepatitis A Virus
HBV	Hepatitis B Virus
HCV	Hepatitis C Virus
HepC	Hepatitis C
HGV	Hepatitis G Virus
HTS	High Throughput Screening

Hz	hertz
IC$_{50}$	Inhibition Concentration
IRES	Internal Ribosomal Entry Site
JEV	Japanese Encephalitis
kDa	kilodalton
LPC	Ligand Protein Contacts
Me	Methyl
MOE	Molecular Operating Environment
NADP	beta-nicotinamide adenine dinucleotide phosphate
NADPH	beta-nicotinamide adenine dinucleotide phosphate, reduced form
NCI	National Cancer Institute
NMR	Nuclear Magnetic Resonance
NSx	Non Structural (x = number)
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PPS	Polyphosphoric Acid
QSAR	Quantitative Structure Activity Relationship
RDP	Ribavirin-5-diphosphate
RMSd	Root Mean Square deviation
RNA	Ribonucleic acid
RP	Ramachandran Plot
RT	Room Temperature
RTP	Ribavirin-5-triphosphate
SAH	S-Adenosylhomocysteine Hydrolase
SAR	Structure Activity Relationship
SCR	Structurally Conserved Regions
SDS	Sodium dodecyl sulfate
ss	single stranded
SVR	Structurally Variable Regions
TLC	Thin Layer Chromatography
USA	United States of America
UTR	Untranslated Region
UV	Ultra Violet
vs	*versus*

WHO	World Health Organisation
WNV	West Nile Virus
YFV	Yellow Fever Virus

INDEX   PAGE NUMBER

chapter 1
INTRODUCTION

1.1   *Flaviviridae.*

Flaviviridae is a family of viruses that infect vertebrates [1]. Virions of the *flaviviridae* family are enveloped and slightly pleomorphic during their life cycle. They are spherical in shape and usually 40-60 nm in diameter. Their nucleocapsids are isometric and sometimes penetrated by stain. The usual size of the nucleocapsids is 25-30 nm in diameter and they have polyhedral symmetry [2].

Flaviviridae consists of three characterised genera and the unclassified ones [3]. The main representatives of each family are summarised below (Figure 1.1).

Flavivirus

Yellow fever virus [YFV]
Dengue fever virus [DENV]
Japanese encephalitis virus [JEV]
Tick-borne encephalitis virus [TBEV]

Pestivirus

Bovine viral diarrhoea virus [BVDV]
Classical swine fever virus [CSFV]
Border disease virus [BDV]

Flaviviridae

Hepacivirus

hepatitis C virus [HCV]

Unclassified

Hepatitis G virus [HGV]
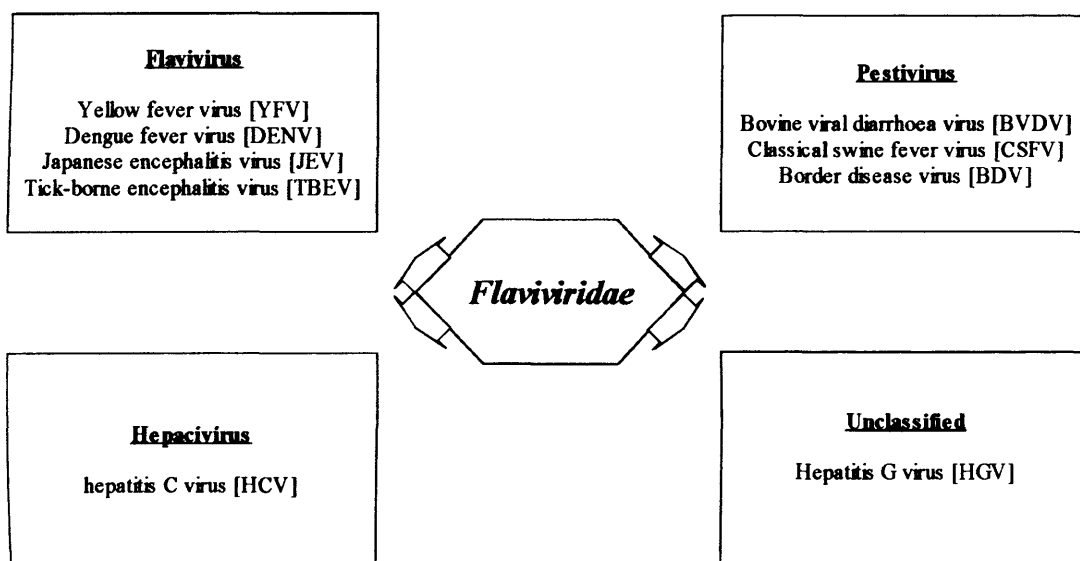
Figure 1.1. The Three genera that constitute the viral family of *Flaviviridae*. Hepatitis C virus was recently discriminated from the rest of the *Flaviviridae*, due to its distinct properties and clinical manifestations. More details about the most representative members of each family can be found in Appendices 1.1 to 1.4.

It is remarkable that even though the above viruses are separated to different genera, and they do not have common biological properties and do not show serological cross-reactivity, they manage to retain high similarity in the morphology of the virion, the organisation of the viral genome, and the estimated life cycles and replication patterns that they follow [4].

1.2   Epidemics of *Flaviviridae*

More than 170 million people worldwide are currently chronically infected with the Hepatitis C virus [5]. They are all considered to be at risk of developing cirrhosis and some of them will develop liver cancer. Hepatitis C has spread all over the world and for every person who has the aids virus, 4 have the Hepatitis C Virus. Today, hepatitis C causes ten thousand deaths per year and is the main cause for more than half of the four thousand liver transplantations that are performed annually [6].

Infections that mosquitoes carry or more generally arthropod-borne *flaviviridae* have reached epidemic dimensions in some parts of the world. Dengue fever infects 50 million people per year in central Africa. According to the World Health Organisation (WHO) there are 6.5 billion inhabitants on this planet that live in areas of high risk of acquiring dengue. For example only for 2006 the Philippines reported 197 deaths and 14,738 cases of dengue fever [7]. Indonesia's

Dengue deaths reached 634 and Malaysia has already confirmed 74 deaths due to Dengue virus in the first 9 months. In Thailand more than 32,000 Thais have been infected with dengue fever and currently Singapore is going through its worst dengue fever outbreak ever on record, since the officially reported Dengue cases are nearing 11,000 [8].

According to the WHO of South-East Asia since November 10th, 2005, a total of 5737 cases of Japanese Encephalitis with 1334 deaths (fatality rate of 23.3%) have been reported from Uttar in India since the outbreak started in July 2005. Moreover, since the 2nd of January 2006, a total of 2824 individuals have been infected with Japanese Encephalitis, 316 of these infections have already resulted in deaths (fatality rate of 11.2%). The Government in despair has employed both anti-larval and anti-adult measures by distributing 200,000 mosquito nets. The government after evaluating the situation decided to establish a law, by which all children between the ages of 1-12 years must be vaccinated and immunised against Japanese Encephalitis. The law will take action in January 2007, in schools and kindergartens [9].

West Nile Virus (WNV) first hit New York with 77 deaths in 1999. Moreover, the United States were alarmed and action was taken to stop the virus from spreading. But this year West Nile virus is again on the front covers of the newspapers. This year WNV has hit Illinois harder than any other state, with 399 cases so far, 21 of which resulted in death. The humid and full of swamps Louisiana comes second with 11 deaths in 2006 [10].

Farming and agriculture have both seriously suffered in the past from the impact

that pestiviruses had on livestock. Many economies depend on primary production

that comes from farming and agriculture and as a result most of these countries

take preventative action against Pestiviruses [11-12].

Disproportionaly to the severity of an infection with almost all members of

Flaviviridae, no specific antiviral therapy is available today. [13-14-15-16]

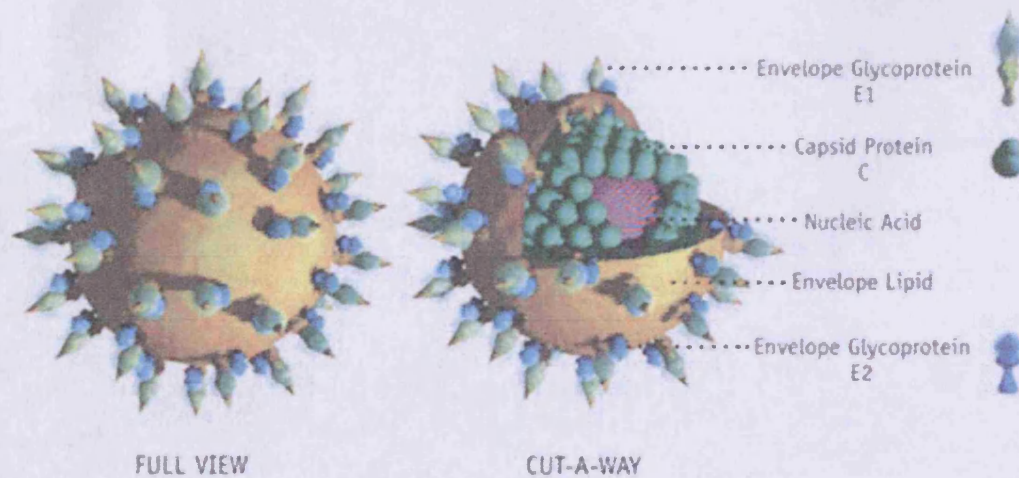### 1.3.1 Hepaciviruses

Figure 1.2. A model of the Human Hepatitis C Virus [17].

Before 1989, when the HCV agent was identified (Figure 1.2), all HCV infections

were referred to as non-A, non-B hepatitis [18]. According to the World's Health

Organization more than 3% of the world's population is currently infected with

HCV [19]. Convert that in numbers and almost 170 million people are chronic

carriers of the Hepatitis C virus (Figure 1.3) [20]. Without exceptions, they all belong to the high risk group for developing initially liver cirrhosis and potentially at a later stage, liver cancer [21]. Statistics reveal that a great portion of the HCV carriers are using or have used in the past drugs intravenously. Another large portion of HCV carriers are those who needed blood transfusion at some stage in their life [22]. The latter is not a problem anymore in developed countries after the law for blood donor screening for HCV was introduced. The HCV infection of post-transfusion patients is under control and has eclipsed in the last few years. Other routes of transmission are needles, body piercing, tattooing, scarification and circumcision in some countries [23]. Still there is always a number of cases where the cause of the infection cannot be identified.

Upon infection with HCV the acute phase will follow. Twenty days later the copies of the Hepatitis C virus have already reached high levels that allow biochemists to detect the viral RNA. It is then that the first symptoms of the disease will become apparent [24]., although in some cases, the infection could be asymptomatic. Nine in ten acutely infected patients will develop chronic HCV infection [25] and eventually liver cirrhosis in 40% of the patients. [26].

If the transmission rate of the HCV infection remains as it is today, it is estimated that the number of patients with chronic hepatitis C will rise dramatically in the next ten years. Today, for example, the number of HCV victims only in the USA is estimated to be 8,800 annually. Statistics reveal that this number will rise to 35,000 in 3 year's time [27].

Hepatitis G virus (HGV) infections have not been so extensively studied as HCV infection. It has been found though that the GB viruses (GBV-A, GBV-B, and GBV-C) are related phylogenetically to the Hepatitis C virus. Humans are the natural host for GBV-C, whereas tamarins are the natural hosts for GBV-A and GBV-B. GBV-C is transmitted in a pattern similar to HCV and is quite often found in co-infection with HCV [28].
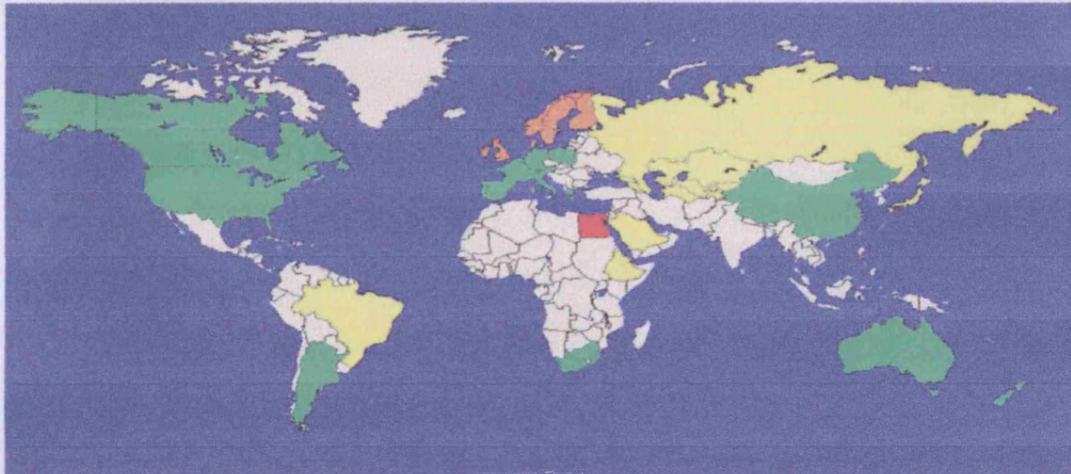
Figure 1.3. The world prevalence of the HCV infection per country

<u>Anti HCV Prevalence</u>

Red	>5%	High
Yellow	1.1 – 5%	Intermediate
Green	0.2 – 1%	Low
Orange	<0.1%	Very Low
White		Unknown

(Adapted from WHO, 1995 [29])

1.3.2 Flaviviruses

Up to date the genera of flaviviruses counts more than seventy members in its ranks. Many of them are very dangerous human pathogens [30]. Human flaviviruses are transmitted primarily with ticks and mosquitoes. This is making it very hard to deal with the disease, especially at some parts of the world, where eliminating mosquitoes is not possible [31].

According to phylogenetic data, there are 72 species of flaviviruses, which have been further sub-divided into 14 groups. Those groups have been characterised and grouped into 3 main categories [32]. First is the mosquito-borne category,

1950   **900**

1960   ⬇   **15497**

1970   ⬇   **122174**

1980   ⬇   **295591**

1990   ⬇   **514139**

The Dengue Virus infects approximately 50.000.000 people per year.

Figure 1.4. The spread of the Dengue Virus from the 50's until today. Even highly-developed areas of the planet are under threat (USA-EU).
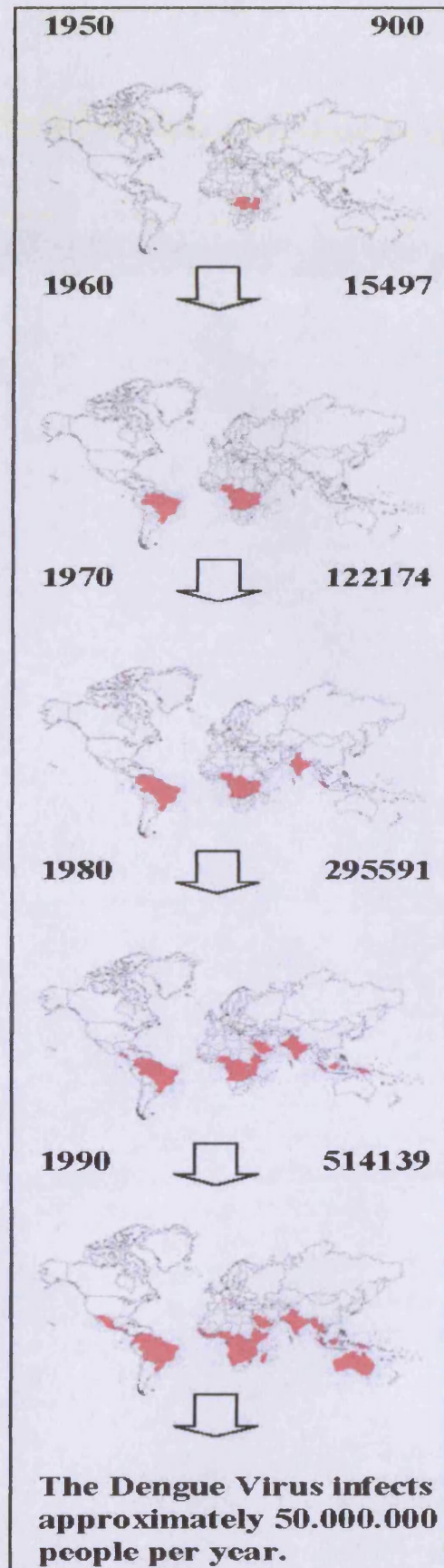
second is the tick-borne, and the final category is the no-vector one. Flaviviruses

that affect humans are grouped in the first two categories [33]. A very dangerous

and representative member of the Flaviviruses genera is the Yellow Fever Virus

(YFV). Even though there is available vaccination against YFV, the fatality rates

are very high [34].. YFV causes hemorrhagic fever upon infection and has a

mortality rate that reaches fifty percent worldwide [34].

Dengue and dengue hemorrhagic fever (DHF) are caused by infection with one of

four antigenically distinct, virus serotypes (DEN-1, DEN-2, DEN-3, and DEN4) [35].

Once infected with one of these serotypes, the individual develops specific

immunity. However, cross-immunity does not develop. It is theoretically possible,

therefore, for an individual to be infected four times, each time with a different

serotype [35]. Dengue is mostly seen in tropical urban areas. As with other members

of the *Flaviviridae* family, the virus is transmitted through mosquito bites,

specifically Aedes aegypti, a domestic, day-biting mosquito that prefers to feed on

humans [35]. Dengue is the most important mosquito-borne viral disease, affecting

humans with a distribution comparable with that of malaria. Approximately 2.5

billion people are living in areas at risk for epidemic transmission [36]. Tens of

millions of cases of dengue fever occur annually along with up to hundreds of

thousands of cases of DHF (Figure 1.4). DHF is the most serious manifestation of

the disease, caused by an immunologic reaction that occurs for the most part in

individuals already sensitized to the disease, either actively through infection or

passively in infants through placental transfer of immunoglobulin from mother to

child [37]. Initially, DHF appears the same as dengue but after several days the patient deteriorates with prostration, restlessness, signs of circulatory collapse (diaphoresis, cold extremities, dyspnea, circumoral and peripheral cyanosis, and hemorrhagic manifestations) [38]. Available laboratory tests cannot identify who will ultimately develop this manifestation [39].

Back in August 1999 an encephalitis outbreak in New York claimed the lives of 6 humans, after 77 officially reported cases of infection [40]. Initially the outbreak was characterised serologically as St. Louis encephalitis virus. Further studies and sequence analysis revealed that the virus had higher similarity with the West Nile virus. West Nile virus is a mosquito-borne virus found most commonly in Africa, France, India, Indonesia, the Middle East, and Soviet countries [41]. In 1999, a West-Nile-like virus was identified in patients living in the Northeast United States. The bird is the primary host and the principal vector is Culex univittatus [42]. However, other mosquitoes are known to carry the virus, including Culex pipiens, Culex antennatus, and Culex tritaeniorhynchus (Asia). Other animal reservoirs are not part of the virus's normal life cycle [43]. West Nile fever is common in the Middle East with most individuals exposed as children. Children experience a nondescript viral illness with fever that is rarely diagnosed. Neighbouring Israel also experiences infection although there, it is more likely to be the young adult than the child who becomes infected. Spread occurs primarily in the summer months when the mosquito population increases [44].

Finally, another very important member of the Flavivirus genus is Japanese

Encephalitis Virus (JEV) [45]. JEV is the major cause of viral encephalitis worldwide.

Almost fifty thousand infections happen in Asia every year. JEV has a very high

mortality rate that reaches thirty percent. Another thirty percent of the infected

patients will develop long-lasting neurological conditions [46-47].

1.3.3 Pestiviruses

Pestiviruses are mainly viruses that affect animals. The most representative members of the Pestivirus genera are the Classical Swine Fever Virus (CSFV), Bovine Viral Diarrhoea Virus (BVDV) and Border Diarrhoea Virus (BDV) [48]. There are two different routes of infection known today [49], the first is the nasal route and the second is the transplacental route. The transplacental route is very dangerous to livestock, since it retains the infection into the livestock and poses a constant thread to the rest of the animals. These viruses cause severe disease that will lead to death, but before that happens, it is very easy to cross species, infect and cause a milder version of the same disease to the other hosts [50-51-52]. Bovine Diarrhoea Virus causes a severe mucosal disease in cattle. Other species, such as swine and various ruminants have been found to be susceptible to the virus [53-54]. There are cytopathic and non-cytopathic biotypes according to their proliferation patterns in cell culture. Infection with BVDV has high morbidity and low mortality rates [55]. This mild disease is marked by ulceration of the nose, mouth, and gastrointestinal mucosa, which causes the virus to spread quickly because of continuous salivation, nasal discharge, coughing, or diarrhoea [56]. The major organs that the viruses attack and use for their replication are the lymphoid tissues, epithelial and all major lymphocyte cells. It has also been found that cells of the gastrointestinal tract, glands, and neurons can provide the host that viruses

need to replicate [57-58]. Changing from the primary infecting non-cytopathic strain

(or biotype) into the cytopathic strain is closely related to the mucosal disease.

There are two scenarios to describe this conversion. The first one is that an

insertion appears in a cleavage site in the NS2-3 protein. The second scenario is

that there is duplication of the NS2-3 gene. As a result the NS3 protein will be

expressed instead of the NS2-3 gene [59-60-61-62-63].

Humans are affected from Pestiviruses in an indirect way. Many countries base

their economies on agriculture and farming. A Pestivirus infection can cause great

damage to economies and as a result, especially in countries with weak economies,

suffering to people [64]. The economic losses upon BVDV infection depends on the

size of the epidemic [65].

CSFV or hog cholera virus is another important and extremely contagious

pathogen of swine [66]. CSFV is transmitted by aerosols, clothes and direct contact

[67-68]. This virus too, has shown in the past that it can lead to great economic losses

[69]. Sheep and goat BDV will develop in a mild clinical disease that will then be

followed by an acute postnatal infection [70]. Following the behaviours of BVDV

and CSFV, BDV too, can be sub-divided into non-cytopathic and cytopathic

biotypes. Here the cytopathic version is due to the production of the NS3 non-

structural domain instead of the NS2-3 protein [71].

Upon pestivirus infection, the infected animal is isolated and slaughtered, which

will eventually prevent further virus transmission. Killing the animal seems to be a

much easier way to eradicate the infection rather than trying to cure it.

Vaccination for BVDV is available and may lead to a much more controlled situation, where no animals have to be killed [72-73-74].

1.4   *Flaviviridae* Genome.

Genus Flavivirus genomes consist of a single piece of linear, single-stranded, positive sense RNA [75], because the viral RNA has positive sense, the nucleic acid itself is capable of causing an infection in the appropriate host cells. The total genome can range from 10 to 12 kilobase (kb) pairs [76]. The 3' terminus of the viral genome is not polyadenylated and the 5' end has a methylated nucleotide cap, which allows the translation process to occur. Sometimes it is possible to have a genome-linked protein (VPg) in place of the methylated nucleotide cap.

The genome of the Pestivirus family, like the Flavivirus gene, is reported to be approximately 12.5 kb in length [77]. The Pestivirus family has no poly-A tail on the 3' end of the RNA and also lack a 5' methylated nucleotide cap. In both genera, structural genes are located towards the 5' end of the RNA (Figure 1.5).

The Pestivirus and the Hepacivirus genus have internal ribosomal entry sites (IRES), which are responsible for providing a site for the initiation of the translation process for host ribosomes [78]. On the other hand the Flavivirus genus does not have IRES, but is capable of scanning the ribosomes to begin protein synthesis.

(A)

(B)

Important flaviviral proteins and their functions

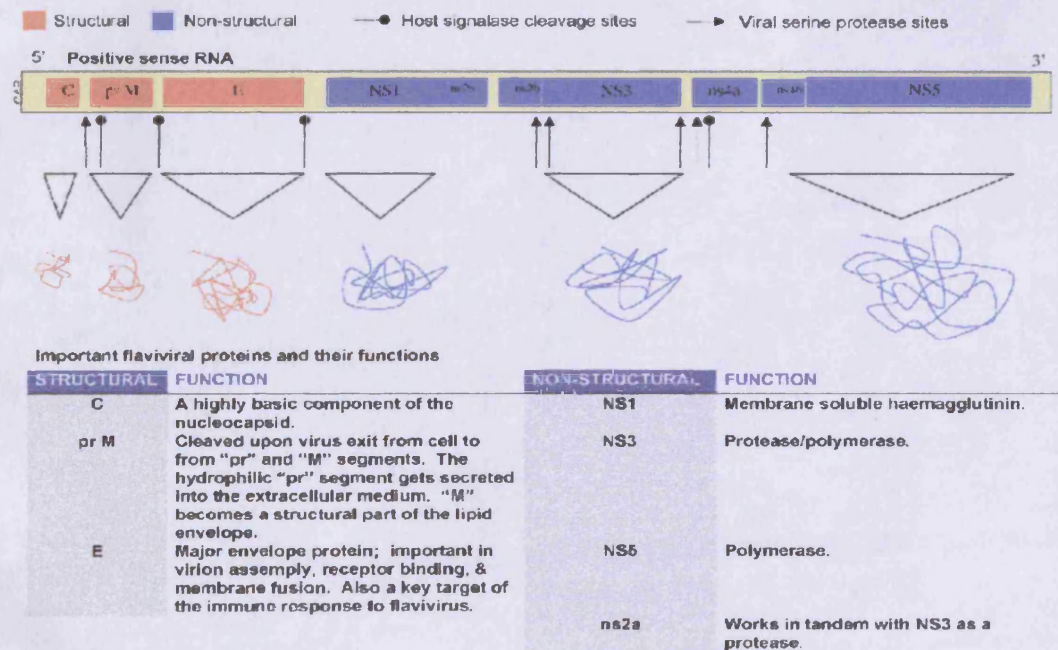| STRUCTURAL	FUNCTION	NON-STRUCTURAL	FUNCTION
C	A highly basic component of the nucleocapsid.	NS1	Membrane soluble haemagglutinin.
pr M	Cleaved upon virus exit from cell to from "pr" and "M" segments. The hydrophilic "pr" segment gets secreted into the extracellular medium. "M" becomes a structural part of the lipid envelope.	NS3	Protease/polymerase.
E	Major envelope protein; important in virion assembly, receptor binding, & membrane fusion. Also a key target of the immune response to flavivirus.	NS5	Polymerase.
		ns2a	Works in tandem with NS3 as a protease.

Figure 1.5. Top (A): The organisation of the genetic material of the three genera that constitute *Flaviviridae*. This family of viruses carry a copy of a linear single-stranded, positive sense RNA. Bottom (B): The generic break down of the *Flaviviridae* genome and the function of each of the proteins that it encodes [79].

1.5   Viral Replication

In order to identify targets for antivirals the life cycle of the *Flaviviridae* has to be thoroughly examined and understood. The replication process takes place in the cytoplasm and lasts approximately 20-30 hours for flaviviruses [80]. First the virus identifies the host cell by a set of special and unique for each virus membrane proteins. Only then is the virus capable of initiating its life cycle and reproducing. Very little is known about how the natural processes of hepatitis C virus infection develops, but with the limited data available the life cycle of *Hepatitis C* can be summarised in 5 different steps (also summarised in figure 1.6):

STEP 1: First the virus has to find and attach itself to a liver cell. Then the Hepatitis C virus will utilise special proteins present on its protective lipid coat in order to help it attach to a receptor site on the host liver cell [81]. For example HCV is looking for the CD81 (a tetraspanin, present on liver cells and lymphocytes) on cell membranes. Dengue virus is looking for heparan sulfate [82-83]. This will enable the virus to concentrate on the outer cell membrane.

STEP 2: The viral protein core will penetrate the host cell's plasma membrane and will enter into the cytoplasm of the liver cell [81]. In order to penetrate the host cell membrane HCV will make use of its protective lipid (fatty) coat [81]. HCV will attempt to merge its lipid coat with the host's outer membrane. As soon as the lipid coat has successfully fused to the plasma membrane, the membrane of the

host will engulf the virus, a process identical to endocytosis, also known as receptor-mediated endocytosis. The difference in pH in the endosome of the cell will initiate the fusion between the envelope of the virus and the membrane of the endosome. The nucleocapsid will be left in the cytosol, where the viral genetic material, the positive single-strand of RNA will be uncoated [81]. When the virus is inside the host its protein coat dissolves. As a result the viral RNA is now released into the cell. It is not clear yet if the virus first enters the host and then dissociation of its protein coat occurs or if during penetration of the cell membrane the protein coat is broken open (after fusion with the liver host cell) and then the contents of the virus are released into the cytoplasm. There is the possibility that there may be enzymes on the liver cells cell membrane that the HCV may utilise to dissolve its protein coat.

STEP 3: Now that the viral RNA is in the cytoplasm it will find the cell's ribosomes and it will begin the process of the production of materials necessary for viral reproduction [81]. Hepatitis C has a positive strand of RNA, so the nucleic acid can be directly read by the host cell's ribosomes – as if it was normal mRNA. It is the 5' untranslated region (5'UTR) of the viral genome that will find its way to the ribosomes, where translation will take place. The viral genome is firstly translated as a single polyprotein, which is later cleaved into individual-mature proteins by viral and host enzymes (signalases [85]). Flaviviruses have a short 5' UTP with a type I m7GpppN1mpN2 cap structure [86]. On the other hand Pestiviruses [87] and Hepaciviruses (HCV [88] and HGV [89]) carry with them an IRES, which will take

the ribosome to the first coding triplet in order to express the viral polyprotein.

During this process two things take place simultaneously. First the virus begins to

produce the materials coded in its RNA, and second it also influences most of the

normal functions of the cell, in an attempt to disorganise the host cell cycle.

Another interesting feature of HCV is that it will also try to push the host cell to

reproduce, in an attempt to create a new host for viral reproduction [81]. This may

be why HCV infection is directly associated with hepatocellular carcinoma. The

first product of the viral RNA is the RNA transcriptase that it will use for

reproduction [81].

STEP 4: As soon as the RNA transcriptase is made, it will make the negative strand

of the viral RNA, to be used later on as a template for the generation of new viral

RNA [81]. The complementary negative strand of the RNA is synthesised by non

structural proteins, the RNA-dependent RNA polymerase, with the aid of various

cofactors [90]. This strand is then used, again by the RNA-dependent RNA

polymerase, as a template for the synthesis of the genomic HCV RNA. Viral RNA

will be copied thousands of times in order to make genetic material for new

viruses and it is certain that this new RNA will contain various point mutations

[90]. This is a powerful weapon of the HCV to escape from the host's immune

system response. The viral RNA will drive the production of protein-based

capsomeres to be used for the new viruses' protein coats. As soon as replication is

over, the viral RNA is encapsidated and moves towards the endoplasmic reticulum

of the host cell. The final product is known as nucleocapsid.

STEP 5: The viral assembly takes place during budding, into cytoplasmic vacuoles and not as it was until recently thought in the cell surface as occurs with its arbovirus relative, the togavirus [81]. The new viruses go to the inside of the plasma membrane, where they interact, attach and finally establish a bud. The plasma membrane surrounds the virus and releases it on the outside of the cell. During the release the virus will take with it a lipid coat from the host's cell membrane. When the virus is about to be ejected from the host cell, the envelope protein will become glycosylated, and the virus will leave the host towards the extracellular matrix [81]. This will help the virus to attach to another liver cell further down. This process will continue until complete lysis of the host cell occurs.
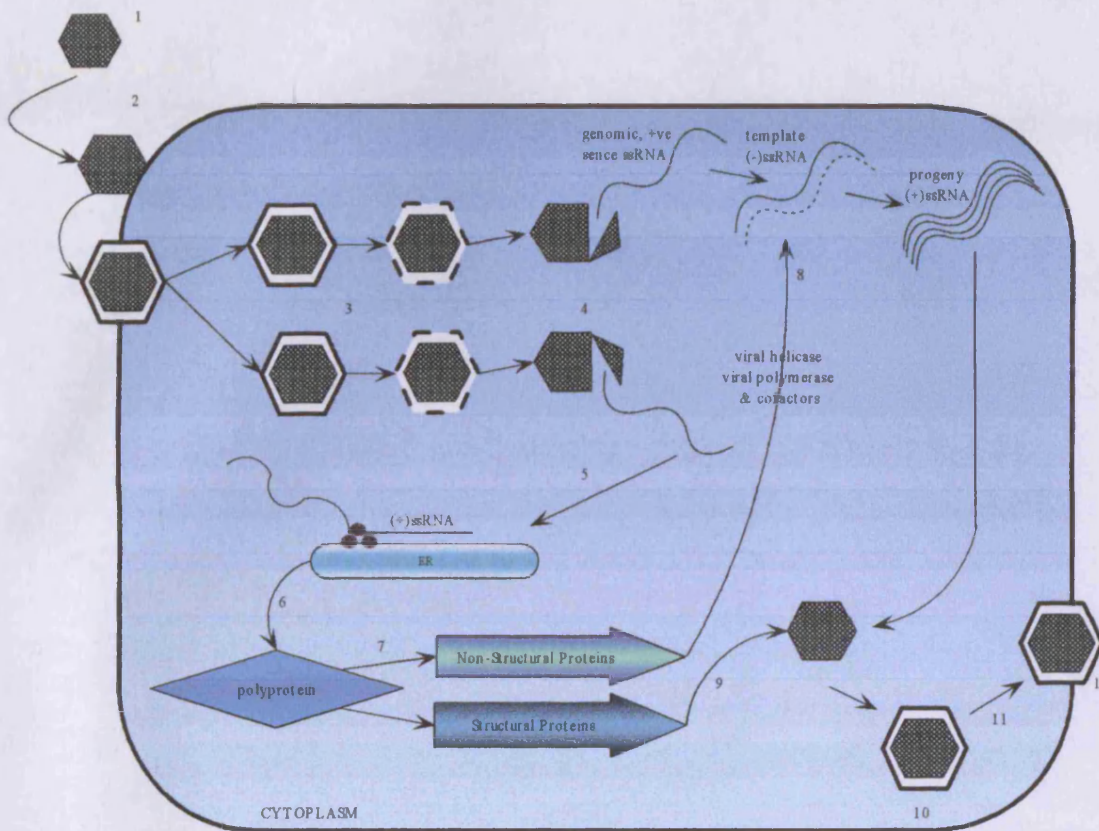
Figure 1.6. The life cycle of the members of *Flaviviridae*, starts with the adsorption (1) then the receptor-mediated endocytosis (2) takes place. The difference in pH in the cytoplasm will trigger the low-pH fusion in lysosomes (3), which will be followed by the uncoating of the genetic material of the virus (4). The linear piece of +ve sense ssRNA will move to the ribosomes to initiate translation (5). Translation will produce the polyprotein the viral genome codes for (6). The polyprotein is going through some proteolytic phases using both viral and host proteases. Eventually a set of structural and non-structural viral proteins is available (7). In the mean time, under the host's cell membrane, the synthesis of copies of –ve sense RNA takes place, followed by +ve sense progeny ssRNA synthesis (8). The nucleocapsid is assembled (9) and the virions bud in the endoplasmatic reticulum (10). They are soon transported to the Endoplasmatic Reticulum (ER) and to the Goldgi Apparatus (GA), where they mature (11). Eventually the mature virions are released (12).

1.6   Targets for Antiviral Research

Significant Research has taken place during the last few years in the area of new antivirals against flaviviruses. New targets have been identified and various experiments have been conducted in the area of drug development, in pursuit of a new and efficient cure to the diseases that flaviviruses are responsible for.

1.6.1   Vaccination

Prevention of the disease is always the best option. Preventing a harmful condition, such as a viral infection, is much more preferred than trying to cure it [91]. Epidemiologically, immunising people against a virus would reduce the number of carriers and therefore the spread of the disease would dramatically decrease after a few generations. The drawback for vaccine development against *Flaviviridae* is that those viruses are highly mutagenic. [92]. Another problem is to which viral strain the vaccine is aimed for. An example here is Dengue virus, where all tested vaccines so far are against a particular serotype and not against all. Efficient broad-spectrum vaccines do not exist, and administering limited vaccination can be even more catastrophic, since any infections will another serotype different to the one that host is immune, could yield a much worse infection [81]. A great number of HCV and Dengue II broad-spectrum formulas have been tested so far, with no much success [93].

1.6.2 Viral IRES Blocking

The replication of the Hepatitis C virus depends on the IRES. IRES will direct the ribosomes to the RNA material of the virus in order for translation to take place and the viral polyprotein to be produced [94]. If the IRES area on the viral RNA was blocked then the process of translation would not be feasible. Merck and Roche have joined in a common attempt to develop an approach of blocking the viral IRES site. Many inhibitory compounds have been suggested and a few very promising antisense oligonucleotides have been produced [95]. Characteristic examples are the peptide nucleic acids that are capable of inhibiting translation at EC_{50} of 50 nM [95]. Biaryl guanidines have been shown to inhibit translation with a dose of 2 μM [95].

1.6.3 Viral Proteases

Viral proteases are employed by the virus for parts of the polyprotein processing [96]. The protease that codes in the NS3 region of the viral genome has been identified as a major target for antiviral research [96]. *Flaviviridae* protease belongs to the chemotrypsin superfamily of proteases. It has a heterodimeric structure and complexes with NS4a proteins in order to be activated [96]. Serine protease activity at four different sites of the viral genome and its crucial contribution to the formation of the viral polymerase give the viral protease high importance and

priority in the list of targets for antiviral agents. Small molecules and peptides

have been tested as potential inhibitors of this enzyme. The most promishing have

been reported to exhibit activities in µM or nM concentrations [97].

1.6.4   Capping Inhibition

Another promising antiviral target is capping inhibition [98]. *Flaviviridae* do not

have a terminal IRES site on their genome. As a result they must achieve 5'

capping in order to retain activity. After RNA triphosphorilation (by an RNA

triphosphatase), guanylyl transferase adds a guanine base on the last phosphate

group. In the next step the site will be methylated by an enzyme called methyl-

transferase [99]. Unfortunately, even though capping inhibition is a promising

approach, no inhibitor has been reported to date.

1.6.5   Inhibition of the Viral Polymerase

Inhibition of the RNA-dependant RNA-polymerase (RdRp) is a very popular

approach that many scientists choose for their research [100]. Viral polymerase is an

enzyme whose structure is very common to all *Flaviviridae* members [101]. The

RNA-dependant RNA-polymerase is coded in the NS5B region of the viral

genome. Research has produced a few active compounds against the HCV RNA-dependant RNA-polymerase.

Ribavirin, which is one of the oldest ones, works via the 5' triphosphate formation; 5-ethynyl-1-β-dribofuranosylimidazole-4-carboxamide (EICAR) is another active derivative of Ribavirin [102].

The *2'-C-methyloadenosine* and *2'-C-methyloguanosine* ribonucleosides are also two potent inhibitors of the HCV RNA replication in vitro [103]. Their inhibitory potency (IC_{50}) was determined to be 1.9 μM and 0.13 μM respectively, in the HCV NS5B enzymatic assay. Unfortunately, 2'-C-methyloadenosine is not orally bioavailable in rats [104].

In an attempt to improve the bioavailability of both 2'-C-methyloadenosine and 2'-C-methyloguanosine a series of nucleosides modified in the purine heterobase were synthesised and biologically evaluated, achieving IC_{50} values as low as 0.12 μM [103].
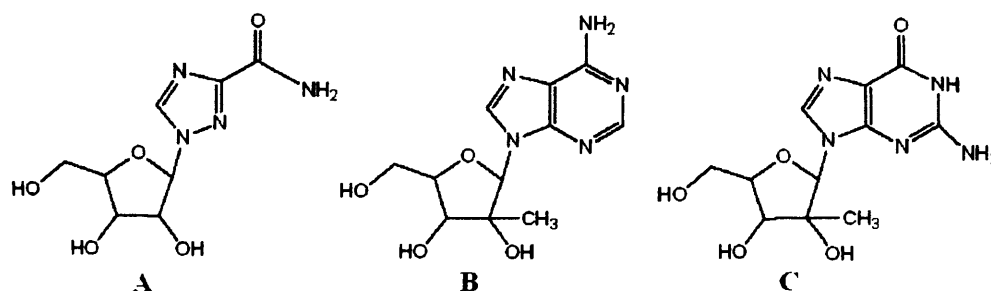

Figure 1.7. The chemical structures of Ribavirin (A), 2'-C-methyloadenosine (B) and 2'-C-methyloguanosine (C).
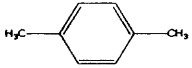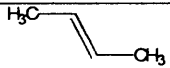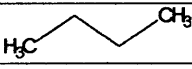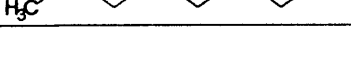
1.6.6   Inhibition of the Viral Helicase

Inhibition of the viral helicase is a very promising approach that is becoming increasingly popular [105]. Helicases are capable of unwinding double stranded DNA and RNA to single strands by breaking the series of hydrogen bonds that keep the two strands together. The unwinding activity of the viral helicase is essential to the virus during its replication process. Mutated inactive helicases in Dengue and Bovine Diarrhea viruses led to reduced proliferation of the virus [106]. It is believed that inhibition of the viral helicase will be an effective tool for the reduction of the replication rates of the *Flaviviridae* viruses. The viral Helicase is coded in the NS3B region of the viral genome next to the NS3A gene, which codes for the viral Protease. A summary of the Flaviviridae Helicase inhibitors is reported below.

1.7   Current Research & Known Inhibitors for *Flaviviridae* Helicases

Virofarma reported a series of compounds with benzimidazole derivatives with

IC_{50} values within 0.7 to 10 µM (Table 1) [107]. The mechanism of action of these

drugs is unclear, but based on their shape one could assume that they should act in

the RNA biding site of the helicase, competing with the single-stranded RNA for

binding to the protein.

Table 1. The 7 compounds reported by Virofarma and their activities.

Compound	Linker (R)	Linker (R) Name	IC_{50}
1		Benzene	10
2		A	0.7
3		C2	0.7
4		C4	0.7
5		C8	0.7
6		C10	0.7
7		C12	0.7

The aminophenylbenzoxazole- and the aminophenylbenzothiazole- containing

compounds showed no inhibitory effect, whereas the aminobenzimidazole-

derived diamides produced a rather moderate 13% inhibition. The

aminophenylbenzimidazole derived diureas showed inhibition in the range

between 20 and 28%. In general it was noted that there is significant decrease of

potency when the benzimidazole moiety is removed and substituted with

benzoxazole or benzothiazole moieties. The same is observed for the removal of

the benzene ring. Activity drops when the benzene ring is removed.

R = C4 or C6   R = C2, C4, C6, C7

R = C4 or C6
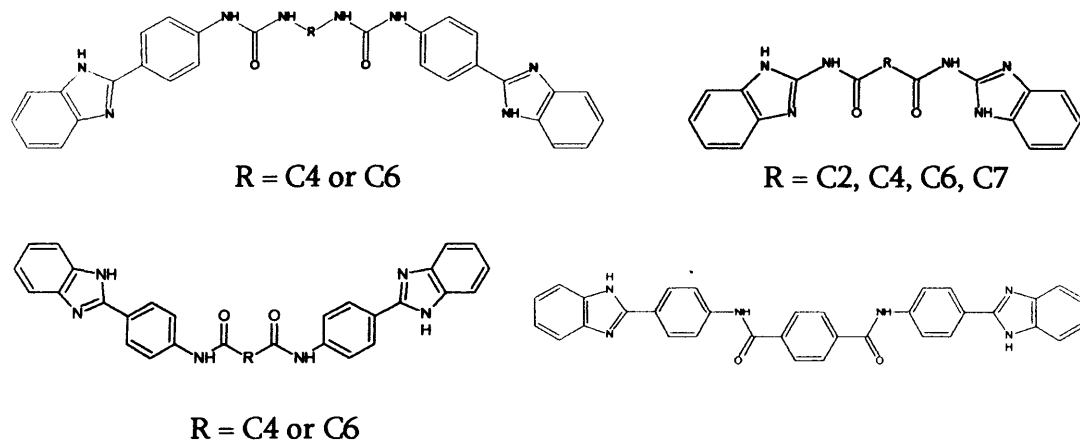
Figure 1.8. Some of the compounds suggested by Virofarma as *Hepatitis C* Helicase Inhibitors.

A series of nucleotide compounds were suggested as potential inhibitors of the

helicase of *Hepatitis C* mainly targeting the ATP site on the protein [110]. The

theory is that since the unwinding process is energy-dependent, even competition

for the ATP, induced by any compound, would result in less energy being

available for the system and as a result reduced unwinding function of the

helicase. So, a wide range of competitive NTPase inhibitors was suggested. These

include ribavirin-5 triphosphate (RTP), ribavirin-5 diphosphate (RDP), adenosine-

5-thiotriphosphate (ATP-S) or ADP. These compounds were all tested against

Hepatitis C helicase and were found to be moderate inhibitors of the unwinding

activity of the HCV Helicase enzyme (summarized in figure 1.9).
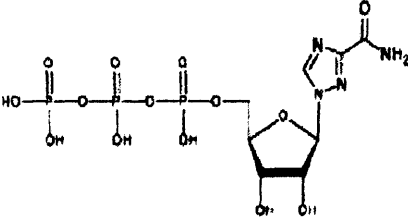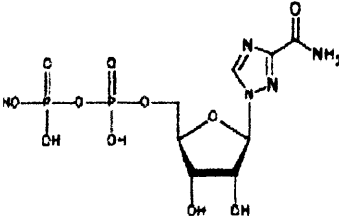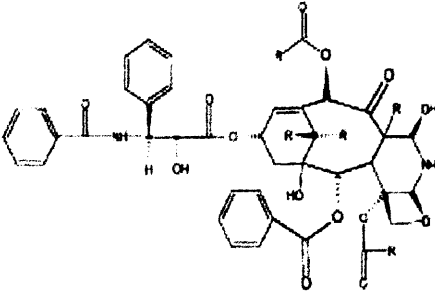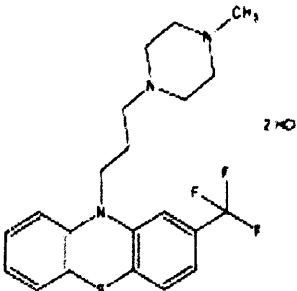
Structure		Inhibition Potency	
		ATPase	Helicase
	RTP	$IC_{50} = 40 \mu M$ Competitive	$IC_{50} = 180 \mu M$
	RDP	$IC_{50} = 90 \mu M$ Competitive	$IC_{50} = 250 \mu M$
	Paclitaxel	$IC_{50} = 17 \mu M$ Competitive	$IC_{50} > 1 \, mM$
	Trifluoperazine	$IC_{50} = 105 \mu M$ Noncompetitive	$IC_{50} = 0.6\text{-}0.7 \, mM$

Figure 1.9. The series of compounds suggested by Borowski [111]

The effect of N7-chloroethylguanine and N9-chloroethylguanine in the helicase activity of the WNV was investigated. These compounds were found to be activators of the helicase activity by 850 and 220% respectively [114]. Similar patterns in activation were observed with the *Hepatitis C* helicase as well (Figure 1.10).
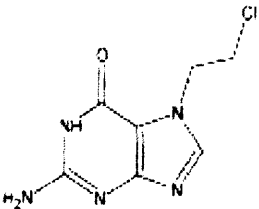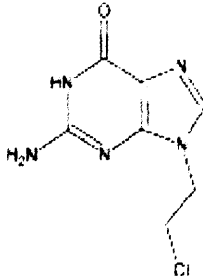
Structure	Activation Potency	
	ATPase	Helicase
N⁷-chloroethylguanine	ED₂₀₀ > 1 mM	ED₂₀₀ = 18 µM
N⁹-chloroethylguanine	ED₂₀₀ > 1 mM	ED₂₀₀ = 120 µM

Figure 1.10. N7 chloroethyl guanine and N9-chloroethyl guanine

A series of ring-expanded ("fat") nucleoside analogues with the 6-aminoimidazo [4,5] [1,3] diazepine-4,8-dione ring system was synthesised and tested against the NTPase/helicase WNV (Table 2) [115]. Some compounds were found to be potent inhibitors of the NTPase/helicase of the WNV. It was also found that the same compound that was active against the helicase activity of the protein was inactive against the NTPase activity. It was supposed that this series of compounds binds to the major or minor groove of dsDNA or dsRNA, affecting the stability of the double helix.
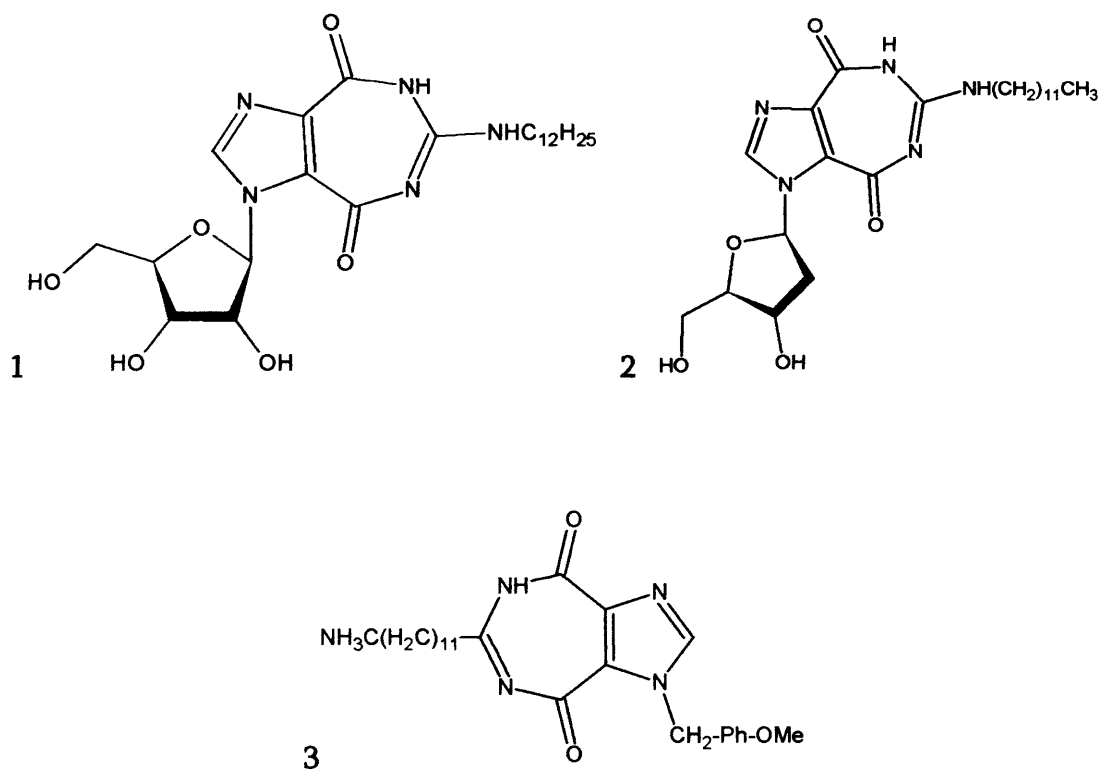
Table 2. The IC$_{50}$ values of the above compounds

Compound Number	WNV - IC$_{50}$ (μM)
1	10
2	1.5
3	1-3

Two series of ring-expanded ("fat") heterocycles, nucleoside and nucleotide

analogues were prepared and tested against the NTPase/helicase activity of the

Flaviviridae family of viruses. The first series contained the *imidazo [4, 5-e][1,3]*

diazepine ring system and the other series contained the *imidazo [4,5-e][1,2,4]*

triazepine ring system. They were tested against the helicases of WNV, HCV and

JEV, with some of these compounds quite potent inhibitors of the helicases (Table

3). Their inhibitory effect is mainly exerted on the ATPase activity of the enzymes

[116].
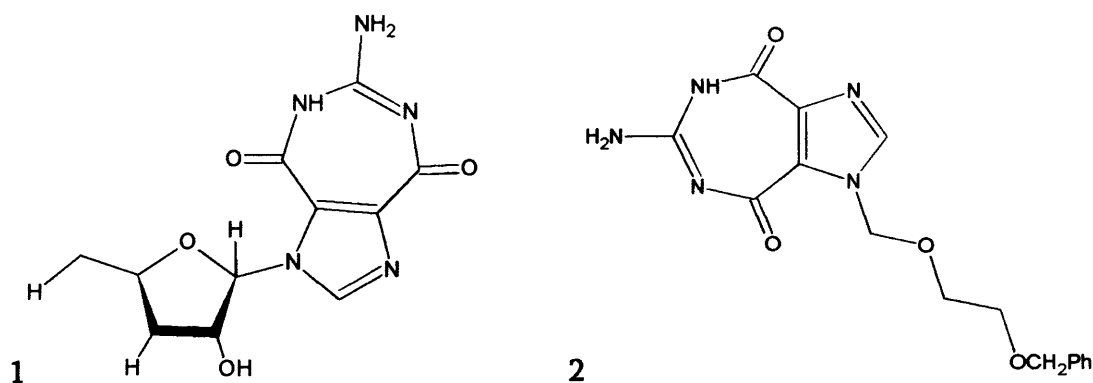
Table 3. The IC50 values of the above compounds

Compound Number	WNV IC$_{50}$ (µM)	WNV IC$_{50}$ (µM)	WNV IC$_{50}$ (µM)
1	5.7	>500	2
2	3.3	5.5	>500

1.8   Aims and Objectives

The main objective of this project is to develop a research methodology to facilitate the design and the discovery of novel antiviral agents. The strategy followed can be broken down into the following steps:

- The first step is the identification of a suitable target. The helicase and the polymerase proteins of *Flaviviridae* were chosen.

- The three-dimensional structures of the new targets were modelled by homology based molecular modelling techniques.

- Then using *de novo* structure – based drug design algorithms a series of lead compounds was generated.

→ At this stage some of those lead compounds will be chemically synthesised by other members of the group.

- The genes of Hepatitis C helicase, Dengue helicase and Dengue NS3 (helicase + protease) domain were cloned into expression vectors and the proteins were produced and purified.

- Finally an enzymatic assay was developed in order to biologically evaluate the potency of the designed inhibitors. The experimental results obtained from the biological assay were fed back to the computer in order to refine and improve the *in silico* model.

chapter 2
Molecular Modelling

2.1   Molecular Modelling

Molecular modelling is very useful for investigating, comparing, analysing and visualizing chemical structures and for giving qualitative and quantitative information about biological systems [117].
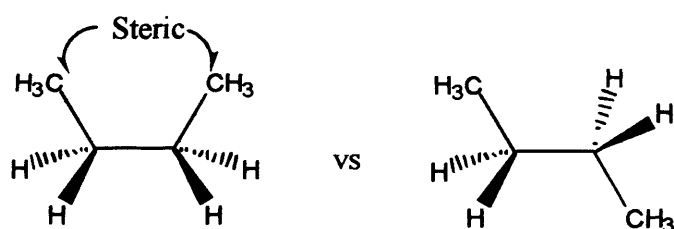
Figure 2.1. Steric hinderance of a small organic compound

Figure 2.1 shows a characteristic example of steric hinderance. Two dimensional models like this only contain qualitative information. Quantitative information can arise through molecular mechanics and in conjunction with a computer, where the physical properties of the molecules can be evaluated and analysed based on a set of predefined criteria concerning various chemical properties (such as bonding, charges, steric hinderance...) [118]. Molecular Modelling can be used to study the geometry, the energy and the chemical properties *in silico* so efficiently that nowadays it is possible to predict the outcome of chemical reactions, design reactions, determine the unknown three dimensional structures of proteins, screen and design new and effective drugs [118]. In this part, the basic principles and theory of the methods that have been used in this chapter will be presented.

2.1.1   Molecular Properties

The geometry and the overall structure of a molecule are described by its bond distances, dihedral angles and bond angle [119]. This unique set of angles and distances create a set of coordinates that define the positioning of each atom in that molecular structure in three dimensional (3D) space. The energy condition of this molecule can also be assessed and evaluated. The energy of a molecule includes all forms of energies, such as kinetic motion (described by vibration, rotation and translation) and forms of the potential energy of the molecule [120]. The potential energy of a molecule can be defined by the analysis of the electrostatic interaction between charges, the magnetic interactions between spinning charges and finally the potential energy of the bonds of the molecule. The total energy is indicative of the reactivity and stability of that a molecule or a system. Below is a reaction coordinate diagram that indicates the energy changes during the course of a chemical reaction [121].
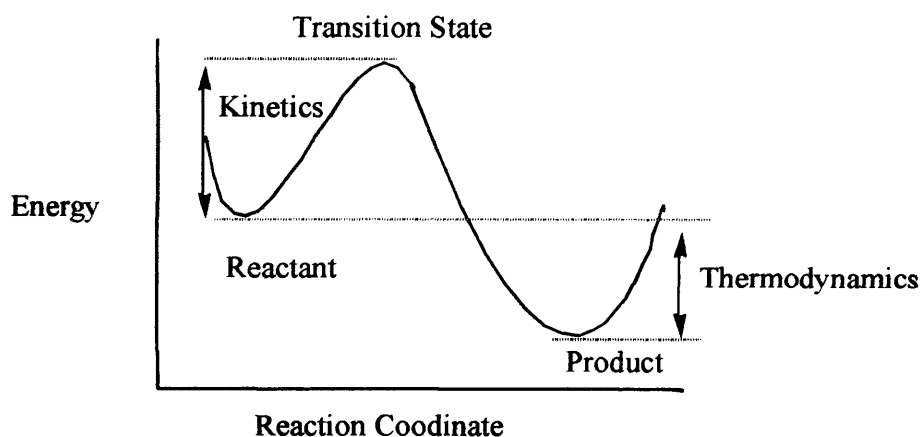
Figure 2.2. Energy changes during the course of a chemical reaction.

Here the products are in the lowest or global minimum, the transition state is at

energy maximum and the reactants are at an energy minimum. The dotted lines in

the above diagram are indicative of the reactivity of the system (its kinetics) and

the thermodynamic stability of the system. Through molecular modelling it is

possible to quantify the above characteristics of the system and, for example,

predict its reactivity. There are two fields in molecular modelling that attempt to

do this: molecular mechanics and quantum mechanics [122].

2.1.2   Molecular Mechanics & Forcefields

Molecular mechanics are based on the ball and spring representation of molecular

systems. Here, the atoms are considered to be little balls, with varying properties

according to the element, and the bonds are considered to be the springs that

make the two interconnecting balls interact with each other. The ball and spring

model is described by Hook's law, which evaluates and quantifies the energy of

the stretching of the spring [123].
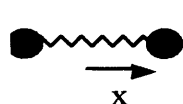
$$F = k\,x \qquad E = 1/2\,k\,x^2$$
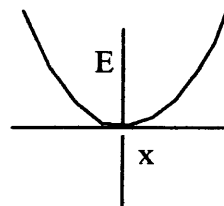
The force constant is the constant k. The energy that is contained in the spring

and the restoring force of the spring are proportional to the force constant. The

force constant will determine the strength of the bond that the spring represents

[124]. The vibrational frequency of the spring is described as:

$$n = \frac{1}{2\pi}\sqrt{\frac{k}{\mu}}$$

The vibrational frequency (n) has been estimated to be proportional to the square

root of the force constant (k) and inversely proportional to the reduced mass of the

atoms that participate in a bond [125].

All of the above can be combined and through potential energy functions of

various structural features, such as bond lengths, bond angles and non-bonded

interactions, can describe a forcefield (figure 2.3) [126]. There are many different

ways to set a forcefield depending on the needs of the system under investigation.

Usually the factors affecting the energy of a molecular system (bonds, angles,

dihedrals, non-bonded, etc), are evaluated separately and they will contribute to

the value of the total energy of the system [127]. The most popular forcefields are

the MM2, which is suitable for small molecules, hydrocarbons and some simple

heteroatom functional groups, AMBER or CHARMM, which are parameterised to

be used for peptides, nucleic acids and generic macromodels [128].
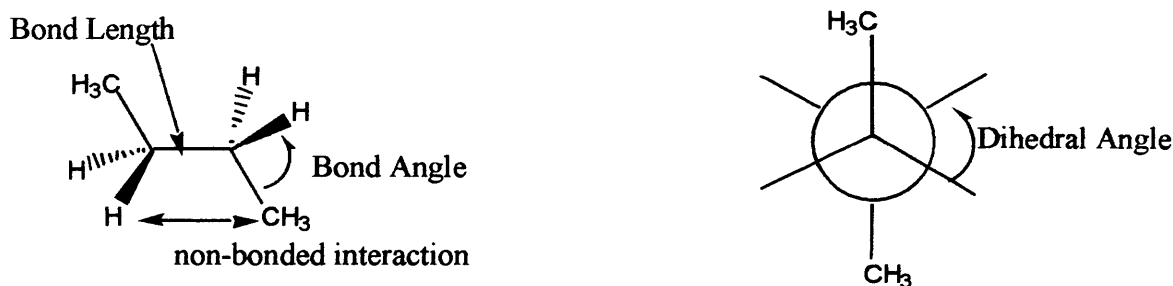
Figure 2.3. Total energy is affected by bond distances, bond angles, dihedral angles
and finally non-bonded interactions

Overall through molecular mechanics the total energy of a molecule is described as a sum of all the contributions that may arise from loss of equilibrium in bond distances, also known as stretching contribution, bond angles, known as bending contribution, dihedral angles, the torsion contribution and finally non-bonded interaction contributions [129].

$$E^{total} = \underbrace{\sum_i E_i^{stretch}}_{\text{bonds}} + \underbrace{\sum_i E_i^{bend}}_{\text{bond angles}} + \underbrace{\sum_i E_i^{torsion}}_{\text{dihedral angles}} + \underbrace{\sum_i \sum_j E_{ij}^{non-bonded}}_{\text{non-bonded atoms}}$$

The energy that is stored in chemical bonds of a molecule can describe the stretch, bend, and torsion energy whereas it is the steric attraction or repulsion that represents the non-bonded energy [130]. The latter is broken down to two different categories: the van der Waals (VDW) and electrostatic interactions [131].
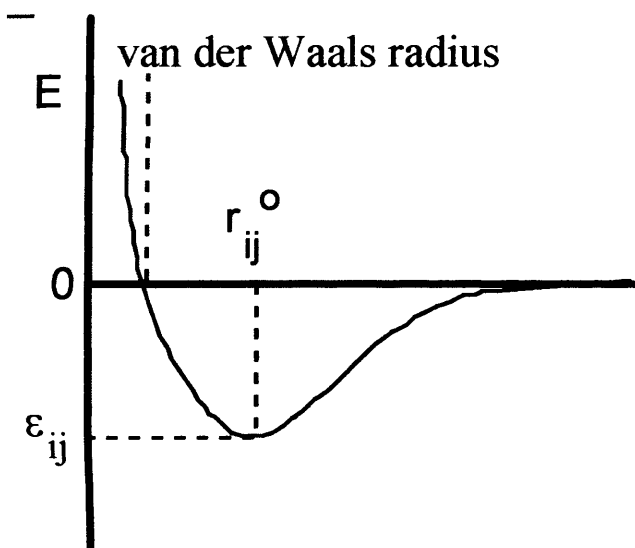
$$E_{ij}^{VDW} = e_{ij}\left((\frac{r_{ij}^0}{r_{ij}})^{12} - 2(\frac{r_{ij}^0}{r_{ij}})^6 \right)$$

where,
r_{ij}^0 is the distance at the minimum
e_{ij} is the energy at the minimum
$s = 2^{-1/6} r_{ij}^0$ is the van der Waals radius

Figure 2.4. The van der Waals interactions plot and formula.

A very steep energy barrier is generated at the van der Waals radius of each atom.

Moreover a very shallow energy well is produced at larger separations (figure 2.4).

The inherent steric size of atoms and elements is dictated by their VDW radii. The

same metric is used to describe weak attractive forces between atoms in close

proximity [132]. A trivial example of the weak van der Waals attractive forces is the

condensation of a gas into liquids. Furthermore it is the van der Waals radii of

each element that is used for its visualisation purposes in space filling models of

the molecule they participate. Steric repulsion takes place only in the case where

two atoms come closer than the sum distance of their VDW radii [133].

2.1.3 Energy Minimisation

As soon as the set of the internal coordinates of a molecular system has been determined, computer algorithms can be used to help find those coordinates which will account for the lowest energy of the system [134]. All bond angles, lengths, dihedral angles and the relative energy between various different conformations of a given system will be evaluated in order to determine the minimum energy conformation [135]. It is crucial to understand that reducing the strain energy of a given molecular system does not mean that the system will reach energy minimum (also known as global minimum). An example is the following figure (figure 2.5) with two different conformations of butane:
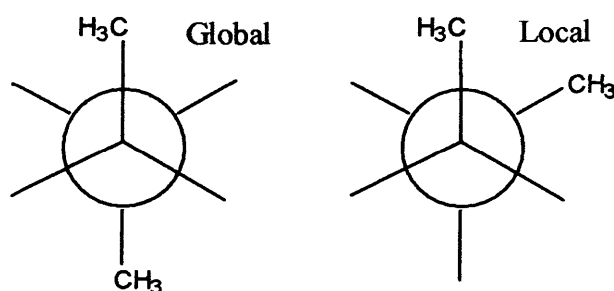
Figure 2.5. Two different conformations of butane

An energy minimisation algorithm will allow the rotation of groups, when their bonding allows. The rotation of the groups will give the molecule the opportunity to explore different conformations that will account for different energy values, thus allowing the compound to move towards its global minimum conformation [136].

2.1.4 Quantum Mechanics

Quantum mechanics are useful for the evaluation of electronic properties that may influence physical and chemical reactions between different groups of a biological system [137]. The estimation of energy conformational profiles and that of intermolecular interactions in various cases can be achieved through quantum mechanical calculations. The quantum mechanics arise from the Schroendinger equation [138]. In a very simplistic representation the latter equation can be described as:

$$H_{psi} = E_{psi}$$

Where: H is the Hamiltonian operator,
           Psi is the wave function
           E is the total energy of the system

Using the Schroendinger equation and the above operators a complete description of any molecular system can be achieved.

At a given molecular system the Schroendinger equation can be solved in two different ways [139]. First is the no approximation or *ab initio* approach [140]. Secondly, the Schroendinger equation could be solved by the introduction of some approximations, which is a method also known as semi-empirical approach [141]. The main advantages of *ab initio* methods is that all electrons are explicitly included and they do not requiring any specific parameterization, thus making *ab initio* calculations a universal tool for all molecular systems [142]. On the other hand in semi-empirical calculations it is the valence electrons that are explicitly

included. Here most of the integrals are neglected or approximated by a set of

parameters that arose from experimental work [143]. Choosing the most suitable

approach to investigate a given system does not solely depend on the size of the

molecular system, but on the type of unique parameters and molecular properties

of the system. For example, the conformation, electrostatic potential and, electron

density of the given molecular system.

The most popular semi-empirical calculation algorithms are the AMI, MNDO,

CNDO and MINDO, whereas for *ab initio* calculations the Gaussian program is

very representative. AMPAC and MOPAC on the other hand are very widely used

for *ab initio* calculations [143].

2.1.5   Molecular Dynamics

Molecular dynamics simulations are used to describe the patterns, strength and properties of drug-receptor interactions, the solvation of molecules, the conformational changes that a molecule may undergo under various conditions and other events that require the systematic evaluation of molecular properties in dynamic molecular systems [144]. Molecular dynamics are concerned with the motion of atoms and molecules [145]. An example is the normal mode analysis, where the dynamic motion of a molecular system is evaluated from its total energy. According to Hook's law, the total energy of a small and simple diatomic molecule should be given by:

$$E(r) = \frac{1}{2} k(r - r^o)^2 \quad \varepsilon \quad F(r) = -\frac{dE}{dr} = -k(r - r^o)$$

Where: E = energy, F = force, r = distance, r^o = initial distance, k = Hook's constant

Here, the forces on the atoms are estimated by the derivative of the energy [146]. When the forces have been assigned the Newton law of motion can be used to solve the molecular motion (F = m x a, force = mass x acceleration). In the case of the small and simple diatomic molecule the displacement from the equilibrium bond length (x) will be given by the formula:

$$x = r - r^o \text{ and } \mu = \frac{m_C \, m_O}{m_C + m_O}$$

Where: μ is the effective mass of the vibrating diatomic molecule.

In molecular dynamics simulations the kinetic energy of the system will depend

on the temperature of the system [147]. The total energy will be the sum of the

kinetic and potential components of the system. The acceleration of each atom is

estimated from the set of forces it accepts, under the given forcefield. The results

generated can be used to estimate the configurational and momentum information

for each atom of the system (i.e. energy, pressure). Molecular mechanics therefore

can be used to further optimise a model generated by homology modelling as well

as docking results, where the protein / ligand interactions can be analysed for their

stability during a specific time [148].

2.1.6   Homology Modelling

Homology modelling is used in order to predict the 3 dimensional structures of proteins with unknown 3D structure, using solved homologous proteins as templates [149].

Homology modelling claims that the biological structure of a protein is more related to its biological properties and function than its sequence [150]. A homologous protein is a protein that belongs to the same family, has the same function and shares more than thirty percent similarity with the protein of interest.

The first step of a homology modelling algorithm is to set up and optimise the sequence alignment between the query protein and its template. Sequence alignment is broken down into four steps [151]. Firstly, it uses rapid alignment methods to calculate all pairwise similarity scores. The second step is the generation of a similarity matrix. Then the sequences are clustered according to the generated similarity matrix with the aid of an algorithm. The next step is the generation of a cluster alignment using a consensus method and finally a multiple-progressive alignment is generated. The groups of the sequences are aligned according to their cluster branch order.

After that the algorithm will perform an initial partial geometry alignment for the sequence of the template protein with the unknown structure [152]. The initial

geometry will be copied from various regions of one or more template proteins. If there is residue identity, between the alignments, then all coordinates are copied. That includes backbone and sidechain. If there is not residue identity but still residue similarity is retained, only the coordinates from the backbone atoms will be passed on. In cases of zero identity or similarity a gap will be left on the model, which is also known as loop. A loop will be modelled by borrowing coordinates from any protein (from the Protein Data Bank) that matches the required sequence. The sidechain is generated automatically using a build-in rotamer explorer module [153].

Finally, the new models must adequately meet and satisfy a scoring function that ensures that the degrees of the non-polar sidechain groups that are buried are within range and that all hydrogen bonding capabilities have been explored [154].

2.1.7   Model Evaluation

The N-Cα and the Cα-C bonds in a protein are capable of rotating. The rotation of

these bonds can be described by the phi and psi torsion angles (figure 2.6 – bottom

right) [155]. The values of these angles define the secondary element that each set of

residues will form. To visualise phi and psi angles in a protein a Ramachandran

plot is constructed. Psi angles are found in the YY' axis, whereas phi angles are

found in the XX' axis (figure 2.6).
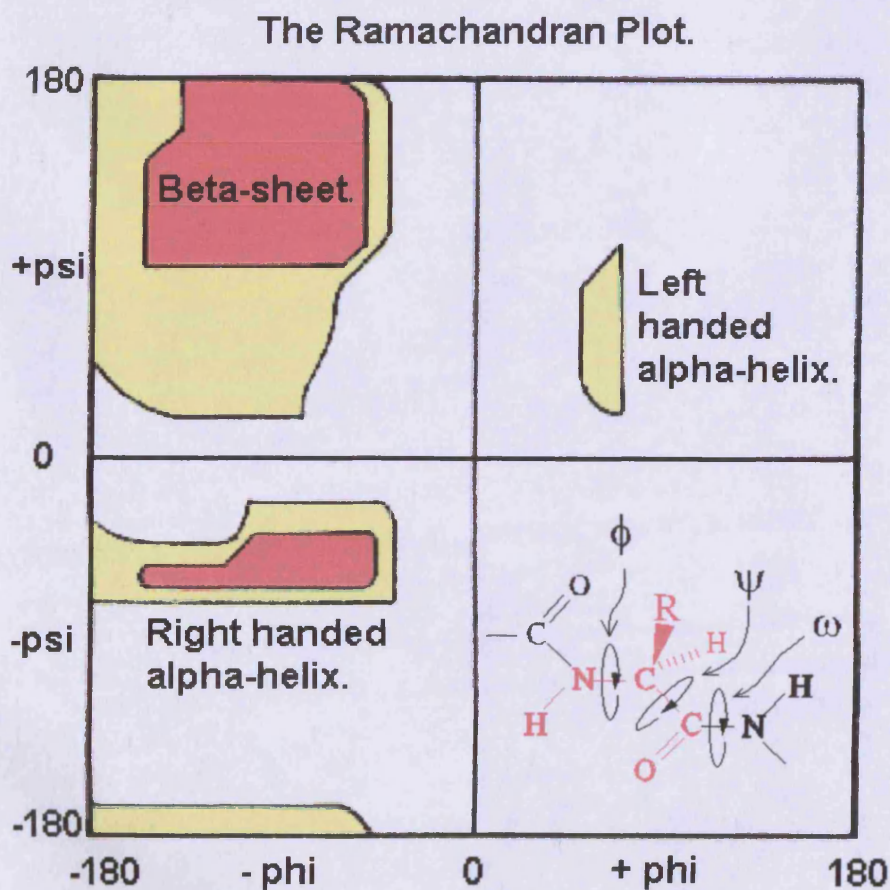
Figure 2.6 The layout of a generic Ramachandran plot. Indicating the major secondary element areas of a protein. Bottom-right: a typical residue with the phi, psi and omega angles designated.

The white space on the diagram above describes the pairs or groups of atoms that have distance smaller than the sum of their van der Waals radii, because of the conformation of the backbone. As a result a sterically unreal conformation will be generated, thus making the combination of those atoms disallowed. The area with no steric clashes is represented in red color. The residues in the red area are considered to be in the allowed regions and this is where the alpha helices and beta sheet conformations are usually found. The yellow areas include the pairs or groups of atoms with radii a fraction less than the sum of their van der Waals radii. This is the area that the motif of the left handed alpha helix would be found, where the atoms are allowed to come a little closer. The generously allowed areas of the Ramachandran plot are areas with pairs or groups of atoms with radii quite less than the sum of their van der Waals radii, but not significantly clashing with each other. Glycine is a versatile amino acid, because it does not have a side chain. Glycine can take phi/psi angles in each of the four quadrants in the Ramachandran plot. As a result glycines are usually encountered in loop regions in the protein, where it would be impossible for any other residue to be, because of the steric hindrance.

2.1.8   Molecular Docking

The last molecular modelling technique that is extensively used in this project is molecular docking [156]. The docking algorithm is basically split into two main parts: the searching algorithm and the scoring algorithm [157].

The searching algorithm will explore all conformations of the ligand within the space available [158]. Practically, it is impossible to perform all these calculations for every compound so most of the rotational and translational states of each compound will be explored within a given threshold of identical conformations. Each compound is not a rigid body, but is a dynamic structure that exists in an ensemble of different conformations. The user can define how fine the docking algorithm will be by altering the various parameters of the task. Very fine calculations are much more accurate, but also much more time consuming. The most popular docking algorithm approaches can involve a coarse grained molecular dynamics simulation or a linear combination of many structures or a genetic algorithm that generates new conformations as it moves along.

The second feature of the docking algorithm is its scoring function [159]. The scoring function must be able to accurately evaluate each different conformation using certain forcefields and rules from physics, and return a value that will describe the energy of the system at the given conformation. Low energies indicate better, more stable interactions.

2.1.9   Drug design

In the last few years the number of biological targets suitable for drug discovery has increased enormously and, at the same time, computer-aided drug design techniques are becoming more popular everyday, sustained by the need to accelerate the drug discovery process [160]. Structure-based methodologies are now widely used to design and optimise new potential drugs and, in particular, *de novo* approaches (structure based novel inhibitor design) remain very attractive to researchers as a time and cost efficient methods to design novel entities, despite the challenges associated with them [161].

Basically, three questions have to be addressed by a *de novo* design program: how to assemble the candidate compounds; how to evaluate their potential quality; and how to sample the search space effectively.

In answering the first of these questions, we can consider two main approaches to build the desired structures [161]. The first (figure 2.7-top) is the linking procedure, where the algorithm selects the most suitable moieties to interact with the active site of the protein and then starts to link them together in a chemically appropriate way. The second approach is the growing procedure (figure 2.7), which involves the determination of a group as a starting point and then the growing of a larger compound that would fit in the active site and that would be capable of establishing interactions with it. The first approach requires that the

user must supply the docking/interaction points of the moieties that will be used

as anchors to the final structure.

The linking approach

The growing appoach
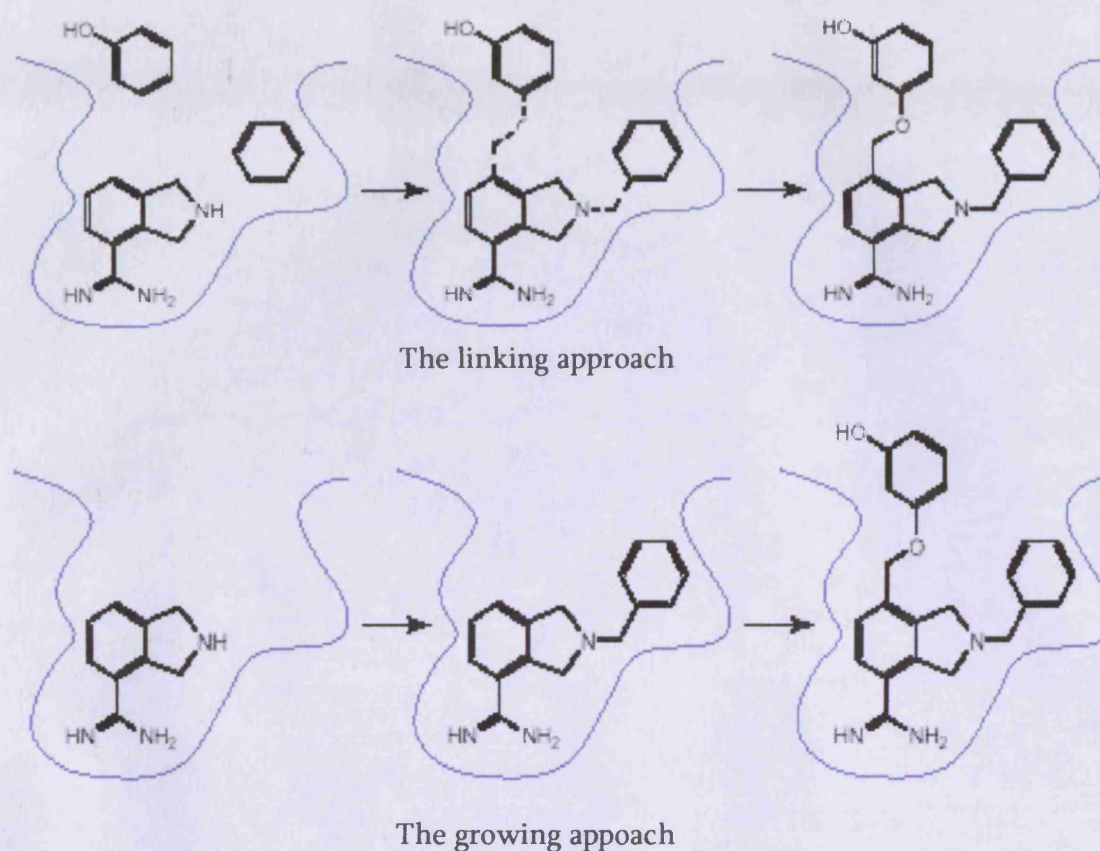
Figure 2.7. LigBuilder's approaches to structure-based drug design.

The second approach requires that the user selects a suitable "seed" as a starting

point, and that the user directs the growing process. The best results will be

obtained from the combination of these two procedures.

Evaluation of the results obtained could be done by different methods [162]:

receptor-based scoring functions: explicit force-field methods; empirical scoring

functions; and knowledge-based scoring functions. All of these approaches attempt

to approximate the binding free energy. Force fields are computationally more costly than the other two types of scoring functions.

The last issue related to the *de novo* drug design methods is the sampling of the chemical space associated with the growing compounds [163]. There are many different search algorithms implemented in the variety of software programs available, and new ones are continuously developed.

It is also important to mention the nature of the "building blocks" used to construct the structures: some programs connect single atoms, others use a library of chemical fragments. The size and the diversity of the fragment library is fundamental for a broad search of the chemical space [164]. The main problem associated with the *de novo* methodologies is represented by the chemical feasibility of the structures generated: unfortunately, computers do not know organic chemistry, yet.

In our project, we have used a *de novo* approach for the design of some helicase inhibitors using a software called LigBuider. This package can use both the "link" and the "grow" approaches and implements a genetic algoritmm based search for the determination of the best structures. Once the compounds are drawn from the population, the new population is formed by combining the information from groups of selected individuals. This is done by randomly splitting the representative strings of the parents and recombining them to form new entities. This operation is called Crossover [165]. Because the parents have been selected for their fitness, the children of these parents will, statistically, be more fit than the

parents. This is done until a new population has been built from the old one. The

final step is the evaluation of the new population, which may result in starting the

whole process over again by changing a single or groups of parameters. The cycle

described in figure 2.8 is continued until a solution that is good enough for the

experiment's criteria has been obtained.
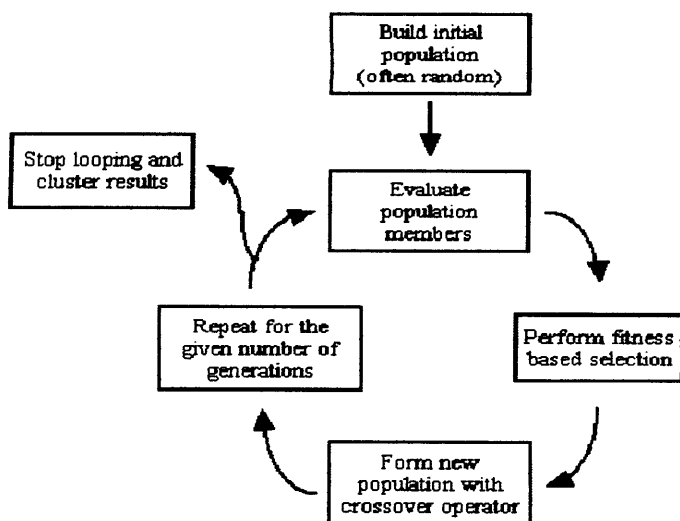
Figure 2.8 The steps of the genetic algorithm of LigBuilder

2.2   Introduction to the Homology Modelling Project

The aim of this project was to design the 3D models of the Helicase and

Polymerase enzymes of the Dengue, Yellow Fever, West Nile and Japanese

Encephalitis virus using homologous proteins as templates, all members of the

flaviviridae viral family. The models are to be used for structure based drug design,

in order to design inhibitors for these enzymes. However, priority was given to

the HCV helicase, firstly due to the available X-ray solved structure (co-

crystallised with ssRNA [166]), and secondly due to the availability of the HCV

helicase gene in the molecular biology lab.

2.2.1   Homology Modelling of the Flaviviridae Helicases

The initial DNA sequences of the genes encoding for the Dengue-type II, West Nile, Yellow Fever and Japanese Encephalitis Viral Helicase proteins were downloaded from GenBank [129]. The DNA sequences were translated on the on-line Expasy DNA/Protein translator [167].

The on-line Blast-P search revealed that the best template for the homology modelling experiments was the corresponding Hepatitis C Helicase, which has been solved by X-ray crystallography (PDB entries:1A1V & 8OHM).

The sequence alignment was performed with the multiple alignment program ClustalW [168]. It revealed an average 30% homology identity between the HCV Helicase and the model sequences. The percentage identity between the Dengue Primary sequence and the HCV sequence was 31.2%, the West Nile virus 29.8%, the Japanese Encephalitis 30% and the Yellow Fever 31%. The sequence alignments are shown in figure 2.9.

The Helicase Homology Experiments were performed by a traditional homology modelling approach (i.e. using a single protein template only), since the 30 % homology identity is marginally acceptable by the algorithm. The homology modelling algorithm would divide the sequence into Structurally Conserved Regions (SCRs) and Structurally Variable Regions (SVRs).

```
                            207-210
HCV     --P-PAVPQSFQVAHLHAPTGSGKSTKVPAAYAAQGYKVLVLNPSVAATLGFGAYMSKA-
DEN     I-E-DDIFRKRRLTIMDLHPGAGKTKRYLPAIVREAIKRGLRTLILAPTRVVAAEMEEAL
YEL-F   LQEIPTMLKKGMTTVLDFHPGAGKTRRFLPQILAECARRRLRTLVLAPTRVVLSEMKEAF
JAP-E   AYT-PNMLRKRQMTVLDLHPGAGKTRKILPQIIKDAIQQRLRTAVLAPTRVVAAEMAEAL
WEST-N  GFE-PEMLRKKQITVLDLHPGAGKTRKILPQIIKEAINKRLRTAVLAPTRVVAAEMSEAL

                                                       290-293
HCV     HGVDPNIRT-GVRTITTGSPIT----YSTYGKFLADGGXSGGAYDIIICDECHSTDATSI
DEN     RGLPIRYQTPAIRAEHTGREIVDLMCHATFTMRLLSPIRVPN-YNLIIMDEAHFTDPASI
YEL-F   HGLDVKFHTQAFSAHGSSGREVIDAMCHATLTYRMLEPTRVVN-WEVIIMDEAHFLDPASI
JAP-E   RGLPVRYQTSAVQREHQGNEIVDVMCHATLTHRLMSPNRVPN-YNLFVMDEAHFTDPASI
WEST-N  RGLPIRYQTSAVHREHSGNEIVDVMCHATLTHRLMSPHRVPN-YNLFIMDEAHFTDPASI


HCV     LGIGTVLDQAETAGARLVVLATATPPGSVTVPHPNIEEVALSTTGEIPFYGKAIPLEVIK
DEN     AARGYISTRVE-MGEAAGIFMTATPPGS-RDPFPQSNAPIMDEEREIPERSWNSGHEWVT
YEL-F   AARGWAAHRAR-ANESATIIMTATPPGT-SDEFPHSNGEIEDVQTDIPSEPWNTGHDWIL
JAP-E   AARGYIATKVE-LGEAAAIFMTATPPGT-TDPFPDSNAPIHDLQDEIPDRAWSSGYEWIT
WEST-N  AARGYIATKVE-LGEAAAIFMTATPPGT-SDPFPESNAPISDMQTEIPDRAWNTGYEWIT

                      370                 393                      411-412
HCV     G--GRHLIFCHSKKKCDELAAKLVALGINAVAYYRGLDVSVIPT--SGDV-VVVATDALF
DEN     DFKGKTVWFVESIKTGNDIAACLRKNGKRVIQLSRKTFDSEYVKTRTNDWDFVVTTDISE
YEL-F   ADKRPTAWFLESIRAANVMAASLRKAGKSVVVLNRKTFEREYPTIKQKKPDFILATDIAE
JAP-E   EYAGKTVWFVASVKMGNEIAMCLQRAGKKVIQLNRKSYDTEYPKCKNGDWDFVITTDISE
WEST-N  EYVGKTVWFVESVKMGNEIALCLQRAGKKVIQLNRKSYETEYPKCKNDDWDFVITTDISE

            422-424                            457     460-467
HCV     TG-DF--DSVIDCNTXVTQTVDFSLDPTFTIETTTLPQDAVSRTQRRGRTGRGKPGIYRF
DEN     MGANFKAERVIDPRRCMKPVILTDGEERVILAGP-MPVTHSSAAQRRGRIGRN-------
YEL-F   MGANLCVERVLDCRTAFKPVLVDEGR-KVAIKGP-LRISASSAAQRRGRIGRN-------
JAP-E   MGANFGASRVIDCRKSVKPTILEEGEGRVILGNP-SPITSASAAQRRGRVGRN-------
WEST-N  MGANFKASRVIDSRKSVKPTIIEEGDGRVILGEP-SAITAASAAQRRGRIGRN-------

                                        501
HCV     VAPGERPSGMFDSSVLCECYDAGXAWYELTPAETTVRLRAYMNTPGLPVCQDHLEFWEGV
DEN     --PRNENDQYIYMGEPLENDEDCAHWKEAKMLLDNINTPEGIIPSMFEPEREKVDAIDGE
YEL-F   --PNRDGDSYYYSEPTSENNAHHVCWLEASMLLDNMEVRGGMVAPLYGVEGTKTPVSPGE
JAP-E   --PNQVGDEYHYGGATSEDDSNLAHWTEAKIMLDNIHMPNGLVAQLYGPEREKAFTMDGE
WEST-N  --PSQVGDEYCYGGHTNEDDSNFAHWTEARIMLDNINMPNGLVAQLYQPEREKVYTMDGE


HCV     FTGLTHIDAHFLSQTKQSGENFPYLVAYQATVCARAQAPP------PSWDQMWKCLIR-L
DEN     YRLRGEARKTFVDLMRRG--DLPVWLAYKVAAEGINYADRRWCFDGTRNNQILEENVE-V
YEL-F   MRLRDDQRKVFRELVRNC--DLPVWLSWQVAKAGLKTNDRKWCFEGPEEHEILNDSGETV
JAP-E   YRLRGEEKKNFLELLRTA--DLPVWLAYKVASNGIQYTDRRWCFDGPRTNAILEDNTE-V
WEST-N  YRLRGEERKNFLEFLRTA--DLPVWLAYKVAAAGISYHDRKWCFDGPRTNTILEDNNE-V


HCV     KPTLHGPTPLLYRLGAVQNEVTLTHPITKYIMTCMS---
DEN     EIWTKEGERKKLKPRWLDARIYSDPLALK---EFAAGRK
YEL-F   KCRAPGGAKKPLRPRWCDERVSSDQSALSEFIKFAEGRR
JAP-E   EIVTRMGERKILKPRWLDARVYADHQALKWFKDFAAGKR
WEST-N  EVITKLGERKILRPRWADARVYSDHQALKSFKDFASGKR
```

Figure 2.9. Multiple sequence alignment among the 5 *Flaviviridae* Helicases. The four known motifs to the *Flaviviridae* helicases are colour coded and enclosed in boxes. Key residues to homology modelling are also highlighted.

The SCRs are obtained from the coordinates of the template and the SVRs are obtained from direct access of the algorithm to its built-in PDB database. The initial models were obtained from COMPOSER [169], a module of the Tripos Sybyl [170] suite running on a SGI workstation [171]. The primary sequence of each of the unknown structures was imported in COMPOSER and the structure of the corresponding HCV template protein was added in the database of the software. The database of COMPOSER at that time numbered 1850 structures. In order to assist the database of COMPOSER to become more efficient with the two helicases, a group of 125 helicases was downloaded from the PDB web site and were imported into COMPOSER's database. All loops were eliminated and the four models for the four *flaviviridae* virus members were built. To further optimise the models, the DNA from the template was added to the structures and the resulting complexes were initially energetically minimised and then subjected to short molecular dynamics simulations.

Energy minimisation was done in MOE (Molecular Operating Environment suite [172]) initially using the Amber forcefield implemented into the same package, up to a RMSd gradient of 0.0001. Molecular dynamics were performed after that at 300K, 1 atm for 1000 picoseconds with 2 fsecond step size, using the NVT ensemble in a canonical environment. NVT stands for Number of atoms, Volume and Temperature that remain constant throughout the calculation. The results of the molecular dynamics simulation were collected into a database by MOE and can be further analysed.

Dengue Helicase Model

West Nile Helicase Model

Jap. Encephalitis Helicase Model

Yellow Fever Helicase Model
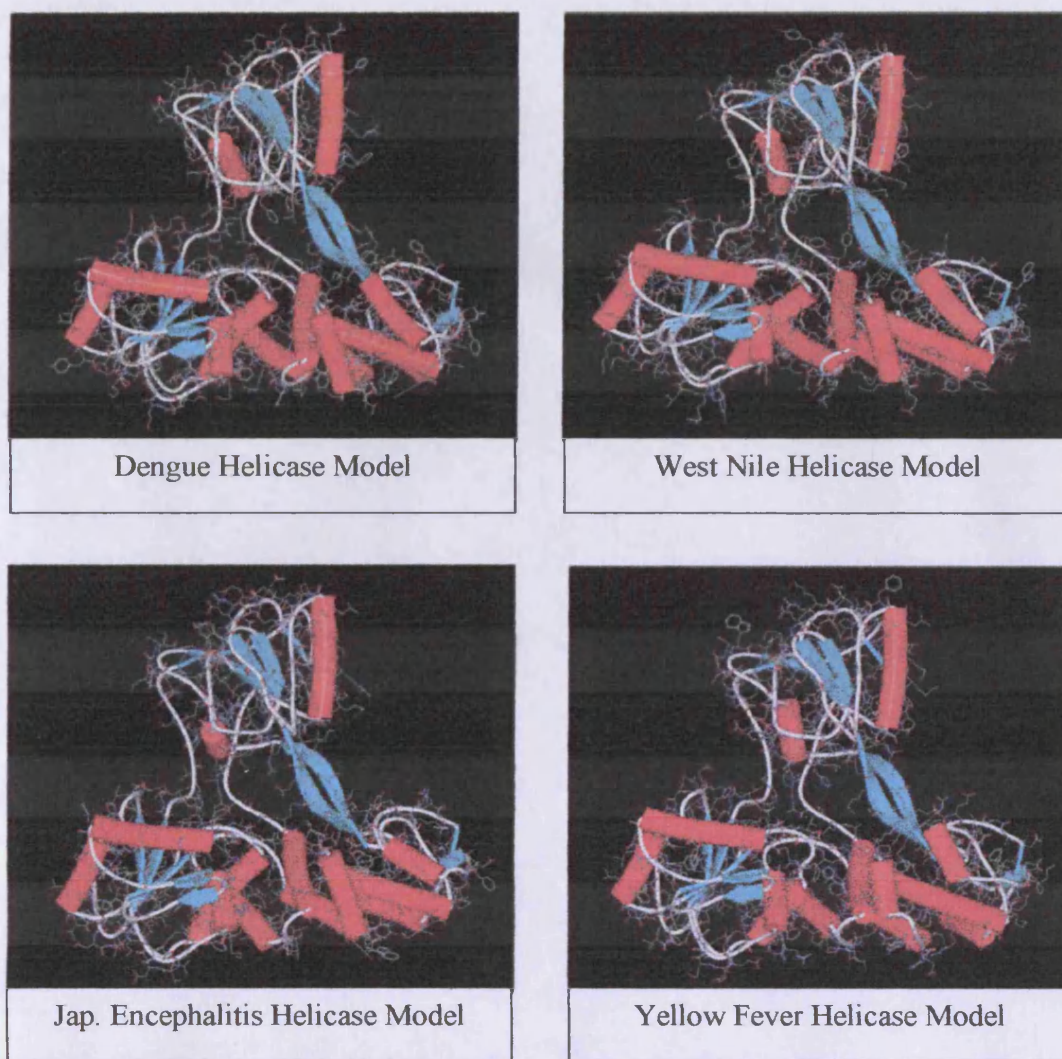
Figure 2.10. The four Helicase Models: the Dengue virus (top left), the West Nile virus (top right), the Japanese Encephalitis virus (bottom left) and the Yellow Fever virus (bottom right). The above conformations were obtained after the 1 nanosecond molecular dynamics simulation.

In order to assess and evaluate in more detail the quality of the four Helicase models they were initially superimposed with the HCV template and the coordinates of the ssRNA were copied to the models and subjected to another course of molecular dynamics for 1000 picoseconds with the oligonucleotide

present. This way the cooperativity between each model and the ssRNA would be

assessed. After the MD simulation, the ssRNA fragment of all four models did not

move significantly (low RMSD < 0.5 Å). Indicating that the ssRNA fragments

established interactions with each model, which stabilised them in the same

pattern as in their template.
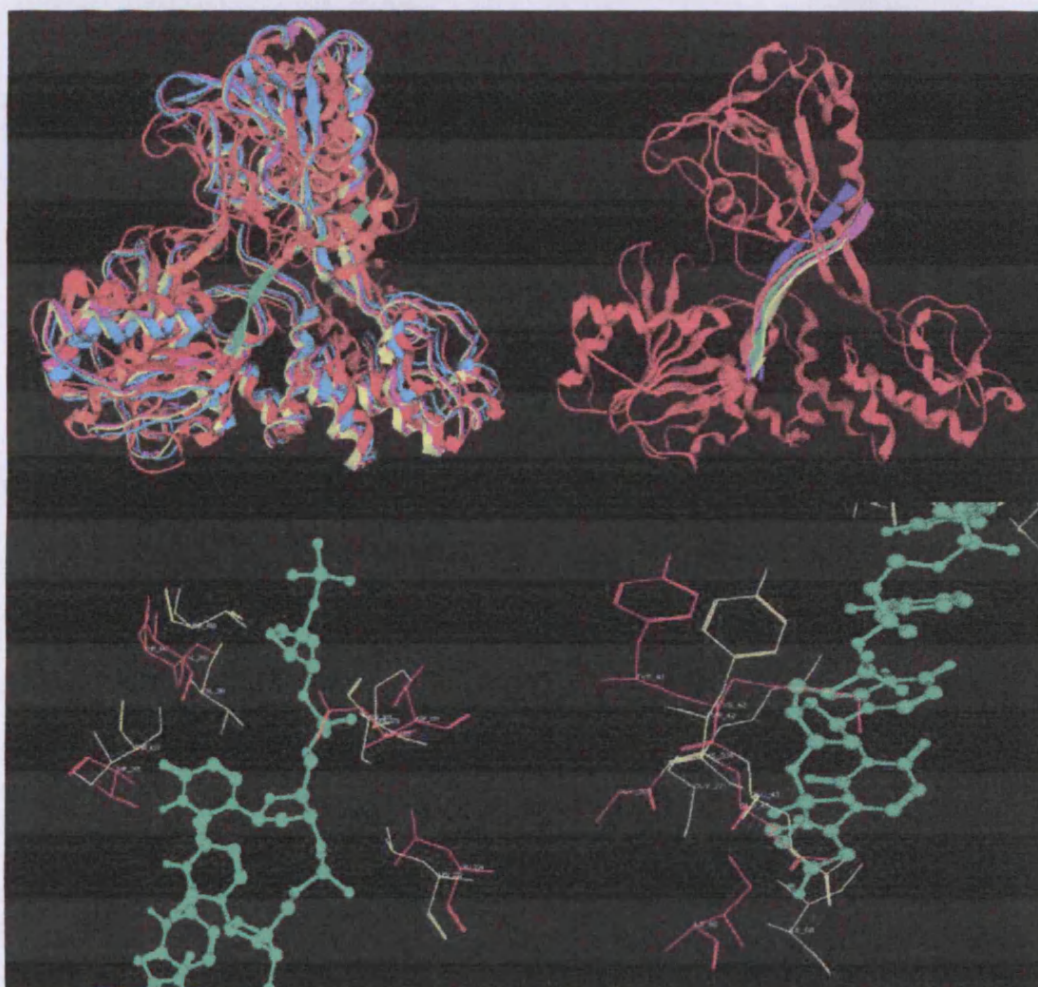
Figure 2.11. The superimposition of the final helicase models with the HCV +
ssRNA structure (top – left) and that of the ssRNA conformation in each of the
models upon MD simulation (top – right). Red is the HCV, in Yellow the Dengue,
in Blue the Yellow Fever, in Magenta the West Nile and in Orange the Japanese
encephalitis. Their corresponding RMSd (all atoms) with HCV are 1.9, 2.2, 1.1 and
1.4. In Green is the ssRNA from the HCV helicase. The ssRNA interacting residues
have been conserved from the template to the model (bottom – left/right).

2.2.2   Evaluation of the Helicase Models

In a homology modelling experiment the fact that structure is more conserved than sequence is always an assumption that has to be taken into consideration. Here, the structural elements of the models are not significantly different to those of HCV helicase. The quality of the models was tested with PROCHECK [173]. The templates 1A1V and 8OHM were found to have 100% and 99.7% of their residues in the allowed regions of the Ramachandran plot, respectively.

Ramachandran plot results for the Dengue model before minimisation and molecular dynamics showed that 82.2 % of its amino acids were in the allowed area of the Ramachandran plot (RP), 13.9% were in the additionally allowed regions and 2.5% were in the generously allowed regions, whereas 1.4% of the residues of the model were found to be in the disallowed area of the Ramachandran Plot (table 2.1). The Procheck evaluation was repeated after the molecular dynamics and in this case the percentage of residues in the allowed area of the RP was increased to 92.5%. Moreover 5.3% were in the additional allowed regions and 2.2% in the generously allowed area. There were no residues in the disallowed area (table 2.1).

The results for the accuracy of the model of the WNV showed that 80.9% of the model's residues were found in the core region of the RP, 13.6% were in the additional allowed regions, 3.7% of the residues were found in the generously

allowed area of the RP and 1.8% of the model's residues were located in the

disallowed region (table 2.1). After Molecular Dynamics the Ramachandran

statistics for the WNV model were 86.1% of residues in the core area of the RP,

11.3% of residues in the additional allowed areas, 2.6% in the generously area and

again the number of residues in the disallowed area was eliminated.

In the same way the Ramachandran statistics for the Japanese Encephalitis

helicase were improved from 79.9% of residues in the core region (rough model)

to 94.2% in the model after molecular dynamics. The percentage was also

decreased from 15.1% to 5.1% in the allowed regions, from 3.3% to 0.7% in the

generously allowed regions of the RP and from 1.6% to 0% in the disallowed

region.

Finally, the Ramachandran statistics for the Yellow Fever helicase were improved

from 81.5% of residues in the core region (initial model) to 96.7% in the model

after molecular dynamics. The percentage was also decreased from 13.4% to 3.2%

in the allowed regions, from 3.2% to 0.1% in the generously allowed regions of the

RP and from 1.9% to 0% in the disallowed region.

Table 2.1. The Ramachandran plot values of the each model after homology and
after optimization and re-evaluation.

Project	Model	After Homology				After Energy Minimisation & Model Optimisation			
		Core	Allowed	Generous	Disallowed	Core	Allowed	Generous	Disallowed
Helicase Project	DenV	82.2	13.9	2.5	1.4	92.5	5.3	2.2	0
	WNV	80.9	13.6	3.7	1.8	86.1	11.3	2.6	0
	JEV	79.9	15.1	3.3	1.6	94.2	5.1	0.7	0
	YF	81.5	13.4	3.2	1.9	96.7	3.2	0.1	0

2.2.3   Discussion of the Helicase Models

Ramachandran plot evaluation is not enough to confirm the practical quality of the model. Even if the geometry, phi, psi and omega angles of a model are perfect, the model could be unreliable if it does not satisfy the known functional criteria of the family of proteins it comes from. All of the known functional motifs of the *flaviviridae* helicases are present in the HCV template and have been inherited by all models (see sequence alignment - figure 2.9).

There are certain motifs conserved among the different helicases of the *flaviviridae* family.   The Hepatitis C, Dengue, West Nile, Yellow Fever and Japanese Encephalitis helicases belong to the superfamily II of helicases and share seven common motifs within their domains. RNA binds to the helicase at the Arginine-rich site of the 2nd domain. There are two supposed mechanisms of action for the *flaviviridae* helicases. The "sliding theory", suggests that the enzyme binds to the edge of the double stranded RNA of the virus and slides along, separating the two strands, by breaking all the hydrogen bonds between corresponding bases [174]. According to this theory the helicase is using its leading edge for unwinding. On the other hand another theory suggests that the helicase may be using its trailing edge [175]. Movement of the helicase along RNA takes place from 3' to 5' and is described by its periodicity (i.e. unwinding is not continuous, but occurs between brakes). The flexibility of the enzyme allows it to undergo significant conformational changes that are essential for its motion. The

energy requirement for the motion of the helicase is supplied by the hydrolysis of

ATP. The ATP binding site is located between domains I & II, where it initiates an

allosteric effect that will eventually change the conformation of the helicase thus

allowing it to move on the viral RNA. ATP must complex with a divalent cation in

order to produce its effect. The extreme importance of the helicase enzymes in the

survival of the viruses was demonstrated by mutagenesis experiments that led to

inactive helicases. No virus that was carrying an inactive helicase survived [175-176].
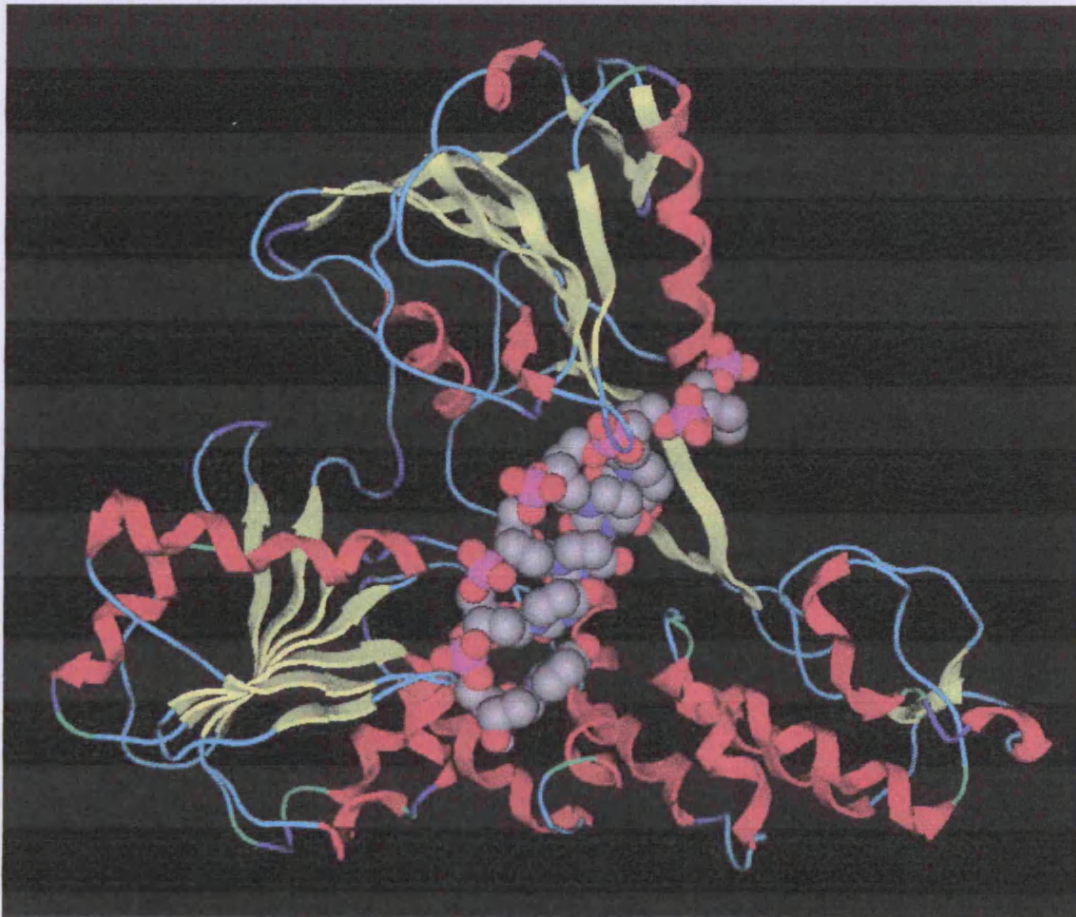
Figure 2.12. The HCV Helicase co-crystallised with a ssRNA fragment [177]

The *flaviviridae* helicases have three domains in total, which are separated by two channels (Figure 2.13). The first and third domains are interacting much more together than they do with domain two. The outcome of this is that the channel between domains 1-2 and 2-3 is larger than the channel between domains 1-3 and 2-3. Domain two is supposed to undergo significant movements compared to the other two domains, during the unwinding of double-stranded nucleic acids.
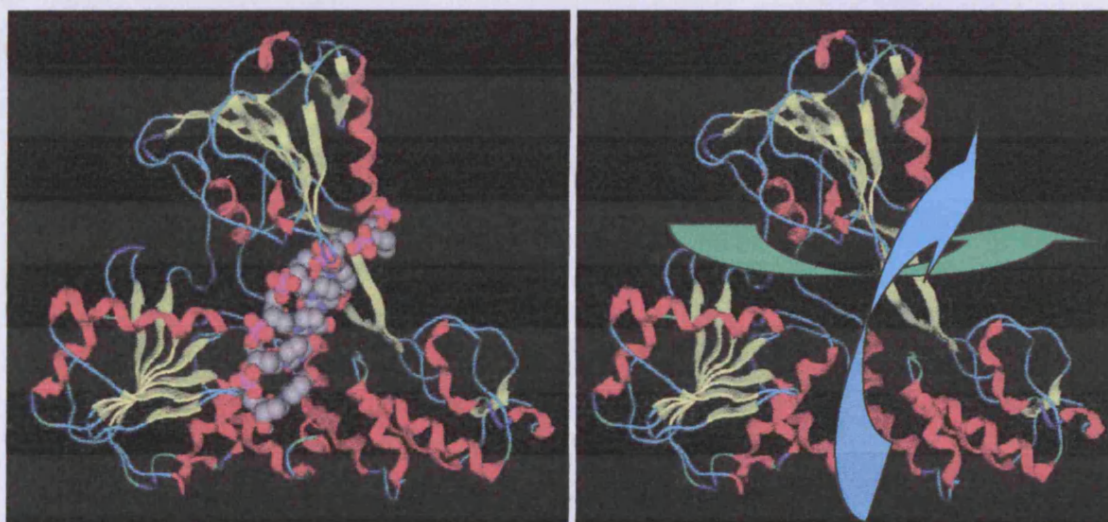
Figure 2.13. The NS3 Helicase domain from HCV complexes with ssRNA. There are two suggested channels for the ssRNA.

So, the positioning of domain two is very flexible relative to that of the other two domains. As a result the Helicase acquires the form of a dynamic "hinge" that moves according to the needs of the protein and the process it is involved in [177].

The topology of the first and the second domains is very similar. These two domains contain the structurally conserved regions of helicases of this family. This

is confirmed by the superimposition of the two domains, which gives an RMSd of

2.0 Å for 76 Cα atoms.

Domain 3 includes a 40 amino-acid long region, just before the helix in the C

terminus that does not contain any secondary structures [177]. This may contribute

to the flexibility required by the protein during its cleavage from the NS4A

domain during polyprotein processing. On the other hand towards the N-terminus

of the protein there is the highly conserved "Walker A box" or "P-loop" motif.

This motif is very often found among helicases and consists of a glycine rich

region of the protein that provides a quite flexible loop between strands and

helices [177]. The "Walker A box" has phosphate-binding properties and is found in

most ATPases. In the HCV helicase crystal structure the sulphate ion interacts

with the nitrogens Gly207 and Gly209, and the side-chains of Ser208, Lys210 and

Ser211. Lys210 establishes a H_2O mediated interaction with As290 of the DExH

motif (Asp-Glu-x-His). The position of the sulphate ion was found to be very

similar to the position that the β-phosphate of ADP would take in the PcrA

helicase–ADP complex. So, it is suggested that this is the space that β-phosphate

should take when NTP or nucleotide diphosphate (NDP) is bound to the HCV

helicase. The residues Gln460, Arg464 and Arg467 are highly conserved residues

from domain 2 that are exposed to solvent in the major channel of the Helicase.

Arg461 and Arg462 are buried amino-acids in the core of the 2nd domain [177].
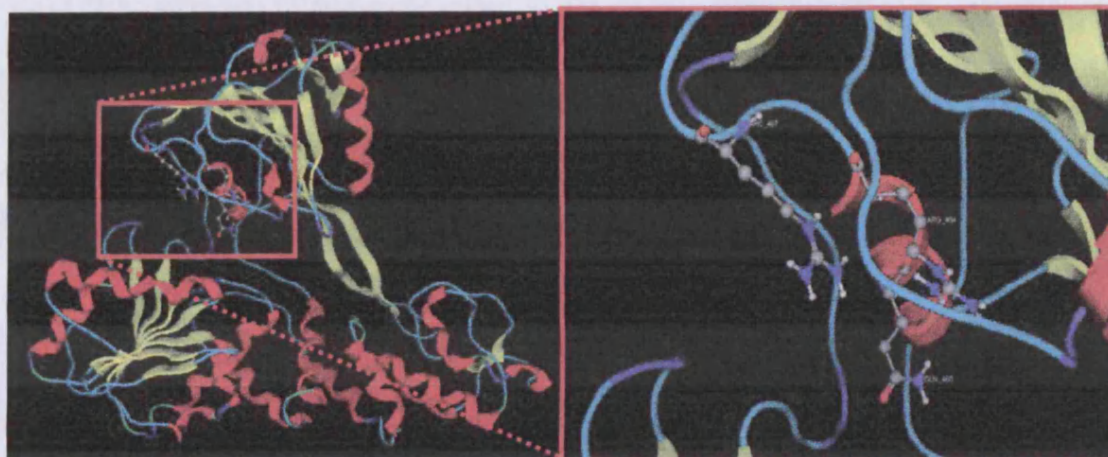
Figure 2.14. The residues Gln460, Arg464 and Arg467, which are highly conserved residues from domain 2.

The single strand of RNA is located in the main channel of the helicase between domains 3 and 1-2 (figure 2.15). The size of the channel is approximately 16 Å in diameter. The 5′ end of the oligonucleotide is towards the part of the channel between domains 2 and 3 and the 3′ end of the oligonucleotide is towards the part of the channel between domains 1 and 3 [178].The ssRNA and the Helicase interaction occur mainly between the backbone of the RNA, since it is a non-specific protein-RNA interaction. The majority of the established interactions are located towards the two ends of the ssRNA. Most interactions arise from regions lacking secondary structures in domains one and two. The positioning of the interaction-participating amino acids is symmetric and as a result it appears to be a symmetric
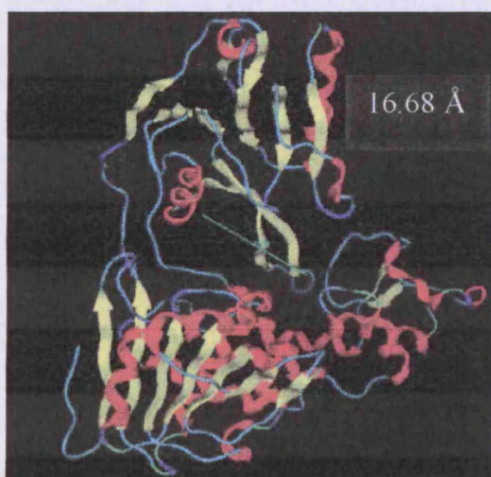
Figure2.15. The size available for the ssRNA in the HCV helicase is approx. 16 Å.

distribution of the interactions between the RNA and the protein. Quick superimposition of the first and the second domains revealed that the residues involved in the phosphate contacts are structurally equivalent. Furthermore the phosphate-binding amino acid series of Ser231, Thr269, Ser370 and Thr411 are conserved in NS3 domains and this is evidence that these two domains may be derived from a gene duplication event. Val432 and Trp501 are also highly conserved residues among HCV NS3 sequences, nevertheless neither seems to play any role in nucleic acid binding or duplex unwinding [177].

The second domain has two extended antiparallel strands (residues 430-452), which interact with the 5' of the oligonucleotide. This is also known as the L-45 loop and it belongs to the family of nucleic-acid binding motifs. The positioning of the domains that make the channel for the RNA in the helicase is very similar to that of the domains in the replication protein A (RPA) [177]. In both cases interactions are better formed towards the ends of the oligonucleotide and only minor interactions occur with the nucleotides in the middle.

There are no sequence-specific interactions with the RNA bases and the helicase. This was anticipated from the biological activity of the helicase and its behavior during enzymatic assays [177]. Any differences in the binding affinity between different nucleotides may be due to differences in RNA distortion and base stacking.

HCV strains appear to have very high sequence conservation among them. The percentage identity of the sequence alignments is calculated to be greater than

eighty percent. Comparison of the sequences of *Bacillus stearothermophilus* [179] and *Escherichia coli* DNA helicase [180] showed that there is an overall structure similarity between the domains 1A-2A of the above strains and 1-2 domains on the HepC helicase. The alignment of the primary amino-acid sequence may not be identical, but the alignment of the motifs is evidence that the arrangement of these different proteins in space and function must be similar to each other [181]. Site-specific mutagenesis has revealed that the function of the residues in these motifs of various helicases (including HepC) is crucial to the function of the helicases. Any change of these amino-acids will result in a protein mutant with lower affinity in unwinding dsRNA and dsDNA [182].

The fold of domain 1 is very similar structurally to that of common ATP transphosphorylases (for example adenylate and thymidine kinases).
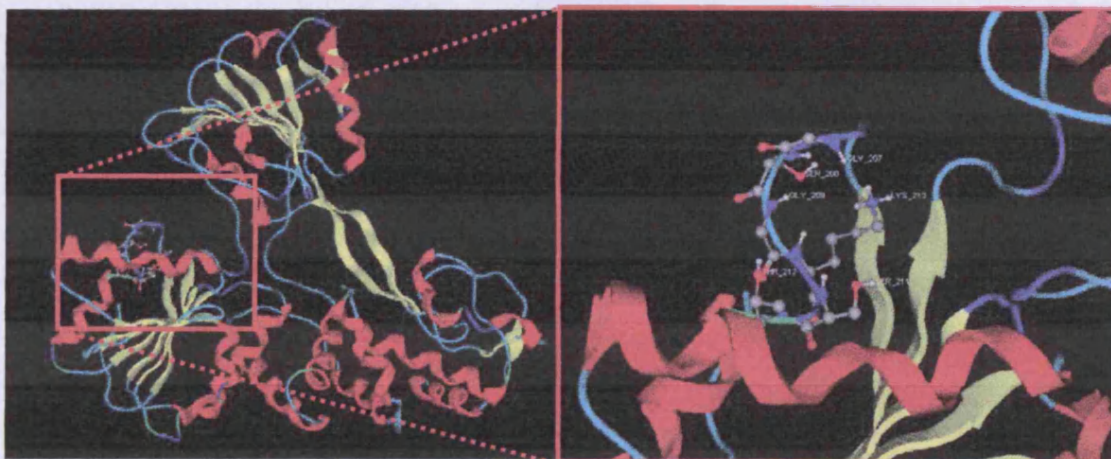
Figure 2.16. The GSGKST motif in domain 1 is conserved in kinases.

The GSGKST motif in domain 1 is conserved to the same loop in kinases, where its role involves binding of the β-phosphate of ATP [183]. Site mutagenesis studies of that motif have reported that the mutant protein is inactive.

Another crucial motif for the helicase is the DExH motif (motif II). The DExH motif is responsible for the binding of the Mg^{2+}–ATP substrate. Studies in adenylate and thymidine kinases revealed that an aspartate binds the Mg^{2+} and helps to establish the optimum orientation of ATP for nucleophilic attack [184]. Mutating this aspartate to any other amino acid will produce a helicase incapable of hydrolysing ATP. Also His293 is involved in the hydrolysis of ATP. Mutation of His293 will lead to an inactive helicase, but still capable of hydrolysing ATP. It is suggested that this residue plays a key role between the ATP hydrolysis and nucleotide binding process [185].
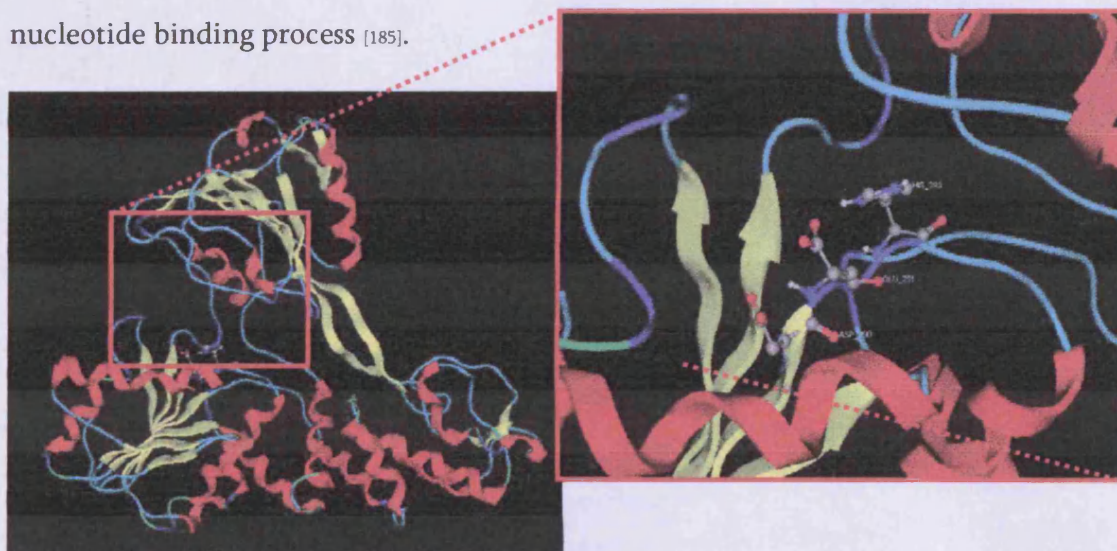
Figure 2.17. The DExH motif, which is responsible for the binding of the Mg^{2+}– ATP substrate.

The role of the QRxGRxGR motif is not clear yet, but this motif appears in most of the helicases of this family. Site mutagenesis studies revealed that mutation of this motif in vaccinia virus helicase will produce a helicase that posses much less ATPase activity [187].

Motif VI consists of three conserved arginines [188]. All three arginines are involved in the binding of ssRNA in the helicase's channel between domains 1 and 2. Kim suggested that Arg461 is hydrogen bonded to Asp412 and Asp412 interacts with ssRNA. So, there is an indirect importance of the Arg461, which is to keep the Asp412 in the correct orientation for the RNA to be able to interact with it. Mutagenesis studies of the arginines revealed that the resulting helicase has decreased RNA binding affinity [177].

Arg464 and Arg467 are expected to interact with ATP from mutation studies. Mutating the arginines to alanine or glutamine in vaccinia NPH-II or eIF-4A reduced the ATPase activity by 20% [189]. Arg467 is found to be conserved among all the superfamilies of helicases [190].

Motif III is located in-between the first and the second domains. Its role is to operate as a "hinge" offering the all important flexibility to the helicase protein [191]. The Ia motif constitutes part of the beta sheet in the 1st domain, while interacting with the oligonucleotide too. Motif V is also in contact with the oligonucleotide. Thr411 hydrogen bonds to a phosphate of an oligonucleotide. The usually conserved motif IV (from known helicases) is absent from the HepC helicase though.

There has been extensive work done on the motif IV of the superfamily I and II of helicases, but it was either done using weak criteria for the alignments or the HepC helicase evolved in a different fashion than the rest of the members of the two superfamilies. In the rest, of the DNA helicases of these superfamilies of helicases, motif IV is involved in the binding of ATP [192].

Residues in the HepC helicase that would be expected to constitute part of motif IV are the residues Ser370 and Lys371. Ser370 interacts with the oligonucleotide via a water-mediated hydrogen bond and Lys371 establishes a backbone interaction [193].

2.2.4   The Proposed function of the HCV Helicase

The conserved motif VI is found across the inner part of the channel in *flaviviridae* helicases and it extends into the ATP site. The same pattern is seen in the structures of the adenylate kinases [194]. In the group of adenylate kinases the conserved motif starts from the channel, where the oligonucleotide binds and extends into the ATP binding site. The purpose for this is that the binding of ATP or ATP analogues will have an immediate and direct allosteric effect on the helicase [195]. So, the energy of the hydrolysis of ATP is utilised directly for the purposes of the enzyme, without losing any of it in exchanges between different motifs or "loose connections" [196].

The binding of the ATP (or its analogues) will result in a conformational change in the structure of the enzyme. The most significant movements include the burial of the phosphates that were exposed to the solvent. Mutation studies on the ATP interacting residues, which lead to helicases incapable of ATP hydrolysis, have resulted in a wider configuration of the helicase molecule with much lower affinity for unwinding double stranded nucleotides [197].

Upon ATP binding, the domains one and two get together by the conformational change initiated originally in motif VI. This pattern would be expected to be common in all superfamily I and II members due to the consistency of the conserved motif VI [198].

The residues Gln460 and His293 belong to the motif VI and stand on opposite sites

in the oligonucleotide-binding channel. These two residues may be the ones responsible for regulating the equilibrium between the "tense" and "relaxed" states upon binding of the polynucleotide. The importance of these two amino acids was suspected by the fact that helicases with the DExH motif II often have a glutamine residue in motif VI, whereas helicases that have a DEAD motif often come with a histidine residue in this position [198].
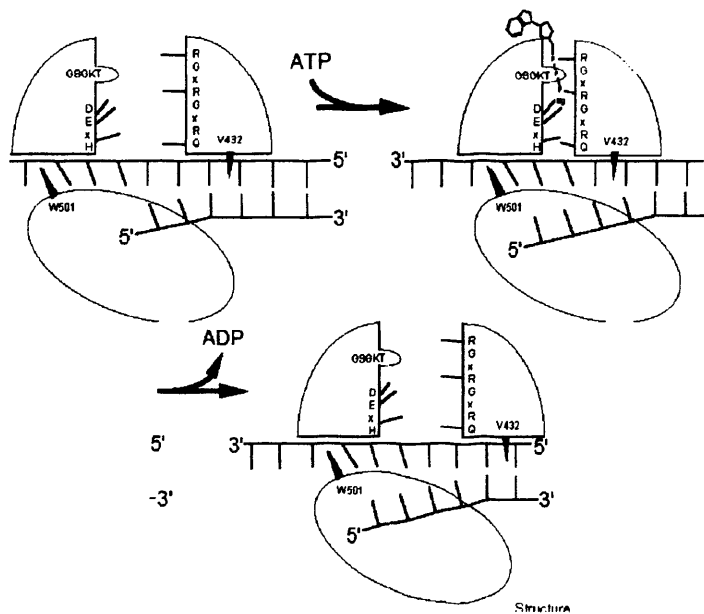
Figure 2.18. The schematic mechanism of RNA unwinding with the HepC Helicase. The binding of ATP will initiate the movement of domains one and two and the opening of the RNA binding channel, so that the ssRNA can translocate in the 5' to 3' direction [177].

The suspicion that the 2nd domain is flexibly linked to the rest of the protein was confirmed from the work of Yao *et al* [199]. The 2nd domain of the hepatitis C virus interacts directly with the polynucleotide. The movement of the above domain could influence a relative movement of the polynucleotide as well (compared to the rest of the protein). The residues Val432 and Thr448 may interact with the nucleotide bases at the 5' end of the single-stranded nucleic acid, and this can result in a movement of the polynucleotide from the 5' towards the 3' direction – during the closure of the 2nd domain [200]. Trp501 helps the single stranded nucleic acid to give up on its interaction with its surrounding residues. The positioning of Trp501 will only help the movement of the polynucleotide towards the 5' direction. The hydrolysis of ATP and the release of ADP will result in a slight opening of the polynucleotide channel and a coordinated movement of the whole domain 2 of the helicase towards the 5' direction [201].

So the hydrolysis of ATP will result in the movement of the polynucleotide relative to the rest of the helicase. Apparently, it has been found that helicases are capable of moving several bases from the polynucleotide per ATP molecule that is hydrolysed to ADP [202].

As mentioned before, the HepC helicase is very similar to other helicase (structurally), from superfamilies I and II [203]. So, instead of just sharing sequence similarities amongst them, the HCV helicase shares many conserved common motifs in its structure. There are two motifs that have been found to be extremely vital for the winding of the double stranded nucleic acids in the helicases [204]. The

first motif is the YRG*X*DV structurally conserved motif and the second one is the DFSLDPTF structurally conserved motif. The YRG*X*DV motif is the link between the IV and the V motifs in the superfamily II helicases. At the end of this loop lies the residue Arg393. Mutagenesis studies have shown that upon mutation of the residue of Arg393 to Ala, the recombinant protein is incapable of unwinding double stranded polynucleotides [205]. Combined with the fact that this arginine residue (Arg393) is fully exposed to the solvent and the nature of the arginine amino acid, makes it a very good target for drug design experiments in an attempt to design compounds that will inhibit the function of the HCV helicase. The affinity of RNA winding is not completely lost to the mutant helicase, but it is extensively reduced [206].

Motif DFSLDPTF is the link of the two anti-parallel beta sheets between the V and VI motifs. Mutagenesis studies for this motif involve the manipulation and mutation of the residue Phe444. Mutating Phe444 to Ala will produce a protein that is capable of hydrolysing ATP, but incapable of unwinding dsRNA. The affinity for unwinding dsDNA has dropped to half compared to the original helicase [207].

These two motifs are known as Arg-clamp and Phe-loop respectively among all helicases of HCV (including various genotypes and quasispecies). What makes these two motifs unique, is the fact that they only appear on HepC helicases and not on the other helicases of the superfamily I and II [208].
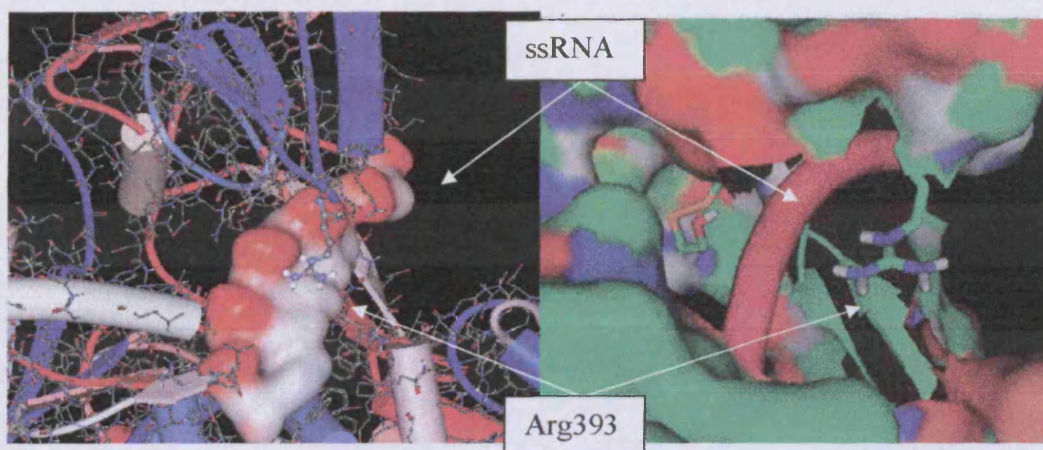
Figure 2.19. The HepC helicase with the Arg393 residue shown in stick representation. It can be seen that the function of this Arginine residue is to hold the oligonucleotide in place.

This characteristic gives these two motifs special value, since they constitute ideal drug targets for structure-based drug design. The position of the two key residues (Arg393 and Phe444) is shown in figure 2.19. The side chain of the residues Arg393 is only 2.5 Å away from the phosphate backbone of the RNA. On the other hand the distance of Phe444 is almost 15 Å away from the phosphate backbone of the RNA. It was not clear whether the RNA strand that is located in the channel of the HepC helicase (figure 2.19) was the translocating or the complementary strand. The interactions between the Arg393 residue and the oligonucleotide (figure 2.19) reveal that the binding is too tight and this implies that the strand in the channel of the helicase is probably the translocating one [209]. The helicase in figure 2.20 is in its ground-state configuration with the single-stranded oligonucleotide located in the major channel of the helicase protein. The oligonucleotide is picking many interactions from the protein and this does not seem to favour the unwinding process.
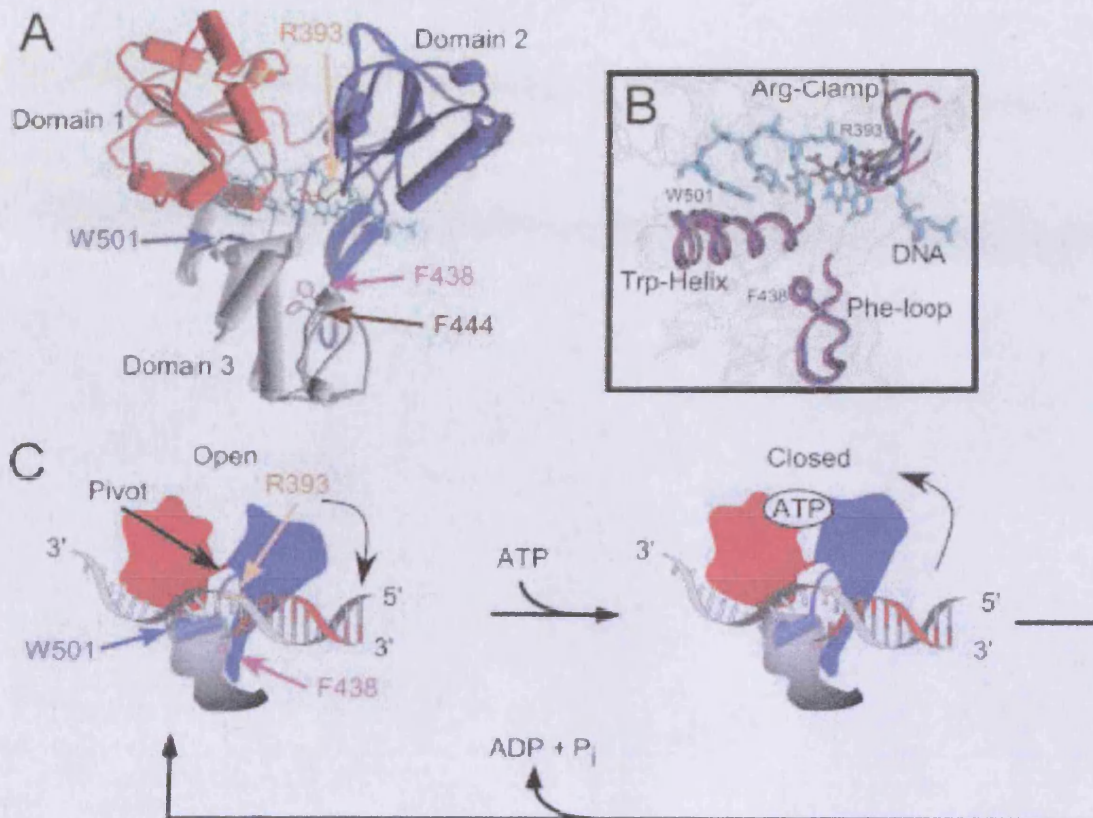
Figure 2.20. A and B show the locations of the Arg–clamp and the Phe loop in the helicase structure of HCV. C is another attempt to explain the mechanism of action of the HCV helicase (Domain 1 is red, domain 2 is blue and domain 3 is gray). When ATP is not present the helicase is interacting with the oligonucleotide so strongly that it is impossible for it to slide through. After ATP binding though, Domains 1 and 2 come closer and the channel between domains 1-2 and 3 becomes bigger and this allows the oligonucleotide to slide towards the 3' direction [210].

The problem is that the residues of the surrounding amino acids are too close to the ssRNA and this probably stabilises it in the protein [211]. The loosening of the interactions between the protein and the ssRNA is achieved by the contribution of ATP. ATP or any ATP analogues bind to the ATP binding site, which is on the same motif with the RNA binding site, the energy released by the hydrolysis to

ADP will initiate an allosteric effect in the helicase that will eventually result in the movement of the second domain and the loosening of the interaction with the oligonucleotide. This will last until the effect from the hydrolysis of the ATP is finished. Then the helicase will go back to its "closed" or "ground" state. To sum up it appears that offering energy to the helicase drives the protein to its "excited" state, and enables it to slide along the polynucleotide chain [212].

The structures of x-ray data of two helicases, one with the bound oligonucleotide and the other without it, were compared. Both structures came from the same HCV genotype and the sequence alignment proved that they are identical. It was proven that the channel of the helicase without the bound oligonucleotide is tighter than the one with the oligo. Moreover it was found that in the helicase with the bound oligonucleotide, Arg393 is further away from the ssRNA and the Trp501 is further too. This proves that there is a coordinated action and movement of amino acids in the helicase during the oligonucleotide binding with a general loosening of interactions during the sliding of the helicases [213].

It is demonstrated in figure 2.19 that the Arg393 residue interacts with the bound ssRNA molecule both when ATP is present and not. It has also been shown that the ΔG of the RNA binding to the helicase in the presence of ADP is weaker when Arg393 is absent (mutated to Ala). Focusing on residue 393 in the wild type helicase and in the mutated one (Arg393 to Ala393), the ΔG contribution of the 393 residue was found to be ten times stronger in the wild type one when the

Arg393 is present. That is another proof of the fact that the Arg393 is an excellent

target for drug design experiments [214].

The Phe loop is basically focused on the amino acids Phe438 and Phe444, but it

also involves the following four amino acids, which are located in close proximity:

His528, Phe531, Trp532 and Phe536. The vitality of the existence of these amino

acids is confirmed by mutagenesis studies on the helicase. The mutation of any of

these residues will lead to a mutant with no helicase unwinding capabilities [215].

Neighbouring amino acids should also be considered. For example Arg393 next to

the conserved Tyr392, which is thought to work in a similar function with Trp501

[216]. Tyr392 is thought to hold nucleotide chain from backsliding towards the 3'

direction. Mutating Tyr392 to Ala though does not change the unwinding

capabilities of the helicase at all [216].

Another vital amino acid that was found after examining the binding mode of

ssRNA in the HepC helicase is the Cys431. This amino acid was shown to have

bonded with a small molecule, probably from the crystallisation process. The

capability of this residue to interact with a molecule alien to the protein-ssRNA

complex is proof that Cys431 is a good target for drug design. Furthermore the

accessible area of the residue (to the solvent) is large (~75%) and the position of

the residue in the helicase's channel is strategic. As will be described later on, the

positions of residues Arg393 and Cys431 will be exploited by structure-based drug

design experiments.

2.2.5 *De novo* drug design for the HCV helicase

Cys431 has established a S-S bond with a mercaptoethanol (HS-CH$_2$-CH$_2$-OH) in the HCV helicase x-ray structure. That can only mean that the Cys431 is accessible to the solvent, even when ssRNA is present, and could potentially establish interactions with a future inhibitor. Cys431 is located in a very strategic position for the blocking of the passage of the ssRNA through the helicase (figure 2.21).
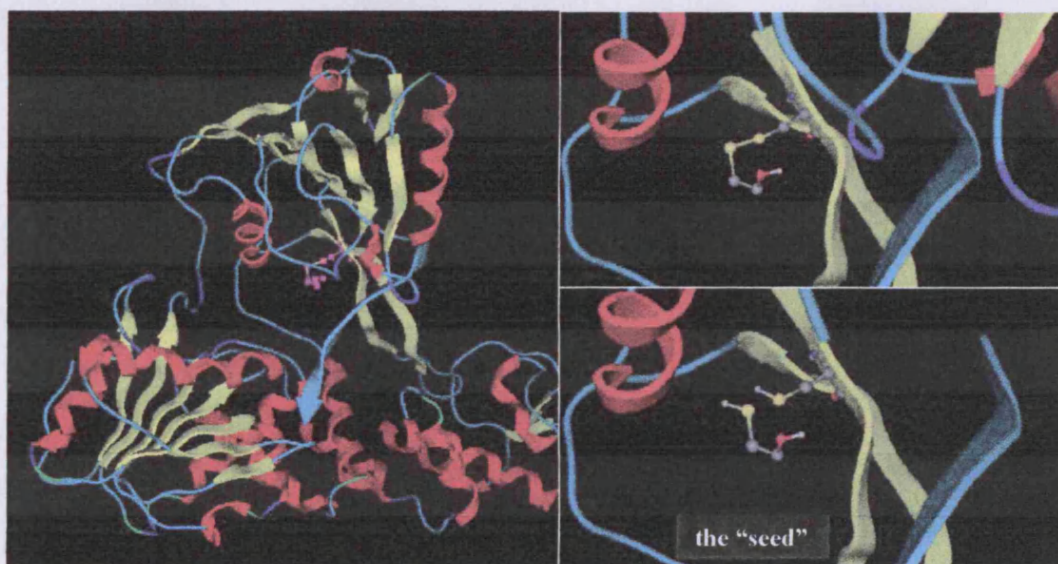
Figure 2.21. Choosing the right starting point for the growing algorithm is a very crucial step. Here the seed was included in the PDB file (1A1V).

Further examination of the helicase for exposed to the solvent residues revealed two arginine residues (Arg393 and Arg481), positioned in such a way that the ssRNA was crossing the space between them. These residues were set to define a possible active site of the helicase that could be targeted for the design of novel inhibitors: the arginine residues could establish H-bonds, whereas the Cys431

residue could possibly react with the compound, maybe establishing a S-S or a Hydrogen bond. Ideally, such an inhibitor would interact with the two Arg393, Arg481 and the Cys431, thus forming a bridge in the middle of the RNA channel in the helicase. If the compound covalently bonds to the receptor, then it is expected to be strong enough to block the passage of the ssRNA thus inhibiting the helicase.

A variety of different seeds was tested, the most suitable one was a small compound attached to the Cys431. First the S-S bond was broken and missing hydrogens were restored on both sulphur atoms (one of the compound and the other of the Cys431 residue). Then the oxygen was removed, since its existence would significantly reduce the number of fragments suitable for that seed-receptor arrangement. The remaining compound was used as a starting point for the "growing" algorithm of LigBuilder.
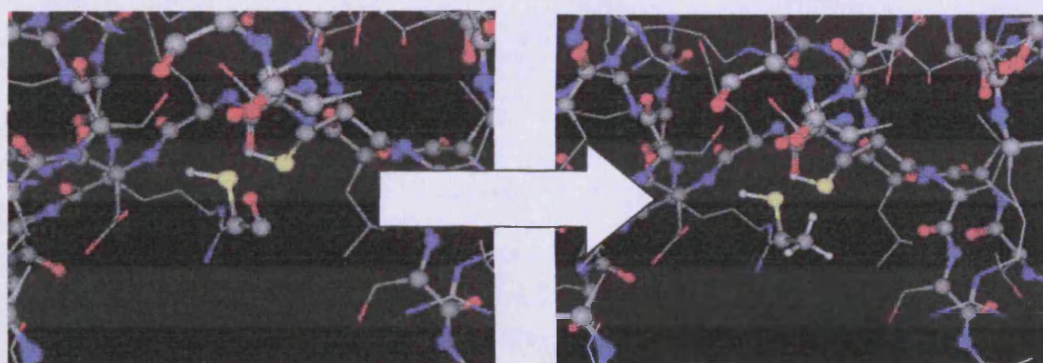
Figure 2.22. The co-crystallised fragment was firstly released from Cys431 by breaking the S-S bond and then the β-mercaptoethanol OH was removed, since an –OH group would dramatically limit the diversity of the generated compounds.

The complex was energetically minimised using a molecular mechanics algorithm, having fixed the backbone of the protein. The detached compound was entered as the starting point of the drug design algorithm and thus it was expected that this moiety would be present in this position on all the new compounds.
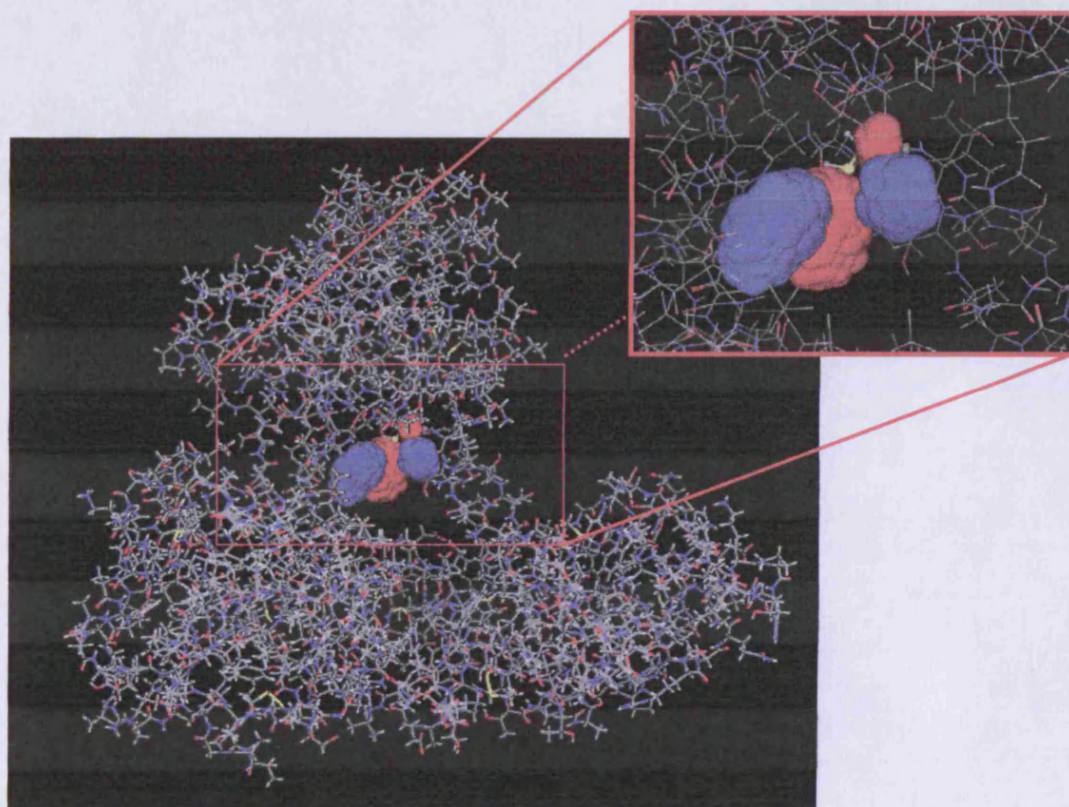
Figure 2.23. The Active site analysis of LigBuilder prior to the analysis of the seed. The receptor is examined and probable (predicted) electron-donating regions are coloured blue whereas probable electron-accepting regions are coloured red. In the next steps the algorithm will position the "seed" in space and depending on the area that it falls in, the algorithm will attempt to combine the suitable fragments from its database, trying to satisfy all user and built-in criteria.

The algorithm used the small compound as a starting point and started to grow structures by combining different chemical fragments that it stores in its database. The criteria are to optimally utilise the available space of the receptor and to

establish the maximum amount of interactions with the adjacent residues of the

helicase (Figure 2.23). All the different compounds that were designed were

deposited in a folder for further investigation. A similarity cut-off of 90% was used

in order to make sure that structurally all the different compounds in that folder

would be at least 90% different. The space available (Figure 2.24) was filled with

newly designed compounds, with the only size-limiting parameter being the pre-

defined molecular weight of the compound. For all drug design experiments the

Lipinsky's drug-likeness rules were applied [217].
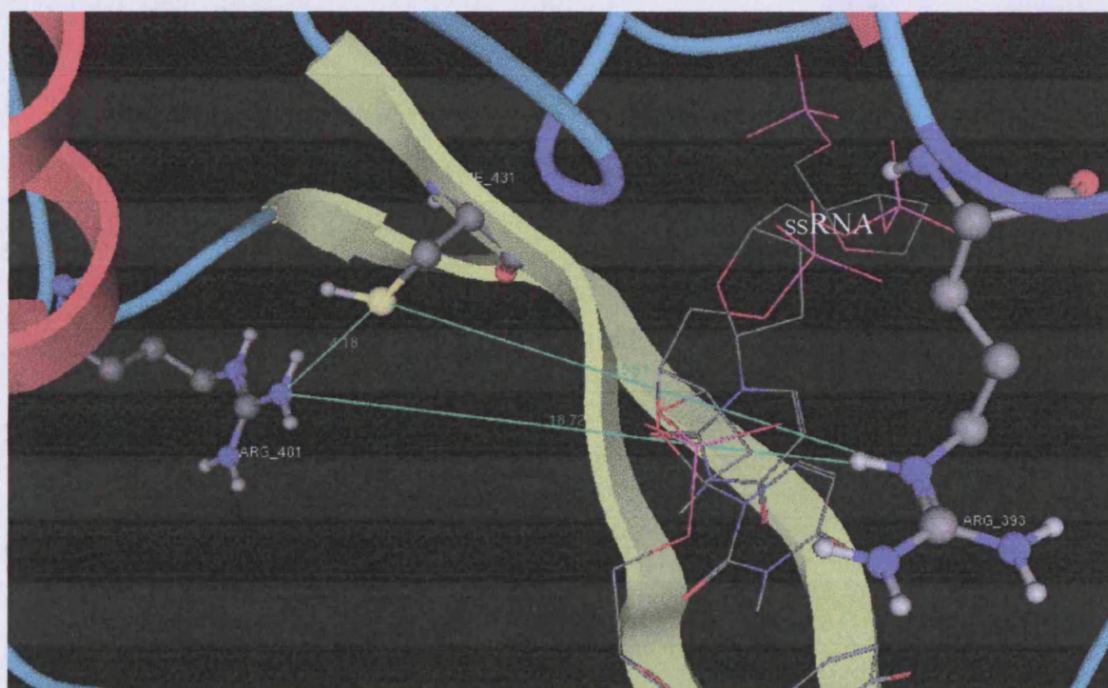
Figure 2.24. Distances and the available space in the Helicase's active site

After many iterations of the genetic algorithm of LigBuilder a possible lead was

obtained, which had all the moieties that could interact with the given active site

on the helicase and met all algorithm's criteria, but, unfortunately, it cannot be

considered a suitable drug. The lead compound from LigBuilder had 16 chiral

centres and would not be feasible to synthesise (Figure 2.25).
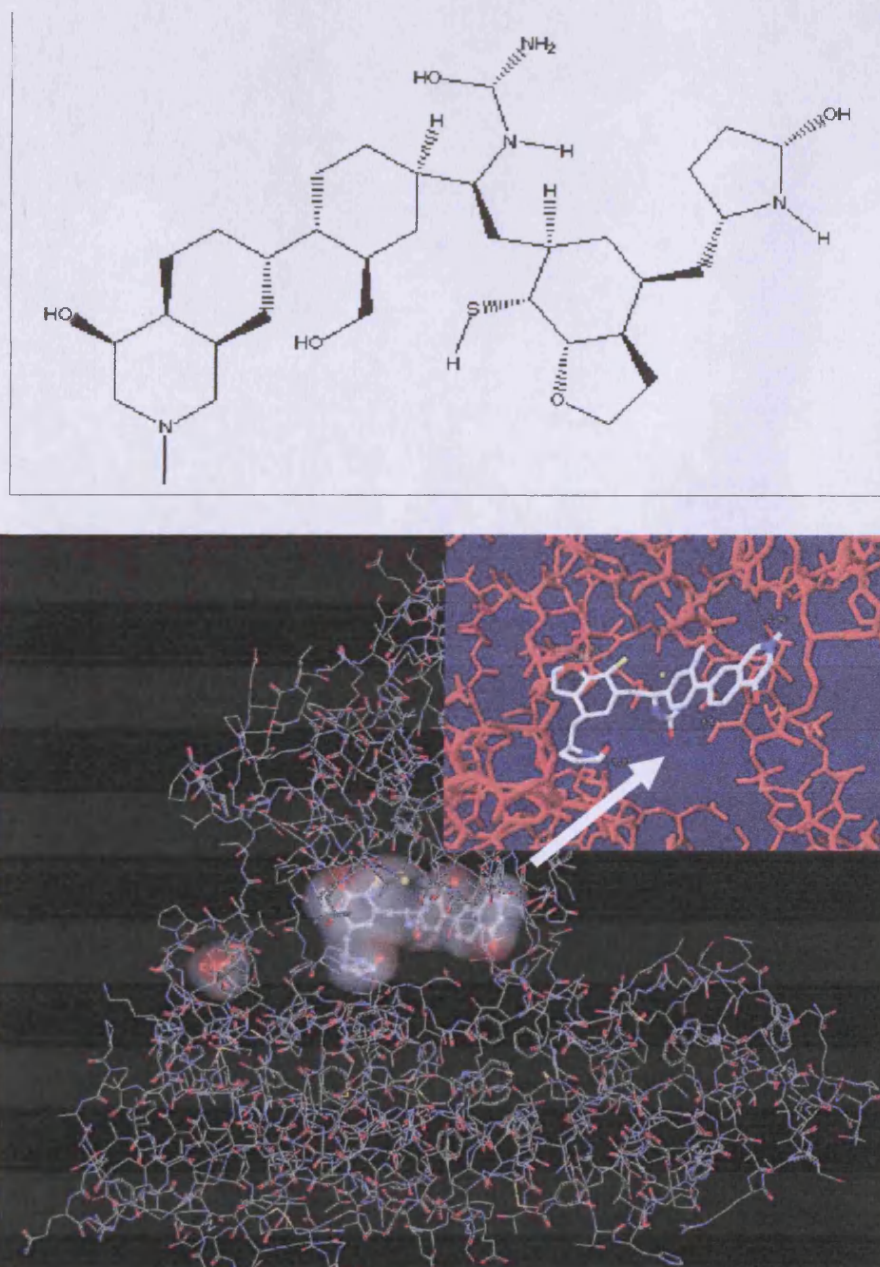
Figure 2.25. The lead compound generated by LigBuilder docked into the helicase

From the observation of the ligand/protein complex it is evident that the computer program tried to fill all available space and, in order to obtain useful results, we had to force the software to use only the space defined by the three target residues. That was achieved by physically incorporating a tube-like structure that encloses all key residues and restricts considerably the search space (figure 2.26). The tube was originally made from carbon atoms and the PDB file [218] was edited converting the carbon atoms to "Dummy" atoms (Du). Du atoms do not have any atomic properties, they are incapable of establishing interactions of any nature and they have no charge. In this way the tube acts as an inert wall, limiting noticeably the search space.
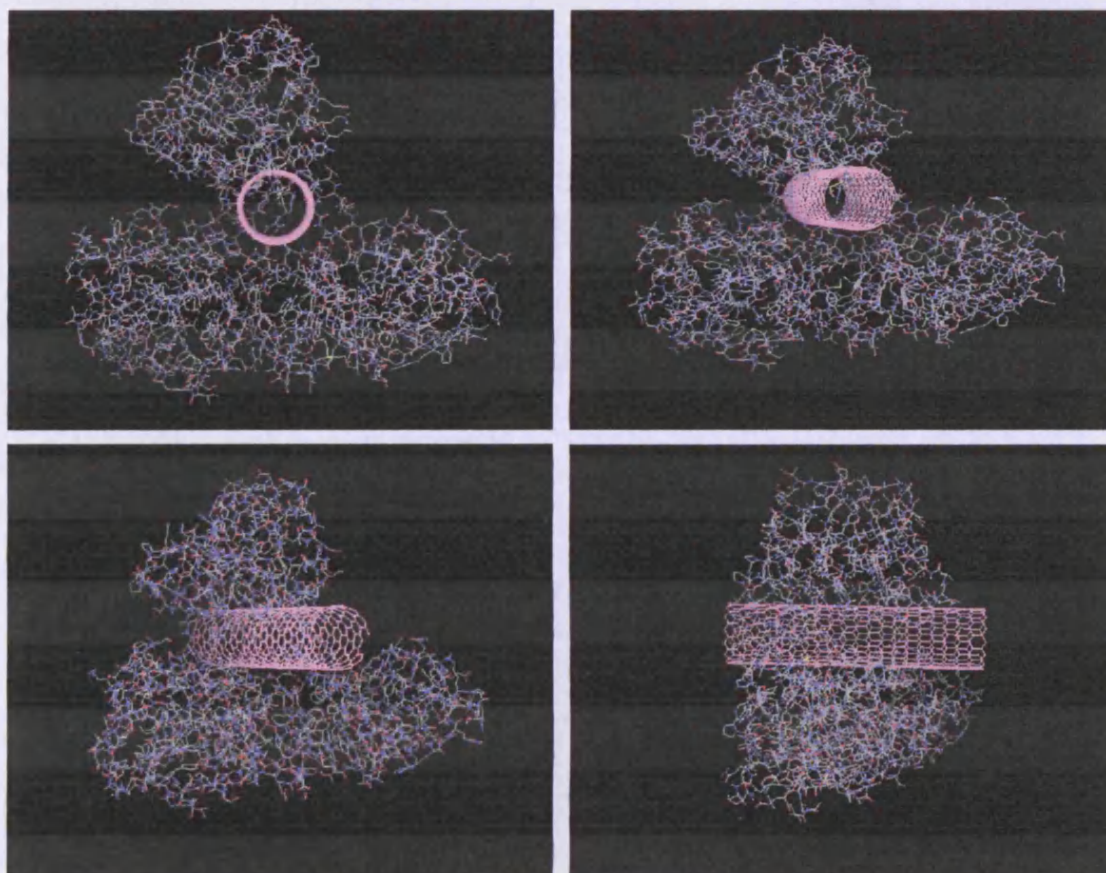
Figure 2.26. The application of the tube simplified the task of LigBuilder.

The structures that LigBuilder generated (Figure 2.27) proved to be considerably simpler and more drug-like. The novel structure was isolated and then docked again into the original receptor, where it originated from. Only this time the "tube" was not present and the full receptor was available to be explored by the docking algorithm.

The result of the docking confirmed that this compound, that had been specifically designed for the particular area (between Arg481-Cys431-Arg393) on the helicase found its way to the suggested site and managed to establish the interactions that it was supposed to (figure 2.27). The second best lead from LigBuilder was the same compound with an extra CH₃ substitution on one of the phenyl rings. That CH₃ was able to interact with the sulphur of the Cys431 and further stabilise the docking.

It is obvious that the presence of the extra carbon has pushed the compound a bit lower, which is evidence that the extra carbon successfully established an interaction with the nearby available sulphur from the Cys431 residue.

Following these promising results, we decide to use this structure as our basic scaffold for the design of a potential inhibitor capable of reacting with the desired cysteine (figure 2.27). To achieve this, our initial step was to include a chemical moiety able to react with the sulphur atom of Cys431. Michael acceptors, like the vinyl ketone, are known to react quickly with thiols and we decide to include this moiety in our compound replacing one of the acid groups (Figure 2.27b). The

ketone should be still able to form the hydrogen bond with Arg481 while the

cysteine should be close enough to react with the double bond.
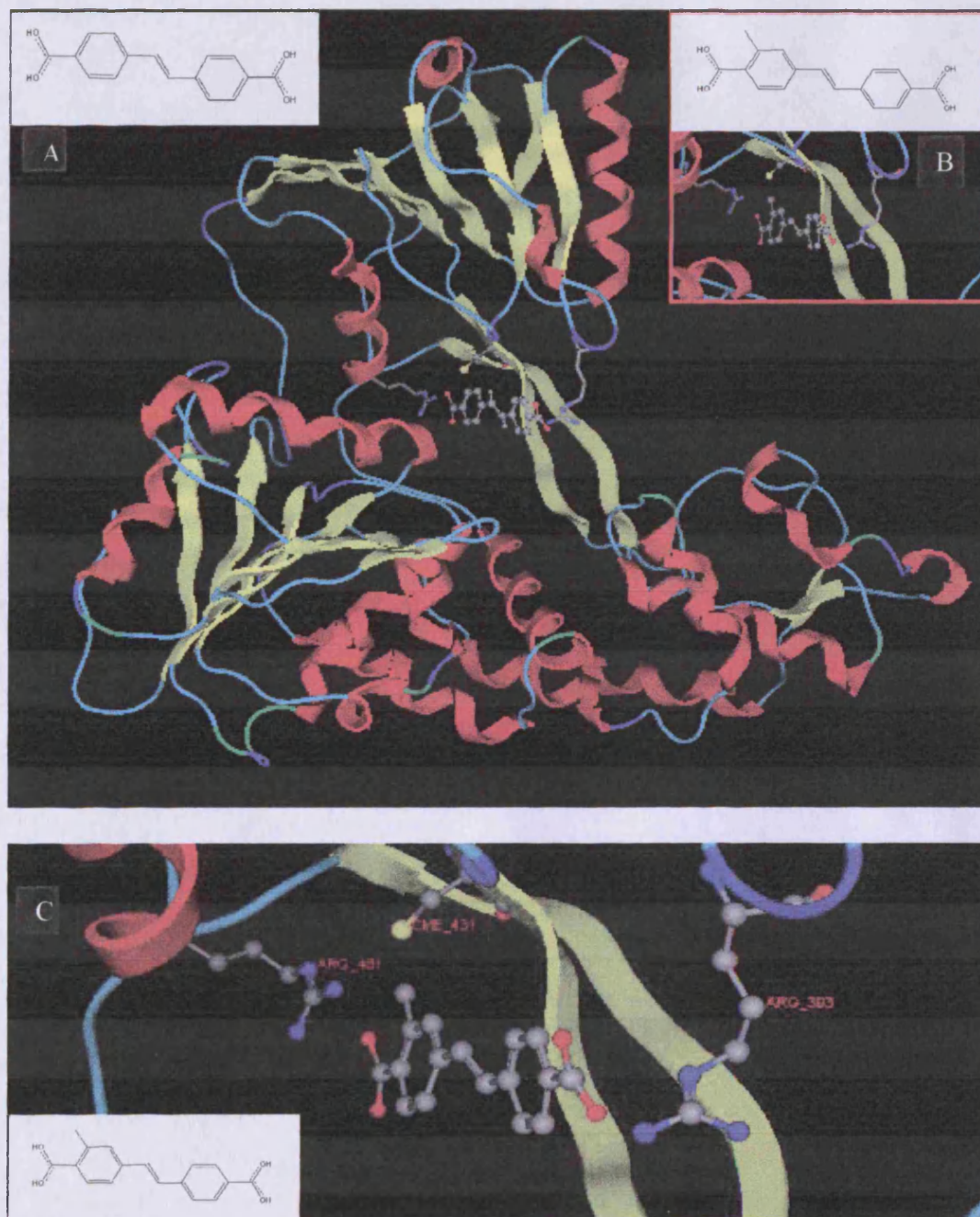
Figure 2.27.   The new lead compounds using LigBuilder and the "tube"

We were not able to dock successfully compound B (Figure 2.28) into the active

site, simply because the structure was now too big to fit between the two

arginines, but a simple change in the position of the vinyl ketone to the meta

position led to compound C which was able to dock in the desired position.

Although the obtained molecule has all the desired properties, we have further

modified the structure by replacing one of the carbon atoms of the linking double

bond with a nitrogen atom, to simplify the synthetic procedure (Figure 2.29).
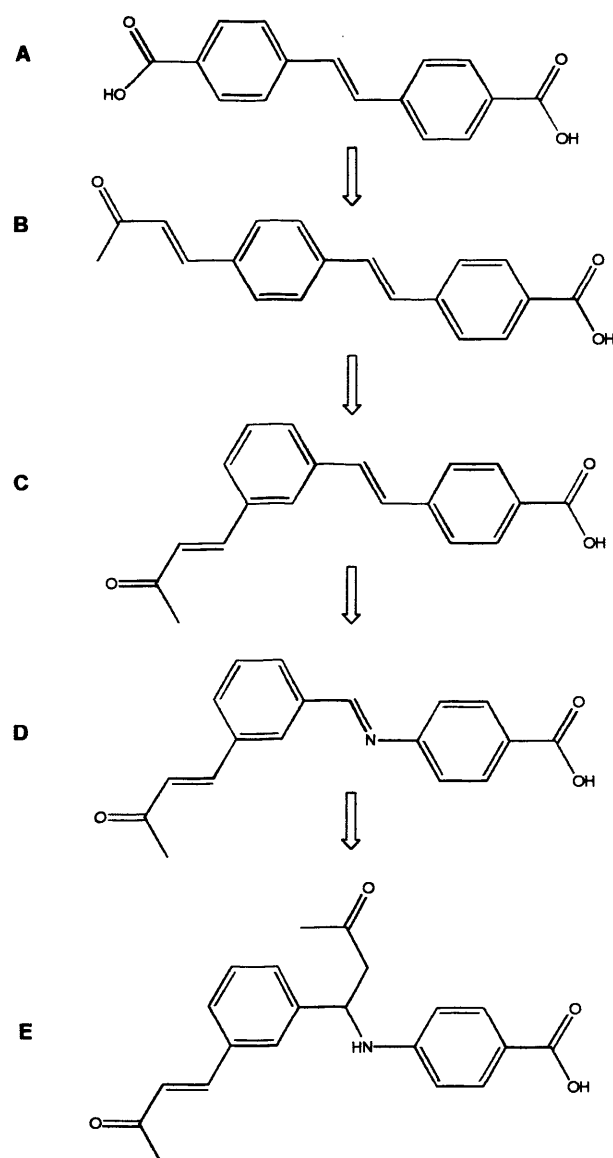
Figure 2.28. Route followed for the design of the final compound.

Compound D can be further functionalised by adding a side chain that could stabilise the ligand / protein complex by establishing new interactions with the enzyme. Those interactions mainly involve new hydrogen bonds, possibly with Cys431. It is important to mention that the last two modifications were "chemistry driven" and an already established synthetic methodology could be applied to obtain compound E in 3 steps [219-221]. (figure 2.29).
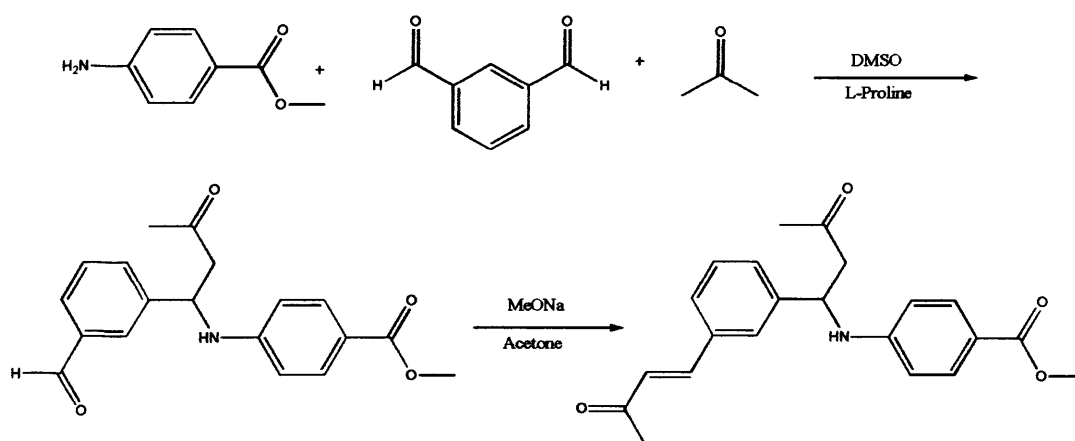
Figure 2.29. Proposed synthetic scheme.

Observing the minimised docking result of the compound E/helicase complex (figure 2.30), we can appreciate how the carboxylic acid moiety nicely interacts with Arg393 while the ketone forms a hydrogen bond with Arg482. The sulphur atom is just 4Å away from the carbon atom and it should react with it, since it is correctly placed for a nucleophilic attack to the double bond of the vinyl ketone moiety (Figure 2.30).
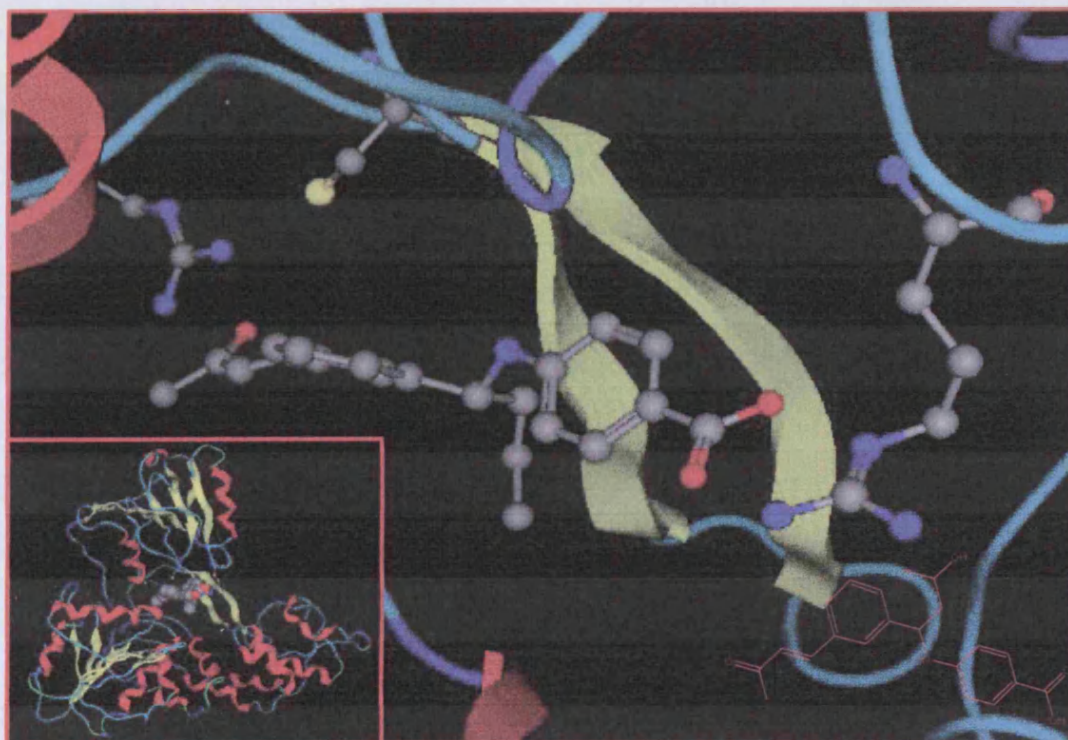
Figure 2.30. The combined lead from all different LigBuilder Experiments.

In conclusion, we have designed a series of compounds (figure 2.28) as potential inhibitors of the HCV helicase. These molecules have all the desired properties: they interact with the proposed key residues (Arg393, Cys431 and Arg481) and their structural simplicity should allow a simple synthesis. At this point, all compounds in figure 2.28 are currently being prepared by other members of the group.

2.2.6   A Comparison between two X-ray Resolved Polymerases

The HCV and BVDV polymerases have both been resolved by x-ray crystallography, and their 3D coordinates have been deposited at the PDB web site. Initially, the obtaining homology models for the Dengue, West Nile, Japanese Encephalitis and Yellow Fever viruses seemed impossible due to the low sequence identity between the available templates and the model sequences, which is approximately 17% with both HCV and BVDV as templates (Figure 2.31). Generally, such low identity does not allow homology modelling but considering that our aim is to design models that could be used later on as templates for the drug design experiments, the overall accuracy of the generated structures is not as significant as the local reliability of the RNA-binding site, which is going to be used as the target area for the inhibitor design experiments. Furthermore, structural comparison between the HCV and BVDV polymerase proteins revealed that even though their overall sequence alignment score is very low (11% homology), the area around the co-crystallised oligonucleotide is very much conserved (figure 2.32). Structural superimposition between the polymerases of HCV and BVDV reveals that structure is more conserved than sequence. This also means that even though the overall alignment score is not satisfactory, the HCV and BVDV could be used as templates for the rest of the *Flaviviridae* models focusing on an accurate modelling of the active site area and the conservation of all known structural motifs vital for the function of the *Flaviviridae* polymerases.

```
HCV    ---MV--YATTSRSASLRQKKVTFDRLQ-VLDDHYRDVLKE-MKAKASTVKAKL--LSIE
BVDV   KLPIV--RAQTDTKTF---HEAIRDKIDKSENRQNPELHNKLLEIFHTIAQPTL--KHTY
DEN    VVPMVTQMAMTDTTPF-GQQRVFKEKVD-TRTQEPKEGTKKLMKITAEWLWKELGKKKTP
WNL    TITNVTTMAMTDTTPF-GQQRVFKEKVD-TKAPEPPEGVKYVLNETTNWLWAFLAREKRP
YEF    KIEEVTRMAMTDTTPF-GQQRVFKEKVD-TRAKDPPAGTRKIMKVVNRWLFRHLAREKNP
JEN    AIANVTTMAMTDTTPF-GQQRVFKEKVD-TKAPEPPAGAKEVLNETTNWLWAHLSREKRP

HCV    EAC---KLTPPHSAKSKFG--YGAKDVRNLSSRAVNHIRSVWEDLLEDTE-------TPI
BVDV   GEVTWEQLEAGVNRKGAAG--FLEK---KNIGEVLDSEKHLVEQLVRDLKAGRKIKYYET
DEN    RMCTREEFTKKVRSNAALGAIFTDENKWKSAREAVEDSR-FWELVDKERNLHLEGKC-ET
WNL    RMCSREEFIRKVNSNAALGAMFEEQNQWRSAREAVEDPK-FWEMVDEEREAHLRGEC-HT
YEF    RLCTKEEFIAKVRSHAAIGAYLEEQEQWKTANEAVQDPK-FWELVDEERKLHQQGRC-RT
JEN    RLCTKEEFIKKVNSNAALGAVFAEQNQWSTAREAVDDPR-FWEMVDEERENHLRGEC-HT

HCV    DTT-IMAKSEVFG---VQPEKGGRKPARLIVFPDLGVRVC-EKMALYDVVS--TLEQAVM
BVDV   A-IHKNEKRDVSDDWQAG-DLVVEKRPRVIQYPEAKTRLAITKVMYNWVK-----QQPVV
DEN    CVYNMMGKREKKL---GEF--GKAKGSRAIWYMWLGARFL-EFEALGFLNEDHWFSRENS
WNL    CIYNMMGKREKKF---GEF--GKAKGSRAIWFMWLGARFL-EFEALGFLNEDHWLGRKNS
YEF    CVYNMMGKREKKL---SEF--GKAKGSRAIWYMWLGARYL-EFEALGFLNEDHWASRENS
JEN    CIYNMMGKREKKF---GEF--GKAKGSRAIWFMWLGARYL-EFEALGFLNEDHWLSRENS

HCV    GSSYGFQYS-PKQRV-EFLVNTWKSKKCPMGFSYDTRCFDSTVTESDIRVEESI-YQCCD
BVDV   IP--GYEGKTPLFNIFDKVRKEWDSFNEPVAVSFDTKAWDTQVTSKDLQLIGEI--QKYY
DEN    LS--GVEGE-GLHKL-GYILREVSKKEGGAMYADDTAGWDTRITIEDLKNEEMITNHMAG
WNL    GG--GVEGL-GLQKL-GYILREVGIRPGGKIYADDTAGWDTRITRADLENEAKVLELLDG
YEF    GG--GVEGI-GLQYL-GYVIRDLAAMDGGGFYADDTAGWDTRITEADLDDEQEILNYMSP
JEN    GG--GVEGS-GVQKL-GYILRDIAGKQGGKMYADDTAGWDTRITRTDLENEAKVLELLDG

HCV    LAPEARQAIRSLTER-LYIGGPLTNSKG----QNCGYRRCRASGVLTTSCGNTLT---CY
BVDV   YKKEWHKFIDTITDH--MTEVPVITADG----EVYIRNGQRGSGQPDTSAGNSMLNVLTM
DEN    EHKKLAEAIFKLTYQNKVVRVQRPTPRG-TVMDIISRRDQRGSGQVGTYGLNTFTNMEAQ
WNL    EHRRLARAIIELTYRHKVVKVMRPAADGRTVMDVISREDQRGSGQVVTYALNTFTNLAVQ
YEF    HHKKLAQAVMEMTYKNKVVKVLRPAPGGKAYMDVISRRDQRGSGQVVTYALNTITNLKVQ
JEN    EHRMLARAIIELTYRHKVVKVMRPAAEGKTVMDVISREDQRGSGQVVTYALNTFTNIAVL

HCV    LKATAACRA-----------------------AKLQDCTMLVNGDDLVVICESAG-TQ
BVDV   MYAFCESTGVPYK--------------------SFNRVARIHVCGDDGFLITEKGLGLK
DEN    LIRQMEGEG-IFK-SIQHLTASEEIAVQDWLARVGRERLSRMAISGDDCVVKP-LDD---
WNL    LVRMMEGEGVIGPDDVEKLTKGKGPKVRTWLFENGEERLSRMAVSGDDCVVKP-LDD---
YEF    LIRMAEAEMVIHHQHVQDCDESVLTRLEAWLTEHGCNRLRRMAVSGDDCVVRP-IDD---
JEN    LVRLMEAEGVIGPQHLEQLPRKTKIAVRTWLFENGEERVTRMAISGDDCVVKP-LDD---

HCV    EDAAALRAFTEA-MTRYSAPPGDPPQPEYDLELITSCSSNVSVAHDASGKRVYYLTRDPT
BVDV   FANKGMQILHEAGKPQK-ITEGEKMKVAYRFEDIEFCSHTPVPVRWSDNTSSHMAGRDTA
DEN    RFARALTALNDMGKVRKDIQQWEPSRGWNDWTQVPFCSHHFHELIMKDGRTLVVPCRNQD
WNL    RFATSLHFLNAMSKVRKDIQEWKPSTGWYDWQQVPFCSNHFTELIMKDGRTLVVPCRGQD
YEF    RFGLALSHLNAMSKVRKDISEWQPSKGWNDWENVPFCSHHFHELQLKDGRRIVVPCREQD
JEN    RFATALHFLNAMSKVRKDIQEWKPSHGWHDWQQVPFCSNHFQSIVMFDGRSIVVPCRGQD

HCV    TPLARAAWETARHTPINSWLGNIIMYAPTLWARMILMTHFFSILLAQEQLEKALDCQIYG
BVDV   VILSKMATRLDSSGE-RGTTAYEKAVA-FSFL--LMYSWNPLVR----RI-----CLLVL
DEN    ELIGRARISQGAGWSLRETACLGKSYA-QMWS--LMYFHRRDLRLAANAI-----CSAVP
WNL    ELVGRARISPGAGWNVRDTACLAKSYA-QMWL--LLYFHRRDLRLMANAI-----CSAVP
YEF    ELIGRGRVSPGNGWMIKETACLSKAYA-NMWS--LMYFHKRDMRLLSLAV-----SSAVP
JEN    ELIGRARISPGAGWNVKDTACLAKAYA-QMWL--LLYFHRRDLRLMANAI-----CSAVP

HCV    ACYSIEPLDLPQIIERLHGL--SAFTLHSYSPGEINRVASCLRKLGVPPLRTWRHRARSV
BVDV   S----QQPETDPSKHATYYY--KGDPIGAYKDV-IGRNLSELKRTGFEKLANLNLSLSTL
DEN    SHW-VPTSRTTWSIHASHEWMTTEDMLTVWNKVWILENPWMEDKTPVESWEEIPYLGKRE
WNL    VNW-VPTGRTTWSIHAGGEWMTTEDMLEVWNRVWIEENEWMEDKTPVEKWSDVPYSGKRE
YEF    TSW-VPQGRTTWSIHGKGEWMTTEDMLEVWNRVWITNNPHMQDKTMVKEWRDVPYLTKRQ
JEN    VDW-VPTGRTSWSIHSKGEWMTTEDMLQVWNRVWIEENEWMMDKTPITSWTDVPYVGKRE

HCV    RAKLLSQGG----RAATCGRYLFNWAVRTKLKLTPIPAASQL-DLSG          motif G
BVDV   GVWTKHTSKRIIQDCVAIGKEEGNWLVKPDRLISSKTGHLYIPD-KG          motif F
DEN    DQWCGSLIG-LTSRATWAKNIQTAINQVRSLIGNEEYTDYMPSMKR           motif A
WNL    DIWCGSLIG--TRARATWAENIQVAINQVRAIIGDEKYVDYMSSLK           motif C
YEF    DKLCGSLIG--MTNRATWASHIHLVIHRIRTLVGQEKYTDYLTVMD           motif D
JEN    DIWCGSLIG--TRSRATWAENIYAAINQVRAVIGKENYVDYMTSL-R          motif E
                                                             Motif I
```

Figure 2.31. Sequence alignment of the NS5 polymerase of six members of the *flaviviridae* family
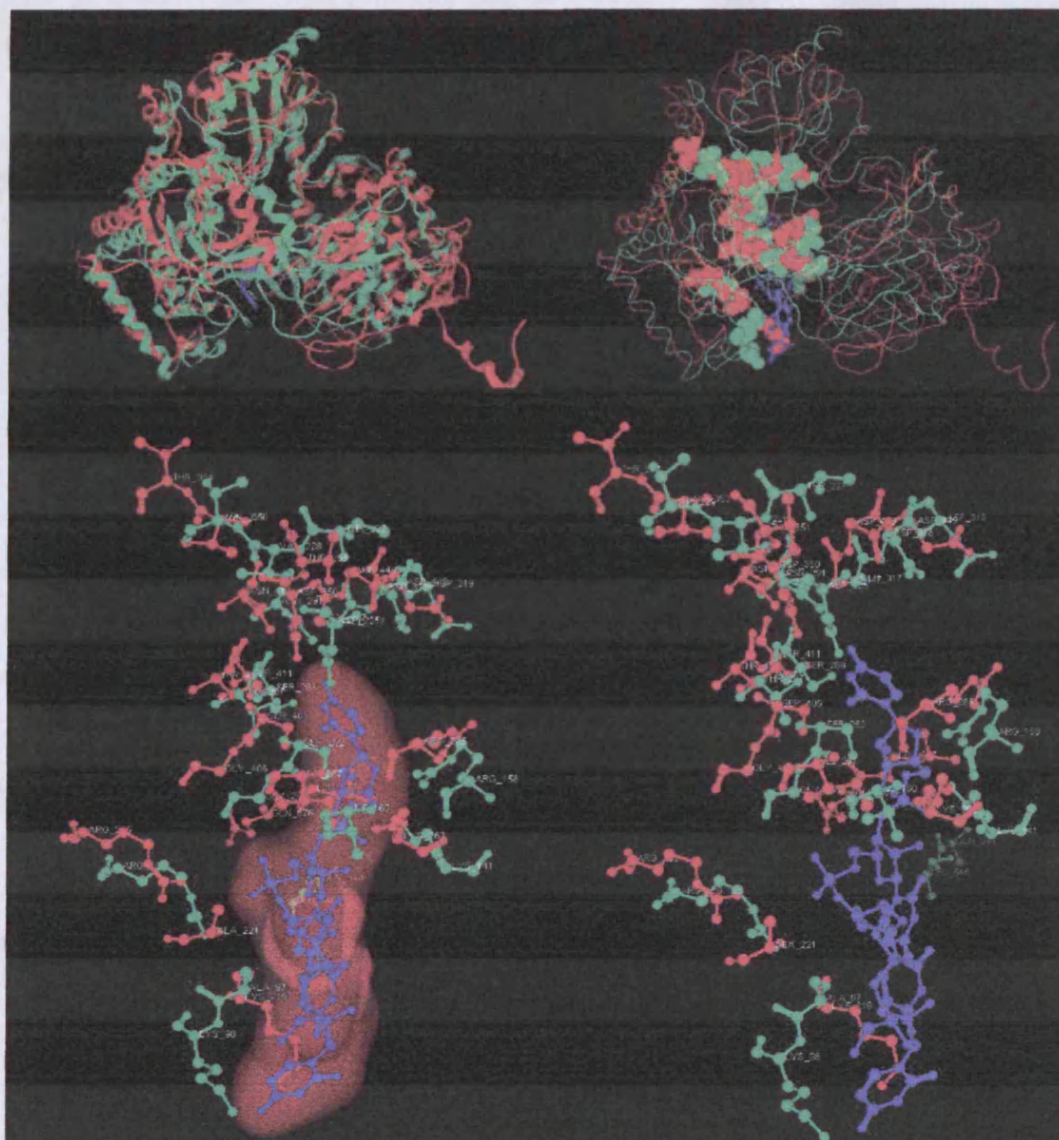
Figure 2.32. The 3D structures of the HCV and BVDV polymerases. Green is the HCV and Red the BVDV polymerase. Blue is the RNA oligonucleotide. Both structures have been resolved by x-ray crystallography. Although they only share 10% sequence identity, their structure, and structural elements are much more conserved. Superimposition of the two structures yields an RMSD of 1.9 Å. It becomes apparent that it is not the alignment score that mostly influences the accuracy of the homology models, but the conservation of the vital motifs for the enzyme to be functional. Here is can be seen that the RNA interacting residues are conserved between the HCV and BVDV polymerases. The RNA binding site should be conserved to all new models.

The theoretical structure of the polymerase proteins for each of the species of Dengue, West Nile virus, Japanese Encephalitis and Yellow Fever were modelled using the HCV and BVDV polymerases as a template structures. The alignment identity of the models against the template was calculated to be approximately 17%. It can be seen from the structural alignment of figure 2.31 that even though the sequence alignment between Dengue, West Nile Virus, Japanese encephalitis, yellow fever and either of the HCV or BVDV templates is very low, the alignment score between the four sequences to be modeled exceeds 50%.

All four polymerase models were expected to bind the ssRNA oligonucleotide in the same way that HCV polymerase does. In order to ensure that the models will have by default a properly shaped active site and to further provide a scaffold to the algorithm, the ssRNA fragment found in the HepC X-Ray file was incorporated as part of the template structure during the building of the homology model.

## 2.2.7   The ligand supported homology modelling using MODELLER

The design of the initial model was done using MODELLER [222], which is a complete homology modelling software package that generates 3D structures of models based on the information from the initial sequence alignment. Spatial restraints are attempted to be satisfied according to the rules imposed by the CHARMM22 built-in forcefield. The 3D protein model is obtained by optimising the molecular probability density function while simultaneously minimising input restraint violations [222]. To guarantee sufficient conformational sampling of each active-site residue, several homology models are generated in this step. Preliminary tests showed that a number between 10 and 100 models provides a satisfactory sampling. To optimise the local interactions, all models obtained are subjected to a crude simulated annealing refinement protocol available in MODELLER (Figure 2.33). Positioning the "ligand" ssRNA into the homology model was done from the information obtained from the Hepatitis C template polymerase. Here it is assumed that ssRNA will bind the model in a mode similar to the one in the template protein. As a result, the ssRNA fragment was copied and moved into the resulting model whilst keeping its orientation and initial conformation to be used as a restraint for the model building process (Figure 2.34). The ssRNA fragment was kept fixed in all steps of the homology modelling experiment using user-defined restraints (Figure 2.35).

The Dengue Virus Polymerase

The West Nile Virus Polymerase
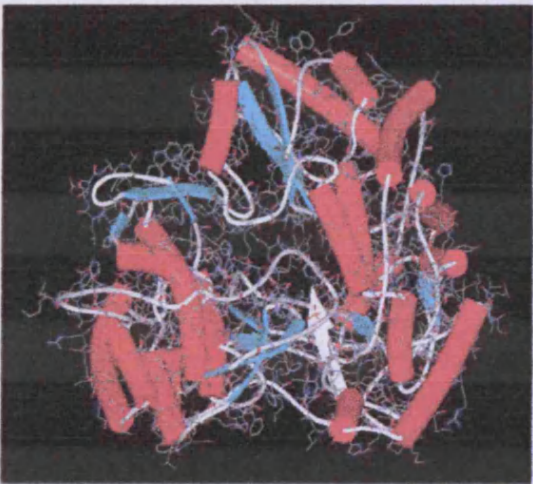
The Jap. Encephalitis Polymerase

The Yellow Fever Polymerase

Figure 2.33. The four Helicase Models: the Dengue virus (top left), the West Nile virus (top right), the Japanese Encephalitis virus (bottom left) and the Yellow Fever virus (bottom right).
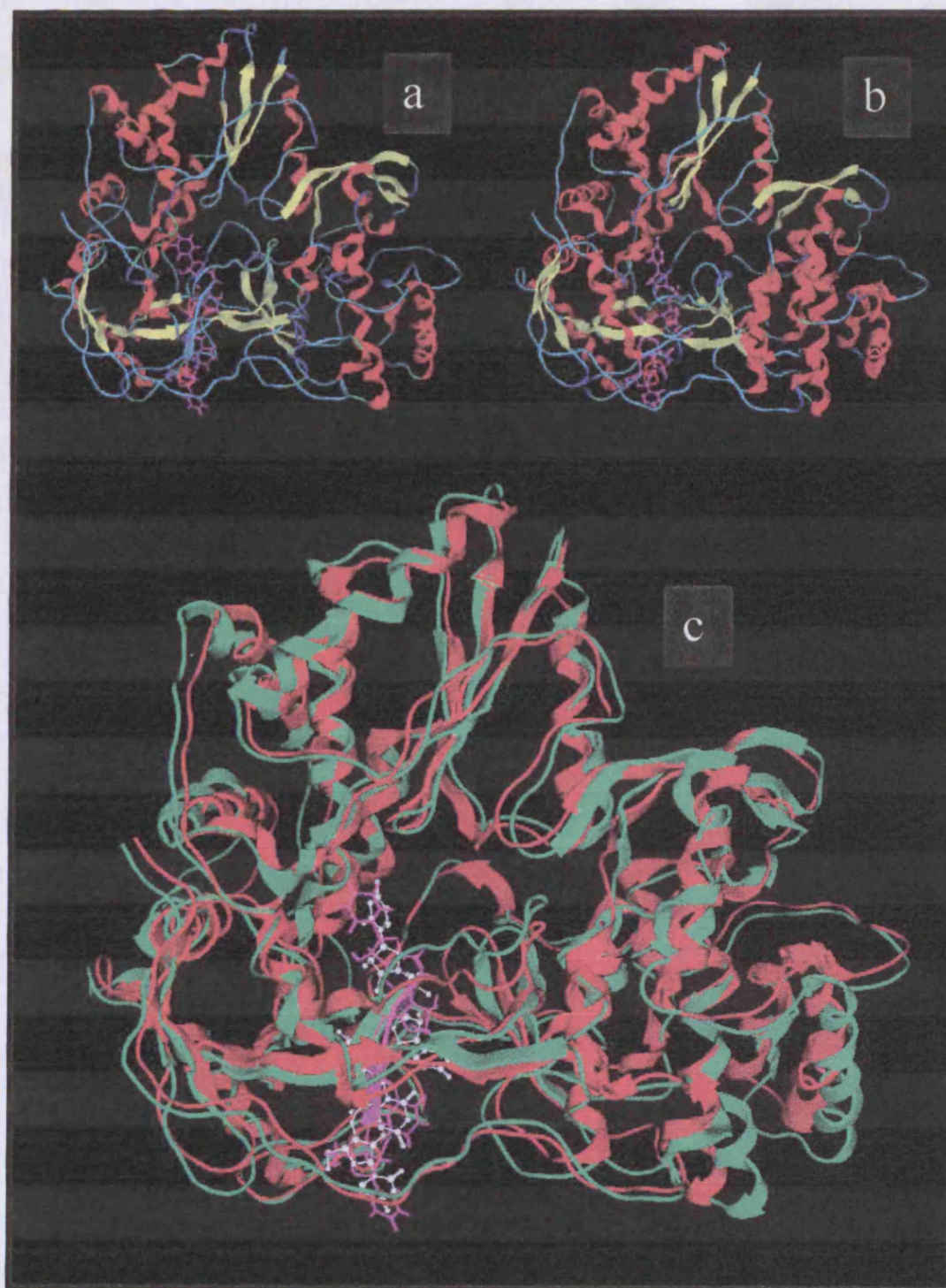
Figure 2.34. (a) The three dimensional structure of HepC polymerase determined by X-ray crystallography (pdb code: 1HB7). (b) The predicted three dimensional structure of the dengue polymerase by homology modelling and (c) the superimposition of (a) and (b). 1HB7 is in Green and the dengue model is in Red.

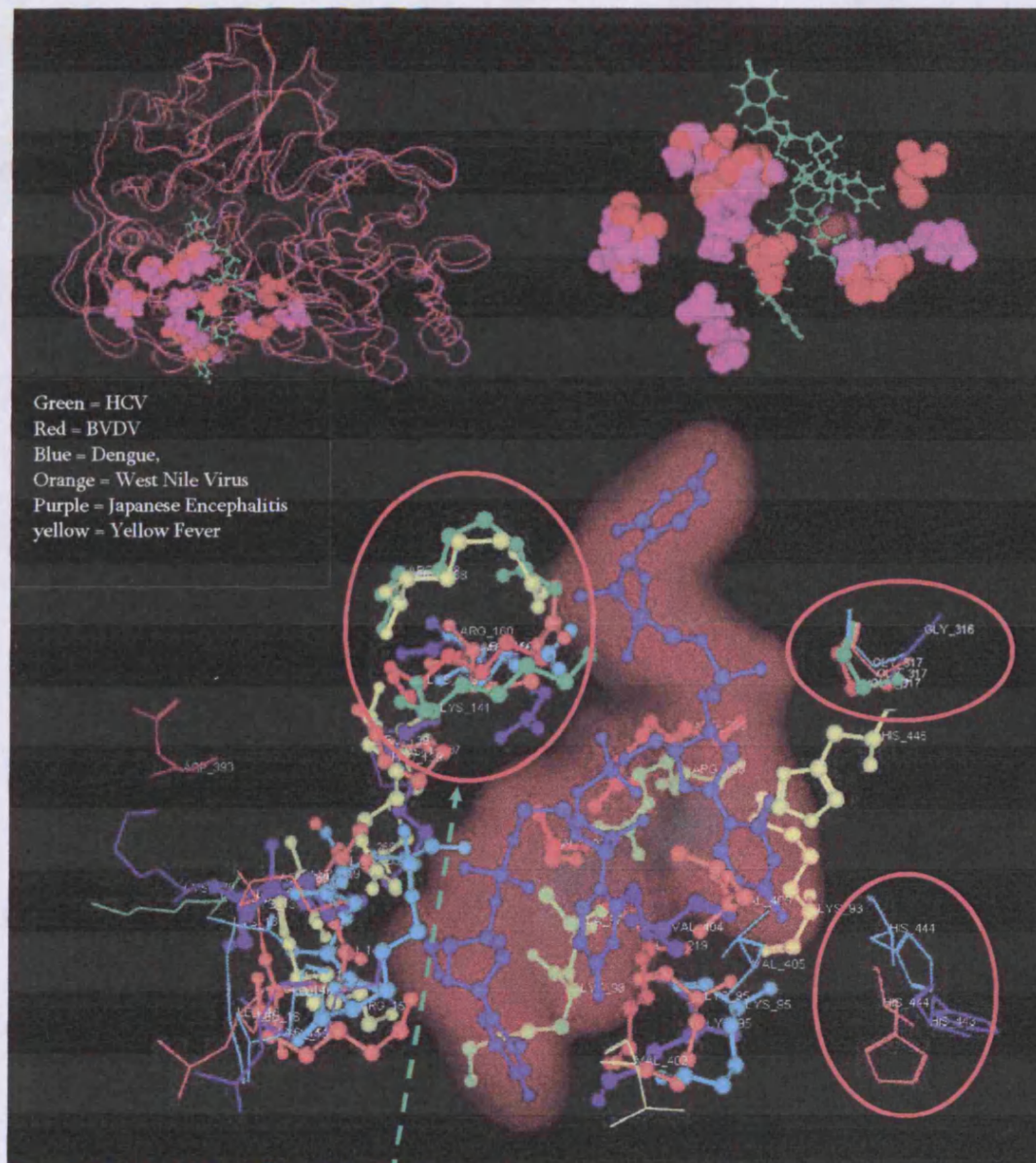**Figure 2.35.** The superimposition of the key residues that interact with the ssRNA. For example this Arg (R) residue is located at the exact same position, with the exact same geometry and orientation on all 4 models as in the template protein. The key residues of a *Flaviviridae* polymerase have been highlighted in circles. Those ssRNA interacting residues have been conserved during homology modelling from template to models.

2.2.8   Model Evaluation for the Polymerase

The models were initially checked with the built-in model evaluation function of the MOE package and was found to be perfect in dihedrals, angles and bond lengths. Minor atom-atom proximity contact errors were reported, but none of them was disturbing the 3D structure of the model or was contributing to an elevation of the overall energy of the complex, thus destabilising it. The polymerase models were then checked with a Ramachandran plot evaluation. The same energy minimisation and molecular dynamics process as in the Helicases section was performed. The results are summarised in the table below (table 2.2):

Table 2.2. The Procheck values of the polymerase models after homology and after optimization and re-evaluation.

Project	Model	After Homology				After Energy Minimization & Model Optimization			
		Core	Allowed	Generous	Disallowed	Core	Allowed	Generous	Disallowed
Polymerase	DenV	79.7	16.9	2.0	1.4	80.9	16.7	2.4	0
	WNV	68.2	24.6	6.1	1.0	89.8	8.4	1.8	0
	JEV	68.7	24.1	6.4	0.8	92.1	4.9	3.0	0
	YF	67.5	25.3	6.6	0.6	89.4	8.5	3.1	0

2.2.9   Homology Modeling of the Dengue Polymerase using MOE

The homology modelling of the Dengue polymerase was also performed using the

homology function of MOE. MOE homology algorithm allows the incorporation

of multiple templates and this time both BVDV and HCV templates were used.

Figure 2.36 shows the new alignment from MOE

Figure 2.36. The sequence Alignment of Dengue, HCV and BVDV used in MOE:

```
DENGUE        ESETPNLDIIGKRIEKIKQEHETSWHYDQDHPYKTWAYHGSYETKQTGSASSMVNGVVRL
HCV-HCV-1NB   -----------------------------SMSYTWT-GALITPCAAEESKLPINPLSNS
BVDV-1S4F     VIREHNKWIL-KKI-RFQGNLNTK---KMLNPGKLSE-QLDREGRKRNIYNHQIGTIMSS

DENGUE        LTKPWDIIPMVTQMAMTDTTPFGQQRVFKEKVDTRTQEPKEGTKKLMKITAEWLWKELGK
HCV-1NB7      LLRHHNMV-YAT--TSRSASLR-QKKVTF---DRLQVLDDHYRDVLKEMKAKAS------
BVDV-1S4F     AGIRLEKLPIVR--AQTDTKTF-HEAIRDKIDKSENRQNPELHNKLLEIFHTIAQPTL--

DENGUE        KKTPRMCTREEFTRKVRS-NAALGAIF-TDENKWKSAREAVEDSGFWELVDKERNLHLEG
HCV-1NB7      TVKAKLLSIEEACKLTPPHSAKSKFGYGAKDVRNLSSRAVNHIRSVWEDLLEDTETPIDT
BVDV-1S4F     KHTYGEVTWEQLEAGVNR-KGAAGFLEKKNIGEVLDSEKHLVEQLVRDLKAGRKIKYYET

DENGUE        KCETCVYNMMGKREKKLGEFGKAKGSRAIWYMWLGARFLEFEALGFLNEDHWFSRENSLS
HCV-1NB7      TIMAKSEVFC----VQPEKGGRKP-ARLIVFPDLGVRVCEKMALYDVVSTLPQAVMGSSY
BVDV-1S4F     AIPKNEKRDVSD-DWQAGDLVVEKRPRVIQYPEAKTRLAITKVMYNWVKQQPVVIP--GY

DENGUE        GVEGEGLHKLGYILRDVSKKEGGAMYADDTAGWDTRITLEDLKNEEMVTNHMEGEHKKLA
HCV-1NB7      GFQYSPKQRVEFLVNTWKSKKCPMGFSYDTRCFDSTVTESDIRVEESIYQCCD--LAPEA
BVDV-1S4F     EGKTPLFNIFDKVRKEWDSFNEPVAVSFDTKAWDTQVTSKDLQLIGEIQKYYY--KKEWH

DENGUE        EAIFKLTYQNKVVRVQRPTPRGTVMDIISRRDQRGSGQVVTYGLNTFTNMEAQLIRQMEG
HCV-1NB7      RQAIRSLTERLYIGGPLTNSKG---QNCGYRRCRASGVLTTSCGNTLTCYLKATAACRAA
BVDV-1S4F     KFIDTITDHMTEVPVITAD--G---EVYIRNGQRGSGQPDTSAGNSMLNVLTMMYAFCES

DENGUE        EGV-FKSIQHLTVTEEIAVKNWLVRVGRERLSRMAISGDDCVV---KPL---DDRFASAL
HCV-1NB7      KLQ----------------------DCTMLVNGDDLVVICESAGTQEDAAALRAF
BVDV-1S4F     TGVPYKSFN-----------------RVARIHVCGDDGFLITEKGL---GLKFANKG

DENGUE        T-ALNDMGKVRKDIQQWEPSRGWNDWTQVPFCSHHFHELIMKDGRVLVVPCRNQDELIGR
HCV-1NB7      TEAMTRYSAPPGDP----PQPEY-DLELITSCSSNVSVAHDASGKRVYYLTRDPTTPLAR
BVDV-1S4F     MQILHEAGKPQKITEGEKMKVAY-RFEDIEFCSHTPVPVRWSDNTSSHMAGRDTAVILSK

DENGUE        ARIS----QGAGWSLRETACLGKSYAQMWSLMYFHRRDLRLAANAICSA-VPSHWVPTSR
HCV-1NB7      AAWETARHTPINS----------WLGNIIMYAPTLWARMILMTHFFSILLAQEQLEKAL
BVDV-1S4F     MATR----LDSSGE-RGTTAYEKAVAFSFLIMYSWNPLVRRICLLVLS--QQPETDPSKH

DENGUE        TTWSIHATHEWMTTEDMLTVWNRVWIQENPWMEDKTPVESWEEIPYLGKREDQWCG--SL
HCV-1NB7      DCQIYGACYS-IEPLDLPQIIERL-HGLSAFTLHSYSPGEINRVASCLR----KLGVPPL
BVDV-1S4F     ATY-------YYKGDPIGAYKDV-IGRNLSELKRTGFEKLANLNLSLSTLGVWTKHTSK

DENGUE        IGLTSRATWAKNIQTAINQVRSLIGNEEYTDYMPSMKRFRREEEEAGVL----W------
HCV-1NB7      RTWRHRARSVRAKLLSQGGRAATCGRYLFNWAVRT-KLKLTPIPAASQLDLSGWFVAGYS
BVDV-1S4F     RIIQDCVAIGKEEGNWLVKPDRLISSKTGHLYIPD-KGFTLQGKHY--------------
```

In yellow are the conserved residues among the Dengue, HCV and BVDV. The main template for the homology experiment is the BVDV and the HCV template has only been used for the RNA binding site alignment (HCV residues: 218-226 corresponding to BVDV residues: 532-540). The Mn^{++} atoms were pasted from the HCV template into the model. The same was done with the ssRNA oligonucleotide. MOE generated a series of possible conformations for the model, which was saved in a database, and then averages were saved in a single pdb file. The best model as ranked by the algorithm was chosen to be further optimized by molecular dynamics.

The molecular dynamics calculations were done using GROMACS 3.2 [www.gromacs.com]. There are two important parameters that define the molecular system in a dynamics simulation. Those are the coordinates and the topology of the system. The coordinates are the ensemble of the Cartesian (x, y, z) position of each atom in the system. The topology of the system contains all the atomic parameters, such as charges, bond lengths, bond angles and dihedrals. The GROMACS suite contains a database with all amino acid and nucleotide topologies. In this study, not only the topology of the protein residues had to be defined, but the RNA too. The ssRNA of the template consists of four uridines and although GROMACS has the residue "URA" (a standard uridine) already defined in its database, two new residues had to be defined: the "URD", which is the first uridine and the "URF", which is the last one. "URD" is a uridine without its phosphate group at the 5' position and "URF" is a uridine monophosphate, but

with a hydrogen in the position of the 3' –OH. In GROMACS the URA nucleotide

is defined as N-O-Uridine-PO$_3$*N, where N is the next heterocyclic base of the

RNA. It means that there is no clear topology for terminal bases, but only for those

in the middle of the nucleic acid strain.

Another issue was that after defining the URD and URF nucleotides the molecular

dynamics program would not connect the 4 bases, but position them as individual

nucleotides in space (URD URA URA URF). The topology was edited again and

the correct type of bonds, angles and dihedrals were defined. Finally the four

nucleotides were bonded together forming a small oligonucleotide chain (URD-

URA-URA-URF), corresponding to the ssRNA taken from the template.

Furthermore, the correct parameters for Mn^{++} where not included in GROMACS.

Thus the ions where substituted with Mg^{++} ions, which are correctly parameterised

by GROMACS. Having defined everything in the molecular system, the latter was

ready to proceed to a geometry and energy optimisation by molecular dynamics.

The molecular dynamics were initiated by generating a periodic box enclosing the

system, which was then filled with water and the system was neutralised, by

exchanging water molecules with the appropriate ions. The system was

energetically minimised using the steepest descent algorithm, in an attempt to give

the system the opportunity to optimise its conformation in the ssRNA area. The

molecular dynamics simulation was run having constrained all the bond lengths of

the molecular system for 1 nanosecond in an NVT environment. After that time,

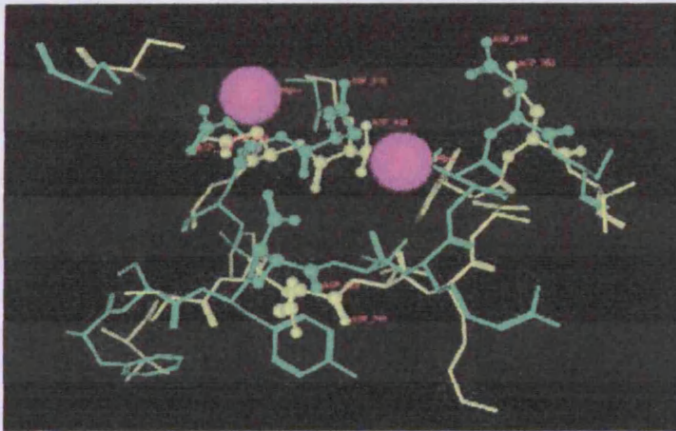the water was removed and the system was energetically minimised using a

Figure 2.37. The three Aspartic Acids coordinated around the Mn^{++} atoms in the Dengue Polymerase Model.

steepest descent initially and a conjugate gradient algorithm later from within GROMACS. The restrains were used in an attempt to allow the system to improve its geometry without moving too far away from its original conformation. Figure 2.37 shows the two models and the 3 Asp residues coordinated against the two Mg^{++} atoms. Figure 2.38 shows the Dengue polymerase model and the ssRNA (in Green) in very close proximity to the three Asp residues (in spacefill representation).
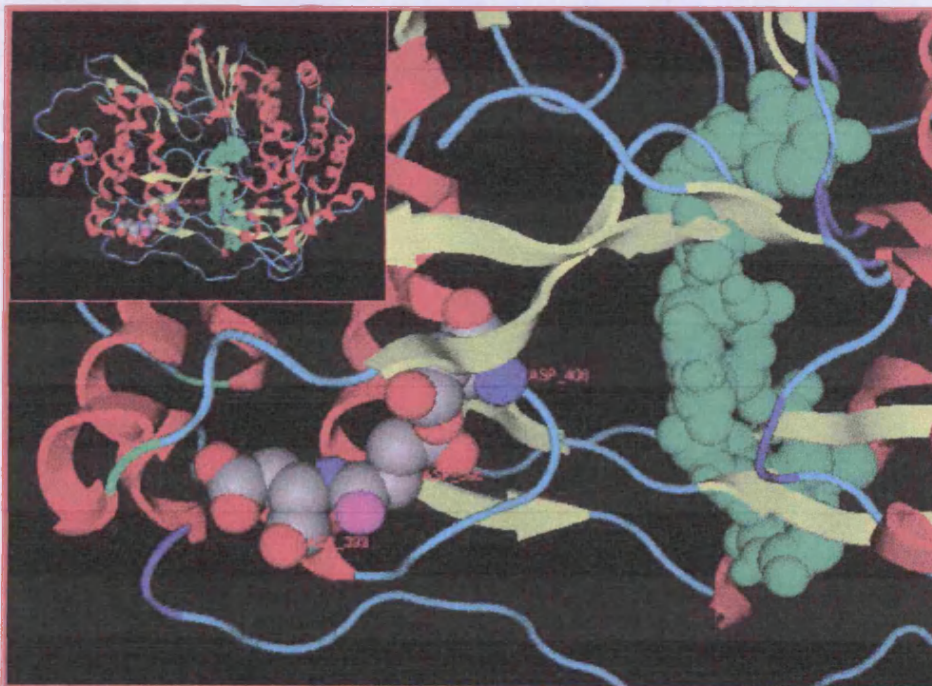

Figure 2.38. The model of the Dengue Polymerase by GROMACS (in green is the ssrna and in spacefill representation are the three Asp residues.

2.2.10 Discussion of the Helicase Models

There are 5 conserved motifs that are highly conserved in all *Flaviviridae* Polymerases [223]. Those motifs are found in the two x-ray resolved models and were also inherited by the four models.

Motif I is used for Magnesium coordination or sugar selection. There are three highly conserved ASPs in this motif. Motif II is supposed to participate in the sugar selection process with the sequence GxxxTxxxNTxT. Motif III is used to coordinate the magnesium atoms as well, and is made of GDD. Motif IV completes the palm structure and must contain either an R or a K. Motif V establishes hydrophobic interaction with the thumb and is made of a conserved CS and an overall hydrophobic patch around it.

The function of the *Flaviviridae* polymerase N-terminal domain is not known. The C-terminal residues are highly hydrophobic and predicted to be membrane-anchoring region. The *Flaviviridae* polymerases present a deep cleft in the middle with the palm at the base. The fingers and thumb domains are both important for correctly positioning the substrates for catalysis by the palm domain. Like poliovirus polymerase and calicivirus RNA polymerase, their structure shows an elaborate arrangement of polymerase domains that have been termed "fingers," "palm," and "thumb," on the basis of its resemblance to a right hand. The overall topology of the *flaviviridae* polymerase models has a typical 'right hand' polymerase structure, with catalytic sites in the base of the palm domain,

surrounded by thumb and finger domains (Figure 2.39). These latter domains fully

encircle the active site, creating a channel for binding to a ssRNA template. In

addition, a β-hairpin structure protrudes from the thumb toward the active site

and is likely to be involved in correct positioning of the template. The overall

structure of NS5B is remarkably similar to the RNA dependent RNA polymerases

(RdRp) of bacteriophage PHI 6. NS5B also has a low-affinity GTP-binding site,

distal from the active site, which is thought to be an allosteric regulator of the

finger–thumb interaction (Figure 2.40). NS5B is tethered to membranes by a C-

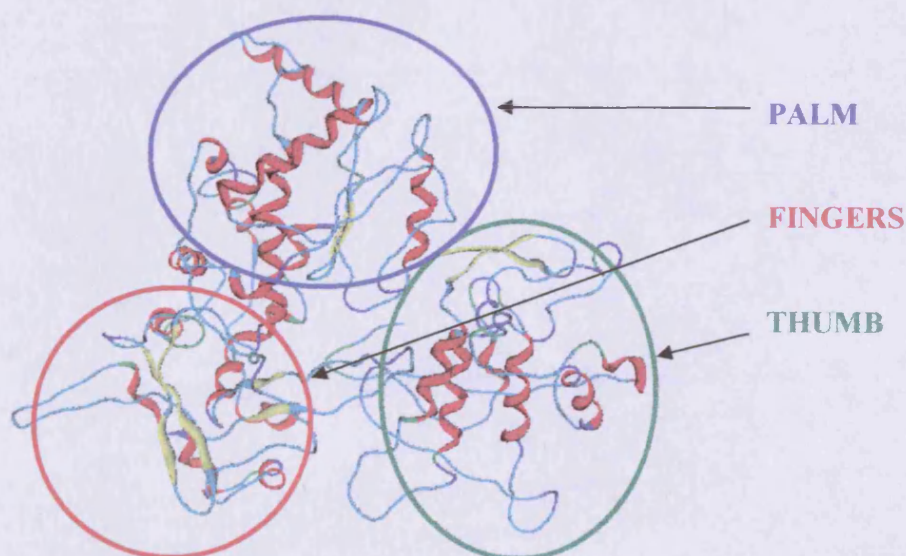terminal peptide anchor and interacts with itself to form higher-order RdRP

complexes.

Figure 2.39. Ribbon diagram of predicted three dimensional structure of Dengue polymerase showing the conserved fingers, palm and thumb domains common to all *Flaviviridae* polymerases.

The fingers domain contributes to binding both the incoming NTP and the template-strand. The fingers domain of the *flaviviridae* polymerases can be divided into a palm-proximal region that is an α-helix-rich subdomain called α fingers, and a distal region that is a β-strand-rich subdomain, which we have named the "fingertips" or β fingers. The fingertips subdomain is composed of seven β-strands and three α-helices. There are three conserved sequence motifs (F–H) shared by all RdRps in the fingers domain that play an important functional roles in the mechanism of polymerization [223]. Motif F contains several conserved positively charged residues (KxxRxxxxxxxxR). It forms a β-strand and two α-helixes, and combines the fingers and the thumb to help build the nucleotide import tunnel and help position the incoming templates. Motif G also contains several conserved positively charged residues (xxxKxExxx). Motif H contains an α-helix and a β-strand in Dengue RdRp, while in most RdRp structures, the motif H forms a loop and an α-helix. This structural element lies at the gate of the template tunnel, and is also predicted to be involved in orientation of incoming template [223].

The palm is the catalytic domain and contains a folding motif that is highly conversed among polymerases. It consists of a three-stranded antiparallel β-sheet, a small helix, three supporting α-helices and a β-strand. The antiparallel β-sheet exists in all RdRps and is the catalytic core of the palm domain. At the interface with the thumb domain, a long loop followed by the pair of β-strands belonging to the thumb domain completes the palm domain. This pair of β-strands region is similar in all RdRps and in HIV-1 Reverse Transcriptase (RT). The palm domain,

the catalytic domain of RdRp, contains the four-amino acid sequence motifs found in all classes of polymerases, named A–D, plus a fifth motif, E. The A–D motifs are highly conversed in RdRp, and motif A and C are also found in the DNA polymerase, whereas motif D, like motif B, is conserved in sequence only in the RNA-directed polymerases. Motif E occurs only in RdRps and RTs. Motif A contains three aspartic acid residues (DTxxxDxxxTxxDxxxxxxx), which are responsible for binding the catalytic metal ions. The Aspartic acid near the end of the β-strand is completely conserved in all classes of polymerases. Motif C contains the highly conserved GDD motif that coordinates the catalytic metal ions (Figure 2.40). This structure is very similar in all classes of polymerases and positions the two aspartates (xxxxxGDDxxx) close to the conserved aspartate of motif A. In motif E, the hydrophobic residues are important for the interactions with the palm core structure and account for the conservation of several hydrophobic residues in motifs A, C and D of RNA-dependent polymerases.

The thumb domain of the *flaviviridae* RNA-dependent RNA polymerases contain motif I, consisting of an α-helix and a β-strand. An arginine is a strictly conserved residue in the motif, and is essential for enzymatic efficiency. The corresponding thumb region in HCV is mainly α-helical, containing eight α-helices and four β-strands. In Dengue RdRp there is a long loop rich in positively charged residues near the gate of the template tunnel. Moreover, Dengue RdRp, similar to HCV RdRp, has a long β-hairpin connecting two α-helices and occluding the active site cleft. This hairpin acts as a discriminator for distinguishing the single-stranded

RNA templates from double-stranded RNA. Without the β-hairpin structure,

calicivirus, poliovirus and Φ6 enzymes are consistent with the ability of these

enzymes to utilise double-stranded RNA as templates [223]. Another function of this

domain is to form a hydrophobic binding pocket near the domain core with the

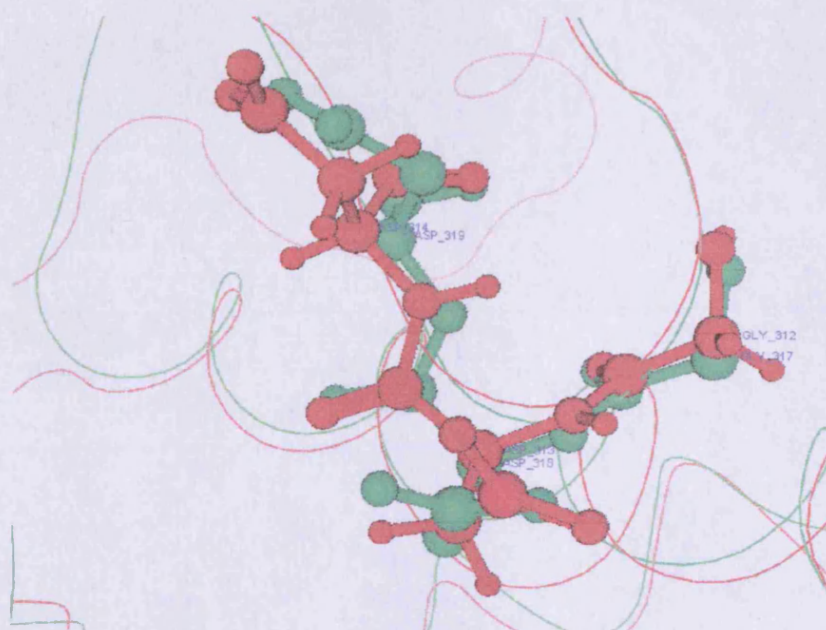help of the palm domain and two long loops of the fingers domain [223].

Figure 2.40. Catalytic residues (drawn in ball and stick conformation) of predicted three-dimensional structure of Dengue Polymerase and crystal structure of HCV Polymerase.

chapter 3
Molecular Biology

3.1   Introduction

The genes encoding for the HCV helicase protein, Dengue type 2 NS3 domain (full length – 1.8kb) and helicase protein (alone – 1.3kb) were obtained from a cDNA library, from the Pasteur Institute in Paris, in order to produce the corresponding recombinant proteins. The cDNA product was amplified by PCR and ligated into the pGEM-T plasmid [224] and later on the pET28a and pET23b+c plasmids [225], which encode ampicillin resistance. The plasmid was transformed into E.Coli bacteria (DH5-α) strain [226] and the bacteria were plated out on ampicillin containing plates. The only surviving bacteria the next day should be those carrying the pGEM-T plasmid. Single colonies were picked and cultured overnight in liquid suspensions. Then the cells were lysed and their plasmid was extracted (mini-prep). The molecular weight of the plasmids was confirmed by gel electrophoresis. The DNA was then purified and concentrated to be sequenced with custom-made primers that had to be designed for complete sequencing of the insert. Then the inserts were re-cloned into the expression vectors pET28a and pET23b+c, in order to prepare for protein expression.

3.1.1 Polymerase Chain Reaction

The sequences encoding (i) the full length NS3 and (ii) the helicase domain alone were amplified by PCR. Gene specific (custom-made) forward and reverse primers were used:

NS3F: AAGCTTGCCGGAGTATTGTGGGATG
NS3R: AAGCTTCTACTTTCTTCCGGCTGCAAATTC
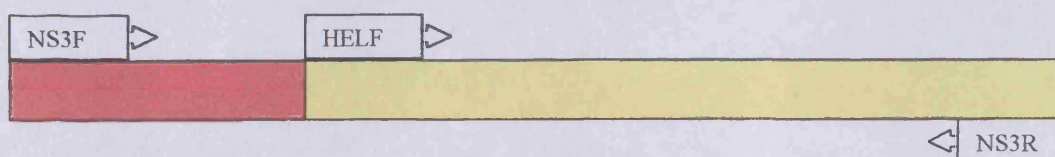HELF: AAGCTTAGTGCTATAGCCCAGACTG

Figure 3.1. The annealing of the custom-made primers for the two genes in PCR.

The two primers annealed at each end on the NS3 & the helicase genes and the PCR reaction went on for 30 cycles. The cycle used was the default one involving the denaturing of DNA at 95 °C for one minute then the temperature was decreased to 58 °C for one minute in order to initiate primer annealing and then increased to 72 °C for five minutes to allow the DNA polymerase to extend the DNA (Figure 3.2).
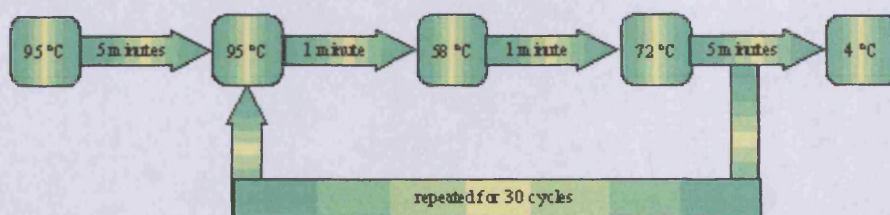

Figure 3.2. The PCR conditions for the amplification of the NS3 & Helicase Domains.

3.1.2 Plasmid Preparation & Phosphatase Reaction

Restriction enzymes are the main tools used for cloning. These enzymes are able to cut DNA at specific recognition sites. Here, the plasmid of pET23b was cut at its unique *Hind* III [227].

To stop the plasmid religating to its original circular form a phosphatase reaction was performed straight after digestion. The phosphatase reaction removed the phosphate groups from the 5' ends of the restricted plasmid so that the overhanging ends could not get back together. The insert has phosphate groups at both of its ends and this way when the insert was applied there was no problem for it to ligate to the plasmid. The insert was capable of ligating to the vector, since it was carrying its own phosphate groups at both its 5' ends. Self-ligation of the insert is possible, but since it would not encode Amp resistance it will not be able to survive the antibiotics.

3.1.3 Gene Insertion - Ligation

The process of joining DNA fragments together with covalent bonds is known as ligation. In a ligation reaction a phosphodiester bond is formed between the 3' hydroxyl end of one nucleotide chain and the 5' phosphate end of another. The enzyme used in the experiment was the T4 DNA ligase (from T4 bacteriophage [228]). The plasmid and the inserts were digested with the appropriate restriction endonucleases. Gel electrophoresis was used to separate the fragments and they were isolated by gel purification (Figure 3.3). The mix of the ligation reaction was 1:4 molar ratio (vector : insert). The mixture contained 1 × DNA ligase buffer (Promega Ltd., Southampton, UK) in a final reaction volume of 10μL and 3 units of DNA ligase (Promega Ltd., Southampton, UK). The ligation mixture was incubated for 16 hours at 17°C.

Initially the PCR products of the genes were ligated into the pGEM-T vector (Promega Ltd, Southampton, UK.). The pGEM-T vector is a linearised vector with a 5' thymidine overhang at each end. The 5' thymidine overhangs will improve the efficacy and success of ligation experiment. This is achieved firstly by preventing the re-circularisation of the vector and secondly, by providing an overhang that is compatible for the ligation of PCR products, that have been made by thermostable polymerases.
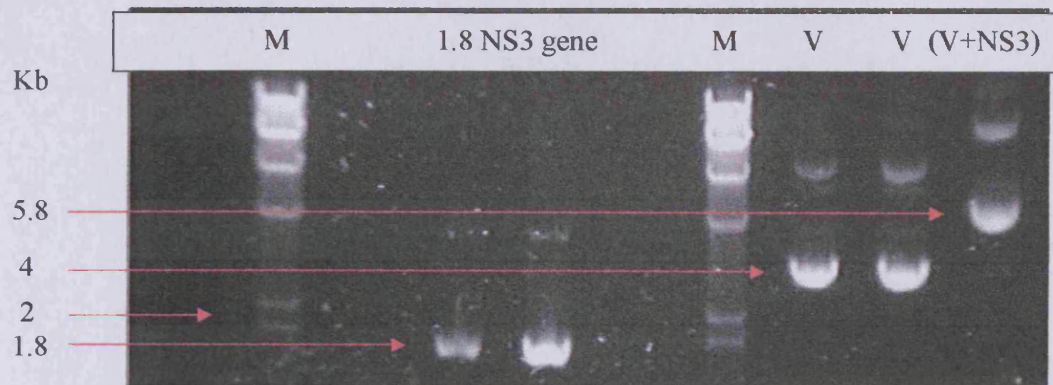
Figure 3.3. Gel electrophoresis of the NS3 gene before and after ligation. The plasmid was also run in a lane to confirm its size (M=Marker, V=Vector).

3.1.4 Transformation – mini prep

Bacterial transformation is the process where competent bacteria take up DNA (usually in the form of plasmids) from their environment (Figure 3.4). The term competent means that bacteria have been treated and their cell membranes are more permeable to genetic material than before. Here the *E. Coli* strain DH5-α was used, because of its versatility,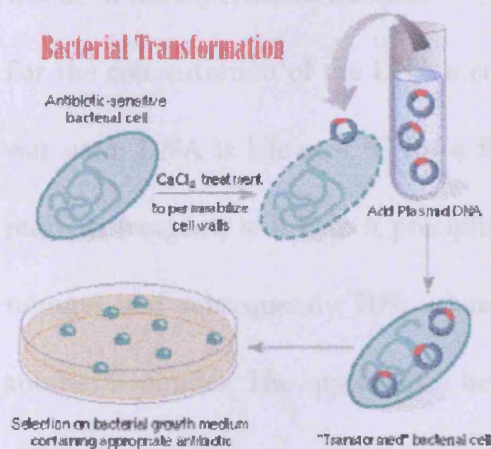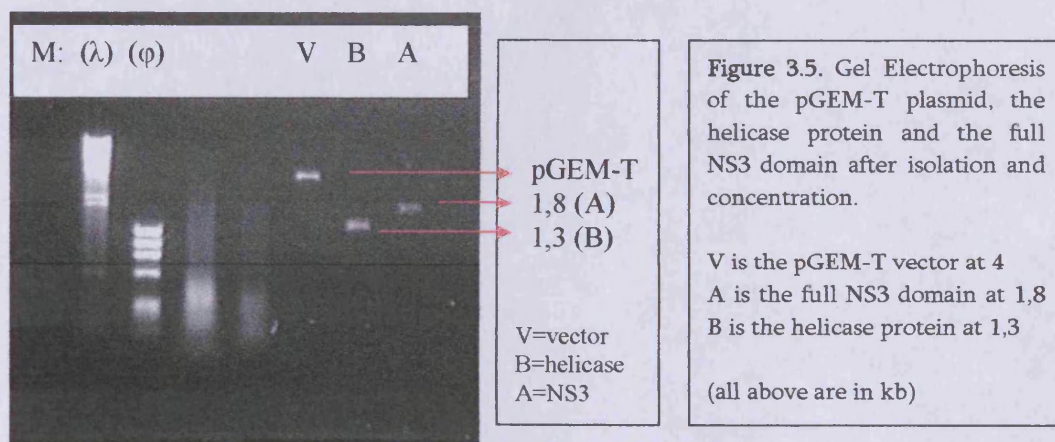 short replication times, that it can sustain a high plasmid content and most importantly for its compatibility with the chosen plasmids. The transformation was done by the "heat shock" method [229]. According to this method the plasmid is added to thawed cells (4 °C) and they are kept on ice for 45 minutes, and then they are heat-shocked for 2 minutes at 42 °C. This will permealize

Figure 3.4. The process of bacterial transformation [234]

their membranes and will allow to the plasmids to enter the cytoplasm. Finally, LB medium is added and the cells are placed in a 37 °C waterbath for approximately sixty minutes. After a short incubation the cells are inoculated overnight on a petri dish with LB agar and ampicillin antibiotic (at 100 µg/mL). The cells were lysed and their plasmids were isolated.

3.1.5 DNA Purification & Concentration

For the purification and cleaning of the isolated plasmid DNA, the DNA / plasmid purification kit from Qiagen was used (see experimental). The DNA precipitant is a combination of phenol chloroform and isoamyl alcohol. After the addition of a solution of phenol-chloroform the DNA solution was centrifuged at 13000 rpm for 10 minutes. Cell debris should stay at the bottom of the centrifuge tube, while pure DNA will be in the supernatant fraction.

For the concentration of the DNA a combination of isopropanol and sodium acetate was used. DNA is left at 4 °C for a few hours in a solution containing the above reagents (reagents will help it precipitate faster). Then it is spinned at 13000 for 20 minutes and subsequently 70% ethanol is added and the solution is spinned for another 5 minutes. The eppendorf is heated in a PCR block to evaporate ethanol at 52 °C.

M: (λ) (φ)   V B A

→ pGEM-T
→ 1,8 (A)
→ 1,3 (B)

V=vector
B=helicase
A=NS3

Figure 3.5. Gel Electrophoresis of the pGEM-T plasmid, the helicase protein and the full NS3 domain after isolation and concentration.

V is the pGEM-T vector at 4
A is the full NS3 domain at 1,8
B is the helicase protein at 1,3
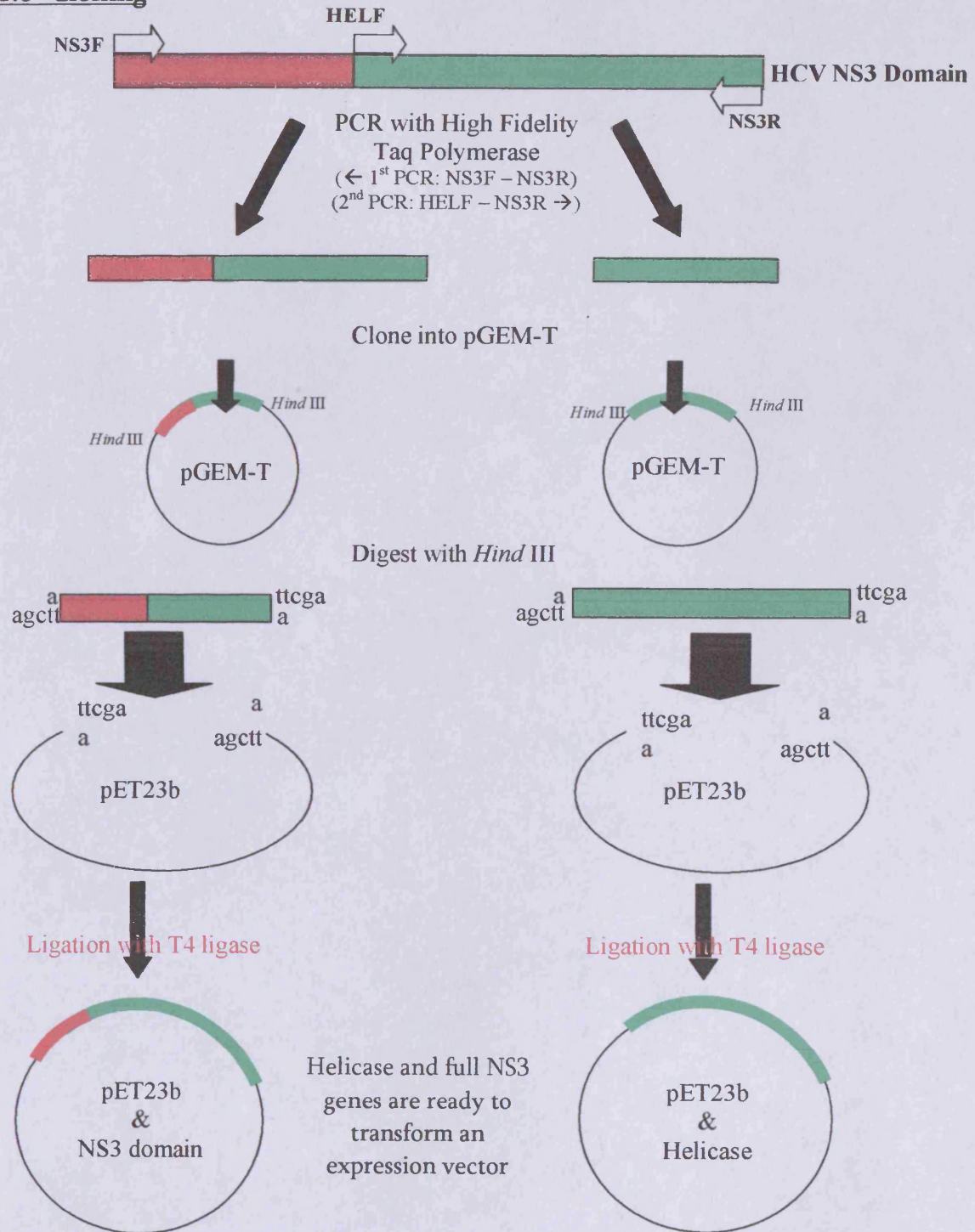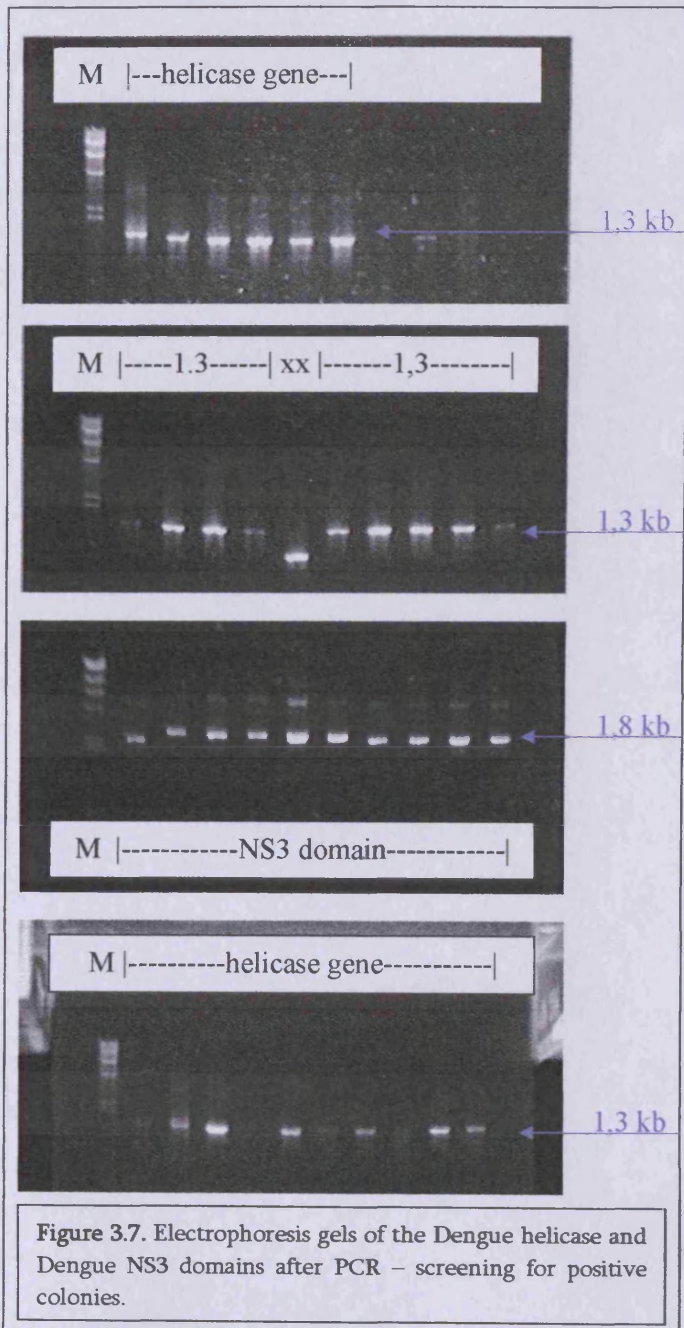
(all above are in kb)

3.1.6 Cloning

Figure 3.6. Illustration of the method of cloning the Dengue genes in pET23b.

The genes for the Dengue helicase and Dengue NS3 domain were cloned into the pGEM-T vector using the single cut *Hind* III restriction site. As a result, orientation of the insert had to be confirmed by a PCR screen (Figure 3.7). The PCR was setup using 50 μL of 10x Taq buffer, 50 μL $MgCl_2$, 20 μl dNTPs, 5 μl of each primer, 365 μl of H_2O and 5 μl of Taq polymerase. The PCR mix was enough for 20 different PCR reactions. The PCR cycles were set as: an initial 95 °C denaturing step for 5 min and then 95 °C for 1 min, 58 °C for 1 min and 72 °C for 2 min. Each cycle was repeated 30 times. Final temperature was set to 4 °C.

Figure 3.7. Electrophoresis gels of the Dengue helicase and Dengue NS3 domains after PCR – screening for positive colonies.

3.1.7 Primer Design

The sequencing reactions are only accurate for 400-500 bases one way. There are two

available primers for both genes:

NS3F =   AAGCTTGCCGGAGTATTGTGGGATG
HelF =   AAGCTTAGTGCTATAGCCCAGACTG
NS3R =   AAGCTTCTACTTTCTTCCGGCTGCAAATTC

The full NS3 gene that is 1.8 kb long requires an internal primer for its sequencing.

The region of interest was identified and the primer was designed using the

approximation: Tm = [4 (G + C) + 2 (A + T)] and making sure that there is a C≡G base

pair at the 3' end. The annealing temperature of the newly designed primer was

estimated to be 60°C.
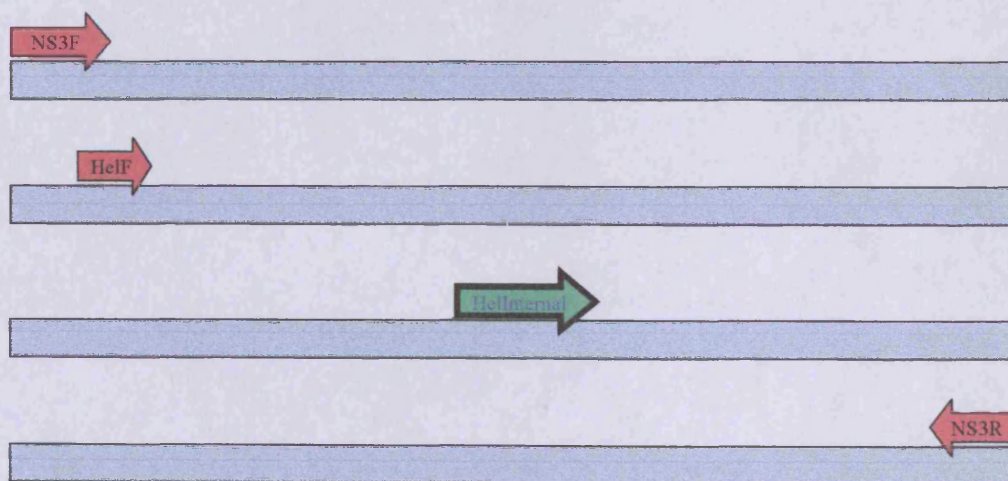
The designed primer sequence was: **gcagagacccatttcctcag**

Figure 3.7. The primer setup for the NS3 gene used for cloning (in red). The primer in green is the primer that was manually designed to allow the sequencing of the centre of the helicase gene.

3.1.8 Site Directed Mutagenesis (Quick Exchange)

Sequencing of the transformed pET vector revealed that both the NS3 domain and the helicase genes contain a stop codon at the end. There is a TAG triplet in frame that stops the transcription machinery prematurely (see below). By that it is assumed that the inserts were not prepared having protein isolation in mind. Site Directed Mutagenesis was used to mutate the T base of the TAG triplet to A. This way the specific triplet would not code for a stop codon anymore, but for a lysine. The choice for lysine was made because of the existence of the two neighbouring lysines on either side of the TAG triplet, thinking that a third lysine would not change significantly the amino-acid environment of the area.

The original seq:	cagccggaagaaagTagaagcttgcggcc
The mutated seq:	cagccggaagaaagAagaagcttgcggcc

In red are the gene's bases before the stop codon
In blue are the stop codon triplets (top) and the mutated one (bottom)
In green are the vector's bases after the stop codon

The side-directed mutagenesis was performed following the quick-change approach as described by Nelson and McClelland [230]. Specific oligonucleotide primers were designed in a manner that they encoded the desired mutation, along with 14 nucleotides complementary to the region around the desired mutation point. These oligonucleotide primers (or cassettes) should not exceed 20 nucleotides on either side of the mutation point. The oligonucleotides will amplify DNA around the desired

location and will induce the desired mutation. About 100ng of the NS3 domain gene

in pGEM-T vector was used as a template for the PCR reaction (Figure 3.8). The

reaction mix contained 1µg of each oligonucleotide primer, and 0.1mM dNTPs in a

final reaction volume of 50µL. The reaction was performed in 1× Pfu Turbo reaction

buffer (containing 10mM $MgCl_2$) with 1 unit of Pfu Turbo DNA polymerase

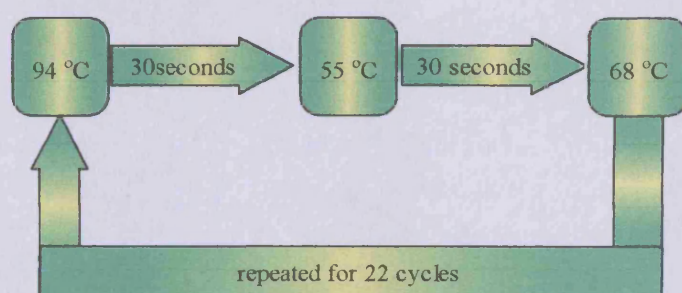(Stratagene, La Jolla, CA, USA).

Figure 3.8. The conditions for the Polymerase Chain Reaction for the site directed
mutagenesis experiment

After PCR, the sample was finally used to transform competent DH5∝ cells. Positive

clones were isolated by sequencing after mini-prep. Site directed mutagenesis is a

PCR based application that involves the extension of a mismatched oligonucleotide,

thus incorporating a mutation. The primer designed for SDM was prepared with the

criteria that its melting temperature would be well above the 76 °C, that it would span

at least 12 bases either way of the base to be replaced and finally that there would be

a CCG base pair rather than an A=T.

SDM primer1 = AAGCTTAGTGCTATAGCCCAGACAG
SDM primer2 = GGATCCAGTGCTATAGCCCAGACAG

3.1.9 Re-Cloning of the stop-free genes

The new genes were digested with *Hind* III and then re-ligated in phosphatase-treated pGEM-T vector. The mixture of the overnight ligation at 17 °C was then used to transform DH5a *E.Coli* cells using the heat-shock method. The colonies that grew on ampicillin (100 μg/mL) plates were PCR screened for the existence and the directionality of the insert (see gel below). Positive colonies were picked up and inoculated for overnight cultures with 10 mL LB. The cells were then lysed and their plasmids were isolated (mini-prep). Sequencing results confirm the correct mutation to lysine and the removal of the stop codon. The genes were then restricted out of the pGEM-T vector with *Hind* III and ligated into the pET23b vector. The ligation was used to transform BL21 DE3 plysS *E.Coli* that were screened for insert-directionality using vector-forward and gene -reverse primers.
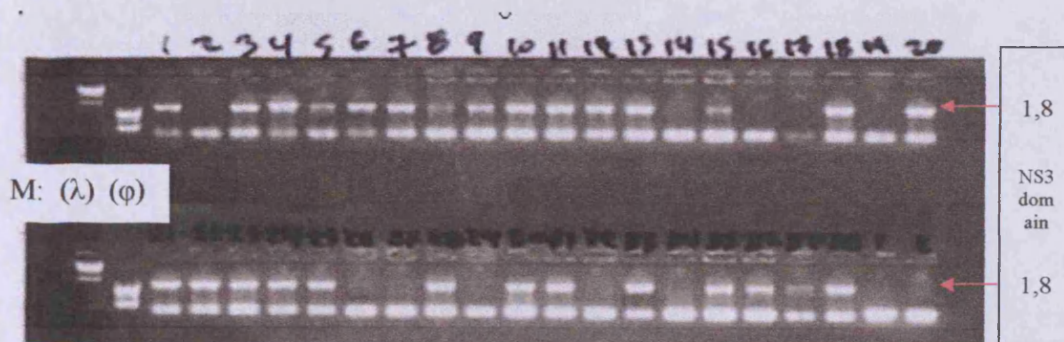

Figure 3.9. The following figure shows the electrophoresis gel of the final screen for directionally correct insert of the full NS3 domain on pET23b in BL21 DE3 plysS *E.Coli* cells. The helicase gene can be extracted by PCR from any of the above positive colonies. Then the amplified helicase genes are re-ligated into the pET23b vector and re-transformed into BL21 DE3 plysS *E.Coli* cells. (Experimental Data in Appendix)

3.2   Protein Expression

Protein expression is a general term to describe how information encoded in a segment of DNA (a gene) is converted into a functioning protein in a cell. As such, protein expression covers the processes of transcription (converting the DNA sequence of a gene into a messenger RNA molecule) and translation (converting the information in the messenger RNA into an amino acid sequence) as well as all of the methods a cell uses to regulate these processes.

A very common and representative protein expression system in E. *coli* is the pET vector system [231]. The pET system is based on the T7 phage promoter (T7 RNA polymerase expression system). T7 polymerase has the unique characteristic that it is specific for T7 promoters and will not recognise any other promoters within the cell. The latter combined with the facts that those T7 promoters are very rare and that termination sequences for T7 promoters are also rare, make it possible to have very long transcripts with no interruption. Finally it has been found that the T7 RNA polymerase is roughly five times faster than the host (*E.Coli*) RNA polymerase. As a result, genes controlled by the T7 RNA polymerase can be overexpressed.

Special *E.Coli* strains have been developed that contain the T7 RNA polymerase. Responsible for the operation of the T7 RNA polymerase gene is the lacUV5 promoter. This promoter can be induced by the addition of isopropyl beta D

thiogalactopyranoside (IPTG) [232].

So, if the system was perfect the insert (gene of interest) should only be expressed in the presence of IPTG, by the induction of the lacUV5 promoter that will initiate the production of the T7 RNA polymerase.

Figure 3.10. 750 ml cultures in 2 L flasks were used at 280 rpm in expression flasks. A lot of different expression conditions were tested in order to achieve the optimal expression conditions and establish an expression protocol.

3.2.1 Cloning and expression vectors

The pGEM-T vector that was purchased from Promega Ltd (Southampton, UK), was used for cloning. For the expression part of the Dengue proteins, pET23b vector was purchased from Novagen (AMS Biotechnology, Milton Keynes, UK). The vector pET28a, was a kind gift from Dr Klump, Roche Biosciences, Palo Alto, USA.

3.2.2 Preparation of competent *E. coli* cells

A glycerol stock of each of the BL21 (DE3) plysS and C41 (DE3) plysS *E. coli* strains were used to inoculate 10mL of LB medium with and without the corresponding antibiotics. The culture was grown overnight in a shaking incubator at 250 rpm, at 37°C for 16 hours. A 1:200 fold dilution was made in fresh LB medium and that was grown too in a shaking incubator at 250 rpm at 37°C. This time the culture was grown until its optical attenuance at 600nm reached 0.4. Cells were centrifuged at 2000 g for 10 minutes and the pellets were re-suspended in half the culture-volume of 100mM $CaCl_2$/ 15 % (v/v) glycerol. After that they were incubated on ice for 30 minutes. The cells were centrifuged as before, only this time they were re-suspended in 1/10th culture-volume of a mixture of $CaCl_2$/glycerol. The cell suspension was aliquoted and snap frozen in liquid nitrogen. The frozen aliquots were stored at –80°C.

3.2.3 SDS-polyacrylamide gel electrophoresis (SDS-PAGE)

The SDS-PAGE protocol is described by Laemmli [233]. The mix for the running gel is 12.5% (w/v) acrylamide solution (acrylamide and N, N'- methylene bisacrylamide in a ratio of 37:1 (w/w)). Moreover, the gel contains 375mM Tris-HCl buffer, pH8.8, 0.1% (w/v) SDS, 0.225% (w/v) ammonium persulphate, 13.2 mM TEMED and water so that the final volume is adjusted to 10mL. On the other hand the stacking gel contains 5% (w/v) acrylamide, 0.14% (w/v) N-N' methylene bisacrylamide, 0.1% (w/v) SDS in 125mM Tris-HCl buffer, pH 6.8, 0.45% (w/v) ammonium persulphate, 13.2mM TEMED and was made to a final volume of 5 mL. The electrolyte running buffer contains 0.1% (w/v) SDS, 192mM glycine and 25mM Tris-HCl at pH 8.3.

The protein sample buffer consists of 80mM Tris-HCl buffer at pH 6.8. It also contains 0.2% SDS, 8% glycerol, 0.003% (w/v) Bromophenol Blue, 10% v/v β-mercaptoethanol. β-mercaptoethanol must be added freshly just before use. The gels were run at 200 V. Coomassie Blue stain solution was used (0.03% w/v in 50% v/v methanol and 10% v/v acetic acid). The stain phase lasts for 1–2 hours and the de-stain phase follows in a wash solution containing methanol (25% v/v) and acetic acid (7% v/v). Pre-stained protein markers (New England BioLabs) were used in electrophoresis along with the samples in order to determine molecular weights of samples.

3.2.4 Recombinant protein immunological detection (Western blotting)

After SDS-PAGE, the resolved proteins were transferred to a moistened nitrocellulose membrane (Amersham International plc, Bucks, UK) [235]. The protein transfer was done by electro-blotting the membrane at 145mA for 45 minutes using a semi-dry blotter (Sartorius Ltd., Epsom, Surrey, UK). After the completion of the transfer, the membrane was blocked in tris buffer solution (TBS), (20mM Tris-HCl, pH 8.0, 50mM NaCl), containing 1% (w/v) skimmed milk powder for over an hour at room temperature. Then the membrane was washed 3 times for 5 minutes each in TTBS (TBS with 0.05% (v/v) Tween 20). After that they were left shaking for 1 hour at room temperature in 5ml of TTBS containing 1% (w/v) skimmed milk powder and the primary antibody. The membrane was washed 3 times for 5 minutes each in TTBS and then agitated for 1 hour with TTBS containing 1% (w/v) skimmed milk powder and the second antibody. The second antibody is an alkaline phosphatase conjugate (Sigma Chemical Company, Poole, UK) that is applied at a dilution of 1/5000. The membrane was washed 3 times for 5 minutes each in TTBS and was stained with Nitro blue tetrazolium and 5-bromo-4-chloro-3-indoyl phosphate (Bio-Rad Laboratories, Watford, Herts, UK). Continuous shaking will initiate the colour reaction (usually within 2-5 minutes). Washing the developed membrane with distilled water will stop the colour reaction.

The immunological detection of proteins the anti hexa-histidine conjugated antibody, requires washing 3 times for 10 minutes each in TBS (10mM Tris-HCl pH 7.5, 150mM NaCl) of the membrane after transfer. The membrane is then blocked for 1 hour in TBS-His containing 3% BSA. It is then washed 3 times for 10 minutes each in TTBS-His (20mM Tris-HCl pH 7.5, 500mM NaCl, 0.05% Tween 20). Then the anti hexa-histidine conjugated antibody (Qiagen Ltd., Crawley, West Sussex, UK) is added for 1 hour at a 1:1000 dilution (in TTBS-His). Excess antibody is removed by washing 3 times for 10 minutes in TTBS-His after the one-hour incubation. The membrane is stained as before using nitro blue tetrazolium and 5-bromo-4-chloro-3-indoyl phosphate.

3.2.5 The Expression of the Dengue full NS3 protein

The full-length Dengue NS3 domain (including the helicase and protease coding regions) was incorporated into the pET23b vector (Novagen, Madison, WI, USA). The difference between this recombinant protein and the native HCV NS3 protein consists of the presence of the six-histidine residues at the C-terminus of the expressed protein. The nucleotide sequence was verified prior to protein production. The primers used for PCR in order to amplify the NS3 domain of the Dengue virus were:

NS3F: AAGCTTGCCGGAGTATTGTGGGATG

NS3R: AAGCTTCTACTTTCTTCCGGCTGCAAATTC.

The recombinant *E. Coli* strain BL21 (DE3) plysS (Novagen) was used for protein expression. Cultures were induced at an attenuance (600 nm) of 0.4, upon addition of isopropyl-β-D-thiogalactopyranoside (IPTG, Sigma, St-Louis, MO, USA) at a concentration of 0.25 mM in Luria Bertani medium containing 100 µg ampicillin / ml and 34 µg chloramphenicol / mL. Protein expression was allowed to proceed for 5 h at 16°C. All subsequent steps were performed at 4°C. The total pellet from a 750 mL culture was resuspended in 10 mL of 20 mM sodium phosphate pH 7.5, 300 mM NaCl, 100 µg/mL lysozyme and Triton X-100 was added to 0.1% final concentration. Following a 60-min incubation on ice, the suspension was sonicated four times for 10

s with 30 s intervals (Branson sonifier 450, Branson Ultrasonics, Danbury, CT, USA).

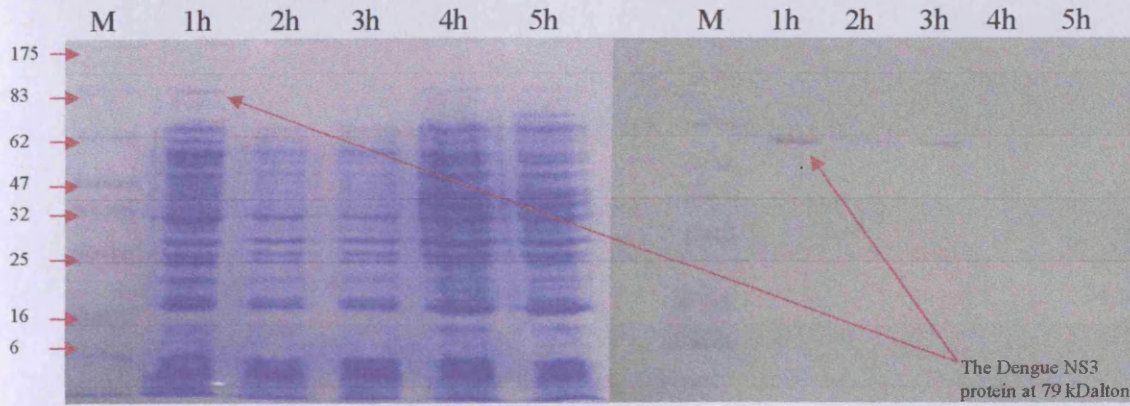Then it was centrifuged at 17 000 × g for 20 min.

Figure 3.11. The expression of the full-length Dengue NS3 domain after elution. The Left gel is coomasie stained whereas the right one is a western blot. Lanes represent hours after induction with IPTG.
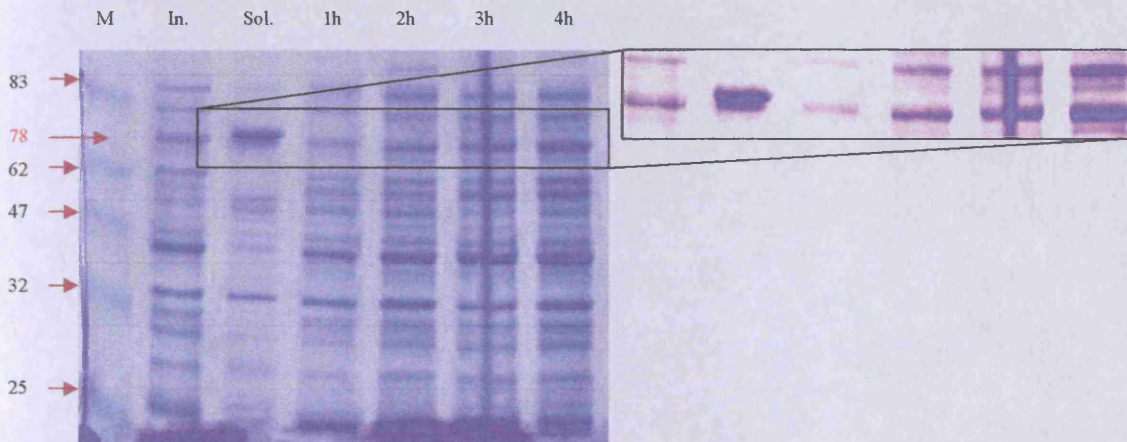

Figure 3.12. Dengue NS3 domain (at 78 kDalton). First lane is the marker, second the insoluble fraction, third the soluble fraction and forth-seventh lanes are the 1st to 4th hours of incubation after the addition of IPTG (0.5 mM) at 16 °C

3.2.6 The Expression of the Dengue helicase

The Dengue helicase coding region was incorporated into the pET23c vector (Novagen, Madison, WI, USA). The difference between this recombinant protein and the native Dengue NS3 protein consists of the presence of the six-histidine residues at the C-terminus of the expressed protein. The insert was copied from the previous Dengue full NS3 domain using the primers: AAGCTTAGTGCTATAGCCCAGACTG and AAGCTTCTACTTTCTTCCGGCTGCAAATTC. The recombinant *E. Coli* strain BL21 (DE3) (plysS) (Novagen) was used for protein expression. Luria Bertani medium was prepared containing 100 µg/mL ampicillin and 34 µg/mL chloramphenicol. The inoculated cultures (from overnights 1:50 v/v) were induced with 0.5 mM IPTG at an attenuance (600 nm) of 0.4. The protein expression was allowed to proceed for 4 h at 18°C. Then the culture was centrifuged at 8000 rpm for 10min and the cell pellet was resuspended in phosphate buffer.
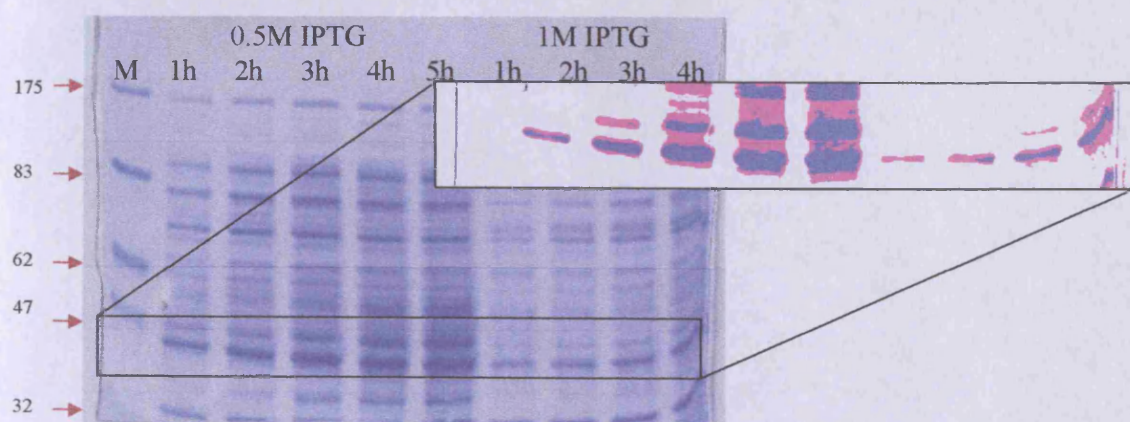

Figure 3.13. The SDS gel from the expression of the Dengue helicase. Here the hours are reported after inoculation of IPTG.

3.2.7 The Expression and purification of HCV helicase

The full-length HCV helicase coding region was incorporated into the pET28a vector

(Novagen, Madison, WI, USA). The difference between this recombinant protein and

the native HCV NS3 protein consists of the presence of the six-histidine residues at

the N-terminus of the expressed protein. The nucleotide sequence was verified prior

to protein production. The HCV helicase gene was generously provided by Roche

laboratories in Palo Alto, USA (through dr Klump).

750 mL cultures containing LB, chloramphenicol 34 μg/mL and kanamycin 25 μg/mL

were prepared. The initial inoculation was taken from an overnight culture with a

1/50 ratio. The recombinant *E. Coli* strain C41 (DE3 - plysS) (Novagen) was used for

protein expression. Cultures were induced at an attenuance (600 nm) of 0.45 with

0.5mM IPTG. The culture was allowed to grow for 3 hours upon induction at 18 °C.

The total pellet from 4 x 750 mL cultures was resuspended in 30 mL of 20 mM sodium

phosphate pH 7.5, 300 mM NaCl, 100 μg/mL lysozyme and Triton X-100 was added to

0.1% final concentration. Following a 30-min incubation on ice, the suspension was

sonicated four times for 20 s with 15 s intervals (Branson sonifier). Then it was

centrifuged at 15 000 × g for 20 min. Clarified homogenates were adjusted to 10 mM

imidazole (Sigma), filtered through a 0.45 μm membrane and loaded twice on a nickel

affinity column. After washing the column with five column volumes of buffer S (20

mM sodium phosphate pH 7.4, 500 mM NaCl) containing 10 mM imidazole, NS3 was

eluted with buffer S containing 300 mM imidazole. To avoid precipitation, and

immediately after elution, the buffer in helicase-containing fractions was exchanged

for 25 mM Tris–HCl pH 7.5, 0.05% CHAPS (3-[(3-cholamidopropyl)-

dimethylammonio]-1-propane sulphonate), 20% glycerol, 5 mM DTT (dithiothreitol)

(buffer P). Protein concentration was evaluated using the Bradford protein assay (Bio-

Rad Laboratories, Hercules, CA, USA) with bovine serum albumin (BSA) as standard.
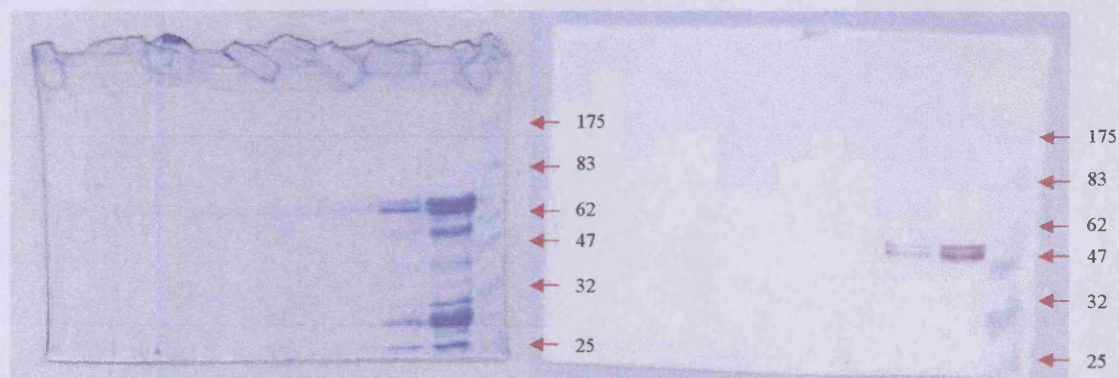
Figure 3.14. The SDS (gel on the left) with the HCV helicase after a quick binding experiment in order to run a western blot (gel on the right) to confirm the existence of the His-tagged protein

Purified NS3 protein was aliquoted and stored at – 80°C. This protein preparation was

estimated to be greater than 75% pure by sodium dodecyl sulphate-polyacrylamide

gel electrophoresis and Coomassie blue staining. The purification procedure yields

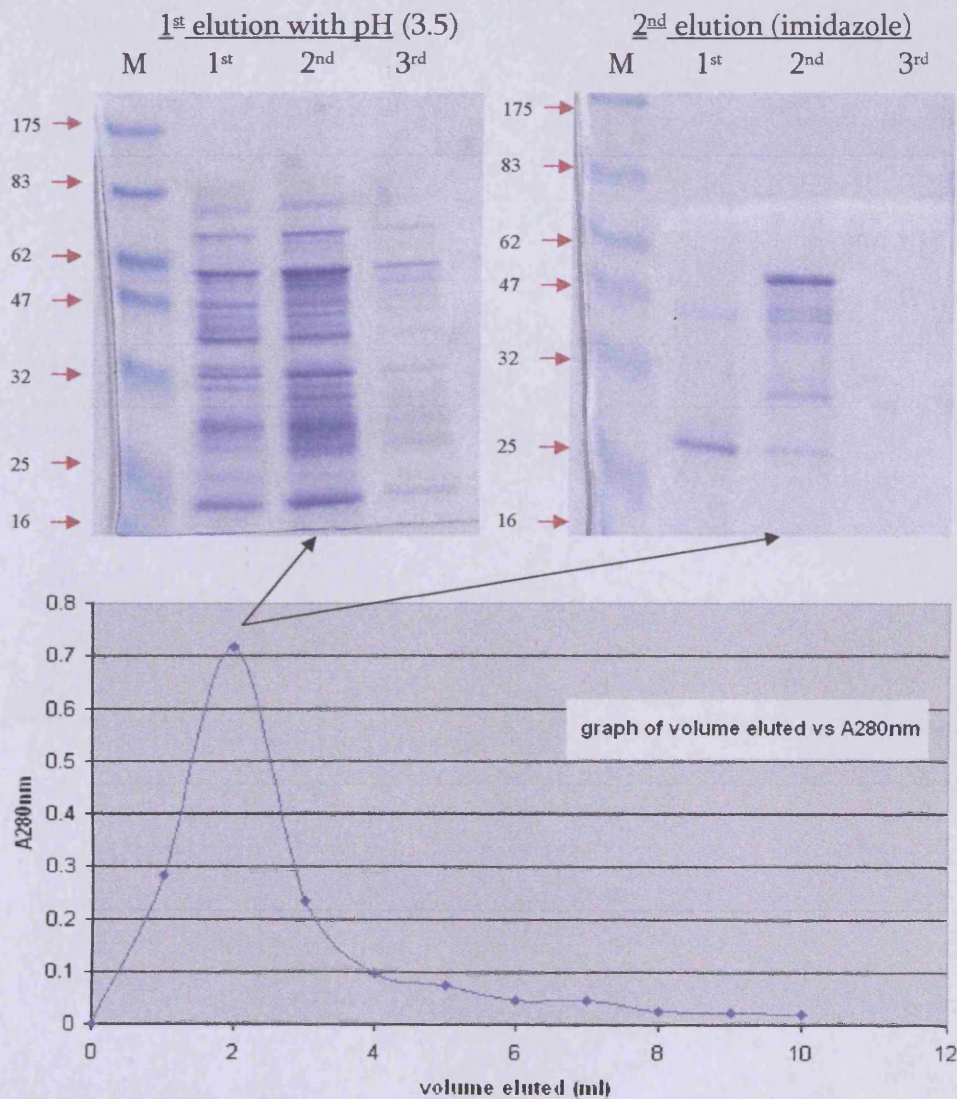approximately 1.6 mg of HCV NS3 per 3 litres of *E. Coli* cultures.

Figure 3.15. The SDS gels from the first and second elutions for the HCV helicase. Only the 2nd tube of the 2nd elution contains the HCV helicase. The contamination could not be removed and it was decided to proceed with the assay

3.3   The Enzymatic Assay

A novel assay for measurement of Hepatitis C virus (HCV) NS3 helicase activity was

developed using modified Flashplate™ technology [236]. This assay involves the use of a

DNA duplex substrate and recombinant HCV NS3 produced in *E. Coli.* The DNA

duplex consisted of a pair of oligonucleotides, one biotinylated, the other DIG-

labelled at their respective 5' termini. This DNA duplex was immobilised, via the

biotin molecule, on the surface of a neutravidin-coated 96 well plate (Pierce) [237].

Helicase activity results in the release of the DIG labelled oligonucleotide, which

translates in signal (luminescence) reduction with respect to control wells.

Biochemical characterisation of the HCV NS3 helicase activity was performed using

this assay. It was demonstrated that the NS3-mediated unwinding is proportional to

both the amount of DNA substrate in the well, and to the NS3 concentration in the

reaction. Most of the NS3-mediated unwinding was achieved in the initial 60 min of

incubation. As expected the reactions were ATP-dependent. We found this assay to

be highly reproducible since only slight variation was observed when a total of 98

helicase reactions (including controls) were performed on one plate. Therefore, this

Flashplate™ helicase assay is fast, convenient and reproducible. These criteria make it

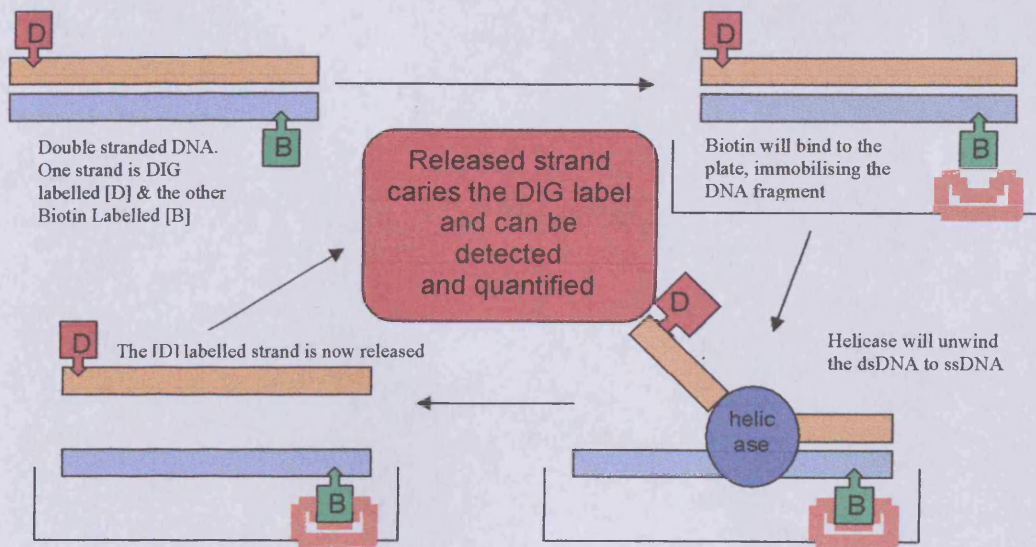suitable for high throughput screening of potential NS3 helicase inhibitors.

Double stranded DNA.
One strand is DIG
labelled [D] & the other
Biotin Labelled [B]

Released strand
caries the DIG label
and can be
detected
and quantified

Biotin will bind to the
plate, immobilising the
DNA fragment

The [D] labelled strand is now released

Helicase will unwind
the dsDNA to ssDNA

helic
ase

Figure 3.16. The enzymatic Assay setup for testing the activities of novel compounds against the HCV helicase.

3.3.1 The HCV Helicase Assay

Two DNA oligonucleotides were prepared. The 5'-biotin-GCTGA-

CCCTGCTCCCAATCGTAATCTATAGTGTCACCTA (39-mer template strand) and

the 5'-DIG-CGATTGGGAGCAGGGTCAGC (20-mer release strand). The release

strand was DIG-labelled at the 5'-end.

The helicase substrate was prepared by annealing the template to the release strands.

Both strands were mixed in a 1:1 molar ratio in 2 mM HEPES (N-[2-hydroxyethyl]

piperazine-N-[2-ethanesulphonic acid]), 0.05 M NaCl, 0.1 mM EDTA, and 0.01% SDS

(w/v) and subjected to a denaturation–renaturation process in which the

oligonucleotides were heated at 100°C for 5 min., followed by a 30 min incubation at

65°C and then a slow renaturation step at 22°C for 4 h. The hybridised NS3 helicase

substrate was kept at – 20°C. A stock solution of neutravidin (Pierce, Rockford, IL,

USA) was prepared at a final concentration of 1 mg/mL in phosphate buffered saline

(PBS). The Flash-plate™ (NEN Life Science Products Inc., Boston, MA, USA) was

coated overnight at 4°C with 100-μL/well of a 5-μg/mL neutravidin solution in a 0.5-

M sodium carbonate buffer pH 9.3. Wells were subsequently blocked upon addition

of 100 μL of a 0.1% (w/v) BSA solution followed by an incubation at 22°C for 2 h.

Plates were then washed three times with 200 μL/well of PBS, air-dried at room

temperature and stored at 4°C with desiccant. For standard assays, 75 μL of 1 M phos-

phate buffer pH 7.0, containing 1 M NaCl and 2.5 ng of the partially annealed DNA duplex was applied to each well, followed by an incubation at 22°C for 4 h. The wells were then washed twice with 200 µL PBS and once with 200 µL of 50 mM Tris HCl pH 7.5, 50 mM NaCl solution, pre-warmed at 37°C.

Helicase reactions were initiated upon addition of 90 µL of a reaction mix consisting of 11 nM of purified full-length HCV NS3 protein, 25 mM 4-morpholine-propanesulphonic acid (MOPS) pH 7.0, 5 mM ATP, 2 mM DTT, 3 mM $MnCl_2$, and 100 µg/mL of BSA to the wells in which 2.5 ng of DNA substrate was previously applied. For negative controls, the reaction mix contained no ATP. Moreover, in experiments where the effect of metal cations was investigated, either $MgCl_2$ or $MnCl_2$ was used as the metal co-factor. Reactions were allowed to proceed for 60 min at 37°C. Wells were then washed twice with 200 µL of a 150-mM NaCl solution and dried at room temperature for 15 min.

The wells of the multi-well plate were washed for 5 min with detection washing buffer (0.1 M maleic acid, 0.15 M NaCl, pH 7.5, 0.3% tween20). Then each well was incubated in 100 µl of Blocking Solution (Blocking reagent: 10% BSA (w/v) in Maleic acid buffer -same as washing buffer without tween20) for 30 min followed by a 30 min incubation in 20 µL Antibody solution (1:10000 solution of the antibody -75 mU/mL- in Blocking solution). The wells were washed twice with 100 µL of detection washing buffer (0.1 M Tris-HCl, 0.1 M NaCl, pH 9.5). Then 20 µL of detection buffer were applied for equilibration. 1 µL of chemiluminescence substrate
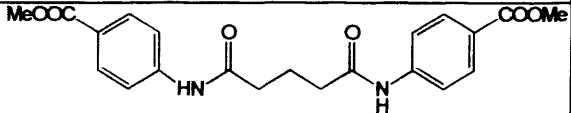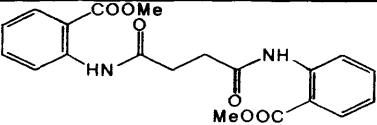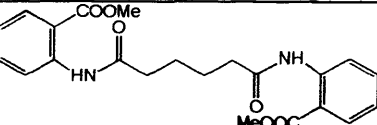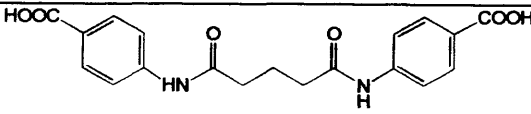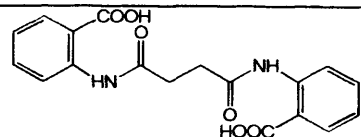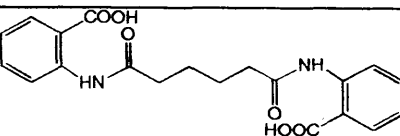
(CSPD) working solution was applied to each well and the plates were incubated for 5 minutes at 17 °C. The wells were drained and after a quick incubation at 37 °C, in order to dry the wells completely, the plate was exposed to a luminescent imager for 10 minutes. The luminescence continues for almost 48 hours. The constant-intensity phase lasts for the first 24 hours.

Remaining DIG label in each well of the 96 well-plate was counted against controls (one with no protein and one with no ATP) in a luminescence multi-well plate reader (luci II).

3.3.2 Results from the Assay & Compounds Tested

Each activity test was repeated three times on different plates (different batches of coating). The results were then all averaged and the average luminescence reading is used as an indication of the inhibitory action of each compound. The following table summarises the results only from the active compounds.

Table 3.1 The results from the enzymatic assay are summarised in the following table Here, active compounds are considered to be all compounds that have at least slightly blocked the unwinding efficiency of the HCV helicase.:

	STRUCTURE	NAME	M.W.
A		CF-AB 14	398.41
B		CF-AB 21	384.38
C		CF-AB 23	412.44
D		CF-AB 25	370.36
E		CF-AB 26	356.33
F		CF-AB 27	384.38

A

	1st	2nd	3rd	Average	% inhibition
10 mM	0.168	0.169	0.175	0.171	3.2
1 mM	0.164	0.165	0.171	0.167	2.6
No Inhibitor	0.152	0.153	0.154	0.153	---

B

	1st	2nd	3rd	Average	% inhibition
10 mM	0.218	0.208	0.207	0.211	8.4
1 mM	0.175	0.171	0.170	0.172	1.1
No Inhibitor	0.165	0.166	0.166	0.166	---

C

	1st	2nd	3rd	Average	% inhibition
10 mM	0.247	0.247	0.253	0.249	16.2
1 mM	0.217	0.223	0.218	0.219	10.6
No Inhibitor	0.165	0.159	0.164	0.163	---

D

	1st	2nd	3rd	Average	% inhibition
10 mM	0.371	0.360	0.361	0.364	36.1
1 mM	0.240	0.241	0.235	0.239	12.6
1 µM	0.182	0.187	0.185	0.185	2.5
No Inhibitor	0.173	0.168	0.175	0.172	---

E

	1st	2nd	3rd	Average	% inhibition
10 mM	0.401	0.388	0.406	0.398	47.3
1 mM	0.264	0.255	0.276	0.265	22.3
1 µM	0.165	0.167	0.169	0.167	3.9
No Inhibitor	0.148	0.144	0.147	0.146	---

F

	1st	2nd	3rd	Average	% inhibition
10 mM	0.455	0.446	0.461	0.454	56.7
1 mM	0.305	0.309	0.311	0.308	29.3
1 µM	0.169	0.171	0.172	0.171	3.6
No Inhibitor	0.154	0.151	0.152	0.152	---

The % inhibition for each compound is produced by the equation:

$$\% \text{ inhibition} = \frac{(a - b) * 100}{c}$$

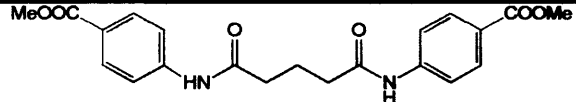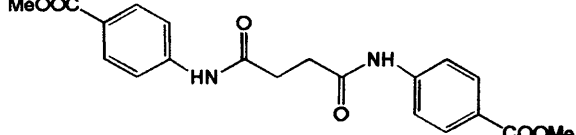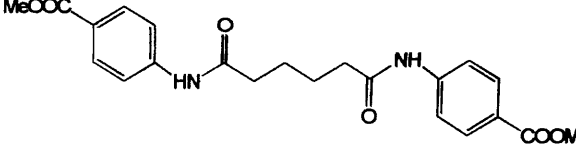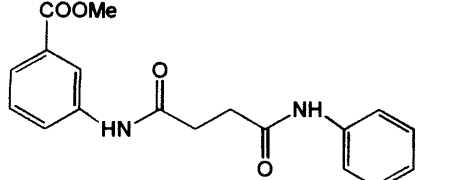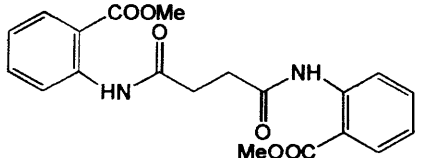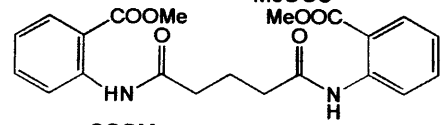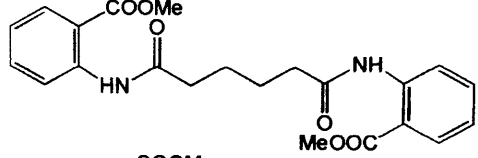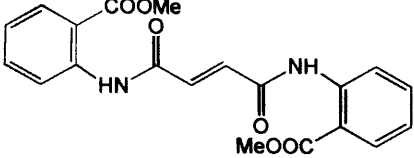where, a = Average luminescence at given concentration
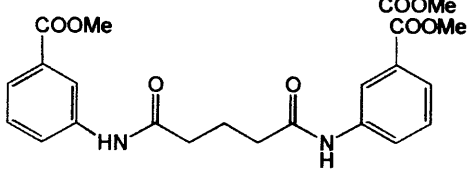
b = Average luminescence without compound

c = Average luminescence without protein (lumi-max)

The above luminescence values are correlated to the calibration runs of the luminecensce plate reader before the assay. So, maximum luminescence averages 0.532 (0.525-0.537) over 10wells. Wells without DNA substrate have an average of 0.048 (0.047-0.05), wells without ATP substrate have an average of 0.052 (0.045-0.056) and wells without anything else but the neutravidin coating have an average luminescence of 0.039 (0.037-0.041). All compounds were subjected to a 60-min assay.

The following table summarises all compounds that were tested against the HCV helicase in comparison with the results obtained from the replicon assay for each compound.

Table 3.2 Comparison of Active Compounds in the Enzymatic and Replicon Assays

Structure	Name	Enzymatic Activity	Replicon Activity
	CF-AB 14	3.2% (10mM)	YES
	CF-AB 16	NO	NO
	CF-AB 18	NO	NO
	CF-AB 19	NO	NO
	CF-AB 20	NO	NO
	CF-AB 21	8.4% (10mM)	YES
	CF-AB 22	NO	NO
	CF-AB 23	16.2% (10mM)	YES
	CF-AB 24	NO	NO

	CF-AB 17	NO	NO
	CF-AB 25	36.1% (10mM)	YES
	CF-AB 26	47% (10mM)	?
	CF-AB-27	57% (10mM)	?
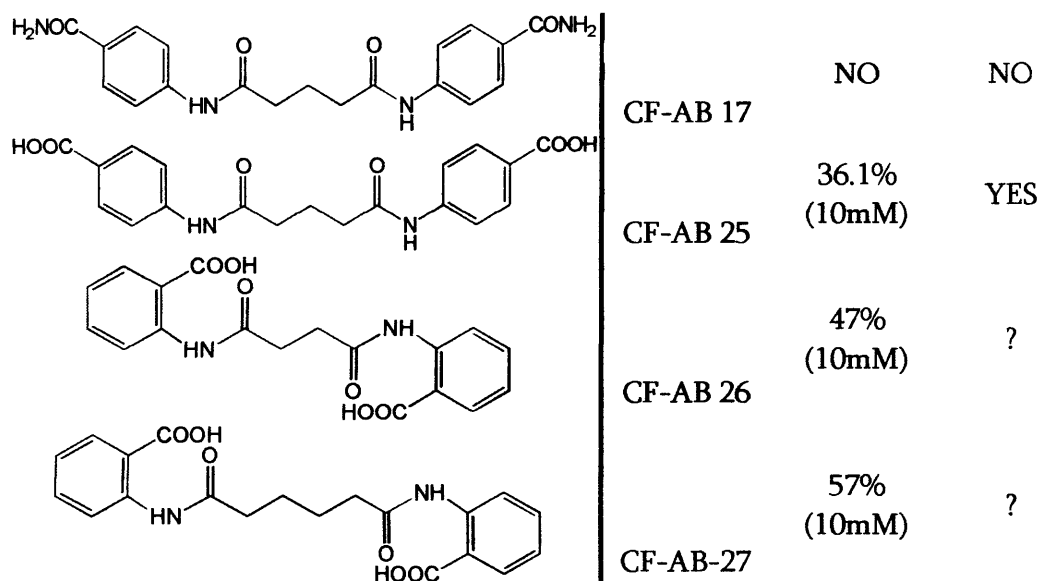
3.3.3 Discussion

Altogether 13 different compounds were tested in this assay. Even though the number of compounds does not allow positive conclusions to be drawn and is only a limited basis for SAR analysis, they are satisfactory for some basic observations.
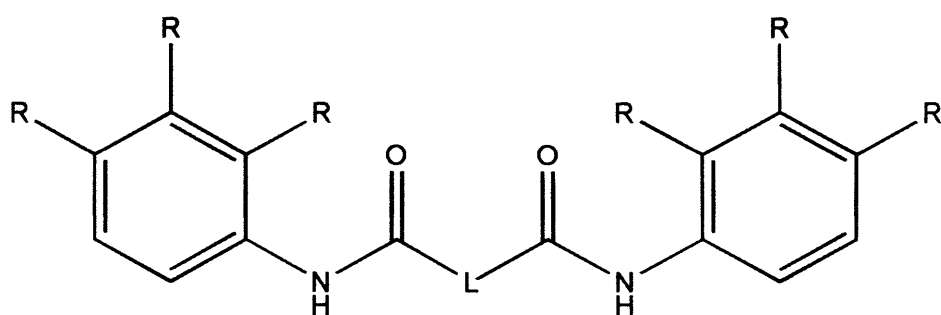
Figure 3.17. The structural scaffold of the family of compounds tested is shown here. Side Group (R) – linker (L) – Side Group (R). Given the symmetry Rleft=Rright. Multiple substitutions have not been tested.

All of the above compounds belong to the same family of compounds described by the symmetry: side group – linker – side group. Looking at the linker it was found that all three aliphatic chain combinations that were used, carrying single bonds gave both active and non-active results. In this experiment the linkers that were used were the C2, C3, C4 and once a C2 double bond. The latter was not active and that may be an indication that rigid compounds are not favourable for the helicase. CF-AB24 with the double bond is not very different to CF-AB23 compound, which was found to be active (16.2% at 10 mM). That is something that was expected from the docking of the compounds from previous *in silico* experiments. The supposed site of interaction of this family of compounds on the helicase seems to be very sensitive in efficiently accommodating a more restrained compound such as CF-AB24 (due to its double bond in the linker region). Enough flexibility must be allowed to the linker region of all helicase potential inhibitor compounds. The linker region must be capable of full rotation, giving the compound optimal geometries that would optimize the interactions established with the enzyme.

It was also observed that the substitution on the phenyl ring of the side-groups should not contain a $CONH_2$. That is demonstrated from the comparison between CF-AB17, CF-AB25 and CF-AB14, where the only structural difference is the $CONH_2$ substitution instead of the COOH or COOMe and this is responsible for the loss of activity in the assay (generally: $CO_2H > CO_2Me > CONH_2$).

It is difficult to interpret a comparison of CF-AB21, CF-AB22 and CF-AB23. These

compounds have the same side groups and different linkers. Although, both C2 and

C4 linkers produced active compounds C3 did not. CF-AB23 is twice as active as CF-

AB21, and it was expected that C3 would be even more active (based on the C3

compounds CF-AB25 and CF-AB27 that are all active). The assay was repeated again,

and the same results were obtained. Finally, it was found that the acids of the active

esters (CF-AB21 and CF-AB23) increased activity. Docking results favoured acids

anyway, but since esters were easier to be chemically synthesised they were tested

first. The conversion of the CF-AB23 to CF-AB27 (from ester to acid) increased the

activity of the compound by 3.5-fold.

3.3.4 Conclusion

The setup of the assay makes it a very quick and efficient tool to evaluate the activities of novel compounds as potential inhibitors against the HCV helicase. It was found that the results from the above HCV assay were in full agreement with the results from the replicon assay, suggesting that the activities observed in the latter are the effect of the given compounds on the helicase proteins.

To sum up, it was confirmed that the HCV helicase enzymatic assay is reliable and is suitable for use as a quick testing method for the identification of novel inhibitors for the HCV helicase.
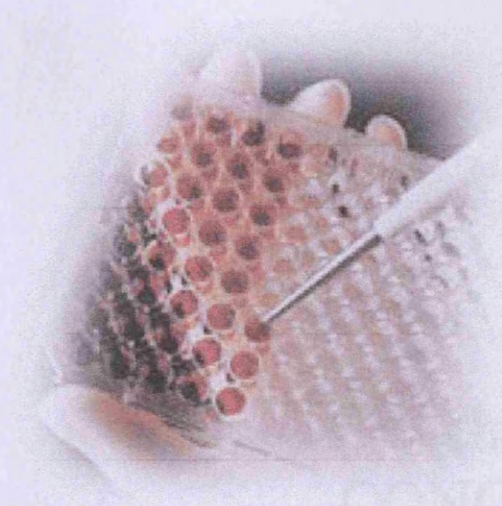

Figure 3.18. Up to 96 compounds can be simultaneously evaluated for their biological activities against the HCV helicase.

chapter 4
CONCLUSIONS
& FUTURE WORK

4.1   General Conclusions & Future Work

The homology models of both the Helicase and the Polymerase family series have been established, tested and evaluated for the Dengue, West Nile, Japanese Encephalitis and Yellow Fever viruses. After the development of the protein models a series of compounds, derived from a single lead compound with few modifications (based on *de novo* structure-based drug design methods), were screened and tested by docking, molecular mechanics & molecular dynamics simulations. From this study, conclusions were drawn as to which moiety was more favourable that are, the type of substitutions that are preferred and the binding mode of these compounds. This data was essential for selecting compounds for synthesis. It was found *in silico* that long, flexible linkers and side-group substitutions with benzimidazole or carboxylic groups increased the binding affinity of the compound candidates.

The verification of the *in silico* experiments came from the molecular biology and biochemistry experiments that followed. The lead and some of the most promising compounds were prepared. These new compounds were then subjected to an enzymatic inhibition test in the biological assay for the HCV helicase that was developed. The genes of the Dengue helicase, NS3 domain and HCV helicase were cloned and expressed. The HCV helicase was purified and an enzymatic assay was established that would estimate the efficacy of the HCV helicase to unwind double stranded DNA. The compounds were tested against the enzymatic assay of the HCV

helicase and the results compared to data from the replicon assay. Data from both assays was fed back into the model in an attempt to optimise the accuracy of the prediction of activity/inhibition of candidate compounds from the drug design algorithms.

Future work on both projects will involve the use of new algorithms for compound docking. Virtual screening of large databases is now feasible. The process of virtual screening has already started using the National Cancer Institute (NCI) database and even though there are no results available yet, it will be very interesting to see the results of this experiment as this will give new ideas for ligand design.

Synthesis of new leads and promising candidate compounds has been planned to start soon. Results from the biological assays of these compounds will provide vital information that can be used in molecular modelling and lead optimisation.

The proteins of the Dengue helicase and the full NS3 domains have already been expressed and purified. An enzymatic assay similar to the HCV helicase one should be established using the Dengue proteins. The data from the Dengue assays can then be correlated to the ones from the HCV and replicon assays, in order to provide more information for the molecular modelling experiments.

The ultimate future aim would be to establish a parallel *in silico* / *in vitro* testing model that will be capable of accurately predicting and evaluating the activities of novel compounds against the above viral helicases. Moreover, another very

interesting future modification would be to replace dsDNA with dsRNA from the

HCV assay and repeat the test of the same compounds.

chapter 5
Appendices

A. Viruses

B. Ramachandran results

B.1 The helicase PROCHECK results

B.2 The polymerase PROCHECK results

C. Protein – DNA contacts in 1A1V

D. Protein – DNA contacts by LigPlot

E. Molecular Biology Methodology

F. Sequencing

G. Virtual screening – lead optimization

A. Viruses

The Hepatitis C Virus

The genome of HCV was first cloned in 1989. The genome of the HCV is 9.4 kb long, contains a single long open reading frame which is responsible for encoding a polyprotein of 3,010 amino acids [15]. There is a non-coding region (NCR) of 324-341 nucleotides at the 5' end & on the 3' end there is a NCR of variable length including a poly (U) tract. The 5' NCR contains an IRES that is very similar in function (not structure) to that of picornaviruses. The nucleotidic sequence of the Hepatitis C virus is highly variable; the most distant strains have only 60% nucleotide sequence homology. Strains from around the world have now been classified into 6 main categories, each with several subtypes, based on sequencing properties. Categories 1, 2 and3 account for almost all infections in Europe. Category 4 is prevalent in Egypt & Zaire, category 5 is prevalent in South Africa finally category 6 is found in Hong Kong. It is not clear yet if immunity to one category automatically defends infection with another strain from another category. There is though evidence that various genome types are pretty much different in their biological properties.

The non-structural region of the NS3 domain of the HCV ranges between 1027 and 1657 of the polyprotein [38]. The genome of the HCV is a positive-stranded RNA virus [13]. The 3010 amino

acid long polyprotein of the virus is processed by cellular and virus-
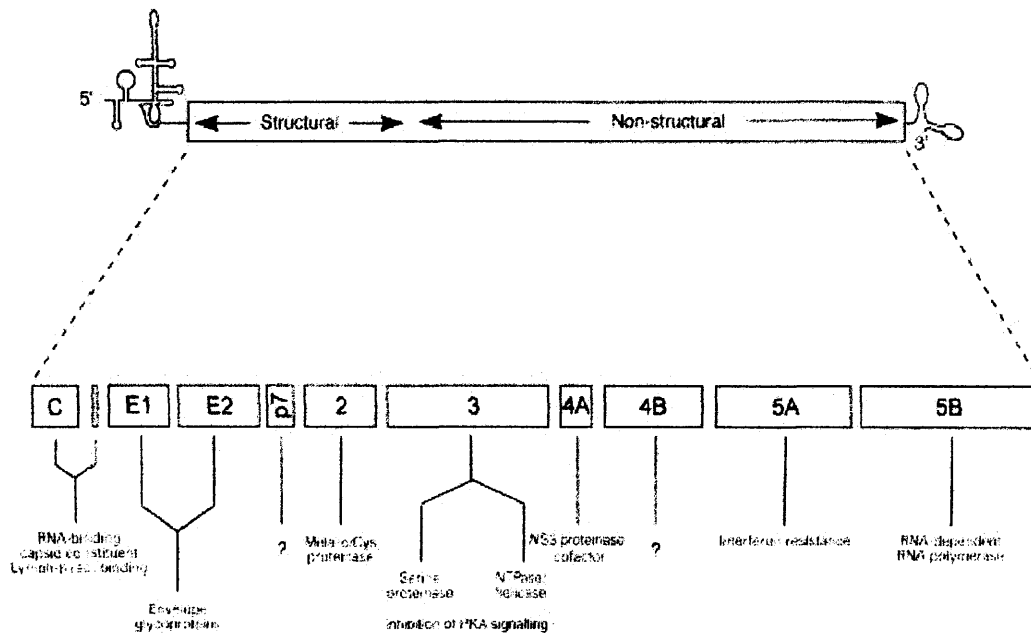
encoded proteases.

Figure App1. Hepatitis C genome organization, polyprotein processing and protein properties. On the top there is a schematic representation of the HCV genome, below are the polyprotein cleavage products and defined functions as described underneath (reference: © 1999 Blackwell Science ltd. *Journal of Viral Hepatitis,* 6, 165-181).

The West Nile Virus

West Nile fever is caused by a virus that is part of the Flaviviridae family. There are nearly 70 different viruses in this group, formerly termed group B arboviruses, of which nearly half are known to cause illness in humans. The World Health Organization defines arboviruses (arthropod-borne viruses) as a group as those "which are maintained in nature principally, or to an important extent, through biological transmission between susceptible vertebrate hosts by hematophagous arthropods; they multiply and produce viremia in the vertebrates, multiply in the tissues of arthropods, and are passed on to new vertebrates by the bites of arthropods after a period of extrinsic incubation" (Sanford, 1991). Common viruses in this classification, in addition to West Nile, include yellow fever, dengue, Japanese encephalitis, St. Louis encephalitis, and tick-borne encephalitis viruses. These viruses are generally spread by mosquitoes or ticks; human-to-human spread does not occur. Infection with these viruses does not produce a unique clinical picture. Therefore, travel to an endemic area and laboratory tests are important for identifying a specific infection.

West Nile virus is a mosquito-borne virus found most commonly in Africa, France, India, Indonesia, the Middle East, and Soviet countries. In 1999, a West-Nile-like virus was identified in patients living in the Northeast United States. The bird is the primary host

and the principal vector is *Culex univittatus*. However, other mosquitoes are known to carry the virus, including *Culex pipiens, Culex antennatus*, and *Culex tritaeniorhynchus* (Asia). Other animal reservoirs are not part of the virus's normal life cycle.

West Nile fever is common in the Middle East with most individuals exposed as children. Children experience a nondescript viral illness with fever that is rarely diagnosed. Neighboring Israel also experiences infection although there, it is more likely the young adult than the child who becomes infected. Spread occurs primarily in the summer months when the mosquito population increases.

The Dengue Virus

Dengue fever is caused by a virus that is part of the Flaviviridae family. There are nearly 70 different viruses in this group, formerly termed group B arboviruses, of which nearly half are known to cause illness in humans. Other common viruses in this classification include yellow fever, West Nile, Japanese encephalitis, St. Louis encephalitis, and tick-borne encephalitis viruses. The most common infection in humans is caused by the dengue virus, of which there are four types. Flaviviruses are generally spread by mosquitoes or ticks; human-to-human spread does not occur. Infection with these viruses does not produce a unique clinical picture. Therefore, travel to an endemic area and laboratory tests are important for identifying specific infection.

Dengue and dengue hemorrhagic fever (DHF) are caused by infection with one of four antigenically distinct, virus serotypes (DEN-1, DEN-2, DEN-3, and DEN4). Once infected with one of these serotypes, the individual develops specific immunity. However, cross-immunity does not develop. It is theoretically possible, therefore, for an individual to be infected four times, each time with a different serotype.

Dengue is mostly seen in tropical urban areas. As with other members of the Flaviviridae family, the virus is transmitted through mosquito bites, specifically *Aedes aegypti*. This mosquito, a domestic, day-biting mosquito, prefers to feed on humans (Gubler and Clark, 1995). In some parts of the world (mostly

Asia and Oceania) other vectors have been implicated: *A. albopictus, A. scutellaris, and A. polynesiensis.*

Dengue is the most important mosquito-borne viral disease, affecting humans with a distribution comparable to that of malaria. Approximately 2.5 billion people are living in areas at risk for epidemic transmission (Gubler and Clark, 1995). Tens of millions of cases of dengue fever occur annually along with up to hundreds of thousands of cases of dengue hemorrhagic fever.

Dengue hemorrhagic fever is the most serious manifestation of the disease. This process, an immunologic reaction, occurs for the most part in individuals already sensitized to the disease, either actively through infection or passively in infants through placental transfer of immunoglobulin from mother to child. Initially, dengue hemorrhagic fever appears the same as dengue but after several days the patient deteriorates with prostration, restlessness, signs of circulatory collapse (diaphoresis, cold extremities, dyspnea, circumoral and peripheral cyanosis, and hemorrhagic manifestations). Available laboratory tests cannot identify who will ultimately develop this manifestation.

The Yellow Fever Virus

Yellow fever is a viral hemorrhagic fever which strikes an estimated 200 000 persons worldwide each year and causes an estimated 30 000 deaths. The case fatality rate may reach 20% to 80%; however, these figures are based on the most severe cases that are hospitalized and the overall case fatality rate is lower.

The yellow fever virus is small (35 to 45 nm) and consists of a core containing single-stranded RNA surrounded by a lipid envelope. The genome has been completely sequenced and found to contain 10 862 nucleotides *(Rice et al. 1985)*. The envelope contains a single glycoprotein with type and group-specific antigenic determinants. Yellow fever virus can be inactivated with lipid solvents (ether, chloroform), heat (56°C for 30 minutes), and ultraviolet light *(Monath 1990)*.

Antigenic differences have been shown between strains of yellow fever virus. Antibody-absorption techniques can distinguish between strains from South America and Africa *(Clarke 1960)*. Strains can also be differentiated on the basis of virulence characteristics for mice *(Fitzgeorge, Bradish 1980)*. RNA oligonucleotide mapping has shown three genetically distinct geographical variants in Africa:

Senegal-Gambia; Ivory Coast-Burkina Faso-Nigeria; Central and East Africa *(Deubel et al. 1986)*.

The disease yellow fever was first distinguished from malaria, dengue, and other tropical diseases during the 1647 to 1649 epidemics in Barbados, Cuba, Guadeloupe and Mexico *(Bres 1986)*. Since then, it has raged as periodic epidemics in the Americas and Africa. In 1900, a commission headed by Walter Reed confirmed that the disease was transmitted from human to human by the mosquito *Aedes aegypti,* a finding earlier theorized by the Cuban physician Carlos Finlay in 1881. This information led to efforts at mosquito control in the Americas, with excellent results in eliminating the disease from many areas.

There are two epidemiologic patterns of yellow fever virus transmission: the urban cycle and the forest cycle (also known as the jungle or sylvatic cycle). The two epidemiologic patterns lead to clinically identical disease, since they are produced by the same virus. In some instances, spread from forest to urban cycles has been documented. Today, the yellow fever virus circulates in an endemic, forest cycle in the Americas, resulting in up to 500 reports of infections of unimmunized forest workers per year. In Africa, yellow fever virus circulates in both urban and forest cycles, and the disease

periodically explodes out of its endemic pattern to infect large number of unimmunized persons during major epidemics.

In the urban pattern, the virus is transmitted by mosquito from infected humans to susceptible humans. For the urban cycle, the mosquito vector is usually A. *aegypti,* a domestic mosquito that breeds near houses, with the female preferring to lay eggs in water collected in water jars, old tires, gutters, or discarded tin or plastic containers. In 1978, it was found that A. *aegypti* females could transmit yellow fever virus transovarially to a small proportion of their offspring and these eggs can thus maintain the virus during the dry season *(Aitken et al. 1979).* The virus multiplies in the mosquito vector. About 12 to 21 days after biting an infected person or monkey the mosquito becomes infective and it remains infective for the rest of its life.

The disease in humans is characterized by sudden onset of fever, headache, backache, general muscle pain, nausea, and vomiting. As the disease continues, albuminuria, oliguria (even anuria), and jaundice occur. Hemorrhagic symptoms may include epistaxis, hematemesis, and melena.

B. Ramachandran results

Table app1. Ramachandran Plot stats for HepC Helicase (1A1V) & Dengue Model.

Ramachandran Plot statistics for HepC Helicase (1A1Va)

		No. of residues	%-tage
Most favoured regions	[A,B,L]	316	89.5%*
Additional allowed regions	[a,b,l,p]	37	10.5%
Generously allowed regions	[~a,~b,~l,~p]	0	.0%
Disallowed regions	[XX]	0	.0%
		----	------
Non-glycine and non-proline residues		353	100.0%
End-residues (excl. Gly and Pro)		7	
Glycine residues		38	
Proline residues		31	
Total number of residues		429	

Ramachandran Plot statistics for Dengue Virus (Model)

		No. of residues	%-tage
Most favoured regions	[A,B,L]	296	82.2%*
Additional allowed regions	[a,b,l,p]	50	13.9%
Generously allowed regions	[~a,~b,~l,~p]	9	2.5%
Disallowed regions	[XX]	5	1.4%*
		----	------
Non-glycine and non-proline residues		360	100.0%
End-residues (excl. Gly and Pro)		2	
Glycine residues		40	
Proline residues		27	
Total number of residues		429	

Table app2. Ramachandran Plot stats for the Dengue Model after Molecular Dynamics

Ramachandran Plot statistics for Dengue Virus (Model) after miminisation

		No. of residues	%-tage
Most favoured regions	[A,B,L]	333	92.5%
Additional allowed regions	[a,b,l,p]	19	5.3%
Generously allowed regions	[~a,~b,~l,~p]	8	2.2%
Disallowed regions	[XX]	0	0%
		----	------
Non-glycine and non-proline residues		360	100.0%
End-residues (excl. Gly and Pro)		2	
Glycine residues		40	
Proline residues		27	

Total number of residues		429	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20.0 a good quality model would be expected to have over 90% in the most favoured regions [A,B,L].

Table app3. Ramachandran Plot statistics for HepC Helicase (8OHM) & West Nile Virus (Model)

```
Ramachandran Plot statistics for the HepC Helicase (8OHM)

                                  No. of residues      %-tage
Most favoured regions     [A,B,L]          323         88.5%*
Additional allowed regions [a,b,l,p]        39         10.7%
Generously allowed regions [~a,~b,~l,~p]     2          .5%
Disallowed regions         [XX]             1          .3%*
                                           ----        ------
Non-glycine and non-proline residues       365        100.0%
End-residues (excl. Gly and Pro)             1
Glycine residues                            39
Proline residues                            30
                                           ----
Total number of residues                   435
```

```
Ramachandran Plot statistics for West Nile Virus (Model)

                                  No. of residues      %-tage
Most favoured regions     [A,B,L]          309         80.9%*
Additional allowed regions [a,b,l,p]        52         13.6%
Generously allowed regions [~a,~b,~l,~p]    14          3.7%
Disallowed regions         [XX]             7          1.8%*
                                           ----        ------
Non-glycine and non-proline residues       382        100.0%
End-residues (excl. Gly and Pro)             1
Glycine residues                            27
Proline residues                            25
                                           ----
Total number of residues                   435
```

Table app4. Ramachandran Plot statistics for the improved West Nile Virus Models

```
Ramachandran Plot statistics for West Nile Virus (Model)

                                  No. of residues      %-tage
Most favoured regions     [A,B,L]          329         86.1%
Additional allowed regions [a,b,l,p]        43         11.3%
Generously allowed regions [~a,~b,~l,~p]    10          2.6%
Disallowed regions         [XX]             0           0%
                                           ----        ------
Non-glycine and non-proline residues       382        100.0%
End-residues (excl. Gly and Pro)             1
Glycine residues                            27
Proline residues                            25
                                           ----
Total number of residues                   435
```

Table App5a. The Yellow Fever Helicase Model's Ramachandran plot analysis summary.

```
+----------<<< P R O C H E C K    S U M M A R Y  >>>----------+
| YF   2.2                                         435 residues |
| Ramachandran plot:   81.5% core   13.4% allow    3.2% gener   1.9% disall |
| Gly & Pro Ramach:     5 labelled residues (out of  60)        |
| Chi1-chi2 plots:      1 labelled residues (out of 242)        |
+-------------------------------------------------------------+
```

Table App5b. The Yellow Fever Helicase Model's Ramachandran plot analysis summary. AFTER MOLECULAR DYNAMICS

```
+----------<<< P R O C H E C K    S U M M A R Y  >>>----------+
| YF   2.2                                         435 residues |
 | Ramachandran plot:   96.7% core    3.2% allow    0.1% gener    0% disall |
| Gly & Pro Ramach:     5 labelled residues (out of  60)        |
| Chi1-chi2 plots:      1 labelled residues (out of 242)        |
+-------------------------------------------------------------+
```

Table App6a. The Japanese Encephalitis Helicase Model Ramachandran plot analysis summary.

```
+----------<<< P R O C H E C K    S U M M A R Y  >>>----------+
| je   2.2                                         435 residues |
| Ramachandran plot:   79.9% core   15.1% allow    3.3% gener   1.6% disall |
| Gly & Pro Ramach:    11 labelled residues (out of  69)        |
| Chi1-chi2 plots:      2 labelled residues (out of 233)        |
+-------------------------------------------------------------+
```

Table App6b. The Japanese Encephalitis Helicase Model Ramachandran plot analysis summary. AFTER MOLECULAR DYNAMICS

```
+----------<<< P R O C H E C K    S U M M A R Y  >>>----------+
| je   2.2                                         435 residues |
 | Ramachandran plot:   94.2% core    5.1% allow    0.7% gener    0% disall |
| Gly & Pro Ramach:    11 labelled residues (out of  69)        |
| Chi1-chi2 plots:      2 labelled residues (out of 233)        |
+-------------------------------------------------------------+
```

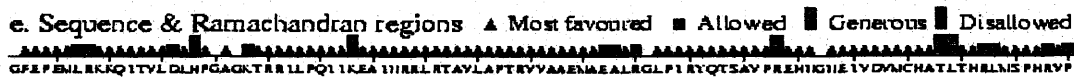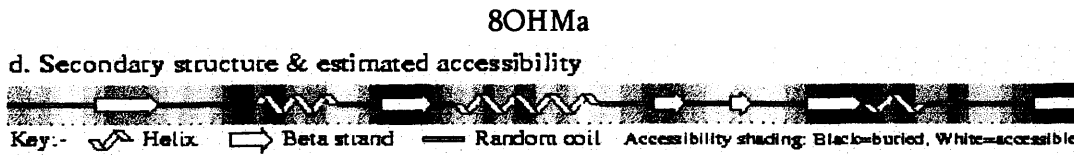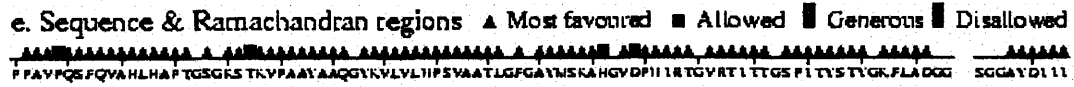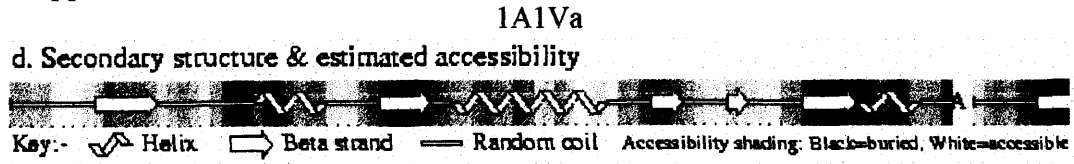**Figure App2.** Structural motifs, accessibility and PROCHECK summary.

1A1Va

d. Secondary structure & estimated accessibility

Key:- 〜⋀ Helix   ▭⟩ Beta strand   ══ Random coil   Accessibility shading: Black=buried, White=accessible

e. Sequence & Ramachandran regions   ▲ Most favoured   ■ Allowed   ▌ Generous   ▌ Disallowed

FFAVPQSFQVAHLHAPTGSGKSTKVPAAYAAQGYKVLVLIIPSVAATLGFGAYYNSKAHGVDPIIIRTGVRTITTGSFITYSTYGKFLADGG   SGGAYDIII

8OHMa

d. Secondary structure & estimated accessibility

Key:- 〜⋀ Helix   ▭⟩ Beta strand   ══ Random coil   Accessibility shading: Black=buried, White=accessible

e. Sequence & Ramachandran regions   ▲ Most favoured   ■ Allowed   ▌ Generous   ▌ Disallowed

FFAVPQTFQVAHLHAPTGSGKSTKVPAAYAAQGYKVLVLIIPSVAATLGFGVYNSKAHGIDPIIIRTGVRAITTGGFITYSTYGKFLADGGCSGGAYDIIIC

Dengue

d. Secondary structure & estimated accessibility

Key:- 〜⋀ Helix   ▭⟩ Beta strand   ══ Random coil   Accessibility shading: Black=buried, White=accessible

e. Sequence & Ramachandran regions   ▲ Most favoured   ■ Allowed   ▌ Generous   ▌ Disallowed

AGVLWDVFSPFFYGKAELEDGAYRIKQKGILGYSQIGAGVYREGTFHTWWHVTRGAVLNHKGKRIEFSWADVKKDLISYGGGWKLEGEWKEGEEVQVLAL

West Nile virus

d. Secondary structure & estimated accessibility

Key:- 〜⋀ Helix   ▭⟩ Beta strand   ══ Random coil   Accessibility shading: Black=buried, White=accessible

e. Sequence & Ramachandran regions   ▲ Most favoured   ■ Allowed   ▌ Generous   ▌ Disallowed

GFEFENLEKKQITVLDLHFGAGKTRRILPQIIKEAIIIRRLRTAVLAFTRVVAAENAEALEGLPIRYQTSAVPREHIIGIIEIVDVWCHATLTHELLNISPHRVF

Japanese Encephalitis

d. Secondary structure & estimated accessibility

Key:- 〜⋀ Helix   ▭⟩ Beta strand   ══ Random coil   Accessibility shading: Black=buried, White=accessible

e. Sequence & Ramachandran regions   ▲ Most favoured   ■ Allowed   ▌ Generous   ▌ Disallowed

GGVFWDTPSPKPCSKGDTTTGVYRIMARGILGTYQAGVGVAEENVFHTLWHTTRGAAIMSGEGRKLTPYWGSVKEDRIAYGGPWRFDRKWNGTDDVQVIVV

Yellow Fever

d. Secondary structure & estimated accessibility

Key:- 〜⋀ Helix   ▭⟩ Beta strand   ══ Random coil   Accessibility shading: Black=buried, White=accessible

e. Sequence & Ramachandran regions   ▲ Most favoured   ■ Allowed   ▌ Generous   ▌ Disallowed

DVLWDIPTPKIIEECEHLEDGIYGIFQSTFLGASQRGVGVAQGGVFHTWHVTRGAFLVRNGKKLIPSWASVKEDLVAYGGSWKLEGRWDGEEFVQLIAA

Table App7a. The Dengue Polymerase Model's Ramachandran plot analysis summary.

```
+----------<<< P R O C H E C K   S U M M A R Y >>>----------+
| den   2.2                                      566 residues |
| Ramachandran plot:   79.7% core   16.9% allow    2.0% gener    1.4% disall |
| Gly & Pro Ramach:    4 labelled residues (out of  56)         |
| Chi1-chi2 plots:     0 labelled residues (out of 360)         |
+-----------------------------------------------------------+
```

Table App7b. The Dengue Polymerase Model's Ramachandran plot analysis summary. AFTER MOLECULAR DYNAMICS

```
+----------<<< P R O C H E C K   S U M M A R Y >>>----------+
| den   2.2                                      566 residues |
| Ramachandran plot:   80.9% core   16.7% allow    2.4% gener     0% disall |
| Gly & Pro Ramach:    4 labelled residues (out of  56)         |
| Chi1-chi2 plots:     0 labelled residues (out of 360)         |
+-----------------------------------------------------------+
```

Table App8a. The West Nile Polymerase Model's Ramachandran plot analysis summary.

```
+----------<<< P R O C H E C K   S U M M A R Y >>>----------+
| c:\dimitris\pdbs\wnv   2.0                      566 residues |
| Ramachandran plot:   68.2% core   24.6% allow    6.1% gener    1.0% disall |
| Gly & Pro Ramach:    4 labelled residues (out of  76)         |
| Chi1-chi2 plots:     8 labelled residues (out of 346)         |
+-----------------------------------------------------------+
```

Table App8b. The West Nile Polymerase Model's Ramachandran plot analysis summary. AFTER MOLECULAR DYNAMICS

```
+----------<<< P R O C H E C K   S U M M A R Y >>>----------+
| c:\dimitris\pdbs\wnv   2.0                      566 residues |
| Ramachandran plot:   89.8% core   16.7% allow    2.4% gener     0% disall |
| Gly & Pro Ramach:    4 labelled residues (out of  76)         |
| Chi1-chi2 plots:     8 labelled residues (out of 346)         |
+-----------------------------------------------------------+
```

Table App9a. The Japanese Encephalitis Polymerase Model's Ramachandran plot analysis summary

```
+----------<<< P R O C H E C K   S U M M A R Y >>>----------+
| je   2.2                                       566 residues |
| Ramachandran plot:   68.7% core   24.1% allow    6.4% gener    .8% disall |
| Gly & Pro Ramach:    6 labelled residues (out of  78)         |
| Chi1-chi2 plots:    11 labelled residues (out of 341)         |
+-----------------------------------------------------------+
```

Table App9b. The Japanese Encephalitis Polymerase Model's Ramachandran plot analysis summary AFTER MOLECULAR DYNAMICS

```
+----------<<< P R O C H E C K   S U M M A R Y >>>----------+
| je   2.2                                       566 residues |
| Ramachandran plot:   92.1% core    4.9% allow    3.0% gener     0% disall |
| Gly & Pro Ramach:    6 labelled residues (out of  78)         |
| Chi1-chi2 plots:    11 labelled residues (out of 341)         |
+-----------------------------------------------------------+
```

Table App10a. The Yellow Fever Polymerase Model's Ramachandran plot analysis summary.

```
+----------<<< P R O C H E C K    S U M M A R Y >>>----------+
| c:\dimitris\pdbs\yf   2.0                          566 residues |
| Ramachandran plot:   67.5% core   25.3% allow    6.6% gener    .6% disall |
| Gly & Pro Ramach:     5 labelled residues (out of  63)              |
| Chi1-chi2 plots:      4 labelled residues (out of 348)              |
+----------------------------------------------------------------+
```

Table App10b. The Yellow Fever Polymerase Model's Ramachandran plot analysis summary. AFTER MOLECULAR DYNAMICS

```
+----------<<< P R O C H E C K    S U M M A R Y >>>----------+
| c:\dimitris\pdbs\yf   2.0                          566 residues |
| Ramachandran plot:   89.4% core    8.5% allow    3.1% gener    0% disall |
| Gly & Pro Ramach:     5 labelled residues (out of  63)              |
| Chi1-chi2 plots:      4 labelled residues (out of 348)              |
+----------------------------------------------------------------+
```

C. Protein – DNA contacts in 1A1V

Analytic Protein-DNA/RNA Contacts

The list of contacting atom pairs:

```
41 PRO (230) A  O    <>  435 OURA ( 7) B  C5' D= 2.82 (B-S)
43 VAL (232) A  CG2  <>  435 OURA ( 7) B  P   D= 3.97 (S-B)
43 VAL (232) A  CG2  <>  435 OURA ( 7) B  C3' D= 3.57 (S-B)
43 VAL (232) A  CG2  <>  435 OURA ( 7) B  O5' D= 3.18 (S-S)
43 VAL (232) A  CG2  <>  436 OURA ( 8) B  P   D= 3.70 (S-B)
43 VAL (232) A  CG2  <>  436 OURA ( 8) B  OP1 D= 3.10 (S-B)
66 GLY (255) A  N    <>  436 OURA ( 8) B  P   D= 3.60 (B-B)
66 GLY (255) A  N    <>  436 OURA ( 8) B  OP2 D= 3.06 (B-S)
66 GLY (255) A  CA   <>  436 OURA ( 8) B  P   D= 3.93 (B-B)
80 THR (269) A  CB   <>  436 OURA ( 8) B  OP2 D= 3.28 (S-S)
80 THR (269) A  OG1  <>  436 OURA ( 8) B  OP2 D= 2.42 (S-S)
82 GLY (271) A  CA   <>  436 OURA ( 8) B  C5' D= 3.39 (B-S)
82 GLY (271) A  C    <>  436 OURA ( 8) B  C5' D= 3.00 (B-S)
82 GLY (271) A  C    <>  436 OURA ( 8) B  C4' D= 3.59 (B-S)
83 LYS (272) A  N    <>  436 OURA ( 8) B  C5' D= 2.94 (B-S)
83 LYS (272) A  CA   <>  436 OURA ( 8) B  C5' D= 3.52 (B-S)
86 ALA (275) A  CB   <>  436 OURA ( 8) B  O3' D= 3.34 (S-S)
179 HIS (369) A  CD2  <>  431 OURA ( 3) B  C5' D= 3.64 (S-S)
179 HIS (369) A  NE2  <>  431 OURA ( 3) B  P   D= 3.61 (S-B)
179 HIS (369) A  NE2  <>  431 OURA ( 3) B  OP2 D= 2.76 (S-S)
179 HIS (369) A  NE2  <>  431 OURA ( 3) B  C5' D= 3.73 (S-S)
179 HIS (369) A  O    <>  432 OURA ( 4) B  C5' D= 2.96 (B-S)
179 HIS (369) A  O    <>  432 OURA ( 4) B  C4' D= 3.37 (B-S)
180 SER (370) A  CA   <>  432 OURA ( 4) B  OP2 D= 3.17 (B-S)
180 SER (370) A  CA   <>  432 OURA ( 4) B  C5' D= 3.67 (B-S)
180 SER (370) A  CB   <>  432 OURA ( 4) B  OP2 D= 3.12 (S-S)
181 LYS (371) A  N    <>  432 OURA ( 4) B  OP2 D= 2.93 (B-S)
203 ARG (393) A  N    <>  433 OURA ( 5) B  P   D= 3.90 (B-B)
203 ARG (393) A  N    <>  433 OURA ( 5) B  OP2 D= 2.80 (B-S)
203 ARG (393) A  CB   <>  433 OURA ( 5) B  P   D= 3.66 (S-B)
203 ARG (393) A  CB   <>  433 OURA ( 5) B  OP1 D= 3.36 (S-B)
203 ARG (393) A  CB   <>  433 OURA ( 5) B  OP2 D= 3.41 (S-S)
203 ARG (393) A  NE   <>  433 OURA ( 5) B  C2' D= 3.08 (S-S)
203 ARG (393) A  NE   <>  434 OURA ( 6) B  OP1 D= 3.06 (S-B)
203 ARG (393) A  NE   <>  434 OURA ( 6) B  C5  D= 3.68 (S-S)
203 ARG (393) A  CZ   <>  434 OURA ( 6) B  OP1 D= 3.30 (S-B)
203 ARG (393) A  CZ   <>  434 OURA ( 6) B  C6  D= 3.47 (S-S)
203 ARG (393) A  CZ   <>  434 OURA ( 6) B  C5  D= 3.48 (S-S)
203 ARG (393) A  NH2  <>  434 OURA ( 6) B  P   D= 3.71 (S-B)
203 ARG (393) A  NH2  <>  434 OURA ( 6) B  OP1 D= 2.92 (S-B)
203 ARG (393) A  NH2  <>  434 OURA ( 6) B  O5' D= 3.28 (S-S)
203 ARG (393) A  NH2  <>  434 OURA ( 6) B  C2' D= 3.50 (S-S)
203 ARG (393) A  NH2  <>  434 OURA ( 6) B  C6  D= 3.00 (S-S)
203 ARG (393) A  NH2  <>  434 OURA ( 6) B  C5  D= 3.52 (S-S)
221 THR (411) A  CB   <>  432 OURA ( 4) B  O3' D= 3.43 (S-B)
221 THR (411) A  OG1  <>  433 OURA ( 5) B  P   D= 3.46 (S-B)
221 THR (411) A  OG1  <>  433 OURA ( 5) B  OP2 D= 2.54 (S-S)
223 ALA (413) A  N    <>  433 OURA ( 5) B  C5' D= 3.48 (B-S)
223 ALA (413) A  CA   <>  433 OURA ( 5) B  C3' D= 3.56 (B-S)
223 ALA (413) A  CA   <>  433 OURA ( 5) B  C5' D= 3.76 (B-S)
223 ALA (413) A  CA   <>  433 OURA ( 5) B  C4' D= 3.77 (B-S)
223 ALA (413) A  CB   <>  433 OURA ( 5) B  C3' D= 3.82 (S-B)

223 ALA (413) A  CB   <>  433 OURA ( 5) B  OP2 D= 3.03 (S-S)
238 VAL (432) A  CG1  <>  431 OURA ( 3) B  C4' D= 3.73 (S-S)
238 VAL (432) A  CG2  <>  431 OURA ( 3) B  C5' D= 3.55 (S-S)
238 VAL (432) A  CG2  <>  431 OURA ( 3) B  C4' D= 3.54 (S-S)
240 GLN (434) A  CG   <>  432 OURA ( 4) B  C4  D= 3.82 (S-S)
240 GLN (434) A  CG   <>  432 OURA ( 4) B  O4  D= 2.72 (S-S)
240 GLN (434) A  CD   <>  432 OURA ( 4) B  O4  D= 3.02 (S-S)
240 GLN (434) A  NE2  <>  432 OURA ( 4) B  O4  D= 2.89 (S-S)
254 THR (448) A  CG2  <>  432 OURA ( 4) B  C5  D= 3.75 (S-S)
254 THR (448) A  CG2  <>  432 OURA ( 4) B  C4  D= 3.83 (S-S)
254 THR (448) A  CG2  <>  432 OURA ( 4) B  O4  D= 3.35 (S-S)
256 THR (450) A  OG1  <>  430 OURA ( 2) B  O3' D= 2.95 (S-B)
306 TRP (501) A  CG   <>  436 OURA ( 8) B  C2  D= 3.81 (S-S)
306 TRP (501) A  CD2  <>  436 OURA ( 8) B  N1  D= 3.67 (S-S)
306 TRP (501) A  CD2  <>  436 OURA ( 8) B  C2  D= 3.53 (S-S)
306 TRP (501) A  NE1  <>  436 OURA ( 8) B  N3  D= 3.45 (S-S)
306 TRP (501) A  CE2  <>  436 OURA ( 8) B  C4  D= 3.55 (S-S)
306 TRP (501) A  CE2  <>  436 OURA ( 8) B  N3  D= 3.33 (S-S)
306 TRP (501) A  CE2  <>  436 OURA ( 8) B  C2  D= 3.55 (S-S)
306 TRP (501) A  CE3  <>  436 OURA ( 8) B  C1' D= 3.73 (S-S)
306 TRP (501) A  CE3  <>  436 OURA ( 8) B  N1  D= 3.59 (S-S)
306 TRP (501) A  CE3  <>  436 OURA ( 8) B  C6  D= 3.85 (S-S)
306 TRP (501) A  CZ2  <>  436 OURA ( 8) B  C6  D= 3.75 (S-S)
306 TRP (501) A  CZ2  <>  436 OURA ( 8) B  C5  D= 3.35 (S-S)
306 TRP (501) A  CZ2  <>  436 OURA ( 8) B  C4  D= 3.21 (S-S)
306 TRP (501) A  CZ2  <>  436 OURA ( 8) B  N3  D= 3.54 (S-S)
306 TRP (501) A  CZ3  <>  436 OURA ( 8) B  N1  D= 3.75 (S-S)
306 TRP (501) A  CZ3  <>  436 OURA ( 8) B  C6  D= 3.48 (S-S)
306 TRP (501) A  CH2  <>  436 OURA ( 8) B  C6  D= 3.44 (S-S)
306 TRP (501) A  CH2  <>  436 OURA ( 8) B  C5  D= 3.29 (S-S)
306 TRP (501) A  CH2  <>  436 OURA ( 8) B  C4  D= 3.72 (S-S)
307 TYR (502) A  OH   <>  436 OURA ( 8) B  O2  D= 2.93 (S-S)
361 ASN (556) A  OD1  <>  433 OURA ( 5) B  O4  D= 2.83 (S-S)
```

D. Protein – DNA contacts by LigPlot

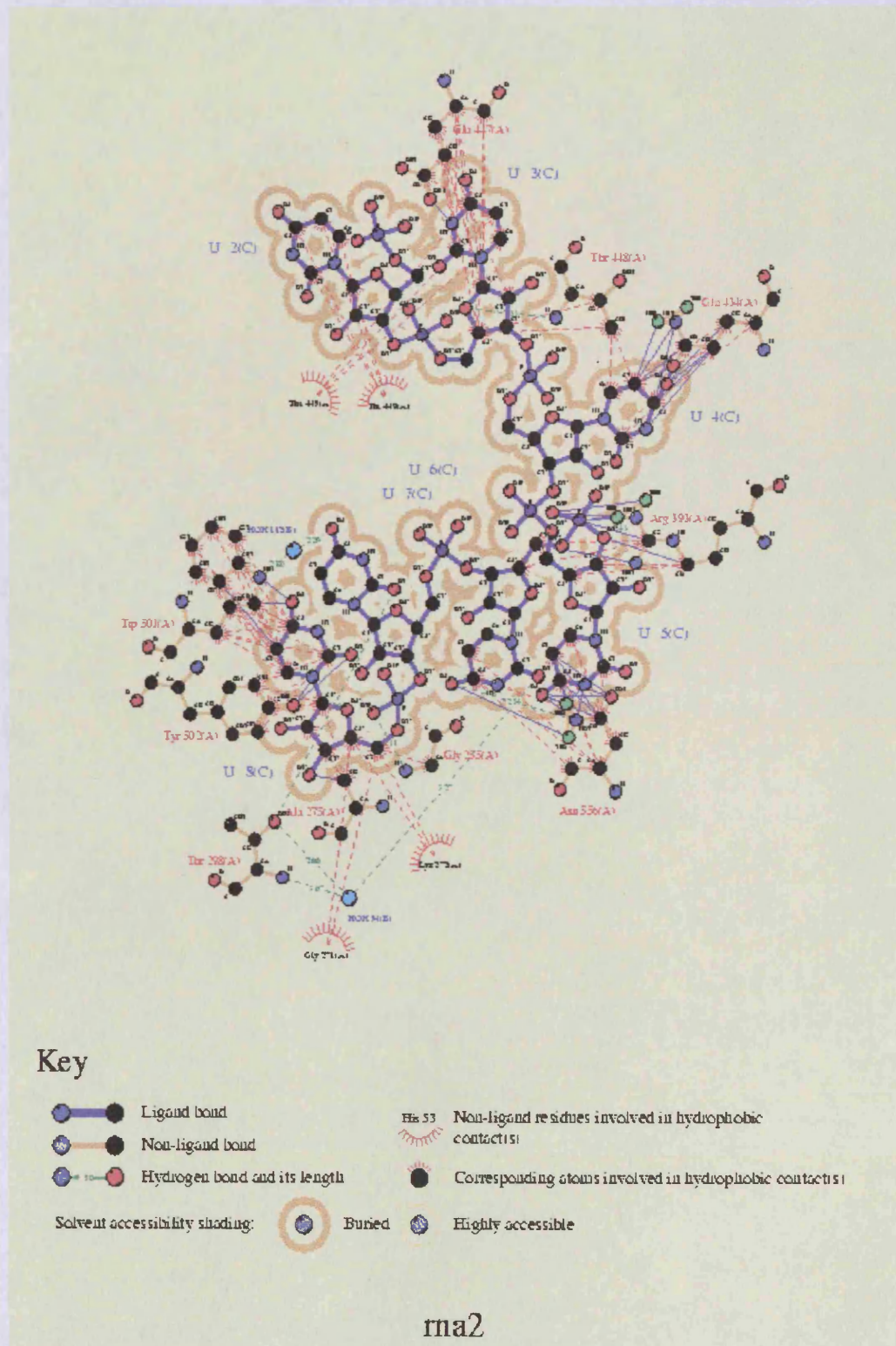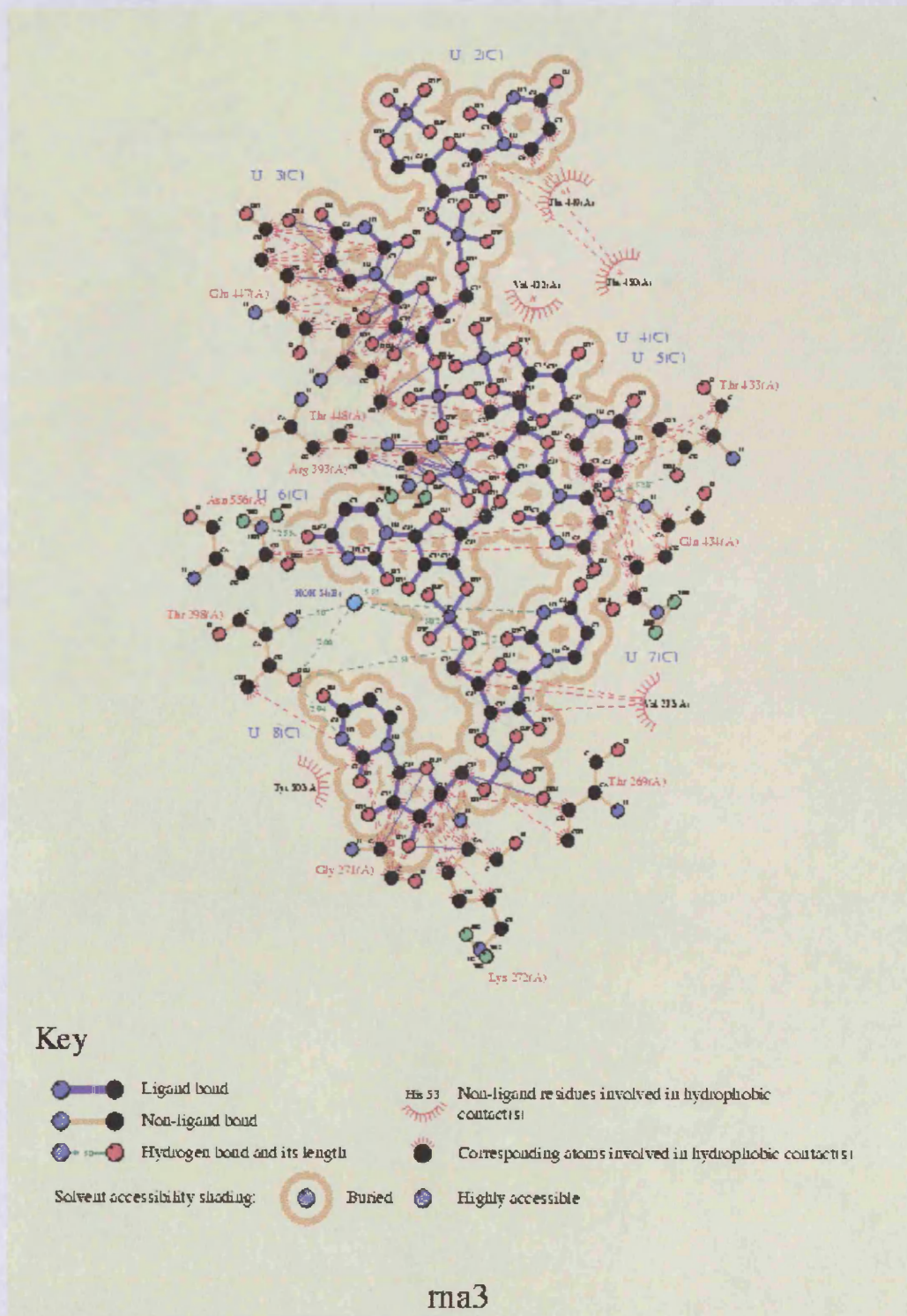**Figure App3.** Interactions between the ssRNA and the helicase' s channel – map1

Figure App4. Interactions between the ssRNA and the helicase's channel – map2

Figure App5. Interactions between the ssRNA and the helicase's channel – map3

Figure App6. Interactions between single nucleic acids of the ssRNA and the helicase' s channel – maps 4-10

E. Molecular Biology Methodology

BACTERIAL TRANSFORMATION:

1. 1 µL of plasmid DNA added to cells

2. Leave on ice for 45 minutes

3. Heat in waterbath at 42 °C for 2 minutes

4. Add 0.5 µL of LB medium with no antibiotic

5. Leave on ice for 1 hour at 37 °C

6. Microcentrifuge at 8000 rpm for 1minute

7. Take off and discard 0.5 mL of supernatant

8. Re-suspend gently in remaining volume

9. Plate on LB agar + AMP Petri dishes

E.COLI CULTURE

1. Pick up a healthy looking (round, consistent) colony with a
   loop

2. Inoculate in a universal tube with 10 mL of LB + AMP

3. Incubate in shaking centrifuge at 37 °C for 13 to 14 hours

MINI-PREP PREPARATION

1. Add 1.5 mL of overnight culture of E.Coli in an eppendorf and
   spin at 13000 for 1 minute

2. Repeat 3 times for each tube → passing a total of 4.5 mL of
   cell suspension from each tube

3. Each time discard supernatant

MINI – PREP

using a microcentrifuge

This protocol is designed for purification of up to 20 μg of high-copy

plasmid DNA from1–5 mL overnight cultures of *E. coli* in LB (Luria-

Bertani) medium.

Note: All protocol steps should be carried out at room temperature.

Procedure

1. Resuspend pelleted bacterial cells in 250 μL Buffer P1 and transfer to a microcentrifuge tube.

2. Ensure that RNase A has been added to Buffer P1. No cell clumps should be visible after resuspension of the pellet.

3. Add 250 μL Buffer P2 and gently invert the tube 4–6 times to mix.

4. Mix gently by inverting the tube. Do not vortex, as this will result in shearing of genomic DNA. If necessary, continue inverting the tube until the solution becomes viscous and slightly clear. Do not allow the lysis reaction to proceed for more than 5 min.

5. Add 350 μL Buffer N3 and invert the tube immediately but gently 4–6 times.

6. To avoid localized precipitation, mix the solution gently but thoroughly, immediately after addition of Buffer N3. The solution should become cloudy.

7. Centrifuge for 10 min at 13.000 rpm (~17.900 x g) in a table-top microcentrifuge.

8. A compact white pellet will form.

9. Apply the supernatants from step 4 to the QIAprep Spin Column by decanting or pipetting.

10. Centrifuge for 30–60 s. Discard the flow-through.

11. (Optional): Wash the QIAprep Spin Column by adding 0.5 mL Buffer PB and centrifuging for 30–60 s. Discard the flow-through.

12. This step is necessary to remove trace nuclease activity when using endA+ strains such as the JM series, HB101 and its derivatives, or any wild-type strain, which have high levels of nuclease activity or high carbohydrate content. Host strains such as XL-1 Blue and DH5α™ do not require this additional wash step.

13. Wash QIAprep Spin Column by adding 0.75 mL Buffer PE and centrifuging for 30–60 s.

14. Discard the flow-through, and centrifuge for an additional 1 min to remove residual wash buffer.

15. IMPORTANT: Residual wash buffer will not be completely removed unless the flow-through is discarded before this

additional centrifugation. Residual ethanol from Buffer PE may inhibit subsequent enzymatic reactions.

16. Place the QIAprep column in a clean 1.5 mL microcentrifuge tube. To elute DNA, add 50 μL Buffer EB (10 mM Tris·Cl, pH 8.5) or water to the center of each QIAprep

17. Spin Column, let stand for 1 min, and centrifuge for 1 min.

DNA CONCENTRATION PROTOCOL

Materials:

1. TE solution

   o 10 mM Tris (pH to 7.5)

   o 1 mM EDTA (pH to 8.0 to dissolve)

2. DNA sample

3. SYBR Green I(R) nucleic acid gel stain (Molecular Probes)

4. Plastic wrap

5. distilled water

6. DNA marker stock (10 mg/mL)

Supplies:

1. Tubes

2. Polaroid setup (with proper filter - SYBR Green/Gold gel stain photographic filter) and UV light box

3. Micropipetter and tips

Procedures:

1. Prepare 6 DNA standards from DNA marker stock:

   o standard I (5 ug/ul): 1:2 dilution of DNA marker stock

   o standard II (2.5 ug/ul): 1:2 dilution of standard I

   o standard III (1.25 ug/ul): 1:2 dilution of standard II

   o standard IV (0.625 ug/ul): 1:2 dilution of standard III

- o standard V (0.313 ug/ul): 1:2 dilution of standard IV

- o standard VI (0.156 ug/ul): 1:2 dilution of standard V

2. Make a 1:5000 dilution of the SYBR Green I(R) with TE solution.

3. Mix 5 ul of the DNA sample and each of the 6 standards with 5 ul of the diluted SYBR Green I(R) dye.

4. Place a sheet of plastic wrap smoothly onto the UV light box.

5. Spot the mixtures individually onto plastic wrap.

6. Spot the set of 6 standards.

7. Turn on the UV light and take a photo (Polaroid 667 black-and-white print film).
8. Compare the brightness of the DNA sample with the DNA standards and estimate concentration.

LIGATION - PCR

- Add 1 μL of the vector

- Add 1 μL of 10x buffer

- Add PCR product (1 μL)

- Add 8 μL of SIGMA water

- Add 1 μL of Ligase

LIGATION – OVERNIGHT

- Add 0.5 μL of the vector

- Add 5 μL of 10x buffer

- Add 3.5 μL Insert DNA

- Add 1 μL of Ligase (last)

- Incubate overnight at 17 °C

LIGATION – BENCHTOP

- Add 0.5 µL of the vector

- Add 5 µL of 10x buffer

- Add 3.5 µL Insert DNA

- Add 1 µL of Ligase (last)

- Leave at room temperature for 2 hours + 30 minutes.

<u>SDS-Page Gel Preparation</u>

	10% Running Gel	Stacking
Solution		
Acrylamide	3.3 mL	696 µl
1.5M Tris-HCl pH8.8	2.5 mL	---
0.5M Tris HCl pH6.8	---	650 µl
10% SDS	100 µl	100 µl
10% APS	50 µl	50 µl
TEMED	20 µl	10 µl
Water (D)	4.0 mL	3.65 mL

→ The Running Buffer is made by preparing 200 mL SDS Page Buffer (containing 288g Glycine and 80g Tris). 20 mL of SDS 10% buffer into 2000mL of dH_2O will give the SDS buffer.

2.1.1 Molecular weight standards

Prestained high range molecular weight protein standards were purchased from Bio-Rad Laboratories, Hertfordshire, U.K or New England BioLabs, Beverly, MA, USA. The λ *Hind* III and φX174 *Hae* III DNA markers were purchased from Promega Ltd., Southampton, UK.

High range protein molecular weight markers (Bio-Rad)

Myosin (H chain)	200.000 Da
Phosphorylase B	97.400 Da
Bovine serum albumin	68.000 Da
Ovalbumin	43.000 Da
Carbonic Anhydrase	31.000 Da
β Lactoglobulin	18.400 Da
Lysozyme	14.300 Da

High range protein molecular weight markers –(New England Biolabs)

MBP-β-galactosidase	175.000 Da
MBP-paramyosin	83.000 Da
Glutamic dehydrogenase	62.000 Da
Aldolase	47.500 Da
Trisosephosphate isomerase	32.500 Da
β-Lactoglobulin A	25.000 Da
Lysozyme	16.500 Da
Aprotinin	6.500 Da

F. Sequencing

>DENGUE--NS3
```
agvlwdvpsp ppvgkaeled gayrikqkgi lgysqigagv ykegtfhtmw hvtrgavlmh
kgkriepswa dvkkdlisyg ggwklegewk egeevqvlal epgknpravq tkpglfrtnt
gtigavsldf spgtsgspiv dkkgkvvgly gngvvtrsga yvsaiaqtek siednpeied
difrkrrlti mdlhpgagkt krylpaivre aikrglrtli laptrvvaae meealrglpi
ryqtpairae htgreivdlm chatftmrll spirvpnynl iimdeahftd pasiaargyi
strvemgeaa gifmtatppg srdpfpqsna pimdeereip erswnsghew vtdfkgktvw
fvpsiktgnd iaaclrkngk rviqlsrktf dseyvktrtn dwdfvvttdi semganfkae
rvidprrcmk pviltdgeer vilagpmpvt hssaaqrrgr igrnprnend qyiymgeple
ndedcahwke akmLldnint pegiipsmfe perekvdaid geyrlrgear ktfvdlmrrg
dlpvwlaykv aaeginyadr rwcfdgtrnn qileenveve iwtkegerkk lkprwldari
ysdplalkef aagrk
```

>Yellow fever virus--NS3
```
sgdvlwdipt pkiieecehl edgiygifqs tflgasqrgv gvaqggvfht mwhvtrgafl
vrngkklips wasvkedlva yggswklegr wdgeeevqli aavpgknvvn vqtkpslfkv
rnggeigava ldypsgtsgs pivnrngevi glygngilvg dnsfvsaisq tevkeegkee
lqeiptmLkk gmttvldfhp gagktrrflp qilaecarrr lrtlvlaptr vvlsemkeaf
hgldvkfhtq afsahgsgre vidamchatl tyrmLeptrv vnweviimde ahfldpasia
argwaahrar anesatilmt atppgtsdef phsngeiedv qtdipsepwn tghdwiladk
rptawflpsi raanvmaasl rkagksvvvl nrktfereyp tikqkkpdfi latdiaemga
nlcvervldc rtafkpvlvd egrkvaikgp lrisassaaq rrgrigrnpn rdgdsyyyse
ptsennahhv cwleasmLld nmevrggmva plygvegtkt pvspgemrlr ddqrkvfrel
vrncdlpvwl swqvakaglk tndrkwcfeg peeheilnds getvkcrapg gakkplrprw
cdervssdqs alsefikfae grr
```

>Japanese encephalitis virus--NS3
```
ggvfwdtpsp kpcskgdttt gvyrimargi lgtyqagvgv myenvfhtlw httrgaaims
gegkltpywg svkedriayg gpwrfdrkwn gtddvqvivv epgkaavniq tkpgvfrtpf
gevgavsldy prgtsgspil dsngdiigly gngvelgdgs yvsaivqgdr qeepvpeayt
pnmLrkrqmt vldlhpgsgk trkilpqiik daiqqrlrta vlaptrvvaa emaealrglp
vryqtsavqr ehqgneivdv mchatlthrl mspnrvpnyn lfvmdeahft dpasiaargy
iatkvelgea aaifmtatpp gttdpfpdsn apihdlqdei pdrawssgye witeyagktv
wfvasvkmgn eiamclqrag kkviqlnrks ydteypkckn gdwdfvittd isemganfga
srvidcrksv kptileegeg rvilgnpspi tsasaaqrrg rvgrnpnqvg deyhyggats
eddsnlahwt eakimLdnih mpnglvaqly gperekaftm dgeyrlrgee kknflellrt
adlpvwlayk vasngiqytd rrwcfdgprt nailedntev eivtrmgerk ilkprwldar
vyadhqalkw fkdfaagkr
```

>west nile--NS3
```
ggvlwdtpsp keykkgdttt gvyrimtrgl lgsyqagagv mvegvfhtlw httkgaalms
gegrldpywg svkedrlcyg gpwklqhkwn ghdevqmivv epgknvknvq tkpgvfktpe
geigavtldy ptgtsgspiv dkngdvigly gngvimpngs yisaivqger meepapagfe
pemLrkkqit vldlhpgagk trkilpqiik eainkrlrta vlaptrvvaa emsealrglp
iryqtsavhr ehsgneivdv mchatlthrl msphrvpnyn lfimdeahft dpasiaargy
iatkvelgea aaifmtatpp gtsdpfpesn apisdmqtei pdrawntgye witeyvgktv
wfvpsvkmgn eialclqrag kkviqlnrks yeteypkckn ddwdfvittd isemganfka
srvidsrksv kptiieegdg rvilgepsai taasaaqrrg rigrnpsqvg deycygghtn
eddsnfahwt earimLdnin mpnglvaqly qperekvytm dgeyrlrgee rknfleflrt
adlpvwlayk vaaagisyhd rkwcfdgprt ntilednnev evitklgerk ilrprwadar
vysdhqalks fkdfasgkr
```

>Hepatitis C--NS3
```
MVDFIPVENLETTMRSPVFTDNSSPPAVPQSFQVAHLHAPTGSGKSTKVPAAYAAQGYKVLVLNPSVAA
TLGFGAYMSKAHGVDPNIRTGVRTITTGSPITYSTYGKFLADGGCSGGAYDIIICDECHSTDATSILGI
GTVLDQAETAGARLVVLATATPPGSVTVPHPNIEEVALSTTGEIPFYGKAIPLEVIKGGRHLIFCHSKK
KCDELAAKLVALGINAVAYYRGLDVSVIPTSGDVVVVATDALMTGFTGDFDSVIDCNTCVTQTVDFSLD
PTFTIETTTLPQDAVSRTQRRGRTGRGKPGIYRFVAPGERPSGMFDSSVLCECYDAGCAWYELTPAETT
VRLRAYMNTPGLPVCQDHLEFWEGVFTGLTHIDAHFLSQTKQSGENFPYLVAYQATVCARAQAPPPSWD
QMWKCLIRLKPTLHGPTPLLYRLGAVQNEVTLTHPITKYIMTCMSADLEVVTGSGSHHHHHH
```

>west nile—NS5

```
GGAKGRTLGEVWKERLNQMTKEEFTRYRKEAIIEVDRSAAKHARKEGNVTGGHPVSRGTAKLRWLVERRF
LEPVGKVIDLGCGRGGWCYYMATQKRVQEVRGYTKGGPGHEEPQLVQSYGWNIVTMKSGVDVFYRPSECC
DTLLCDIGESSSSAEVEEHRTIRVLEMVEDWLHRGPREFCVKVLCPYMPKVIEKMELLQRRYGGGLVRNP
LSRNSTHEMYWVSRASGNVVHSVNMTSQVLLGRMEKRTWKGPQYEEDVNLGSGTRAVGKPLLNSDTSKIK
NRIERLRREYSSTWHHDENHPYRTWNYHGSYDVKPTGSASSLVNGVVRLLSKPWDTITNVTTMAMTDTTP
FGGQQRVFKEKVDTKAPEPPEGVKYVLNETTNWLWAFLAREKRPRMCSREEFIRKVNSNAALGAMFEEQNQ
WRSAREAVEDPKFWEMVDEEREAHLRGECHTCIYNMMGKREKKPGEFGKAKGSRAIWFMWLGARFLEFEA
LGFLNEDHWLGRKNSGGGVEGLGLQKLGYILREVGIRPGGKIYADDTAGWDTRITRADLENEAKVLELLD
GEHRRLARAIIELTYRHKVVKVMRPAADGRTVMDVISREDQRGSGQVVTYALNTFTNLAVQLVRMMEGEG
VIGPDDVEKLTKGKGPKVRTWLFENGEERLSRMAVSGDDCVVKPLDDRFATSLHFLNAMSKVRKDIQEWK
PSTGWYDWQQVPFCSNHFTELIMKDGRTLVVPCRGQDELVGRARISPGAGWNVRDTACLAKSYAQMWLLL
YFHRRDLRLMANAICSAVPVNWVPTGRTTWSIHAGGEWMTTEDMLEVWNRVWIEENEWMEDKTPVEKWSD
VPYSGKREDIWCGSLIGTRARATWAENIQVAINQVRAIIGDEKYVDYMSSLKRYEDTTLVEDTVL
```


\>Yellow fever virus–NS5
```
GTANGKTLGEVWKRELNLLDKQQFELYKRTDIVEVDRDTARRHLAEGKVDTGVAVSRGTAKLRWFHERGY
VKLEGRVIDLGCGRGGWCYYAAAQKEVSGVKGFTLGRDGHEKPMNVQSLGWNIITFKDKTDIHRLEPVKC
DTLLCDIGESSSSSVTEGERTVRVLDTVEKWLACGVDNFCVKVLAPYMPDVLEKLELLQRRFGGTVIRNP
LSRNSTHEMYYVSGARSNVTFTVNQTSRLLMRRMRRPTGKVTLEADVTLPIGTRSVETDKGPLDKEAIEE
RVERIKSEYMTSWFYDNDNPYRTWHYCGSYVTKTSGSAASMVNGVIKILTYPWDKIEEVTRMAMTDTTPF
GQQRVFKEKVDTRAKDPPAGTRKIMKVVNRWLFRHLAREKNPRLCTKEEFIAKVRSHAAIGAYLEEQEQW
KTANEAVQDPKFWELVDEERKLHQQGRCRTCVYNMMGKREKKLSEFGKAKGSRAIWYMWLGARYLEFEAL
GFLNEDHWASRENSGGGVEGIGLQYLGYVIRDLAAMDGGGFYADDTAGWDTRITEADLDDEQEILNYMSP
HHKKLAQAVMEMTYKNKVVKVLRPAPGGKAYMDVISRRDQRGSGQVVTYALNTITNLKVQLIRMAEAEMV
IHHQHVQDCDESVLTRLEAWLTEHGCNRLRRMAVSGDDCVVRPIDDRFGLALSHLNAMSKVRKDISEWQP
SKGWNDWENVPFCSHHFHELQLKDGRRIVVPCREQDELIGRGRVSPGNGWMIKETACLSKAYANMWSLMY
FHKRDMRLLSLAVSSAVPTSWVPQGRTTWSIHGKGEWMTTEDMLEVWNRVWITNNPHMQDKTMVKEWRDV
PYLTKRQDKLCGSLIGMTNRATWASHIHLVIHRIRTLVGQEKYTDYLTVMDRYSVDADLQPGELI
```


\>Japanese encephalitis virus–NS5
```
GRPGGRTLGEQWKEKLNAMSREEFFKYRREAIIEVDRTEARRARRENNIVGGHPVSRGSAKLRWLVEKGF
VSPIGKVIDLGCGRGGWSYYAATLKKVQEVRGYTKGGAGHEEPMLMQSYGRNLVSLKSGVDVFYKPSEPS
DTLFCDIGESSPSPEVEEQRTLRVLEMTSDWLHRGPREFCIKVLCPYMPKVIEKMEVLQRRFGGGLVRLP
LSRNSNHEMYWVSGAAGNVVHAVNMTSQVLLGRMDRTVWRGPKYEEDVNLGSGTRAVGKGEVHSNQEKIK
KRIQKLKEEFATTWHKDPEHPYRTWTYHGSYEVKATGSASSLVNGVVKLMSKPWDAIANVTTMAMTDTTP
FGGQQRVFKEKVDTKAPEPPAGAKEVLNETTNWLWAHLSREKRPRLCTKEEFIKKVNSNAALGAVFAEQNQ
WSTAREAVDDPRFWEMVDEERENHLRGECHTCIYNMMGKREKKPGEFGKAKGSRAIWFMWLGARYLEFEA
LGFLNEDHWLSRENSGGGVEGSGVQKLGYILRDIAGKQGGKMYADDTAGWDTRITRTDLENEAKVLELLD
GEHRMLARAIIELTYRHKVVKVMRPAAEGKTVMDVISREDQRGSGQVVTYALNTFTNIAVQLVRRLMEAEG
VIGPQHLEQLPRKTKIAVRTWLFENGEERVTRMAISGDDCVVKPLDDRFATALHFLNAMSKVRKDIQEWK
PSHGWHDWQQVPFCSNHFQEIVMKDGRSIVVPCRGQDELIGRARISPGAGWNVKDTACLAKAYAQMWLLL
YFHRRDLRLMANAICSAVPVDWVPTGRTSWSIHSKGEWMTTEDMLQVWNRVWIEENEWMMDKTPITSWTD
VPYVGKREDIWCGSLIGTRSRATWAENIYAAINQVRAVIGKENYVDYMTSLRRYEDVLIQEDRVI
```


\>Hepatitis C--NS5
```
SMSYTWTGALITPCAAEESKLPINPLSNSLLRHHNMVYATTSRSASLRQKKVTFDRLQVL
DDHYRDVLKEMKAKASTVKAKLLSIEEACKLTPPHSAKSKFGYGAKDVRNLSSRAVNHIR
SVWEDLLEDTETPIDTTIMAKSEVFCVQPEKGGRKPARLIVFPDLGVRVCEKMALYDVVS
TLPQAVMGSSYGFQYSPKQRVEFLVNTWKSKKCPMGFSYDTRCFDSTVTESDIRVEESIY
QCCDLAPEARQAIRSLTERLYIGGPLTNSKGQNCGYRRCRASGVLTTSCGNTLTCYLKAT
AACRAAKLQDCTMLVNGDDLVVICESAGTQEDAAALRAFTEAMTRYSAPPGDPPQPEYDL
ELITSCSSNVSVAHDASGKRVYYLTRDPTTPLARAAWETARHTPINSWLGNIIMYAPTLW
ARMILMTHFFSILLAQEQLEKALDCQIYGACYSIEPLDLPQIIERLHGLSAFTLHSYSPG
EINRVASCLRKLGVPPLRTWRHRARSVRAKLLSQGGRAATCGRYLFNWAVRTKLKLTPIP
AASQLDLSGWFVAGYSGGDIYHSLSRARPR
```


\>BVDV --NS5
```
AYLKLKDFIEEEEKKPRVKDTVIREHNKWILKKIRFQGNLNTKKXLNPGKLSEQLDREGR
KRNIYNHQIGTIXSSAGIRLEKLPIVRAQTDTKTFHEAIRDKIDKSENRQNPELHNKLLE
IFHTIAQPTLKHTYGEVTWEQLEAGVNRKGAAGFLEKKNIGEVLDSEKHLVEQLVRDLKA
GRKIKYYETAIPKNEKRDVSDDWQAGDLVVEKRPRVIQYPEAKTRLAITKVXYNWVKQQP
VVIPGYEGKTPLFNIFDKVRKEWDSFNEPVAVSFDTKAWDTQVTSKDLQLIGEIQKYYYK
KEWHKFIDTITDHXTEVPVITADGEVYIRNGQRGSGQPDTSAGNSXLNVLTXXYAFCEST
GVPYKSFNRVARIHVCGDDGFLITEKGLGLKFANKGXQILHEAGKPQKITEGEKXKVAYR
```

```
FEDIEFCSHTPVPVRWSDNTSSHXAGRDTAVILSKXATRLDSSGERGTTAYEKAVAFSFL
LXYSWNPLVRRICLLVLSQQPETDPSKHATYYYKGDPIGAYKDVIGRNLSELKRTGFEKL
ANLNLSLSTLGVWTKHTSKRIIQDCVAIGKEEGNWLVKPDRLISSKTGHLYIPDKGFTLQ
GKHYEQLQL
```


```
>Dengue -NS5
gtgntgetlg  ekwknrlnal  gksefqiykk  sgiqevdrtl  akegikrget  dhhavsrgsa
klrwfvernl  vtpegkvvdl  gcgrggwsyy  cgglknvkev  kgltkggpgh  eepipmstyg
wnlvrlqsgv  dvfftppekc  dtllcdiges  spnptveagr  tlrvlnlven  wlnnntqfci
kvlnpympsv  iekmealqrk  yggalvrnpl  srnsthemyw  vsnasgnivs  svnmisrmLi
nrftmrhkka  tyepdvdlgs  gtrnigiese  tpnldiigkr  iekikqehet  swhydqdhpy
ktwayhgsye  tkqtgsassm  vngvvrlltk  pwdiipmvtq  mamtdttpfg  qqrvfkekvd
trtqepkegt  kklmkitaew  lwkelgkkkt  prmctreeft  rkvrsnaalg  aiftdenkwk
sareavedsg  fwelvdkern  lhlegkcetc  vynmmgkrek  klgefgkakg  sraiwymwlg
arflefealg  flnedhwfsr  enslsgvege  glhklgyilr  dvskkeggam  yaddtagwdt
ritledlkne  emvtnhmege  hkklaeaifk  ltyqnkvvrv  qrptprgtvm  diisrrdqrg
sgqvvtygln  tftnmeaqli  rqmegegvfk  siqhltvtee  iavknwlvrv  grerlsrmai
sgddcvvkpl  ddrfasalta  lndmgkvrkd  iqqwepsrgw  ndwtqvpfcs  hhfhelimkd
grvlvvpcrn  qdeligrari  sqgagwslre  taclgksyaq  mwslmyfhrr  dlrlaanaic
savpshwvpt  srttwsihat  hewmttedmL  tvwnrvwiqe  npwmedktpv  esweeipylg
kredqwcgsl  igltsratwa  kniqtainqv  rsligneeyt  dympsmkrfr  reeeeagvlw
```
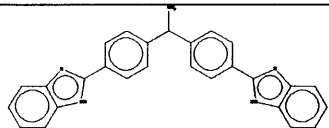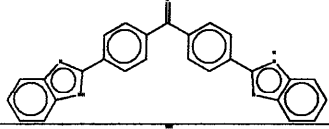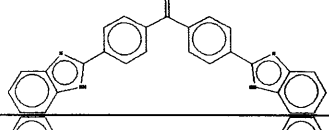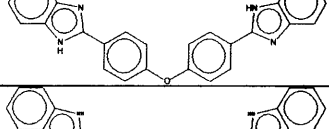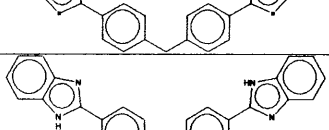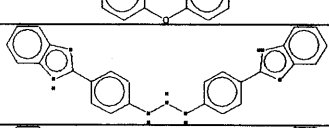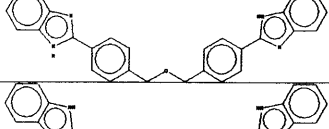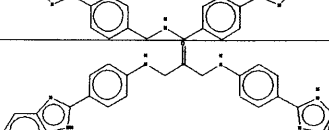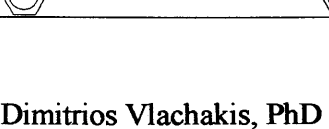
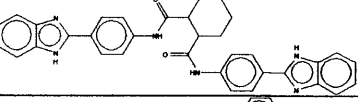## G. Virtual screening – lead optimisation
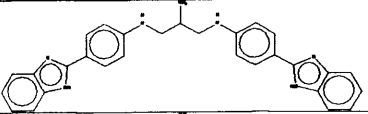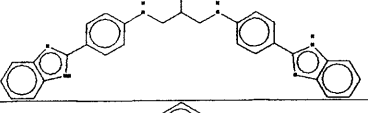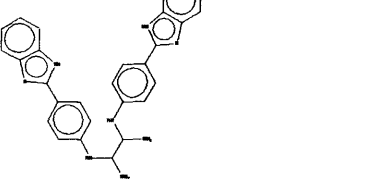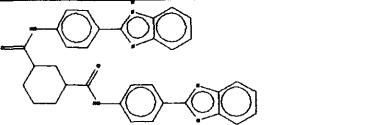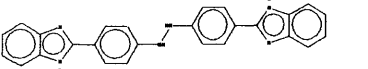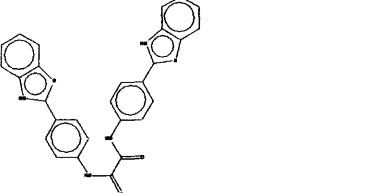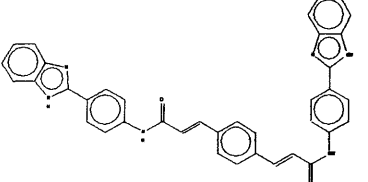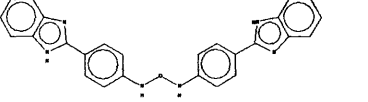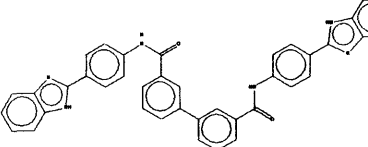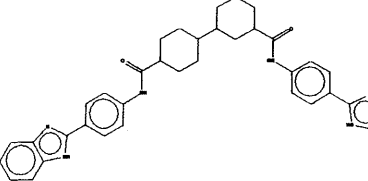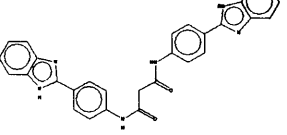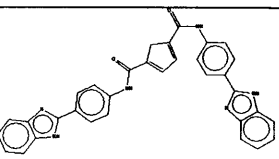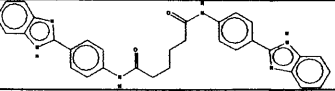
The symbols used stand for:
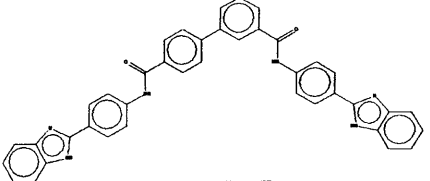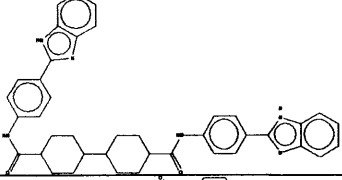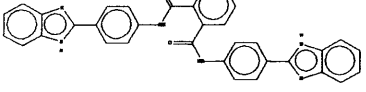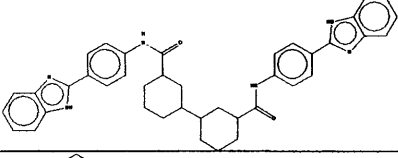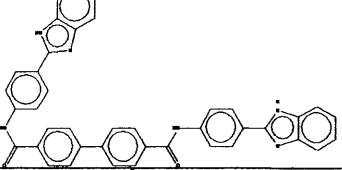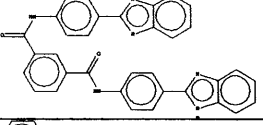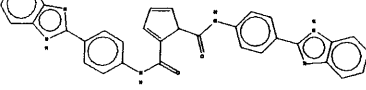
√   → picks up high interaction with the protein (H bonds) and hydrophobic interaction too. It has a good conformation on the protein as well.

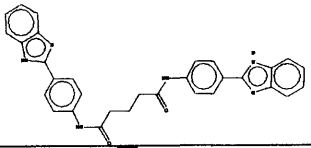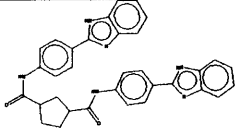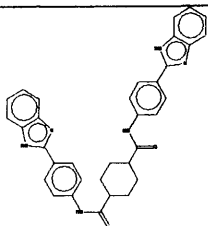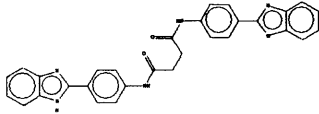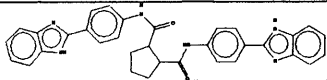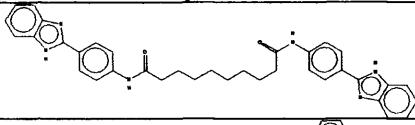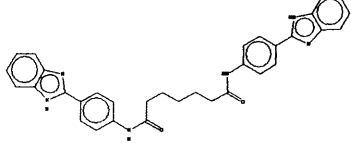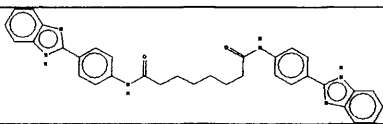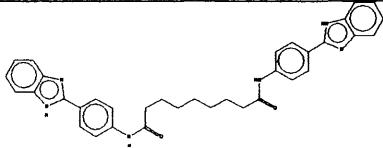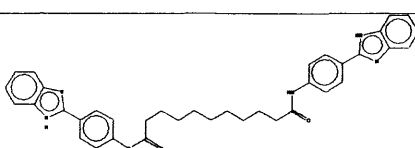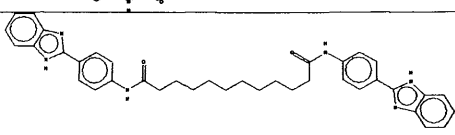~√   →Does not H-bond but it establishes hydrophobic interaction and has a good docking conformation.
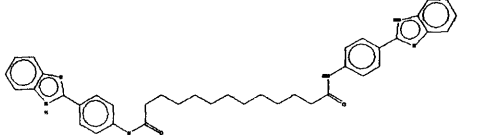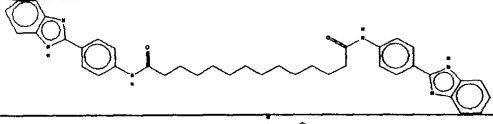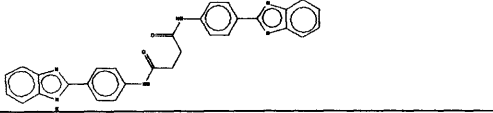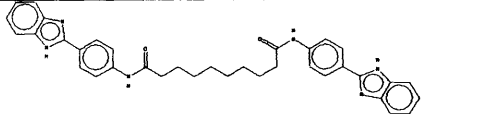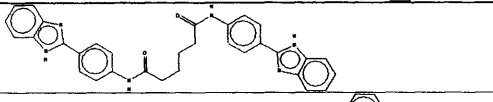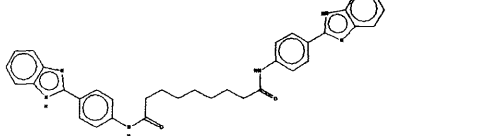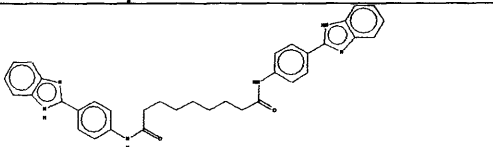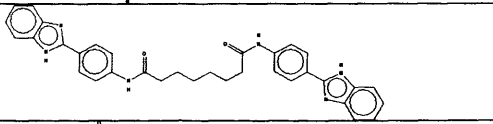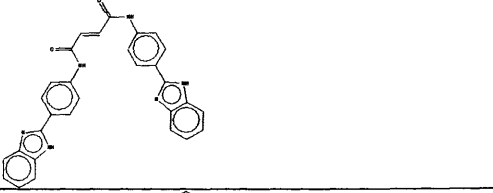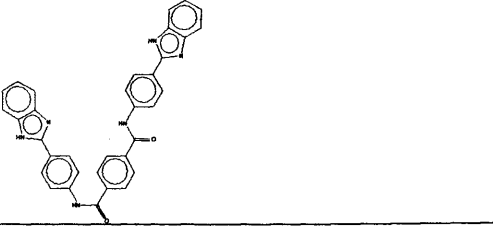
X   →Was impossible to dock the compound in a successful way. No interaction established + the conformation of the compound is bad.

STRUCTURE	Number	HepC	Dengue	WNV
	1	X	X	X
	2	X	X	X
	3	X	X	X
	4	X	X	X
	5	X	X	X
	6	X	X	X
	7	X	X	X
	8	X	X	X
	9	X	X	X
	10	X	X	X

STRUCTURE	Number	HepC	Dengue	WNV
	11	X	X	X
	12	X	X	X
	13	X	X	X
	14	√	X	X
	15	X	X	X
	16	~√	X	X
	17	~√	~√	X
	18	X	X	X
	19	X	X	X
	20	~√	X	X
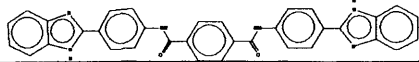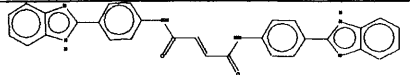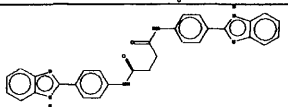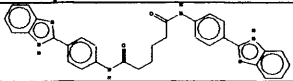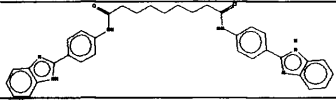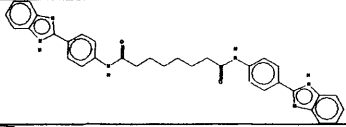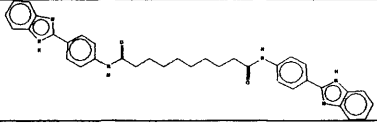
STRUCTURE	Number	HepC	Dengue	WNV
	21	X	X	X
	22	X	X	X
	23	X	X	X
	24	X	X	X
	25	~√	X	X
	26	X	X	X
	27	X	X	X
	28	X	X	X
	29	~√	√	√
	30	X	X	X
	31	√	√	~√

STRUCTURE	Number	HepC	Dengue	WNV
	32	√	√	~√
	33	X	X	X
	34	X	X	X
	35	~√	~√	~√
	36	X	X	X
	37	√	√	√
	38	√	√	√
	39	√	√	√
	40	√	√	√
	41	√	√	√
	42	√	√	√

STRUCTURE	Number	HepC	Dengue	WNV
	43	√	√	√
	44	√	√	√
	45	X	X	X
	46	√	√	√
	47	√	√	~√
	48	X	X	X
	48	X	X	X
	49	X	X	X
	50	X	X	X
	51	X	X	X

Evaluation of the dockings:

The above compounds have been designed by applying a variety of modifications to the following reported compounds by Virofarma Inc.:

CHEMICAL STRUCTURE	No.	IC$_{50}$
	1	10
	2	0.7
	3	0.7
	4	0.7
	5	0.7
	6	0.7
	7	0.7

The Virofarma compounds were subjected to a LigBuilder run using a variety of user defined seeds. A seed is the part of the compound that remains unchanged. The final compound will contain the seed and a variety of different substitutions added by the genetic algorithm of LigBuilder.

The 51 modifications of the original Virofarma compounds were docked using MOE to the HepC X-ray solved helicase and to the Dengue and West Nile helicase models. The docking parameters were kept constant to all dockings (same active site = same aligned residue number, 8000 iterations for 6 cycles). The aim of this study was to generate a wide variety of compounds *in silico*, dock them and look for a consensus pattern and behaviour in those compounds.

The only conclusion that could be drawn was the same as established by the enzymatic assay in chapter 3 of this work: long and flexible "linker" regions behave better than rigid ones. Still though, the number of compounds tested is not large enough to be able to draw more useful conclusions (such as different preferences of the three different receptors). A wider range of compounds has to be designed and a more sensitive and exhaustive molecular docking protocol has to be established in order to be able to distinct small differences in ligand preference between the three viral helicases.

chapter 6
BIBLIOGRAPHY

1.Verh K Acad Geneeskd Belg. (1999). Infections with flaviviridae. 61(6):661-97; discussion 697-9.

2.María G. Guzmán and Gustavo Kourí (2004). Dengue diagnosis, advances and challenges. International Journal of Infectious Diseases, Volume 8, Issue 2, Pages 69-80

3.Charles H. Calisher and Ernest A. Gould (2003). Taxonomy of the virus family Flaviviridae Advances in Virus Research, Volume 59, Pages 1-19

4.Kadare, G. & Haenni, A.L. (1997). Virus-encoded RNA helicases. J. Virol. 71, 2583–2590.

5.Francoise Degos (1994). Epidemiology of hepatitis C virus in Europe. FEMS Microbiology Reviews, Volume 14, Issue 3, Pages 267-271

6.L. Pasta, C. Marrone, M. D'amico, E.A. Bevacqua, G. D'Amico and L. Pagliaro (2005) Familial clustering of HCV-related liver cirrhosis and hepatocellular carcinoma. Digestive and Liver Disease, Volume 37, Issue 9, Pages 716-717

7.Marie-Pierre Courageot , Adeline Catteau, and Philippe Desprès (2003). Mechanisms of Dengue virus-induced cell death. Advances in Virus Research, Volume 60, Pages 157-186

8.Stephen J. Thomas , Daniel Strickman, and David W. Vaughn (2003). Dengue Epidemiology: Virus Epidemiology, Ecology, and Emergence. Advances in Virus Research, Volume 61 , Pages 235-289

9.Sadaf Shadan and BMN News (2004). Unfolding targets for dengue fever. Drug Discovery Today, Volume 9, Issue 8, Page 344

10. Tian Wang and Erol Fikrig (2004). Immunity to West Nile virus. Current Opinion in Immunology, Volume 16, Issue 4, Pages 519-523

11. Marc S. Collett (1992). Molecular genetics of pestiviruses. Comparative Immunology, Microbiology and Infectious Diseases, Volume 15, Issue 3, Pages 145-154

12. Ramiro Avalos-Ramirez, Michaela Orlich, Heinz-Jürgen Thiel and Paul Becher (2001), Evidence for the Presence of Two Novel Pestivirus Species. Virology, Volume 286, Issue 2, 1 August 2001, Pages 456-465

13. Colin W Shepard, Lyn Finelli and Miriam J Alter (2005). Global epidemiology of hepatitis C virus infection. The Lancet Infectious Diseases, Volume 5, Issue 9, Pages 558-567

14. K. Venugopal and E. A. Gould (1994). Towards a new generation of flavivirus vaccines. Vaccine, Volume 12, Issue 11, Pages 966-975

15. Walker, M. 1999. Hepatitis C virus: an overview of current approaches and progress. Drug Discovery Today. Vol.4, No.11

16. Sudo, K. Matsumoto, Y. Matsushima, M. Fujiwara, M. Konno, K. Shimotohno, K. Shigeta, S. and Yokota, T. 1997. Novel Hepatitis C Virus Protease Inhibitors: Thiazolidine Derivatives. Biochemical and Biophysical Research Communications (238)

17. Hepatitis C virus ultrastructure and morphogenesis. Biology of the Cell, Volume 96, Issue 2, March 2004, Pages 103-108

18. Kim, J.L., et al., & 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.

19. Yao, N., Hesson, T., Cable, M., Hong, Z., Kwong, A.D., Le, H.V. & Weber, P.C. (1997). Structure of the hepatitis C virus RNA helicace domain. Nat. Struct. Biol. 4, 463–467.

20. Jean-Pierre Allain (2005). Hepatitis C virus in blood donation. The Lancet, Volume 365, Issue 945, Pages 276-278

21. Lisa K. Gilbert, Jane Bulger, Kelli Scanlon and Linda Moyer (2005) Viral hepatitis prevention education: what do people and providers need to know? Patient Education and Counseling, Volume 59, Issue 1, Pages 46-55

22. Emmet B. Keeffe (2005). Chronic Hepatitis C: Management of Treatment Failures. Clinical Gastroenterology and Hepatology, Volume 3, Supplement 2, Pages S102-S105

23. Marion Riffelmann, Silke Sarr, Antje Ballauff, Helga Meisel and Michael Roggendorf (1997). Transmission of GBV-C/HGV from drug-addicted mothers to their babies. Journal of Hepatology, Volume 27, Issue 1, July 1997, Pages 85-90

24. Beerntsen B. T., Severson D. W., Klinkhammer J. A., Kassner V. A. and Christensen B. M. (1995). Aedes aegypti: A Quantitative Trait Locus (QTL) Influencing Filarial Worm Intensity Is Linked to QTL for Susceptibility to Other Mosquito-Borne Pathogens. Experimental Parasitology, Volume 81, Issue 3, Pages 355-362

25. Daryl T.-Y. Lau and Alex T. Hewlett (2005) Screening for hepatitis A and B antibodies in patients with chronic liver disease. The American Journal of Medicine, Volume 118, Issue 10, Supplement 1, October 2005, Pages 28-33

26. Suzich, J.A., et al., & Collett, M.S. (1993). Hepatitis C virus NS3 protein polynucleotide-stimulated nucleoside triphosphatase and comparison with the related pestivirus and flavivirus enzymes. J. Virol. 67, 6152-6158.

27. Tsukiyama, K. K., N. Iizuka, M. Kohara, and A. Nomoto. 1992. Internal ribosome entry site within hepatitis C virus RNA. J. Virol. 66:1476-1483.

28. Pickering, J. M., H. C. Thomas, and P. Karayiannis. 1997. Predicted secondary structure of the hepatitis G virus and GB virus-A 5$untranslated regions consistent with an internal ribosome entry site. J. Viral Hepat. 4:175-184

29. Miriam J. Alter (2006) Epidemiology of viral hepatitis and HIV co-infection Journal of Hepatology, Volume 44, Supplement 1, Pages S6-S9

30. Chen, Y., T. Maguire, R. E. Hileman, J. R. Fromm, J. D. Esko, R. J. Linhardt, and R. M. Marks. 1997. Dengue virus infectivity depends on envelope protein binding to target cell heparan sulfate. Nat. Med. 3:866-871

31. Pryor, M. J., R. C. Gualano, B. Lin, A. D. Davidson, and P. J. Wright. 1998. Growth restriction of dengue virus type 2 by site-specific mutagenesis of virus-encoded glycoproteins. J. Gen. Virol. 79:2631–2639.

32. Wengler, G., and G. Wengler. 1993. The NS3 nonstructural protein of flaviviruses contains an RNA triphosphatase activity. Virology 197:265–273.

33. Styer LM, Carey JR, Wang JL, Scott TW. (2007) Mosquitoes do senesce: departure from the paradigm of constant mortality. Am J Trop Med Hyg. 76(1):111-7.

34. Khan AM, Heiny A, Lee KX, Srinivasan K, Tan TW, August JT, Brusic V. (2006) Large-scale analysis of antigenic diversity of T-cell epitopes in dengue virus. BMC Bioinformatics. 18;7 Suppl 5:S4.

35. Chareonsirisuthigul T, Kalayanarooj S, Ubol S. (2007) Dengue virus (DENV) antibody-dependent enhancement of infection upregulates the production of anti-inflammatory cytokines, but suppresses anti-DENV free radical and pro-inflammatory cytokine production, in THP-1 cells. J Gen Virol.88(Pt 2):365-75.

36. Suksanpaisan L, Cabrera-Hernandez A, Smith DR. (2007) Infection of human primary hepatocytes with dengue virus serotype 2. J Med Virol. 23;79(3):300-307

37. Ang KT, Ruhaini I, Chua KB. (2006) An epidemiological cluster pattern of dengue outbreak amongst close contacts in Selangor, Peninsular Malaysia. Med J Malaysia. 61(3):292-5.

38. Miranda de Sousa A, Puccioni-Sohler M, Dias Borges A, Fernandes Adorno L, Papais Alvarenga M, Papais Alvarenga RM. (2006) Post-dengue neuromyelitis optica: case report of a Japanese-descendent Brazilian child. J Infect Chemother. 12(6):396-8de la C Sierra B, Garcia G, Perez AB, Morier L, Alvarez M, Kouri G, Guzman MG. (2006) Ethnicity and difference in dengue virus-specific memory T cell responses in Cuban individuals. Viral Immunol. 19(4):662-8.Abhyankar AV, Dash PK, Saxena P, Bhargava R, Parida MM, Jana AM, Sahni AK, Rao PV. (2006) Comparison of a dipstick dot-ELISA with commercial assays for anti-dengue virus IgM antibodies. Viral Immunol. 19(4):630-6.Alter HJ, Stramer SL, Dodd RY. (2007) Emerging infectious diseases that threaten the blood supply. Semin Hematol. 44(1):32-41.Mueller NH, Yon C, Ganesh VK, Padmanabhan R. (2007) Characterization of the West Nile virus protease substrate specificity and inhibitors. Int J Biochem Cell Biol.;39(3):606-14.Upanan S, Cabrera-Hernandez A, Ekkapongpisit M, Smith DR. (2006) A simplified PCR methodology for semiquantitatively analyzing dengue viruses. Jpn J Infect Dis. 59(6):383-7.Neeraja M, Lakshmi V, Teja VD, Umabala P, Subbalakshmi MV (2006) Serodiagnosis of dengue virus infection in patients presenting to a tertiary care hospital. Indian J Med Microbiol. 24(4):280-2.Kularatne SA, Pathirage MM, Medagama UA, Gunasena S, Gunasekara MB. (2006) Myocarditis in three patients with dengue virus type DEN 3 infection. Ceylon Med J. 51(2):75-6.

46. Thomas P Monath (2001). Yellow fever: an update. The Lancet Infectious Diseases, Volume 1, Issue 1, Pages 11-20

47. Stefan Schilling, Diana Ludolfs, Le Van An and Herbert Schmitz (2004). Laboratory diagnosis of primary and secondary dengue infection. Journal of Clinical Virology, Volume 31, Issue 3, Pages 179-184

48. Edwards, S., P. M. Roehe, and G. Ibata. 1995. Comparative studies of border disease and closely related virus infections in experimental pigs and sheep. Br. Vet. J. 151:181–187.

49. Terpstra, C., and G. Wensvoort. 1997. A congenital persistent infection of bovine virus diarrhoea virus in pigs: clinical, virological and immunological observations. Vet. Q. 19:97–101

50. Van Campen, H., E. S. Williams, J. Edwards, W. Cook, and G. Stout. 1997. Experimental infection of deer with bovine viral diarrhea virus. J. Wildl. Dis. 33:567–573.

51. Meehan, J. T., H. D. Lehmkuhl, R. C. Cutlip, and S. R. Bolin. 1998. Acute pulmonary lesions in sheep experimentally infected with bovine viral diarrhoea virus. J. Comp. Pathol. 119:277–292.

52. Pastoret, P. P., and M. Lambot. 1995. Biologie de l'infection des bovins par le Pestivirus responsable de la maladie des muqueuses. Bull. Mem. Acad. R. Med. Belg. 150:244–250. (In French.)

53. Thiel, H.-J., P. G. W. Plagemann, and V. Moennig. 1996. Pestiviruses, p. 1059–1073. In B. N. Fields, D. M. Knipe, and P. M. Howley (ed.), Fields virology, 3rd ed., vol. 1. Lippincott-Raven Publishers, Philadelphia, Pa.

54. Bruschke, C. J., K. Weerdmeester, J. T. Van Oirschot, and P. A. Van Rijn. 1998. Distribution of bovine virus diarrhoea virus in tissues and white blood cells of cattle during acute infection. Vet. Microbiol. 64:23–32.

55. Ellis, J. A., K. H. West, V. S. Cortese, S. L. Myers, S. Carman, K. M. Martin, and D. M. Haines. 1998. Lesions and distribution of viral antigen following an experimental infection of young seronegative calves with virulent bovine virus diarrhea virus-type II. Can. J. Vet. Res. 62:161–169.

56. Donis, R. O., and E. J. Dubovi. 1987. Differences in virus-induced polypeptides in cells infected by cytopathic and noncytopathic biotypes of bovine virus diarrhea-mucosal disease virus. Virology 158:168–173.

57. Kummerer, B. M., D. Stoll, and G. Meyers. 1998. Bovine viral diarrhea virus strain Oregon: a novel mechanism for processing of NS2-3 based on point mutations. J. Virol. 72:4127–4138.

58. Mendez, E., N. Ruggli, M. S. Collett, and C. M. Rice. 1998. Infectious bovine viral diarrhea virus (strain NADL) RNA from stable cDNA clones: a cellular insert determines NS3 production and viral cytopathogenicity. J. Virol. 72:4737–4745.

59. Meyers, G., and H. J. Thiel. 1996. Molecular characterization of pestiviruses. Adv. Virus Res. 47:53–118

60. Tautz, N., G. Meyers, and H.-J. Thiel. 1998. Pathogenesis of mucosal disease, a deadly disease of cattle caused by a pestivirus. Clin. Diagn. Virol. 10:121–127.

61. Makoto Nagai, Michiko Hayashi, Shigeo Sugita, Yoshihiro Sakoda, Masashi Mori, Toshiaki Murakami, Tadashi Ozawa, Naoki Yamada and Hiroomi Akashi (2004). Phylogenetic analysis of bovine viral diarrhea viruses using five different genetic regions. Virus Research, Volume 99, Issue 2, February 2004, Pages 103-113

62. Laevens, H., H. Deluyker, F. Koenen, G. Van Caenegem, J. P. Vermeersch, and A. de Kruif. 1998. An experimental infection with a classical swine fever virus in weaner pigs. II. The use of serological data to estimate the day of virus introduction in natural outbreaks. Vet. Q. 20:46–49

63. Kim, J.L., et al., & 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.

64. Laevens, H., F. Koenen, H. Deluyker, D. Berkvens, and A. de Kruif. 1998. An experimental infection with a classical swine fever virus in weaner pigs. I. Transmission of the virus, course of the disease, and antibody response. Vet. Q. 20:41–45.

65. Vilcek, S., T. Stadejek, P. A. Ballagi, J. P. Lowings, D. J. Paton, and S. Belak. 1996. Genetic variability of classical swine fever virus. Virus Res. 43:137–147.

66. Nettleton, P. F., J. A. Gilray, P. Russo, and E. Dlissi. 1998. Border disease of sheep and goats. Vet. Res. 29:327–340.

67. Becher, P., G. Meyers, A. D. Shannon, and H. J. Thiel. 1996. Cytopathogenicityof border disease virus is correlated with integration of cellular sequences into the viral genome. J. Virol. 70:2992–2998.

68. Bolin, S. R. 1995. Control of bovine viral diarrhea infection by use of vaccination. Vet. Clin. North Am. Food Anim. Pract. 11:615–625.

69. Cortese, V. S., D. L. Grooms, J. Ellis, S. R. Bolin, J. F. Ridpath, and K. V. Brock. 1998. Protection of pregnant cattle and their fetuses against infection with bovine viral diarrhea virus type 1 by use of a modified-live virus vaccine. Am. J. Vet. Res. 59:1409–1413.

70. Cortese, V. S., R. Whittaker, J. Ellis, J. F. Ridpath, and S. R. Bolin. 1998. Specificity and duration of neutralizing antibodies induced in healthy cattle after administration of a modified-live virus vaccine against bovine viral diarrhea. Am. J. Vet. Res. 59:848–850.

71. Poole, T. L., C. Wang, R. A. Popp, L. N. Potgieter, A. Siddiqui, and M. S. Collett. 1995. Pestivirus translation initiation occurs by internal ribosome entry. Virology 206:750–754.

72. R. L. Doherty (1958). Effects of yellow fever (17D) and West Nile viruses on the reactions of human appendix and conjunctiva cells to several other viruses. Virology, Volume 6, Issue 3, Pages 575-583

73. David N. Phalen and Bob Dahlhausen (2004). West Nile virus. Seminars in Avian and Exotic Pet Medicine, Volume 13, Issue 2, Pages 67-78

74. Blaithnead Murtagh, Yasmin Wadia, Greg Messner, Paul Allison, Yadollah Harati and Reynolds Delgado (2005). West Nile Virus Infection After Cardiac Transplantation. The Journal of Heart and Lung Transplantation, Volume 24, Issue 6, June 2005, Pages 774-776

75. Thompson, B. and Finch, R. 2005. Hepatitis C Virus Infection. Clinical Microbiology and Infection 11(2), pp. 86-94.

76. Rosenberg, S. 2001. Recent Advances in the Molecular Biology of Hepatitis C Virus. Journal of Molecular Biology 313, pp. 451-464.

77. Poynard, T. Bedossa, P. Opolon, P.1997. Natural history of liver fibrosis progression in patients with chronic hepatitis C. Lancet 349, pp. 825-832.

78. Daly, M. and Ward, F. 2003. Liver Disease. In: Walker, R. Edwards, C. (Ed) Clinical Pharmacy and Therapeutics 3rd ed. Churchill Livingstone. pp. 209-228.

79. Pontisso, P. Belluco, C. Bertorelle, R. Moliner, L. Bianchi, L. Nitti, D. Lise, M. Alberti, A. 1998. Hepatitis C Virus Infection Associated with Human Hepatocellular Carcinoma. Cancer 83(8), pp. 1489-1494.

80. Triyatni, M. Vergalla, J. Davis, A. Hadlock, K. Foung, S. and Liang, T. 2002. Structural Features of Envelope Proteins on Hepatitis C Virus-like Particles as Determined by Anti-envelope Monoclonal Antibodies and CD81 Binding. Virology 298, pp. 124-132.

81. Keller, P. Gordon, C. 2005. Control of Hepatitis C: A Medicinal Chemistry Perspective. Journal of Medicinal Chemistry 48(1), pp 1-20

82. Steffens, S. Thiel, H. and Behrens, S.1999. The RNA-dependent RNA polymerases of different members of the family Flaviviridae exhibit similar properties in vitro. Journal of General Virology 80, pp. 2583-2590.

83. Heidemarie Holzmann (2003). Diagnosis of tick-borne encephalitis. Vaccine, Volume 21, Supplement 1, Pages S36-S40

84. Lewis Markoff (2000). Points to consider in the development of a surrogate for efficacy of novel Japanese encephalitis virus vaccines. Vaccine, Volume 18, Supplement 2, Pages 26-32

85. Denise S. Olive, Masayoshi Konishi and George Y. Wu (2004). Cell culture and animal models for human viral hepatitis. Hepatology Research, Volume 28, Issue 2, Pages 61-67

86. Caruthers, J. & McKay, D. Helicase structure and mechanism. Curr. Opin. Struct. Biol. 12, 123–133 (2002).

87. Pascal Pineau , Anne Dejean and Pierre Tiollais (2002). Animal hepadnaviruses and their host species: Models for human hepatocarcinogenesis. Perspectives in Medical Virology, Volume 6, Pages 123-141

88. Gwack, Y., Kim, D.W., Han, J.H. & Choe, J. (1996). Characterization of RNA binding activity and RNA helicase activity of the hepatitis C virus NS3 protein. Biochem. Biophys. Res. Commun. 225, 654–659.

89. Behrens, S. E., L. Tomei, and R. De Francesco. 1996. Identification and properties of the RNA-dependent RNA polymerase of hepatitis C virus. EMBO J. 15:12–22

90. Wengler, G. and Wengler, G. 1993. The NS3 Nonstructural Protein of Flaviviruses Contains an RNA Triphosphatase Activity. Virology 197

91. Leyssen, P. Clercq, E. and Neyts, J. 2000. Perspectives for the Treatment of Infections with Flaviviridae. Clinical Microbiology Reviews 13(1), pp. 67-82.

92. Tan, S. He, Y. Huang, Y. and Gale, M. 2004. Strategies for hepatitis C therapeutic intervention: now and next. Current Opinion in Pharmacology 4, pp. 465-470.

93. Walker, M. 1999. Hepatitis C virus: an overview of current approaches and progress. Drug Discovery Today. Vol.4, No.11, pp 518-529

94. Nulf, C. J. and Corey, D. 2004. Intracellular inhibition of hepatitis C virus (HCV)internal ribosomal entry site (IRES)-dependent translation by peptide nucleic acids (PNAs) and locked nucleic acids (LNAs). Nucleic Acids Research 32(13), pp. 3792-3798.

95. Jefferson,E. Seth, P. Robinson, D. Winter, D. Miyaji, A. Osgood, S. Swayze, E. and Risen, L.2004. Biaryl guanidine inhibitors of in vitro HCV-IRES activity. Bioorganic Medicinal Chemistry Letters 14, pp. 5139-5143.

96. Wengler, G. and Wengler, G. 1993. The NS3 Nonstructural Protein of Flaviviruses Contains an RNA Triphosphatase Activity. Virology 197, pp. 265-273.

97. Sudo, K. Matsumoto, Y. Matsushima, M. Fujiwara, M. Konno, K. Shimotohno, K.Shigeta, S. and Yokota, T. 1997. Novel Hepatitis C Virus Protease Inhibitors:Thiazolidine Derivatives. Biochemical and Biophysical Research Communications (238) pp 643–647

98. Ingallinella, P. Altamura, S. Bianchi, E. Taliani, M. Ingenito, R. Cortese, R. De Francesco, R. Steinkühler, C. and Antonello Pessi 1998. Potent Peptide Inhibitors of Human Hepatitis C Virus NS3 Protease Are Obtained by Optimizing the Cleavage Products. Biochemistry 37(25) pp 8906 - 8914.

99. Benarroch, D. Egloff, M. Mulard, L. Guerreiro, C. Romette, J. and Canard, B. 2004. A Structural Basis for the Inhibition of the NS5 Dengue Virus mRNA 2'-OMethyltransferase Domain by Ribavirin 5'-Triphosphate. The Journal of Biological Chemistry 279(34), pp. 35638-35643.

100.   Chang, S. Cheng, J. Kou,Y. Kao,C. Chiu, C. Wu, H. and Chang, M. 2000. Roles of the AX4GKS and Arginine-Rich Motifs of Hepatitis C Virus RNA Helicase in ATP- and Viral RNA-Binding Activity. Journal of Virology 74(20), pp. 9732-9737.

101.   Artsaenko, O. Tessmann, K. Sack, M. Haussinger, D. Heintges, T. 2003. Abrogation of hepatitis C virus NS3 helicase enzymatic activity by recombinant human antibodies. Journal of General Virology 84, pp 2323–2332.

102.   Matusan, A. Pryor, M. Davidson, A. Wright, P. 2001. Mutagenesis of the Dengue VirusType 2 NS3 Protein within and outside Helicase Motifs: Effects on Enzyme Activity and Virus Replication. Journal of Virology, pp 9633–9643

103.   Lam, A. Rypma, R. Frick, D. 2004. Enhanced nucleic acid binding to ATP-bound hepatitis C virus NS3 helicase at low pH activates RNA unwinding. Nucleic AcidsResearch, Vol. 32, No. 13, pp 4060-4070

104.   Bianco, P. 2004. Hepatitis C NS3 helicase unwinds RNA in leaps and bounds. The Lancet. Vol. 364, pp 1385-1387 Borowski, P. Schalinski, S. Schmitz, H. 2002. Nucleotide triphosphatase/helicase of hepatitis C virus as a target for antiviral therapy. Antiviral Research (55) pp 397-412

105.   Phoon, C. Ng, P. Ting, A. Yeo, S. Sim, M. 2001. Biological Evaluation of Hepatitis C Virus Helicase Inhibitors. Bioorganic & Medicinal Chemistry Letters (11) pp 1647–1650

106.   Diana, G. Bailey, T. 1997. Compounds, compositions and methods for treatment ofhepatitis C. United States Patent No. 5,633,388

107.   ViroFarma Patent: Compounds, Compositions and Methods for treatment of Hepatitis C, Patent number: 5.633.388, Date of Patent: May 27, 1997

108.   Steffens, S. Thiel, H. and Behrens, S.1999. The RNA-dependent RNA polymerases of different members of the family Flaviviridae exhibit similar properties in vitro. Journal of General Virology 80

109.   Chang, S. Cheng, J. Kou,Y. Kao,C. Chiu, C. Wu, H. and Chang, M. (2000). Roles of the AX4GKS and Arginine-Rich Motifs of Hepatitis C Virus RNA Helicase in ATP- and Viral RNA-Binding Activity. Journal of Virology 74(20), 9732-9737.

110.   Lam, A. Rypma, R. Frick, D. (2004). Enhanced nucleic acid binding to ATP-bound hepatitis C virus NS3 helicase at low pH activates RNA unwinding. Nucleic Acids Research, Vol. 32, No. 13, 4060-4070

111.   Artsaenko, O. Tessmann, K. Sack, M. Haussinger, D. Heintges, T. (2003). Abrogation of hepatitis C virus NS3 helicase enzymatic activity by recombinant human antibodies. Journal of General Virology 84, 2323–2332.

112.   Matusan, A. Pryor, M. Davidson, A. Wright, P. (2001). Mutagenesis of the Dengue Virus Type 2 NS3 Protein within and outside Helicase Motifs: Effects on Enzyme Activity and Virus Replication. Journal of Virology, pp 9633–9643

113.   Jin, L. & Peterson, D.L. (1995). Expression, isolation, and characterization of the hepatitis C virus ATPase/RNA helicase. Arch. Biochem. Biophys. 323, 47–53.

114.   Preugschat, F., Averett, D.R., Clarke, B.E. & Porter, D.J.T. (1996). A steady-state and pre-steady-state kinetic analysis of the NTPase activity associated with the hepatitis C virus NS3 helicase domain. J. Biol. Chem. 271, 24449–24457.

115.   Morgenstern, K.A., et al., & Thomson, J.A. (1997). Polynucleotide modulation of the protease, nucleoside triphosphatase, and helicase activities of a hepatitis C virus NS3–NS4A complex isolated from transfected COS cells. J. Virol. 71, 3767–3775.

116.   Hong, Z., et al., & Kwong, A.D. (1996). Enzymatic characterization of hepatitis C virus NS3/4A complexes expressed in mammalian cells by using the herpes simplex virus amplicon system. J. Virol. 70,4261–4268.

117.   M.P. Allen and D.J. Tildesley. Computer Simulation of Liquids. Oxford University Press, New York, 1987.

118.   A.D. Bates and A. Maxwell. DNA Topology. In Focus series, Oxford Univeristy Press, New York, 1993.

119.   C. Branden and J. Tooze. Introduction to Protein Structure. second edition, Garland Publishing Inc., New York, 1999.

120.   P. Bratley, B. L. Fox, and L. E. Schrage. A Guide to Simulation. Springer-Verlag, New York, 1987.

121.   C. L. Brooks, III, M. Karplus, and B.M. Pettitt. A Theoretical Perspective of Dynamics, Structure, and Thermodynamics. Wiley Interscience, New York, 1988.

122.   U. Burkert and N.L. Allinger. Molecular Mechanics. American Chemical Society, Washington D.C., 1980.

123.   C.R. Cantor and P.R. Schimmel. Biophysical Chemistry. Vol.1,2,3. W.H. Freeman and Company, San Fransisco, 1980.

124.   N. R. Cohen, Editor. Guidebook on Molecular Modeling in Drug Design. Academic Press, San Diego, 1996.

125.   Computational Molecular Dynamics: Challenges, Methods, Ideas -- Proceedings of the 2nd International Symposium on Algorithms for Macromolecular Modelling, Berlin, May 21-24, 1997, P. Deuflhard, J. Hermans, B. Leimkuhler, A. E. Mark, S. Reich,. R. D. Skeel, Eds., Lecture Notes in Computational Science and Engineering, Vol. 4 (Series Editors M. Griebel, D. E. Keyes, R. M. Nieminen, D. Roose, and T. Schlick), Springer-Verlag, Berlin and New York, 1999.

126.   T.E. Creighton, Editor. Protein Folding. W.H. Freeman & Company, New York, 1992.

127.   D. Eisenberg and D. Crothers. Physcial Chemistry with Applications to the Life Science. Benjamin Cummings, Menlo Park, California, 1979.

128.   A. Fersht. Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding. W. H. Freeman and Company, New York, 1999.

129.   D. Frenkel and B. Smit. Understanding Molecular Simulations. From Algorithms to Applications. Academic Press, San Diego, California, 1996.

130.   L.M. Gierasch and J. King, Editors. Protein Folding, Deciphering the Second Half of the Genetic Code. AAAS, Washington D.C., 1990.

131.   H. Gould and J. Tobochnik. An Introduction to Computer Simulation Metheods: Appkications to Physical Systems. Part 1 and 2 Addison-Wesley, Reading, MA, 1988.

132.   A.Y. Grosberg and A.R. Khokhlov. Giant Molecules. Here, There, and Everywhere... Academic Press, San Diego, California, 1997.

133.   J.M. Haile. Molecular Dynamics Simulations: Elementary Methods. Wiley, New York, 1992.

134.   M. Kalos and P. A. Whitlock. Monte Carlo Methods. John Wiley & Sons, New York, 1986.

135.   A. R. Leach. Molecular Modelling. Principles and Applications. Addison Wesley Longman, Essex, England, 1996.

136.   K.B. Lipkowitz and D.B. Boyd, Editors. Reviews in Computational Chemistry. VCH Publishers, New York, 1990 –

137.   J.A. McCammon and S.C. Harvey. Dynamics of Proteins and Nucleic Acids. Cambridge University Press, Cambridge, 1987.

138.   D.A. McQuarrie. Statistical Mechanics. Harper Collins Publishers, New York, 1976.

139.   National Research Council report. Mathematical Challenges from Theoretical / Computational Chemistry, National Academy Press, Washington D.C., 1995.

140.   D.C. Rapaport. The Art of Molecular Dynamics Simulation. Cambridge University Press, Cambridge, England., 1995.

141.   W. Saenger. Principles of Nucleic Acid Structure. Springer Advanced Texts in Chemistry, Springer-Verlag, New York, 1984.

142.   R.R. Sinden. DNA Structure and Function. cademic Press, San Diego, California, 1994

143.   G.E. Schulz and R.H. Schirmer. Principles of Protein Structure. Springer Advanced Texts in Chemistry, Springer-Verlag, New York, 1990.

144.   L. Stryer. Biochemistry. W.H. Freeman, New York, latest edition (fourth in 1995).

145.   W. Van Gunsteren and P. Weiner, Editors (1989) and W. Van Gunsteren, P. Weiner, and A.T. Wilkinson, Editors (1993, 1996): Computer Simulation of Biomolecular Systems: Theoretical and Experimentational Applications. Vol. 1,2,3. ESCOM, Leiden, The Netherlands, 1989, 1993, 1996.

146.   Ding, J., Das, K., Moereels, H., Koymans, L., Andries, K., Janssen, P.A.J., Hughes, S.H., Arnold, E. Nat. Struct. Biol. 2, 407, (1995).

147.   Eriksson, M.A.L., Pitera, J., Kollman, P.A. J. Med. Chem. 42, 868-881, (1999).

148.   Potter, M.J., Kirchhoff, P.D., Carlson, H.A., McCammon, J.A. J. Comput. Chem. 20, 956-970, (1999).

149.   Gill, P., Murray W., Wright, M.Practical Optimization. Academic Press, London, 1981.

150.   Stewart, J. J. P. MOPAC Manual (Seventh Edition), 1993.

151.   Allen, M.P., Tildesley, D.J. Computer Simulation of Liquids. Oxford University Press (1987).

152.   Verlet, L. Computer 'Experiments' on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. Phys. Rev. 159, 98-103 (1967).

153.   Bond, S.D., Benedict, J.L., Laird, B.B. The Nosé-Poincaré Method for Constant Temperature Molecular Dynamics. J. Comp. Phys. 151, 114-134 (1999).

154.   Sturgeon, J.B., Laird, B.B. Symplectic Algorithm for Constant Pressure Molecular Dynamics Using a Nosé-Poincaré Thermostat. University of Kansas Technical Paper (2002).

155.   Sun, G. Symplectic Partitioned Runge-Kutta Methods. J. Comput. Math. 11, 365-372 (1993).

156.   Berendsen, H.J.C., Postma, J.P.M., Van Gunsteren, W.F., Di Nola, A., Haak, J.R. Molecular Dynamics with Coupling to an External Bath. J. Chem. Phys. 81, 3684-3690 (1984).

157.   Ryckaert, J.P., Ciccotti, G., Berendsen, H.J.C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. J. Comput. Phys. 23, 327-341 (1977).

158.   Hess, B., Bekker, H., Berendsen, H.J.C., Fraaije, J.G.E.M. LINCS: A Linear Constraint Solver for Molecular Simulations J. Comp. Chem. 18, 1463-1472 (1997).

159.   Ferguson, D.M., Raber, D.J. A New Approach to Probing Conformational Space with Molecular Mechanics: Random Incremental Pulse Search. J. Am. Chem. Soc. 111, 4371-4378 (1989).

160.   H.M.Berman, J.Westbrook, Z.Feng, G.Gilliland, T.N.Bhat, H.Weissig, I.N.Shindyalov, P.E.Bourne. The Protein Data Bank. Nucleic Acids Research. 28, 235-242, (2000).

161.   Fechteler T., Dengler, U., Schomberg, D. Prediction of Protein Three-Dimensional Structures in Insertion and Deletion Regions: A Procedure for Searching Databases of Representative Protein Fragments Using Geometric Scoring Criteria. Journal of Molecular Biology. 253, 114-131, (1995).

162.   Jones T.A. & Thirup, S. Using Known Substructures in Protein Modeling and Crystallography. EMBO J.. 5, 819-822, (1986).

163.   LPC: Ligand-protein Contacts, 2003, URL: http://www.weizmann.ac.il/sgedg/lpc/

164.   LigPlot: Schematic diagrams of protein-ligand interactions, 2002, URL: http://www.biochem.ucl.ac.uk/bsm/ligplot/ligplot.html

165.   Raghunathan, S., Ricard, C.S., Lohman, T.M. & Waksman, G. (1997). Crystal structure of the homo-tetrameric DNA-binding domain of Escherichia coli single-stranded DNA-binding protein determined by multiwavelength X-ray diffraction on the selenomethionyl protein at 2.9 Å resolution. Proc. Natl. Acad. Sci. USA 94, 6652–6657.

166.   Gwack, Y., Kim, D.W., Han, J.H. & Choe, J. (1996). Characterization of RNA binding activity and RNA helicase activity of the hepatitis C virus NS3 protein. Biochem. Biophys. Res. Commun. 225, 654–659.

167.   Kanai, A., Tanabe, K. & Kohara, M. (1995). Poly(U) binding activity of hepatitis C virus NS3 protein, a putative RNA helicase. FEBS Lett. 376, 221–224.

168.   Miller, R.H. & Purcell, R.H. (1990). Hepatitis C virus shares amino acid sequence similarity with pestiviruses and flaviviruses as well as members of two plant virus supergroups. Proc. Natl. Acad. Sci. USA 87, 2057–2061.

169.   Sybyl 7.0 Tripos Inc, 1699 South Hanley Road, St. Louis, Missouri, 63144, USA, URL: http://www.tripos.com

170.   Mirzayan, C. & Wimmer, E. (1992). Genetic analysis of an NTPbinding motif in poliovirus polypeptide 2C. Virology 189, 547–555.

171.   MacPherson, P., Thorner, L., Parker, L.M. & Botchan, M. (1994). The bovine papilloma virus E1 protein has ATPase activity essential to viral DNA replication and efficient transformation in cells. Virology 204, 403–408.

172.   MOE, Molecular Operating Environment 2005.04, Chemical Computing Group, Montreal, Quebec, Canada, URL:http://www.chemcomp.com/

173.   Procheck: Roman A Laskowski, Malcolm W MacArthur, David S Moss and Janet M Thornton, J. App. Cryst. 26 283 (1993)

174.   Martinez, R., Shao, L. & Weller, S.K. (1992). The conserved helicase motifs of the herpes simplex virus type 1 origin-binding protein UL9 are important for function. J. Virol. 66, 6735–6746.

175.   Gorbalenya, A. & Koonin, E.V. (1993). Helicases: amino acid sequence comparisons and structure-function relationships. Curr. Opin. Struct. Biol. 3, 419–429.

176.   Subramanya, H.S., Bird, L.E., Brannigan, J.A. & Wigley, D.B. (1996). Crystal structure of a DExx box DNA helicase. Nature 384, 379–383.

177.   Kim,   J.L., Morgenstern,   K.A., Griffith,   J.P., Dwyer,
M.D., Thomson, J.A., Murcko, M.A., Lin, C., Caron, P.R. Hepatitis C
virus NS3 RNA helicase domain with a bound oligonucleotide: the crystal
structure provides insights into the mode of unwinding. Structure v6
pp.89-100 , 1998

178.   Horscroft, N. Lai, V. Cheney, W. Yao, N. Wu, JZ. Hong, Z. Zhong,
W. 2005. Replicon cell culture system as a valuable tool in antiviral drug
discovery against hepatitis C virus. Antiviral Chemistry & Chemotherapy;
16: 1–12

179.   Korolev, S., Hsieh, J., Gauss, G.H., Lohman, T.M. & Waksman, G.
(1997). Major domain swiveling revealed by the crystal structures of
complexes of E. coli Rep helicase bound to single-stranded DNA and ADP.
Cell 90, 635–647.

180.   Bartenschlager, R., Ahlborn-Laake, L., Mous, J. & Jacobsen, H.
(1994). Kinetic and structural analyses of hepatitis C virus polyprotein
processing. J. Virol. 68, 5045–5055.

181.   Lohman, T.M. & Bjornson, K.P. (1996). Mechanisms of
helicasecatalyzed DNA unwinding. Annu. Rev. Biochem. 65, 169–214.

182.   Walker, J.E., Saraste, M., Runswick, M.J. & Gay, N.J. (1982).
Distantly related sequences in the alpha- and beta-subunits of ATP
synthase, myosin, kinases and other ATP-requiring enzymes and a
common nucleotide binding fold. EMBO J. 1, 945–951.

183.   Saraste, M., Sibbald, P.R. & Wittinghofer, A. (1990). The P-loop – a
common motif in ATP- and GTP-binding proteins. Trends Biochem. Sci.
15, 430–434.

184.   Ruff, M., et al., & Moras, D. (1991). Class II aminoacyl transfer RNA
synthetases: crystal structure of yeast aspartyl-tRNA synthetase complexed
with tRNA(Asp). Science 252, 1682–1689.

185.   Bochkarev, A., Pfuetzner, R.A., Edwards, A.M. & Frappier, L. (1997). Structure of the single-stranded-DNA-binding domain of replication protein A bound to DNA. Nature 385, 176–181.

186.   Kim, Y., Geiger, J.H., Hahn, S. & Sigler, P.B. (1993). Crystal structure of a yeast TBP/TATA-box complex. Nature 365, 512–520.

187.   Kim, J.L., Nikolov, D.B. & Burley, S.K. (1993). Co-crystal structure of TBP recognizing the minor groove of a TATA element. Nature 365, 520–527.

188.   George, J.W., Brosh, R.M., Jr. & Matson, S.W. (1994). A dominant negative allele of the Escherichia coli uvrD gene encoding DNA helicase II. A biochemical and genetic characterization. J. Mol. Biol.

189.   Seeley, T.W. & Grossman, L. (1990). The role of Escherichia coli UvrB in nucleotide excision repair. J. Biol. Chem. 265, 7158–7165.

190.   Black, M.E. & Hruby, D.E. (1992). Site-directed mutagenesis of a conserved domain in vaccinia virus thymidine kinase. Evidence for a potential role in magnesium binding. J. Biol. Chem. 267, 6801–6806.

191.   Brosh, R.M., Jr. & Matson, S.W. (1995). Mutations in motif II of Escherichia coli DNA helicase II render the enzyme nonfunctional in both mismatch repair and excision repair with differential effects on the unwinding reaction. J. Bacteriol. 177, 5612–5621.

192.   Pause, A. & Sonenberg, N. (1992). Mutational analysis of a DEAD box RNA helicase: the mammalian translation initiation factor eIF-4A. EMBO J. 11, 2643–2654.

193.   Heilek, G.M. & Peterson, M.G. (1997). A point mutation abolishes the helicase but not the nucleoside triphosphatase activity of hepatitis C virus NS3 protein. J. Virol. 71, 6264–6266.

194.   Gross, C.H. & Shuman, S. (1996). The QRxGRxGRxxxG motif of the vaccinia virus DExH box RNA helicase NPH- II is required for ATP hydrolysis and RNA unwinding but not for RNA binding. J. Virol. 70, 1706–1713.

195.   Pause, A., Methot, N. & Sonenberg, N. (1993). The HRIGRXXR region of the DEAD box RNA helicase eukaryotic translation initiation factor 4A is required for RNA binding and ATP hydrolysis. Mol. Cell Biol. 13, 6789–6798.

196.   Hall, M.C. & Matson, S.W. (1997). Mutation of a highly conserved arginine in motif IV of Escherichia coli DNA helicase II results in an ATP-binding defect. J. Biol. Chem. 272, 18614–18620.

197.   Bilderback, T., Fulmer, T., Mantulin, W.W. & Glaser, M. (1996). Substrate binding causes movement in the ATP binding domain of Escherichia coli adenylate kinase. Biochemistry 35, 6100–6106.

198.   Schulz, G.E. (1992). Induced-fit movements in adenylate kinases. Faraday Discuss. 93, 85–93.

199.   Moore, K.J. & Lohman, T.M. (1994). Kinetic mechanism of adenine nucleotide binding to and hydrolysis by the Escherichia coli Rep monomer. 1. Use of fluorescent nucleotide analogues. Biochemistry 33, 14550–14564.

200.   Chao, K. & Lohman, T.M. (1990). DNA and nucleotide-induced conformational changes in the Escherichia coli Rep and helicase II (UvrD) proteins. J. Biol. Chem. 265, 1067–1076.

201.   Hakansson, K., Doherty, A.J., Shuman, S. & Wigley, D.B. (1997). X-ray crystallography reveals a large conformational change during guanyl transfer by mRNA capping enzymes. Cell 89, 545–553.

202.   Subramanya, H.S., Doherty, A.J., Ashford, S.R. & Wigley, D.B. (1996). Crystal structure of an ATP-dependent DNA ligase from bacteriophage T7. Cell 85, 607–615.

203.   Phillips, R.J., Hickleton, D.C., Boehmer, P.E. & Emmerson, P.T. (1997). The RecB protein of Escherichia coli translocates along singlestranded DNA in the 3'□to 5'□direction: a proposed ratchet mechanism. Mol. Gen. Genet. 254, 319–329.

204.   Ali, J.A. & Lohman, T.M. (1997). Kinetic measurement of the step size of DNA unwinding by Escherichia coli UvrD helicase. Science 275, 377–380.

205.   Chen, Y.Z., Zhuang, W. & Prohofsky, E.W. (1992). Energy flow considerations and thermal fluctuational opening of DNA base pairs at a replicating fork: unwinding consistent with observed replication rates. J. Biomol. Struct. Dyn. 10, 415–427.

206.   Barbara Selisko, Hélène Dutartre, Jean-Claude Guillemot, Claire Debarnot, Delphine Benarroch, Alexander Khromykh, Philippe Després, Marie-Pierre Egloff and Bruno Canard Comparative mechanistic studies of de novo RNA synthesis by flavivirus RNA-dependent RNA polymerases Virology, Volume 351, Issue 1, 20 2006, Pages 145-158

207.   Claudia M. D'Abramo, Luciano Cellai and Matthias Götte Excision of Incorporated Nucleotide Analogue Chain-terminators can Diminish their Inhibitory Effects on Viral RNA-dependent RNA Polymerases Journal of Molecular Biology, Volume 337, Issue 1, 12 2004, Pages 1-14

208.   Niklaus H. Mueller, Changsuek Yon, Vannakambadi K. Ganesh and R. Padmanabhan Characterization of the West Nile virus protease substrate specificity and inhibitors The International Journal of Biochemistry & Cell Biology, Volume 39, Issue 3, 2007, Pages 606-614

209.   Leandro R. Jones, Rubén O. Zandomeni and E. Laura Weber A long distance RT-PCR able to amplify the Pestivirus genome Journal of Virological Methods, Volume 134, Issues 1-2, 2006, Pages 197-204

210.   Stuart D. Blacksell, Syseng Khounsy and Harvey A. Westbury The effect of sample degradation and RNA stabilization on classical swine fever virus RT–PCR and ELISA methods Journal of Virological Methods, Volume 118, Issue 1, 1 2004, Pages 33-37

211.   Herbert Weissenböck, Jolanta Kolodziejek, Karin Fragner, Roland Kuhn, Martin Pfeffer and Norbert Nowotny Usutu virus activity in Austria, 2001–2002 Microbes and Infection, Volume 5, Issue 12, 2003, Pages 1132-1136

212.   R. Quadri and F. Negro Are there any subgenomic forms of hepatitis C virus RNA in the liver? Digestive and Liver Disease, Volume 33, Issue 6, 2001, Pages 480-486

213.   Peter A. White, Zhengqian Li and William D. Rawlinson Sequence diversity in the 5'-UTR region of GB virus C/hepatitis G virus assessed using sequencing, heteroduplex mobility analysis and single-strand conformation polymorphism Journal of Virological Methods, Volume 83, Issues 1-2, 1999, Pages 91-101

214.   Udo Künkel, Marina Höhne, Thomas Berg, Uwe Hopf, Alexader S. Kekulé, Gert Frösner, Georg Pauli and Eckart Schreier Quality control study on the performance of GB virus C/hepatitis G virus PCR Journal of Hepatology, Volume 28, Issue 6, 1998, Pages 978-984

215.   Takanobu Kato, Masashi Mizokami, Masakazu Nitta, Makoto Nakamura, Hideki Hiramatsu, Kanji Sugihara, Atsunaga Kato, Motokazu Mukaide and Ryuzo Ueda Acute GB virus C/hepatitis G virus hepatitis preceding aplastic anemia Hepatology Research, Volume 9, Issues 2-3, 1997, Pages 164-171

216.   J. E. Whitbya, H. Ni, H. E. Whitby, A. D. Jennings, L. M. Bradley, J. M. Lee, G. Lloyd, J. R. Stephenson and A. D. T. Barrett Rapid detection of viruses of the tick-borne encephalitis virus complex by RT-PCR of viral RNA Journal of Virological Methods, Volume 45, Issue 1, 1993, Pages 103-114

217.   Christiane Ramelow, Jochen Süss, Doris Berndt, Michael Roggendorf and Eckart Schreier Detection of tick-borne encephalitis virus RNA in ticks (Ixodes ricinus) by the polymerase chain reaction Journal of Virological Methods, Volume 45, Issue 1, 1993, Pages 115-119

218.   Stefanie Seipp, Rafael Wahl, Hubert Mueller, Wolfgang Stremmel, Lorenz Theilmann and Tobias Goeser Sequence analysis of hepatitis GB virus C (GBV-C) isolates from 14 patients Virus Research, Volume 46, Issues 1-2, 2 1996, Pages 81-88

219.   Thomas Berg, Ullrike Dirla, Uta Naumann, Hans-Gerd Heuft, Stefan Küther, Hartmut Lobeck, Eckart Schreier and Uwe Hopf Responsiveness to interferon alpha treatment in patients with chronic hepatitis C coinfected with hepatitis G virus Journal of Hepatology, Volume 25, Issue 5, 1996, Pages 763-768

220.   Eckart Schreier, Marina Höhne, Udo Künkel, Thomas Berg and Uwe Hopf Hepatitis GBV-C sequences in patients infected with HCV contaminated anti-D immunoglobulin and among i.v. drug users in Germany Journal of Hepatology, Volume 25, Issue 3, 1996, Pages 385-389

221.   Mini Kapoor, Luwen Zhang, P. Maruthi Mohan and R. Padmanabhan Synthesis and characterization of an infectious dengue virus type-2 RNA genome (New Guinea C strain) Gene, Volume 162, Issue 2, 11 1995, Pages 175-180

222.   Modeller,   2004,   URL:http://guitar.rockefeller.edu/modeller/modeller.html

223.   Elisabeth Lampe, Jaqueline M. de Oliveira, João L. Pereira, Felipe L. Saback, Clara F. T. Yoshida and Christian Niel Hepatitis G virus (GBV-C) infection among Brazilian patients with chronic liver disease and blood donors Clinical and Diagnostic Virology, Volume 9, Issue 1, 1998, Pages 1-7

224.   A Tagger, F Donato, M L Ribero, R Chiesa, V Tomasoni, G Portera, U Gelatti, A Albertini, M Fasola and G Nardi A case-control study on GB virus C/hepatitis G virus infection and hepatocellular carcinoma. Brescia HCC Study Hepatology, Volume 26, Issue 6, 1997, Pages 1653-1657

225.   Gabriele Handschuh and Wolfgang H. Caselmann Bacterial expression and purification of hepatitis C virus capsid proteins of different size Journal of Hepatology, Volume 22, Issue 2, 1995, Pages 143-150

226.   M. S. Marin, P. M. De A. Zanotto, T. S. Gritsun and E. A. Gould Phylogeny of TYU, SRE, and CFA Virus: Different Evolutionary Rates in the Genus Flavivirus Virology, Volume 206, Issue 2, 1 1995, Pages 1133-1139