American University in Cairo

AUC Knowledge Fountain

2-1-2014

Identification of T-cell epitopes in the Hepatitis C virus genotype 4 proteome: a step towards epitope-driven vaccine development

Karim Mohamed Ali Abdel-Hady

Follow this and additional works at: https://fount.aucegypt.edu/etds

The American University in Cairo

School of Sciences and Engineering

IDENTIFICATION OF T-CELL EPITOPES IN THE HEPATITIS C VIRUS GENOTYPE 4 PROTEOME: A STEP TOWARDS EPITOPE-DRIVEN VACCINE DEVELOPMENT

A Thesis Submitted to

The Biotechnology Graduate Program

in partial fulfillment of the requirements for
the degree of Master of Science

by Karim Mohamed Ali Abdel-Hady

under the supervision of

Prof. Hassan M.E. Azzazy

Dr. Anne S. De Groot

Fall 2013

The American University in Cairo

IDENTIFICATION OF T-CELL EPITOPES IN THE HEPATITIS C VIRUS GENOTYPE 4 PROTEOME: A STEP TOWARDS EPITOPE-DRIVEN VACCINE DEVELOPMENT

A Thesis submitted by
Karim Mohamed Ali Abdel-Hady

to the Biotechnology Graduate Program
Fall 2013

in partial fulfillment of the requirements for
the degree of Master of Science in Biotechnology

has been approved by:

Prof. Hassan M.E. Azzazy

Thesis Committee Chair and Thesis Supervisor

Affiliation: Professor of Chemistry, School of Sciences and Engineering, the American University in Cairo.
Date _____

Dr. Anne S. De Groot

Thesis Co-Supervisor

Affiliation: Research Professor and Director, Institute for Immunology and Informatics, University of Rhode Island.
Date _____

Dr. Ahmed Moustafa
Thesis Committee Inernal Examiner
Affiliation:
Date _____

Dr. Aishaa Yaseen Abdel-Ghafaar
Thesis Committee External Examiner
Date _____

Dr. Asma Amleh
Thesis Committee Moderator
Date _____

ACKNOWLEDGEMENTS

ABSTRACT

The American University in Cairo

IDENTIFICATION OF T-CELL EPITOPES IN THE HEPATITIS C VIRUS GENOTYPE 4 PROTEOME: A STEP TOWARDS EPITOPE-DRIVEN VACCINE DEVELOPMENT

BY: Karim Mohamed Ali Abdel-Hady

Under the Supervision of Prof. Hassan M.E. Azzazy and Dr. Anne S. De Groot

Hepatitis C is an inflammatory infectious disease of the liver caused by the Hepatitis C Virus (HCV). It is a global pandemic, chronically inflicting 150 million people worldwide, with millions of new infections arising annually. The standard therapy of HCV is expensive, associated with severe side effects, and has variable success rates. Thus far, no HCV vaccine has been developed, owing to the challenges that faced and still face its development. Despite these challenges, several attempts have been taken to develop a vaccine, some of which have progressed to phase II clinical trials. Most of these attempts, however, have focused on HCV genotypes 1 and 2 as vaccine targets, and almost no attention has been given to HCV genotype 4 (HCV-4), the viral genotype most prevalent in the Middle East and Central Africa. In an attempt to fill this gap in HCV-4 vaccine research, this project describes the *in silico* identification of a group of highly conserved and immunogenic T-cell epitopes from the HCV-4 proteome, using the iVAX immunoinformatics toolkit (EpiVax Inc., RI, USA), as a first step towards the development of an epitope-driven vaccine against the viral genotype. Furthermore, it puts forth a fast and inexpensive method for the validation of the results retrospectively using the repository of empirical HCV immune epitope data on the Immune Epitope Database (IEDB). 90 HLA class I and 14 HLA class II epitopes were identified. From those, 20 HLA class I epitopes were found to be previously uncharacterized, while the *in silico* HLA binding predictions for 27 others (class I and class II) have been retrospectively validated. The retrospective validation results for 4 of the identified HLA class II epitopes were confirmed by a pilot HLA class II binding assay. Furthermore, an investigation of the conservancy of a selected set of the identified epitopes in newly re-sequenced HCV strains from the Egyptian population was performed. The identified and retrospectively validated set of epitopes constitutes a good target for further immunogenicity testing and epitope-driven vaccine development against HCV-4.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CCL: Chemokine (C-C motif) ligand

CCR5: C-C chemokine receptor type 5

CD: Cluster of Differentiation

CLDN-1: Claudin-1

CMI: Cell Mediated Immunity

CXCL9: Chemokine (C-X-C motif) ligand 9

CXCR3: Chemokine (C-X-C motif) receptor 3

DC: Dendritic Cell

DNA: Deoxy-ribonucleic acid

ER: Endoplasmic Reticulum

FOXP3: Forkhead box P3

GAG: Glycosaminoglycans

HCV: Hepatitis C Virus

HCV-4: Hepatitis C Virus Genotype 4

HCVpp: Hepatitis C Virus pseudo type particles

HDL: High Density Lipoproteins

HIV: Human Immunodeficiency Virus

HLA: Human Leukocyte Antigen

ICS: Immunogenic Consensus Sequence

IDV: Immunome Derived Vaccine

IEDB: Immune Epitope Database

IFN: Interferon

IFNAR: Interferon-α/β receptor

IFNGR: IFN-γ receptor

IL: Interleukin

IP-10: Interferon gamma-induced protein 10

IRES: Internal Ribosomal Entry Site

ISG: Interferon Stimulated Genes

ISRE: IFN-stimulated response element

JAK/STAT: Janus Kinase-Signal Transducer and Activator of Transcription signal transduction

LANL: Los Alamos National Lab

LDL: Low Density Lipoproteins

LDLR: Low Density Lipoprotein Receptor

MAVS: Mitochondrial antiviral signaling protein

MHC: Major Histocompatibility Complex

NK: Natural Killer Cells

NS: Non Structural

OAS: 2'–5' oligoadenylate synthetase

ORF: Open Reading Frame

PAMP: Pathogen Associated Molecular Patterns

PD-1: Programmed Death-1 receptor

PKR: Protein Kinase R

PP2A: protein phosphatase 2

RIG-I: Retinoic acid-inducible gene I

RNA: Ribonucleic acid

SOCS3: Suppressor of Cytokine Signaling 3

SR-BI: Scavenger Receptor class B type I

STAT-1: Signal Transducer and Activator of Transcription-1

SVR: Sustained Virological Response

TAP1: Transporter associated with antigen processing 1

TBK-1: TANK binding kinase 1

TCR: T-cell Receptor

TGF-β: Transforming Growth Factor Beta

TLR: Toll-like Receptor

TNF: Tumor Necrosis Factor

TRIF: TIR domain-containing adaptor-inducing IFN-β

UTR: Un-translated Region

VLDL: Very Low Density Lipoproteins

WHO: World Health Organization

1. INTRODUCTION

Hepatitis C is an infectious inflammatory liver disease caused by the Hepatitis C virus (HCV); a blood-borne virus transmitted mainly through the percutaneous route from an infected to a healthy individual. According to the World Health Organization (WHO), approximately 150 million people suffer from chronic HCV infection worldwide, with 3-4 million newly infected each year. Around 350,000 patients die each year as a result of HCV-related diseases [1]. There are six major genotypes of the virus whose distribution and prevalence varies widely between different regions and countries [2]. Global infection rates peak in Egypt, where 22% of the population is chronically infected, and an estimate of 500,000 get newly infected each year [1,3]. Over 90% of the Egyptian patients are infected with HCV genotype 4 (HCV-4), the most prevalent genotype in the Middle East and Central and Western Africa [4-6]. The high prevalence of HCV in Egypt has been attributed to the use of contaminated needles in an antischistosomal therapy campaign that took place in the 1950s [4-7]; and currently, the main route of transmission is through clinical procedures in which improperly sterilized instruments are utilized [4,6].

Apart from the percutaneous route of transmission, HCV transmission can also occur through sexual and perinatal exposures; these routs of transmission, however, are much less efficient and are not major contributors to new infections [8]. Following the onset of infection, the disease starts with an acute phase that lasts for about 6 months. During that phase, 15-30% of the patients are able to clear the virus spontaneously; however in 70-85% of the patients the virus manages to evade the host's immune system and persist, leading to the progression of the disease to a chronic phase. The chronic phase can last for decades, eventually culminating in liver cirrhosis and/or hepatocellular carcinoma for 5-20% of the chronically infected patients and death for 1-5% of the patients [1,9,10].

HCV is an enveloped RNA virus belonging to the *Hepacivirus* genus of the family *Flaviviridae*. Its genome consists of a single sense RNA strand of an approximate size of 9600 bases; harboring a single open reading frame (ORF) that is flanked by 5' and

3' untranslated regions (UTR). The ORF (also known as the polyprotein gene) is translated into a precursor polyprotein molecule approximately 3000 amino acids in size. This molecule is processed during and after its synthesis via viral and host cellular machinery into 10 viral proteins, divided into 3 structural and 7 non-structural (functional) proteins (Table 1) [11]. HCV replicates via a viral RNA-dependant-RNA polymerase known as Non-Structural protein 5B (NS5B). The enzyme lacks the proof-reading capability normally present in DNA polymerases, causing the virus to experience a very high mutation rate upon replication of approximately 0.0014 to 0.0019 mutations per nucleotide per year [11,6]. This high rate of mutation is one of the main mechanisms by which the virus manages to evade the host's adaptive immune system; furthermore, it results in the diversification of the virus into a wide array of genetic variants. These variants are classified based on the extent of genetic variation into genotypes, varying by 30-33% of the genome, subtypes, varying by 20-25% of the genome, and quasi-species within each patient that may vary from a single point mutation up to 10% or more of the genome [2,11]. HCV primarily infects and replicates in liver hepatocytes of human and chimpanzee hosts only [12]. Reports also indicate that it is able to infect some peripheral blood mononuclear cells such as B-cells and dendritic cells (DCs); however it does not replicate within those cells [13].

After the virus infects the liver hepatocytes and commences replication, the host's innate immune system in response attempts to protect the healthy liver hepatocytes from infection and kill the infected ones by the actions of Interferons (IFNs) and Natural Killer (NK) cells. The adaptive immune response that is later initiated involves neutralizing antibodies that may block viral entry into healthy hepatocytes, and the killing of the infected hepatocytes by the action of the CD8[+] cytotoxic T-cells [13-15]. Although in 15-30% of the patients the immune system manages to spontaneously clear the virus, in the majority of the patients the virus manages to evade the immune system by several mechanisms; leading to the persistence of infection. Chief amongst those mechanisms is the highly mutable nature of the virus, which allows for the production of escape variants that evade recognition by the cells of the adaptive immune system. Other mechanisms include inhibition of type I IFN production, induction of T-cell exhaustion, and induction regulatory T-cell action [13,14]. The outcome of HCV infection (spontaneous clearance

or persistent infection) is believed to be determined by the strength and nature of the immune response during the acute phase of the infection. A strong base of evidence demonstrates that a strong, multi-specific and sustained T-cell response to the virus is associated with viral clearance. On the other hand, a weak, narrow, and short-lasting T-cell response is associated with viral persistence. The role of neutralizing antibodies in HCV clearance is debatable, while studies show an association between a rapidly-induced production of high titers of cross-neutralizing antibodies and spontaneous HCV clearance; others have shown that viral clearance is possible in the absence of anti-HCV neutralizing antibodies [13,15].

The standard treatment regimen for HCV infection is a combination therapy of pegylated IFN alpha and ribavirin. The aim of treatment is to achieve a sustained virological response (SVR); defined as the absence of HCV RNA from the patient's blood 24 weeks after the cessation of treatment. This treatment, however, is costly, causes severe side effects, and does not guarantee the elimination of infection [10]. Recently, with the advent of protease inhibitor direct-acting antiviral drugs (e.g.Telaprivir and Boceprivir), the optimum treatment regimen for HCV genotype 1 infection has been altered to include a protease inhibitor in a triple therapy with pegylated IFN alpha and ribavirin. This new regimen increases the chances of achieving SVR for genotype 1 infected patients; however it is still expensive, has associated side effects (e.g. dysgeusia, severe fatigue, debilitating depression, and hemolysis), and is less effective for treating infections with other HCV genotypes such as genotype 4 [16]. The high costs, significant side effects, and uncertainty in achieving SVR associated with HCV therapy warrant support for the development of an HCV vaccine, a goal that so far has not been achieved [17].

On top of the persisting challenge of overcoming the various mechanisms by which the virus evades the immune system, HCV vaccine development was hampered by other technical obstacles. Due to the fastidious nature of the virus, HCV could not be produced in tissue culture systems nor was there a readily available animal model of infection for it. The gap created by the lack of cell culture systems was filled by the development of HCV pseudo type particles (HCVpp) in 2003, and then in 2005 the first

tissue culture system of infection was developed [18]. This allowed for the identification of the receptors used by the virus for cell entry; which allowed for the production of humanized mouse models expressing those receptors. These models are still early in development; however their availability could accelerate preclinical vaccine testing [17,19]. In spite of those challenges, several vaccine studies have been conducted on animal models in the past decade yielding promising results; a small number of which have actually progressed to human phase II clinical trials.

As aforementioned, HCV-4 is highly prevalent in the Middle East and Central and Western Africa and is increasingly spreading to countries in Southern and Western Europe and to North America with increasing immigration and travel [20]. In Egypt, more than 90% of HCV patients are infected with HCV-4; more than 500,000 new infections are estimated to occur in Egypt annually [3,5,7]. Despite the direness of the situation, very little attention has been given to the development of therapies and vaccines for HCV-4. According to a meta-analysis study performed on all the existing HCV epitope data on the Immune Epitope Database (IEDB) (a repository of immune epitope data manually curated from reported peer reviewed literature, patents, and direct submissions from companies and institutions [21]), only 13 epitopes for HCV genotypes 4, 5, and 6 combined were reported out of 3444 unique reported HCV epitopes that tested positive in T-cell and B-cell assays [22]. In an attempt to fill this wide gap in HCV-4 immune epitope research, and to address the un-tackled subject of HCV-4 vaccine design, this project aims to identify, *in silico*, a set of highly immunogenic and conserved T-cell epitopes from within the HCV-4 proteome as targets for an epitope-driven vaccine for HCV-4. The epitopes were identified using the validated immunoinformatics iVAX toolkit (EpiVax Inc., RI, USA), a suite of immunoinformatics tools that can be used for the *in silico* design of epitope-driven vaccines derived from protein sequences of interest [23]. This project also describes a fast and inexpensive method for the retrospective validation of the *in silico* results using the repository of HCV immune epitope data on the IEDB, which can accelerate and cut-down the expenses of the development of the vaccine.

2. LITERATURE REVIEW

2.1 *The Life Cycle of HCV*

The HCV life cycle begins with the entry of the virus into the liver hepatocytes. In the blood, HCV circulates in both free and bound forms; where it can be associated with low density lipoproteins (LDL), very low density lipoproteins (VLDL), and immunoglobulins [12,24,25]. Once the virus reaches the liver, its entry into the hepatocytes is mediated by the attachment of the membrane-bound viral E1-E2 heterodimer (formed by the dimerization of the HCV structural proteins E1 and E2) to hepatocyte cell-surface molecules and receptors; these include CD81, a tetraspanin protein expressed on the surface of several cell types, and Scavenger Receptor class B type I (SR-BI), a glycoprotein whose natural ligand is high density lipoproteins (HDL). Also among those molecules and receptors are the Low Density Lipoprotein receptor (LDLR), the tight junction protein Claudin-1 (CLDN-1), glycosaminoglycans (GAG), and others. Although further investigation needs to be carried out, studies show that LDL receptors together with glycosaminoglycans may act as the primary collectors of HCV particles from the blood stream for further targeting to other HCV receptors such as CD81. Studies show that CD81 and SR-BI are both necessary but not sufficient for viral entry [12,25]. The tight junction protein CLDN-1 was found to be an essential co-receptor acting at a later stage of the entry process after the virus has interacted with CD-81. However, like CD81 and SR-BI, it was found insufficient (when co-expressed with CD81 and SR-BI) for mediating viral entry [12]; meaning that there are other molecules required for the completion of the process. Following attachment, HCV enters the cell via Clathrin-dependent endocytosis. The low pH-environment inside the endosome induces the fusion between the viral and endosomal membranes; which is followed by the release of the viral nucleocapsid into the cytoplasm [12,25]. Figure 1 schematically represents the viral entry step of the HCV life cycle as it is currently understood.

After the release of the nucleocapsid into the cytoplasm, the single sense genomic RNA strand of HCV is released, and serves as the messenger RNA for the synthesis of the viral precursor polyprotein molecule. The translation of the RNA strand is controlled

by an Internal Ribosomal Entry Site (IRES), spanning domains II, III, and IV of the 5'UTR (Figure 2) and the first 24-40 nucleotides of the region coding for the Core structural protein. The IRES mediates the initiation of cap-independent translation of the polyprotein gene. [12,25]. After the synthesis of the polyprotein molecule is initiated, the molecule is directed to the Endoplasmic Reticulum (ER) mediated by an internal signal peptide located between the amino acid sequences of the Core and E1 structural proteins; this sequence is responsible for mediating the translocation of the ectodomain of the E1 protein into the ER lumen. At the ER, the polyprotein molecule is processed via host and viral enzymes during and after its synthesis into the structural and non-structural proteins of HCV. First, the three structural proteins (Core, E1, and E2) and the non-structural protein p7 are processed via host ER signal peptidases. The remaining 6 non-structural proteins, namely NS2, NS3, NS4A, NS4B, NS5A, and NS5B, are processed via the *cis*-acting viral auto-protease NS2-3, and the *cis* and *trans* acting NS3-NS4A protease. First, the NS2-3 *cis*-acting auto-protease cleaves the bond between the NS2 and NS3 polypeptides in the polyprotein chain [12,26]. Then, the NS3 protease with the NS4A protein as its co-factor catalyzes *cis*-cleavage of the NS3/NS4A junction and the *trans*-cleavage of the NS4A/NS4B, NS4B/NS5A, and NS5A/NS5B junctions. The processed structural and non-structural proteins become associated with the ER membrane [12,25]. Figure 2 represents schematically the relative positions of the amino acid sequences of the HCV structural and non-structural proteins in the precursor polyprotein sequence, and figure 3 schematically represents the 3D structures of the proteins and their associations with the ER membrane.

Similar to all positive-strand RNA viruses, HCV RNA replication takes place at a membrane-associated replication complex composed of replicating RNA, viral proteins, and rearranged cellular membranes. The site of HCV RNA replication in the cell is called the membranous web; the name given to a specific alteration of the ER whose formation is induced by the NS4B viral protein alone. At the membranous web, HCV RNA replication is catalyzed by the RNA-dependant RNA polymerase NS5B in a two-step, semiconservative, and asymmetric process. In the first step, the positive RNA strand serves as a template for the synthesis of an intermediate negative RNA strand. The negative strand is then used as a template in the second step for the synthesis of several

positive RNA strands; which will later be used for precursor polyprotein synthesis, the synthesis of new negative RNA strand intermediates, and for packaging into new virions for release from the cell. Both steps of the replication process are catalyzed by NS5B. The replication process is modulated by the fluidity of the membranous web which is dictated by the degree of saturation of its fatty acid component molecules; where the abundance of cholesterol and saturated fatty acids stimulates viral replication and that of poly-unsaturated fatty acids inhibits it [12, 25,26].

Following viral RNA replication and protein synthesis, the newly synthesized positive RNA strands are packaged into new viral particles which are later released from the cell. According to available evidence, this packaging process is most likely initiated by the interaction of the newly synthesized positive RNA strands with the Core viral proteins, which in themselves are capable of self-assembly in order to form the nucleocapsid. Virions are then presumably formed by budding into the ER lumen or that of an ER-derived compartment; where they are then released from the cell via the constitutive secretory pathway [12, 25,26].

2.2   *Immunity to HCV*

2.2.1   Effector Mechanisms of the Host Immune Response:

2.2.1.1 Innate Immunity - Interferons

Immediately after HCV infects the liver hepatocytes, the host's innate immune system attempts to combat the infection mainly via the actions of IFNs and NK cells. Interferons are a family of secreted cytokines that act in autocrine and paracrine manners to stimulate intracellular networks that modulate mechanisms of resistance to viral infections, enhance and shape innate and adaptive immune responses, and regulate the survival and death of normal and cancerous cells [27]. They are categorized based on their cell-surface receptors into three types: type I, type II, and type III. Despite their structural and genetic differences, type I and III Interferons share identical induction mechanisms, signal transduction pathways, and biological actions; where they both induce antiviral activity in their target cells [28]. Type I IFNs are structurally subdivided

into IFN-β and several subtypes of IFN-α; they can be expressed in most body cell types, and their receptor is similarly widely expressed. Type III IFNs, also known as λ interferons, are divided into 3 subtypes namely IFN-λ1, IFN-λ2, and IFN-λ3, which are otherwise known as interleukin (IL)-29, IL-28A, and IL-28B respectively; they are expressed in several –but not all- cell types. In contrast to type I and type III Interferons, the type II interferon, comprised in a single molecule denoted IFN-γ, is expressed only in NK cells and activated T-cells, and possesses mainly immunomodulatory actions [29,30]. IFN-γ will be discussed in more detail later in this section within the context of NK cell and T-cell response mechanisms.

The production of type I interferons is most potently stimulated by viral infections; specifically by the double-stranded RNA intermediates produced during the replication of viruses such as HCV [29,31]. After their production, type I interferons induce antiviral responses in their target cells by binding to a heterodimeric cell-surface receptor called the Interferon-α/β receptor (IFNAR), which is composed of two subunits, namely IFNAR1 and IFNAR2. The binding of the interferons to the receptor causes its subunits to dimerize, activating a Janus Kinase-Signal Transducer and Activator of Transcription (JAK/STAT) signal transduction pathway. The activation of the pathway culminates in the activation of transcription of a set of genes, known as the Interferon Stimulated Genes (ISG), whose protein products act by inhibiting viral replication in the cell. This protects paracrine-induced uninfected cells from infection, and limits the spread of the virus from the autocrine-induced infected cells [29,31]. In addition to inhibiting viral replication, type I IFNs also promote adaptive cell-mediated immunity against intracellular microbe infections by increasing the expression of class I Major Histocompatibility Complex (MHC) molecules on the surface of the infected cells. They also stimulate the development of $CD4^+$ T_H1 helper T-cells, and enhance the effector functions of NK cells as will be explained later [31].

Similar to type I IFNs, the production of type III IFNs is also stimulated by viral infections. Studies show that both types of IFNs are co-produced upon induction with all the inducers that were investigated; suggesting that they are both regulated by the same mechanism [28]. Upon binding to their receptor, type III IFNs also activate a JAK/STAT

8

signaling pathway that culminates in the activation of transcription of a set of ISGs similar to those activated by type I IFNs; thereby also inhibiting viral replication in paracrine and autocrine manners. However, the heterodimeric cell-surface receptor to which type III IFNs bind is different, where it is composed of two polypeptide chains called IFNLR1 and IL10R2. This receptor is not expressed as broadly as the IFNAR receptor on different body cells, however it is expressed in liver hepatocytes [28,29].

Studies show that ISGs are generally highly expressed in the livers of HCV-infected patients [29,30]. Available cell culture studies indicate that HCV pathogen associated molecular patterns (PAMPs) responsible for the induction of IFN production, which include double-stranded RNA (dsRNA) and viral proteins, are identified by the PAMP receptors Retinoic acid-inducible gene I (RIG-I) and the Toll-like receptor 3 (TLR3); the PAMP receptors TLR2 and TLR7 may also be implicated in HCV detection [30]. RIG-I is a cytoplasmic RNA helicase that senses non-self, short dsRNA or ssRNA with a free 5'-triphosphate moiety. Toll-like receptors, in general, are membrane-bound, mainly endosomal receptors; TLR3 senses dsRNA and viral DNA, TLR2 senses certain viral proteins, and TLR7 (mainly expressed in plasmacytoid dendritic cells) senses ssRNA. Following HCV detection, these receptors trigger signal transduction pathways that induce the expression of type I and III IFNs, which are then secreted to induce the expression of ISGs in paracrine and autocrine manners. An example of the ISGs expressed in HCV infection is the IFN-stimulated gene 20-kDa protein (ISG20), whose protein product is a $3'–5'$ exonuclease that suppresses viral replication by targeting ssRNA. Another example is the 2'–5' oligoadenylate synthetase (OAS) enzyme gene, whose protein product catalyzes the synthesis of 2'-5'-oligoAdenylic acid, which in turn activates a latent endoribonuclease called RNase-L. RNase-L catalyzes the degradation of viral ssRNA as well as cellular mRNA and rRNA thereby globally inhibiting protein synthesis in the cell and triggering its apoptosis [29,30]. The production of type I and III IFNs triggered by HCV is depicted in detail in figure 4.

2.2.1.2 Innate Immunity – Natural Killer Cells

In addition to the antiviral actions of IFNs, NK cells combat HCV infection by triggering the apoptosis of infected hepatocytes, thereby preventing the production and

spread of new HCV virions to healthy hepatocytes. NK cells are a subset of lymphocytes belonging to the T-cell lineage that recognize and kill pathogen-infected cells, cancer cells, and cells not expressing self MHC class I molecules, without the need for prior sensitization like that required by the T and B cells of the adaptive immune system [31,32]. Their activation is regulated by a balance between signals produced from activating and inhibitory receptors expressed on their surface. The inhibitory receptors bind to self-MHC class I molecules normally expressed in all body cells independent of the T-cell epitope that these molecules are presenting in their binding clefts. The activating receptors bind to pathogen-derived molecules (proteins and carbohydrates) or non-classical MHC molecules that are expressed/upregulated on the surface of infected, cancerous, or otherwise stressed body cells. Usually if the inhibitory receptors are engaged, the NK cell is not activated due to the dominance of the inhibitory signal. If however the inhibitory receptors are not engaged due to the absence of enough normal self-MHC I molecules on the surface of the target cell, and at the same time the activating receptors are engaged to their target ligands, the NK cell then becomes activated [31,32]. It is important to note, however, that if several activatory signals are generated at once, they can collectively overwhelm the inhibitory signal and activate the cell. Furthermore, in some cases, the engagement of even one activatory receptor may be enough to override the inhibitory signal and activate the cell [32].

Following activation, NK cells induce apoptosis in their target cells. One way by which that happens is via enzymes called Granzymes, which enter the target cells through pores in their membrane created by a protein called Perforin. Both Perforins and Granzymes are stored in granules in the cytoplasm of NK cells, and are released upon the cell's activation. In addition, activated NK cells secrete IFN-γ [31,32]. IFN-γ performs its biological actions via binding to a dimeric receptor that is ubiquitously expressed on all nucleated cells called the IFN-γ receptor (IFNGR) [27]. Upon binding, IFN-γ triggers a JAK/STAT signaling pathway that culminates in the expression of a set of ISGs that are overlapping with –but not identical to- the ones induced by type I and III IFNs. In innate immunity, similar to type I and III IFNs, IFN-γ possesses direct antiviral activity, where it induces the expression of ISGs that inhibit viral replication in its target cells (e.g. 2'–5' oligoadenylate synthetase (OAS) and Protein Kinase R (PKR)). It also serves as a

modulator, where it activates macrophages to kill phagocytosed microbes and stimulates the cytotoxic activity of NK cells [33]. In addition to its functions in innate immunity, IFN-γ also enhances MHC-associated antigen presentation and recognition during the cellular adaptive immune response via stimulating the expression of both classes of MHC molecules, co-stimulatory molecules on Antigen Presenting Cells (APCs), and intracellular proteins implicated in antigen processing [31-33].

In acute HCV infection, studies show that NK cells in peripheral blood increase in number and become activated; up-regulating the expression of the activatory receptor NKG2D, and down-regulating the expression of inhibitory receptor NKG2A. Furthermore, the levels of IFN-γ produced from the NK cells and their cytotoxicity was shown to be higher than their counterparts in healthy controls [34]. In the chronic phase of the infection, generally, the number and function of NK cells are down-regulated. The decrease in NK cell number could be attributed to the decrease/lack of IL-15 produced by DCs, the interleukin that is essential for NK cell development and functionality. However, it could also be due to the compartmentalization of NK cells away from the peripheral blood stream (the site which early studies concerning the subject have concentrated on in their analysis). The down-regulation in NK cell function is reflected in the decrease in their IFN-γ producing activity and their ability to activate DCs [34,35]. Furthermore, NK cells exhibit an increase in IL-10 and TGF-β production; producing an immunosuppressive environment that encourages viral persistence [33]. Despite this down-regulation in number and cytokine-secreting function, the natural cytotoxic functionality of NK cells appears to be intact or increased. This could be explained by the fact that there are activatory receptors that are up-regulated on their surface during the chronic phase of the infection (e.g. NKp30, NKp44 and NKp46) [34].

2.2.1.3 Innate Immunity – Other Mechanisms

In addition to the actions of IFNs and NK cells, the production of other inflammatory cytokines and chemokines is also induced in the HCV-infected liver; which are also important mediators bridging the innate and adaptive immune responses, and are critical for T-cell homing to the infected liver. Examples of chemokines produced in the

liver during HCV infection are the C-C chemokine receptor type 5 (CCR5) ligands: Chemokine (C-C motif) ligand 3 (CCL3) and CCL5, and the Chemokine (C-X-C motif) receptor 3 (CXCR3) ligands: Interferon gamma-induced protein 10 (IP-10) and Chemokine (C-X-C motif) ligand 9 (CXCL9). Studies have shown that there is a correlation between the levels of some of those chemokines in the blood of HCV patients and the outcome of infection or the severity of liver inflammation [29].

The innate immune response to HCV buys time for the development of the adaptive immune response to the virus, in addition to being an important factor in shaping the later [29]. Unlike the innate immune response, triggering the cells of adaptive immune system, the B and T lymphocytes, requires the specific recognition of viral antigens. Adaptive immunity is divided into humoral immunity, and cell mediated immunity (CMI). Humoral immunity targets extracellular pathogens via antibodies, the antigen receptors and effector molecules of B-cells, whereas CMI targets intracellular pathogens via the Cytotoxic actions of CD8$^+$ T-lymphocytes [36].

2.2.1.4 Adaptive Immunity – Cell Mediated Response

CMI is the effector response of T-lymphocytes against microbes that infect non-phagocytic cells and those that survive within phagocytes of phagocytic cells. The best defined functional populations of T-lymphocytes involved in the CMI response are the helper CD4$^+$ T-cells, which play an essential role in stimulating and coordinating the adaptive immune response (both humoral and cellular), and the cytotoxic CD8$^+$ T-lymphocytes (CTL), whose role is to kill microbe-infected cells thus inhibiting the spread of the infection to un-infected cells [37]. The key to CD4$^+$ and CD8$^+$ T-lymphocyte activation is their recognition of microbial antigens in association with Major Histocompatibility Molecules via their antigen receptor, namely the T-cell receptor (TCR). Major histocompatibility molecules are specialized proteins encoded by genes in the Major Histocompatibility Complex (MHC) locus found on the chromosomes of all vertebrates, whose role is to display peptide fragments (epitopes) of digested protein antigens on the surface of body cells for recognition by T-lymphocytes. Class I MHC molecules present epitopes for recognition by CD8$^+$ T-cells, and are expressed on almost

all nucleated body cells, while class II MHC molecules present epitopes for recognition by CD4[+] T-cells, and are only present on the surface of antigen presenting cells (APCs) such as DCs, mononuclear phagocytes, and B-lymphocytes [36].

The only professional APCs that are capable of activating naïve T-cells and initiating a primary CMI response are DCs, since they are the only APCs that possess the co-stimulatory capacity required to activate those cells [38,39]. The two main types of DCs isolated from peripheral blood are myeloid/conventional dendritic cells (mDC), which originate from myeloid precursor cells, and plasmacytoid dendritic cells (pDC), which originate from lymphoid precursor cells. mDCs and pDCs differ in their capabilities of antigen presentation, expression of co-stimulatory molecules, and cytokine production. mDC are much more efficient than pDC at antigen capture, processing, and presentation. In the periphery, immature DCs (iDCs) are efficiently capable of capturing exogenous antigens via the cell-internalization processes of receptor-mediated endocytosis, macropinocytosis, and phagocytosis [38,39]. In the case of HCV, immature DCs capture HCV antigens via phagocytosis of apoptotic infected hepatocytes, endocytosis of HCV-immune complexes, and/or macropinocytosis of free virions [40]. Following antigen capture, DCs begin their maturation process triggered by microbial products and inflammatory cytokines, and immigrate towards the secondary lymphoid organs for the activation of antigen-specific naïve T helper and cytotoxic cells. DCs are capable of presenting both MHC class II and class I antigenic epitopes via the exogenous antigen processing pathway and cross-presentation respectively [38]. Figure 5 schematically represents the maturation process of iDCs into mDCs following HCV capture, and the subsequent activation of T lymphocytes.

In chronic HCV infection, studies show that the numbers of DCs (both pDC and mDC) in peripheral blood decrease and significantly increase in the liver; with the ratio of mDCs to pDCs higher than that of peripheral blood [39]. Several studies also pointed out that despite their distribution during chronic infection, DCs are functionally impaired. mDCs have been reported to exhibit impaired maturation, decreased IL-12 production, increased IL-10 production, lower ability to stimulate allogenic CD4[+] T-cells, and higher ability to stimulate T-regulator (T$_{reg}$) cells; thereby, like NK cells, they produce an

immunosuppressive environment that encourages viral persistence [39,41]. However, other studies produced conflicting evidence showing that DCs in chronic HCV infection are not functionally impaired. This contradiction in results is due to several factors such as the limited number of study subjects, differences in host-related factors (e.g. age, gender), source tissue of the analyzed DCs, and others; confirmation of either claim requires further investigations to be carried out [13,39].

Upon reaching the peripheral lymphoid organs, mature DCs expressing the B7 co-stimulatory molecules present the captured and processed HCV antigens to naïve CD8[+] and CD4[+] T-cells on MHC class I and class II molecules respectively. The recognition of those antigens, in addition to co-stimulation by the B7 molecules and stimulation by activator cytokines produced by the DCs, lead to the activation of circulating antigen-specific T-cells. Consequently, the activated T-cells proliferate and differentiate into effector and memory cell subsets [42]. Activated CD4[+] T-helper cells differentiate into the IFN-γ secreting T_H1 effector subset, induced by the cytokines IL-12 and IFN- γ found in the immediate cytokine milieu. Activated CD8[+] T-cytotoxic cells differentiate into CTL effector cells which harbor membrane-bound cytoplasmic granules carrying proteins required for killing infected cells, including perforins and granzymes, and are able to produce certain cytokines including IFN- γ. It is important to note that in addition to the activator signals provided by the mature DCs, CD8[+] T-cells also require stimulation by cytokines from T_H1 helper cells to successfully differentiate into CTL effector cells [31,42]. Following differentiation, part of the effector and memory T-cells migrate to the liver, where they are preferentially retained. In the liver, CTLs that recognize their specific antigen in association with MHC class I molecules on the surface of infected hepatocytes become activated and kill the infected cells by inducing apoptosis. Furthermore, CTLs secrete IFN-γ similar to the effector T_H1 cells, which apart from "helping" the differentiation of CD8[+] T-cells into CTLs, stimulates the production of opsonizing IgG antibodies from B-cells and the microbicidal activity of phagocytes [31].

The evolution of the T-cell response in HCV infected patients can be divided into three phases in correlation with the progression of liver inflammation; namely the pre-acute, acute, and post-acute phases. The pre-acute phase corresponds to the first few

weeks following the onset of primary infection where the virus titers rise to very high levels and T-cell responses are still undetectable [43]. The acute phase starts 4-12 weeks after primary infection onset when HCV-specific CD4$^+$ and CD8$^+$ T-cells become detectable in blood. Concurrent with the rise in effector HCV-specific T-cell titers, HCV RNA levels start to fall, often becoming transiently undetectable, and a rise in liver enzyme titers may be experienced characterizing the development of acute hepatitis. Early in the acute phase, when the viral load is still relatively high, activated HCV-specific CD8$^+$ T-cells demonstrate a unique behavior where they do not secrete IFN-γ and are defective in their cyototoxic activity despite their expression of activation markers. This phenotype has been described by some investigators as a "stunned" phenotype, where the CD8$^+$ T-cells act as if they were anergic, and is considered a hallmark of acute HCV infection regardless of its outcome. In patients who resolve the infection, HCV-specific CD8$^+$ T-cell functionality is later restored concurrent with a significant drop in HCV viral titers [43,44].

The T-cell response to HCV is known to peak on average 6-12 months following infection onset; and it is the nature of that response in the acute phase which determines the subsequent infection outcome, or to say the characteristics of the post-acute phase of the infection [43]. Studies show that patients who spontaneously resolve the infection exhibit a strong T-cell response targeting several viral epitopes. In these individuals, the post-acute phase is characterized by the absence of HCV RNA from peripheral blood and the persistence of stable long-lasting CD8$^+$ and CD4$^+$ memory T-cell pools after the normal decrease in effector T-cell levels following HCV eradication. On the other hand, patients in whom the disease progresses to a chronic phase exhibit a weak, narrow (targeting very few epitopes), and short-lasting T-cell response to the virus. In these individuals, the post-acute (chronic) phase is characterized by persistence of HCV RNA in the blood, along with a permanent, almost complete loss of HCV-specific CD4$^+$ T-cells and a dramatic decline in HCV-specific CD8$^+$ T-cells which become persistently functionally defective. The almost complete absence of CD4$^+$ cells is critical for viral persistence, and may also be responsible for the impaired CD8$^+$ T-cells' function [43,44]. Here it is important to note that the strength of the CD4$^+$ T-cell response in acute infection alone is not an accurate predictor of infection outcome, as studies showed that

some individuals who managed to mount a strong CD4$^+$ T-cell response against HCV still developed chronic infection following an unexplained loss of that strong response [43]. A graphical depiction of the 3 phases of the T-cell response to HCV and their possible outcomes are illustrated in figure 6.

2.2.1.5 Adaptive Immunity – Humoral Response

The second arm of the adaptive immune response against HCV is the humoral immune response, the effector response of B-lymphocytes against extracellular microbes and microbial toxins mediated by neutralizing antibodies secreted from those cells [31]. Unlike CMI where T-cells can only recognize linear epitopes, B-cells can recognize both linear and conformational epitopes on their target antigens via their membrane-bound antibody receptors. Anti-HCV antibodies become detectable in the patients' blood around 4-14 weeks post-infection [15]. The natural targets of these antibodies are epitopes of the structural E2 and E1 HCV envelope proteins; however, anti-HCV antibodies targeting epitopes of non-structural HCV proteins or their incomplete degradation products released from dying hepatocytes are also detected [15,45]. By binding to the E1 and E2 proteins, anti-HCV neutralizing antibodies would block the entry of HCV into uninfected hepatocytes, and facilitate the capture of HCV by macrophages (opsonization) [43]. The antibodies can be either isolate-specific, targeting epitopes only found in the viral strain infecting the patient, or cross-neutralizing, targeting epitopes found in several strains, possibly providing cross-genotype protection [15].

In addition to antigen recognition, the activation of protein antigen-specific naïve B-cells in peripheral lymphoid organs requires help from effector CD4$^+$ T-helper lymphocytes. Upon binding their antigen, naïve B-cells internalize the antigen-antibody receptor complex, process the internalized protein antigen, and present its epitopes on MHC class II molecules for CD4$^+$ T-cell recognition. When antigen-specific effector T-helper cells recognize the presented antigen, they send activatory signals to the antigen-presenting B-cell in the form of cytokines and CD40-Receptor to CD40-Ligand (CD40:CD40L) interactions; these signals stimulate B-cell proliferation, differentiation into antibody-secreting plasma cells, and antibody heavy chain isotype switching.

Furthermore, CD4$^+$ T-cell help is also required for antibody affinity maturation and differentiation of some of the activated B-cells into memory cells [42]. In HCV infection, the activated T$_H$1 CD4$^+$ effector cells play the aforementioned helper role. The IFN-γ secreted by those cells stimulates antibody isotype switching to Immunglobulin G (IgG) antibodies, which, upon binding to the structural viral proteins, block viral entry into the hepatocytes and promote phagocytosis by macrophages [31].

2.2.2   Viral Evasion Mechanisms from the Immune Response:

In order to persist, HCV has developed several mechanisms by which it can evade both arms of the host immune response [13]. First, HCV counteracts the innate immune response by interfering with type I IFN production and function and inciting NK and DC cell dysfunction [13,14]. Later when the adaptive immune response develops, HCV owes its survival mainly to the highly mutable nature of its genome, driven by its RNA-dependant RNA polymerase, NS5B, which lacks proofreading capability. Furthermore, HCV counteracts the effector functions of the cellular immune response by causing the exhaustion of T-effector cells and by inducing the activation of regulatory T-cells. It also evades the neutralizing actions of anti-HCV antibodies by physically shielding the target epitopes of those antibodies by several mechanisms that will be explained in this section [13,15].

When HCV structural and non-structural proteins are synthesized, they interfere with type I IFN production and function by interacting with key components of the signal transduction pathways leading to type I IFN and ISG expression, and later with the products of the ISG genes themselves [14,29,30]. The NS3/4A viral protease and the NS3 protein interfere with the expression of IFN-β, where NS3/4A cleaves the adapter proteins MAVS (mitochondrial antiviral signaling protein, also known as VISA, Cardif, and IPS-1) and TRIF (TIR domain-containing adaptor-inducing IFN-β) that are essential components of the RIG-I and TLR3 signaling pathways respectively, and where NS3 interacts with TBK-1 (TANK binding kinase 1) which is a component of the RIG-I signaling pathway, leading to a decrease in IFN-β expression as they accumulate in the cells [14,30]. The structural Core protein and NS5A later interfere with the expression of

ISGs by inhibiting the activation of the Signal Transducer and Activator of Transcription-1 (STAT-1) protein of the JAK/STAT pathway that leads to the induction of ISG expression. The Core protein was also shown to inhibit the binding of the ISG factor 3 (ISGF3) transcription factor to the IFN-stimulated response element (ISRE), induce the expression of the JAK/STAT pathway inhibitor "Suppressor of Cytokine Signaling 3" (SOCS3), and induce the upregulation of protein phosphatase 2 (PP2A), which then indirectly attenuates the transcriptional activity of ISGF3 [14,29,30].

In addition to interfering with the signal transduction pathways leading to IFN and ISG expression, the HCV proteins NS5A and E2 interfere with the products of the ISG genes themselves. Studies reported that the NS5A and E2 proteins of HCV genotype 1 strains both interact with and inhibit Protein Kinase R (PKR), which originally reduces protein synthesis in the cell and concurrently inhibits viral replication [14,29,30]. These effects on PKR function however have to be validated since reports on them vary depending on the HCV isolate and experimental setting, and furthermore, they were not consistent in the reports of different research groups [29]. In addition to its interaction with PKR, NS5A was also reported to inhibit 2'–5' oligoadenylate synthetase (OAS) and induce the expression of IL-8; which inhibits overall ISG expression [14,29]. Figure 7 demonstrates the interaction of the aforementioned HCV proteins with the different components of the signaling pathways leading to IFN and ISG expression and with the products of the ISGs.

In addition to counteracting type I IFN production and function, as aforementioned, HCV persistence culminates in NK and DC cell dysfunction observed in the chronic phase of the infection. Both NK and DC cells start to release cytokines that produce an immunosuppressive environment where the CD4$^+$ T$_H$1 cell response is weakened; encouraging viral persistence. Till now, the underlying molecular mechanisms causing NK and DC cell dysfunction have not been fully elucidated. With regard to NK cell dysfunction, a study conducted by Nattermann et al. in 2005 concluded that an epitope from the HCV core protein is responsible for the stabilization of expression of the non-classical HLA class I molecule HLA-E which engages the inhibitory receptor NKG2A; this interaction inhibits NK cell-mediated cytotoxicity [41,46]. Another study

conducted by Herzer. et al in 2003 suggests that the core protein is also responsible for the up-regulation of classical MHC class I expression on the surface of the infected cells via the p-53 dependant up-regulation of the transporter associated with antigen processing 1 (TAP1) protein, which consequently increases MHC class I expression. This enhanced expression leads to the inhibition of NK cell mediated killing of the infected hepatocytes [41,47]. With regard to DC cell dysfunction, studies show that the core, NS3, NS4, and NS5 HCV proteins are implicated in down-regulating the expression of HLA and costimulatory molecules, the reduction of cytokine production, and diminishing the DCs' allostimulatory activity [39].

Prior to the development of the adaptive immune response, several genetic variants (quasi-species) of HCV develop and accumulate driven by the faulty replication process of HCV RNA that is catalyzed by the error-prone viral NS5B polymerase [44,48]. When the adaptive immune response develops, selective immune pressure is exerted on the existing viral variants and a significant decline in viral blood titers is recorded. At this point, only those variants whose epitopes cannot be recognized by the effector cells of the adaptive immune system persist, and the ensuing mutation patterns of HCV become directed at altering those epitopes that are amenable to immune recognition. This is achieved by mutating the amino acid residues within those epitopes that interact with and are recognized by the T-cell receptors and B-cell receptors of effector T-cells and B-effector cells respectively, by mutating the amino acid residues responsible for anchoring T-cell epitopes in the host's HLA molecules, and/or by altering the proteosomal cleavage sites of the viral proteins thereby interfering with T-cell epitope processing and presentation [44,48]. The mutability of those epitopes, however, is restricted by the "fitness cost" those mutations will have on the virus; where the mutation of epitopes located within functional proteins (e.g. NS3, NS5B) will have a higher fitness cost than mutating epitopes located within structural proteins (e.g. E1, E2) as those "functional" epitopes could contain key amino acid residues for the functionality of the proteins they are located within. Targeting such functional epitopes gives the host an advantage and increases the chances of clearing the virus as those epitopes are harder to mutate [44,48].

Besides the mutational escape from recognition by HCV-specific T-effector cells, HCV counteracts the cellular immune response against it by causing T-effector cell exhaustion; which is defined as the gradual loss of T-effector cell functions due to the prolonged exposure to high levels of its cognate antigen [14,41,49]. This might occur due to a high viral replication rate exceeding the capacity of containment by the host's immune system [13,49]. One of the major causes of the induction of T-cell exhaustion in T-effector cells is the high and/or prolonged expression of inhibitory receptors; and one of the major receptors involved is the Programmed Death-1 receptor (PD-1) [49]. Studies show that PD-1 expression is upregulated on HCV-specific effector CD8[+] T-cells in chronic HCV infection as a result of prolonged antigenic stimulation. The PD-1 receptor on these cells interacts with its ligand PD-Ligand 1 (PD-L1) which is expressed on infected hepatocytes preferentially, Kupffer cells, stellate cells, and sinusoidal endothelial cells in the liver. This interaction initiates a signaling pathway that culminates in the inhibition of the T-cell effector functions and eventually causes apoptosis [14,50]. The gradual loss of T-cell effector functions starts by the loss of IL-2 (an important T-cell growth factor) secreting activity, followed by the loss of TNF-α secreting activity and CD8[+] T-cell cytotoxicity, then finally by a substantial decrease in IFN-γ secreting activity [14,41]. The lack of help from CD4[+] T-cells aggravates the severity of CD8[+] T-cell exhaustion [14,49]. This correlates with the aforementioned picture of the blood of chronically infected Hepatitis C patients, where there is a permanent, almost complete loss of HCV-specific CD4[+] T-cells and a dramatic decline in HCV-specific CD8[+] T-cells that become persistently functionally defective.

Another mechanism by which HCV counteracts the cellular immune response is the induction of T-regulatory cells [39,51]. T-regulatory cells are characterized by the expression of the transcription factor forkhead box P3 (FOXP3), the constitutive expression of the IL-2 receptor α-chain (CD25), and the cluster of differentiation-4 (CD4), and their actions are one of the major mechanisms by which the body maintains immune homeostasis, moderates inflammation, and prevents autoimmune reactions in the periphery. There are two kinds of T-regulatory (Treg) cells: natural (n) Treg cells that are selectively produced in the thymus as a consequence of their TCRs' high affinity towards self antigens, and induced (i) Treg cells that are derived from conventional CD4[+] T-

effector cells converted in the periphery via induction by nTreg cells. This conversion is induced either directly through cytokine-dependant mechanisms (IL-35, IL-10 or TGF-β) or indirectly through DC-mediated mechanisms [39,51]. Like CD4[+] T-helper cells, Treg cells are activated in an antigen-specific manner; which happens when their TCR recognizes its cognate epitope in association with MHC class II molecules [52]. When they are activated, they suppress the actions of several types of immune cells including effector T-cells and B-cells, NK cells, DCs, and macrophages via several mechanisms; these include: (1) suppression via inhibitory cytokines (IL-35, IL-10 or TGF-β), (2) suppression of T-effector cells by metabolic disruption, such as triggering apoptosis by IL-2 deprivation driven by the high expression of the IL-2 receptor on the Treg surface, (3) triggering apoptosis of T-effector cells via granzymes A and B and perforin, and (4) inhibiting DC maturation and ability to stimulate naïve T-cells [39,51]. A schematic representation of the suppressive mechanisms of Tregs is illustrated in figure 8.

In HCV infection, studies have shown that the number of Treg cells (both nTregs and iTregs) in the liver and peripheral blood of chronically infected patients is higher compared to uninfected individuals or those who have spontaneously cleared the virus [13,39]. *In vitro* studies showed that the depletion of CD25[+] cells (where CD25 is another marker expressed by Tregs) from chronic HCV patient blood samples resulted in the enhancement of HCV-specific T-effector cell responsiveness[13,14]. While it was unclear whether this increased Treg activity is a byproduct of chronic inflammation and liver disease or the direct induction by T-regulatory cell epitopes, further investigations have uncovered the presence of Treg epitopes (Tregitopes) within HCV structural and non-structural proteins, presenting supporting evidence for the specific induction hypothesis. It was also found that some of these epitopes exhibit strong homology with human self antigens, thus explaining the response and expansion of nTreg cells which are normally involved in the suppression of autoimmunity upon encountering self epitopes [39]. Furthermore, a recent study by Cusick et al demonstrated that, within the course of chronic infection, an immunodominant HCV MHC class II epitope was mutated into a Tregitope under the host's immune pressure [39,53]. These studies elaborate the prominent role of Treg cells in the promotion of chronicity development.

In addition to the mutational escape from recognition by neutralizing anti-HCV antibodies, HCV evades the humoral immune response by several other mechanisms [15,54]. First, HCV can induce the production of non-neutralizing antibodies that interfere with the binding of neutralizing anti-HCV antibodies to their target epitopes. Studies proposed that a group of epitopes located between amino acids 434-446 in the E2 protein region, collectively called "epitope II", trigger the production of non-neutralizing antibodies whose binding to their target epitopes blocks the actions of neutralizing antibodies that target another adjacent E2 protein region between amino acids 412-426 called "epitope I" [15,45,54]. Second, HCV can escape the action of neutralizing antibodies by binding to lipoproteins while circulating in blood, as it has been aforementioned; this association masks the HCV epitopes targeted by those antibodies [12,15,54]. Third, the interaction between the hypervariable region 1 (HVR1) -a highly variable region within the E2 protein-, high density lipoproteins (HDL), and the Scavenger Receptor class B type I (SR-BI) provides protection against neutralizing antibodies targeting epitopes outside HVR1. Studies on HCV pseudo-particles (HCVpp) have shown that the deletion or mutation of HVR1, the removal of HDL particles, or the abolishment of the SR-BI lipid transfer activity obliterates the resistance to antibody-mediated neutralization [15,54]. Fourth, HCV can shield the target epitopes of neutralizing antibodies on the envelope glycoproteins (E1 and E2) by N-glycosylation. Studies on HCVpp show that removing glycans from these proteins increases the susceptibility of neutralization of HCVpp by anti-HCV antibodies [15,54,55]. Finally, HCV can evade recognition by neutralizing anti-bodies by infecting neighboring cells via direct cell-to-cell transfer rather than via the receptor-mediated uptake of HCV particles released from the infected cells into its surroundings [15,54]. The aforementioned mechanisms are illustrated schematically in figure 9. An overall summary of the immune response to HCV and the mechanism by which it evades that response is schematically represented in figure 10.

2.3   HCV Vaccine Development

2.3.1   Characteristics of an ideal HCV vaccine:

In light of the knowledge we have so far on the immune response to HCV, it can be inferred that an ideal HCV vaccine would target highly conserved regions in the HCV genome. Targeting such regions which are mainly conserved for their important structural and functional roles in the HCV life cycle would increase the chances of establishing protection against several HCV strains, as it is less likely for the virus to mutate those regions in order to evade recognition at low fitness costs. An ideal HCV vaccine would also be able to elicit a sustained immune response, one which would successfully create a sustained pool of memory cells that can act upon future encounters with the virus. It would also be capable of triggering a multi-specific response, as studies have shown that targeting several conserved viral epitopes is associated with viral clearance as the virus has a further lower chance of immune evasion via mutating several conserved epitopes at the same time. Furthermore, an ideal HCV vaccine should be safe and not trigger any immunopathology; a goal that has proven to be realistic in the light of previous human studies [17].

2.3.2   Challenges facing HCV vaccine development:

The process of HCV vaccine design and development was hindered by several challenges; some of which have recently been overcome. The first of these challenges was to design a vaccine that would overcome the mechanisms by which the virus evades the host's immune response, foremost the highly mutable nature of the virus. The high mutation rate of the HCV genome makes the development of a cross-genotype effective vaccine very difficult; and for this reason, the development of a prophylactic antibody-inducing vaccine that targets the E1 and E2 envelope proteins, which are the obvious targets for such a vaccine, has been a big challenge, owing to the hyper-variable nature of the major antigenic determinants of those proteins [17].

The second challenge that faced vaccine development was overcoming certain practical obstacles; one of these obstacles was the inability to produce HCV particles *in*

vitro, which made it difficult to understand several aspects of the virus's life cycle and mechanisms of interaction with its host. The first solution to this problem was put forth in 2003 when HCV pseudo-particles (HCVpp) were developed, which are retroviral and lentiviral core particles pseudo-typed with the E1 and E2 structural viral proteins of HCV. The study conducted using these particles provided insight on HCV's entry mechanisms into hepatocytes and the antibody-mediated neutralization of HCV particles through targeting E1 and E2 epitopes [17,56]. After that in 2005, the first successful infection tissue model for HCV was developed, where a unique variant of HCV was cultured in human hepatoma cell line Huh-7 cells. This provided further insight into the stages of the HCV life cycle in the host's hepatocytes [17,18]. Another practical obstacle that HCV vaccine development faced was that the only animal model available as a model for infection and for pre-clinical testing of vaccine candidates was the chimpanzee. Although proven useful due to the similarity of its immune response mechanisms to humans, the chimpanzee animal model has its financial, practical, and sometimes ethical limitations which, in previous studies, allowed only the use of small numbers of this model weakening conclusions that could be drawn from those studies' results. This problem has been dealt with by the development and utilization of transgenic humanized mouse models; these include models expressing human HLA class I molecules used for HCV epitope analysis, and others expressing HCV entry receptors on their hepatocytes which allow for the acceleration of the screening process for potential vaccine candidates [17,19].

Another practical obstacle, particularly facing prophylactic vaccine development, is the challenge of designing a clinical trial for testing the efficacy of a candidate prophylactic vaccine. In developed countries, the low incidence of HCV infection limits the availability of test subjects, except if the study was designed to target intra-venous drug users; targeting this test group in the study raises additional practical and ethical obstacles on its own. In developing countries where the incidence of HCV is higher, careful follow up of the test subjects must be ensured owing to the asymptomatic nature of acute infection. Furthermore, since the vaccine in theory may require some time to facilitate the clearance of the virus, the timing of the start of IFN therapy must be

carefully considered; given the delayed efficacy of the vaccine, the optimal time of commencing therapy may be well after that recommended by current guidelines [17].

2.3.3   HCV vaccines in clinical trials:

Despite the challenges that faced and still face HCV vaccine development, several vaccine studies conducted over the past decade on animal models have yielded promising results; a few of these studies have put forth vaccine candidates that have actually entered human clinical testing [17,19]. Most of these studies focused on the evaluation and development of therapeutic vaccine candidates, which aim at preventing/alleviating the complications associated with disease progression and assisting in the eradication of the virus in combination with standard therapy. A smaller number of those studies have focused on the evaluation and development of prophylactic vaccines, which aim at immunizing the host to prevent infection. With regard to pre-clinical testing (i.e. in animal models), most studies showed that the developed vaccine candidates were indeed able to elicit HCV specific immune responses, which were able to alter the course of subsequent challenge infections by allowing better control over viral replication and increasing the likelihood of clearing the virus [17,57]. HCV vaccine candidates currently undergoing clinical trials are extensively reviewed in Halliday et al. 2011 [17]. Examples of those vaccines and the results of their clinical testing are listed in table 2.

2.4   *Computational Vaccinology and Epitope-driven vaccine design*

In conjunction with the ever increasing applications of bioinformatics in biotechnology, immunoinformatics, the branch of bioinformatics concerned with immunology and vaccinology, has allowed researchers to uncover important functional information impacting molecular immunology; furthermore, it has opened the door for the revolutionary field of Computational Vaccinology [58]. Computational vaccinology refers to the computer-aided design of vaccines, where immunoinformatics and bioinformatics tools are used to identify potential vaccine targets within a target genome *in silico* before being empirically tested and validated [59]. There are generally two

approaches for vaccine design computationally, the first being the antigen-driven approach, where the target genome is screened with the aim of identifying protective antigens; for example, a pathogen's genome can be screened for sequence patterns characteristic of cell surface-expressed proteins that can potentially induce humoral immunity. The second approach is the epitope-driven one, where the target genome is scanned with the aim of identifying immunogenic T-cell and B-cell epitopes; these epitopes are then either themselves synthesized in an appropriate form and packaged in an appropriate delivery vehicle to be used as a vaccine (e.g. peptide vaccine or DNA vaccine), or can be used to identify potentially protective whole protein antigens from the target genome which can be used as vaccine targets (referred to as "fishing for antigens using epitopes as bait") [60]. The total repertoire of protective epitopes (T-cell and B-cell) derived from a target genome/proteome is referred to as the Immunome [60].

Over the past decade, the process of epitope-driven vaccine design (also known as Immunome-Derived vaccine (IDV) design) has witnessed several applications in cancer and infectious disease vaccine development, bolstered by the availability of a wide variety of specialized immunoinformatic epitope-mapping (immunome-mining) tools, which are designed for the *in silico* identification of T-cell and B-cell epitopes from the target genome/proteome [23,60]. There is a wide variety of efficient algorithms developed for T-cell epitope mapping which have found application in several research laboratories in the field of drug discovery; however, fewer algorithms have been developed for B-cell epitope mapping which are not very efficient at epitope prediction. This is due to the nature of B-cell and T-cell epitopes, where the linear and shorter form of T-cell epitopes makes it simpler to develop algorithms for their prediction compared to B-cell epitopes which may be conformational in nature [59-61].

2.4.1 <u>Advantages of immunoinformatics-guided epitope-driven vaccine design:</u>

In comparison to traditional vaccines, IDVs possess the potential of being safer and more effective [23]. The higher potential effectiveness of IDVs is due to the fact that the vaccine is designed to focus the immune response on the most immunogenic and essential epitopes of the target pathogen or neoplasm. With regard to safety, IDVs are

considered safer than live attenuated and vector vaccines; furthermore, since IDVs can be designed in a minimalist form to contain only the target epitopes rather than whole protein antigens, they may be able to avoid some of the health hazards associated with some of those antigens. For example, despite its high cross-genotype conservancy, it was found that the Core protein of HCV is able to induce hepatocellular carcinoma in transgenic mice and cause steatosis formation [25]. Accordingly, the use of immunogenic epitopes from the core rather than the core antigen itself in a vaccine would be considered safer.

Compared to the traditional empirical method where scientists used to isolate target proteins (antigens) and process them to extract and test all potentially immunogenic epitopes [60], epitope-mapping tools allow for the *in silico* identification of the most immunogenic of those epitopes before further testing is carried out in the wet lab. This saves the time and costs associated with testing all potentially immunogenic epitopes empirically before selecting the most immunogenic ones that can actually be included in a vaccine. Furthermore, while the empirical method may result in some important epitopes being overlooked due to imprecise protein digestion, epitope-mapping tools may avoid that if an efficient algorithm is used to scrutinize all possible "epitope frames" within a target protein sequence [58,61]. In addition to being able to map epitopes *in silico*, immunoinformatcs/bioinformatics tools also allow for homology testing of those epitopes against the human genome with the aim of avoiding the use of epitopes that have homologous sequences in the human genome. If used, such epitopes may trigger a suppressive immune reaction due to the immune system being tolerant to them, or may trigger an autoimmune reaction against self. Accordingly, avoiding the use of such epitopes potentially also makes the vaccine safer and more effective [23,60].

2.4.2   Types of T-cell epitope-mapping tools:

In the past, T-cell epitope mapping algorithms were designed to predict T-cell epitopes directly, based on the hypothesis that T-cell epitopes are amphipathic structures with a hydrophilic side facing the T-cell receptor (TCR) of T-cells, and a hydrophobic side facing the binding groove of the MHC molecule; such algorithms were referred to as

27

"direct" methods of epitope prediction [58]. Currently utilized T-cell epitope mapping algorithms however are those designed to predict T-cell epitopes based on the binding specificities of the MHC molecules rather than the molecular structures of the epitopes themselves; such methods are referred to as the "indirect" methods of prediction, and can be classified according to their prediction approaches into: motif-based methods, matrix-based methods, machine-learning-based methods, and *ab intio* methods [58].

In motif-based methods, epitope prediction algorithms predict MHC binders by screening for specific amino acid residues, called "anchor" residues, in the target peptide sequence [58,62]. Anchor amino acid residues of T-cell epitopes interact with polymorphic amino acid residues in MHC molecules which are located within certain 'pockets' in the MHC epitope-binding region (explained in detail in the next section); these polymorphic residues differ from one MHC molecule to the other, and consequently the nature of the anchor residues of the T-cell epitopes they bind. The recurrent pattern of anchor residues that bind each MHC molecule is referred to as a "motif" [62]. Although proven valuable, the *in vitro* validation of the *in silico* results generated by these methods showed that they are relatively nonspecific, as they predicted epitopes that did not bind to their respective MHC alleles or were not immunogenic, and insensitive, as they failed to predict true binders that did not bear the anchor residues however bore 2ry anchor non-conserved residues at other locations in the peptides [58,62]. Specificity of a prediction algorithm is a measure of its efficiency in excluding non-binder epitopes, while sensitivity is a measure of its efficiency in including true binders in its predictions [62].

In contrast to the motif-based approaches which assume that only anchor amino acid residues contribute to epitope binding, matrix-based approaches on the other hand account for the contributions of all amino acids of a target epitope in their predictions [62]. When constructing a matrix-based epitope prediction tool, a reliable training set of epitope-binding data is first identified. Then quantitative or frequency matrices for each MHC molecule to be included in the predictions are constructed by analyzing the distribution of amino acid frequencies at each position in the binding cleft of the MHC molecule and/or the influence of each amino acid binding at each position on the overall

binding affinity of the epitope, and then using the outcome of the analysis to assign a numerical value for every one of those amino acids at their respective binding positions. The prediction of the binding affinity of new epitopes is then carried out by looking up the values assigned for each amino acid in the epitope relative to its position in the MHC molecule in the matrix, and then either adding or multiplying these values to obtain an overall score for the peptide that is used to predict its likelihood of binding relative to a predetermined cut-off value [62,63]. Matrix-based approaches perform better compared to motif-based approaches; however, they handle data in a 'linear' manner, where they consider the individual contributions of each amino acid residue in the epitope to epitope binding independently; this is not the case in reality, as epitope binding is not determined by the simple summation of the contributions of each amino acid residue at each position in the target epitope [58].

While matrix-based prediction approaches cannot handle non-linearity in data, machine learning based approaches can. A limitation of such approaches is, however, that they require a relatively large MHC binder training dataset for their construction, which is available for a small, albeit increasing, number of MHC molecules/alleles. There are several examples of machine-learning techniques employed in T-cell epitope prediction algorithms such as artificial neural networks (ANN), Hidden Markov Models (HMM), and Support Vector Machines (SVM) [58]. The last approach for T-cell epitope prediction is the *ab initio* approach, which functions by predicting binders to an MHC molecule using a structural model of that molecule without the requirement for a training dataset. This approach however requires an elucidated 3D structural model of the MHC molecule of interest as well as high computational power for performing the analysis [58].

2.5 *The Human Leukocyte Antigen*

As aforementioned, the Major Histocompatibility Complex (MHC) molecules are specialized proteins encoded by genes in the Major Histocompatibility Complex locus found in the chromosomes of all vertebrates, their role is to display peptide fragments

(epitopes) of digested protein antigens on cell surfaces to be recognized by T-lymphocytes leading to their activation; class I MHC molecules are expressed on all nucleated cells and activate CD8$^+$ T-cytotoxic lymphocytes, while class II MHC molecules are only expressed in antigen presenting cells and activate CD4$^+$ T-lymphocytes. In humans, the major histocompatibility complex is also known as the Human Leukocyte Antigen (HLA), and is present on the short arm of chromosome 6 [36,64].

2.5.1 __HLA molecules:__

HLA molecules are glycosylated heterodimeric transmembrane proteins composed of two non-covalently associated polypeptide chains; in HLA class II molecules, both polypeptide chains are encoded by HLA class II genes and are called chains α and β, while in HLA class I molecules only one of the chains is encoded by HLA class I genes, called the heavy α chain, while the other chain in not encoded by the HLA locus and called the light β$_2$-microglobin chain [36,64]. The overall tertiary structure of HLA molecules can be divided into four parts: an extracellular N-terminal peptide (epitope)-binding region, an extracellular Immuno-globulin (Ig)-like region, a hydrophobic transmembrane region, and finally a small C-terminal cytoplasmic region. The epitope binding region is composed of two α-helices resting on an eight-stranded β-pleated sheet forming a cleft within which epitopes bind for presentation to T-lymphocytes in a linear conformation. In HLA class II molecules, the binding cleft is composed of the N-terminal domains of the α and β chains called α1 and β1 respectively, where each domain is formed of about 90 amino acids and contributes to half of the structure; the ends of the binding cleft are open, which is why HLA class II epitopes tend to have a variable length that can reach 30 amino acids, typically however falling within the range of 13 to 25 amino acids. In HLA class I molecules on the other hand, the binding cleft is composed of two N-terminal domains of the α chain called α1 and α2; in contrast to class II molecules, the ends of the binding-cleft are closed, restricting the size of HLA class I epitopes to 8-11 amino acids[36,64]. A third domain of the α chain, called α3, forms the Ig-like extracellular region; and is responsible for interacting with the CD8

molecule of the T-cytotoxic lymphocytes. In class II molecules, the same function is performed by the β2 region of the β chain, which forms the Ig-like domain that interacts with the CD4 molecule of T-helper lymphocytes [36,64].

With regard to HLA class II epitopes, the affinity and specificity with which they bind to the epitope-binding cleft of the HLA class II molecule is mainly determined by a core region in those epitopes made up of 9 amino acids [65]; the remaining AA residues flanking the core region also assist in peptide binding by enhancing the binding affinity and stability of the molecular complex [66]. The peptide backbone of the core region interacts with conserved residues in the MHC binding-site via hydrogen bonds, which maintains the peptide in a linear conformation with the core region of the peptide having a conserved secondary structure resembling that of a polyproline chain [64,67]. In this conformation, side chains of AA residues 1, 4, 6, and 9 of the core region of the peptide are directed 'into' the binding-cleft; where they interact with four major pockets distinctive of the MHC class II binding-site. The properties of these pockets vary from one MHC class II isotype (molecule) to the other, and determine the specificity with which the epitopes bind to the binding-groove. Furthermore, the interaction of the stated AA residues with the pockets enhances the stability of the molecular complex; which is why these residues are called the 'anchor' residues of the peptide [65-67]. Side chains of other AA residues in the core region of the peptide also interact with different AA residues in the β-pleated sheet floor and α-helix walls of the binding groove via hydrogen bonds and salt-bridges; interactions which also contribute to the binding affinity of the peptide. For the peptide to be identified by its antigen-specific T-helper cell, side chains of certain AA residues from the core region protrude upwards away from the binding cleft to interact with and be recognized by the T-cell receptor (TCR) of the T_H-cell. The TCR, however, does not only interact with those residues, but it also interacts with polymorphic AA residues of the α-helices of the MHC itself [36,64]. A 3D illustration of an epitope associated with an HLA class II molecule binding cleft is presented in figure 11.

HLA class I epitopes are accommodated in the binding groove of HLA class I molecules by anchoring the ends of the epitope and allowing its middle portion to bulge

upwards. Amino acid side chains of anchor residues at the N and C termini of the epitopes –usually the first and last two- are directed "downwards" into the binding cleft where they interact with amino acid residues found within pockets in the HLA binding cleft, similar to the case of the binding of HLA class II epitopes to their binding cleft. Different HLA molecules encoded by different genes have different pockets, and hence can bind to different peptides. Furthermore, different HLA molecules encoded by different alleles of a given HLA class I gene also show minor differences in their specificity towards anchor residues [64]. Typically, the main contribution to the binding affinity of an epitope is provided by the residues at position 2 and at the C-terminus of the peptide, which interact with 2 pockets in the binding cleft called "B" and "F" respectively [68]. Amino acids in the middle of the epitope are directed upwards where they interact with the TCR of T-cytotoxic lymphocytes; determining the specificity and affinity with which the TCR binds to the HLA-peptide complex. These residues also partly contribute to the peptide binding affinity to the HLA molecule. If it is conformationally possible, HLA class I molecules can accommodate epitopes longer than 10 amino acids if their central amino acid residues can bugle upwards [64].

2.5.2 HLA genes:

The α and β chains of the HLA class II molecules are encoded by multiple functional genes found in the MHC locus on chromosome 6; these genes are located within three regions in this locus, namely DR, DQ, and DP. The DP region harbors genes for two α and two β HLA chains; of which only α1 and β1 are functional. The same is true for the DQ region; except that it includes a third non-functional gene for a β HLA chain (β3). The DR region on the other hand is different; where it harbors only one gene for an α HLA chain that is monomorphic in the region encoding for the peptide-binding α1 domain of the α chain, and 9 polymorphic genes for HLA β chains. Of the nine, only genes number 1, 3, 4, and 5 are functional. β chain genes numbers 1 and 9 are present in all individuals, but the presence or absence of genes 2-8 in DR varies from one individual to the other [36,64]. When translated, the α and β chains encoded by genes of the same region (e.g. DR) most usually form the HLA molecule; and on rare occasions do the α

and β chains encoded by genes of different regions (e.g. DR and DQ) form the molecule. Mixing between α and β chains of the same region but encoded on genes from different parental origin is also common. Both paternal and maternal chromosomes are co-dominantly expressed [36,64].

As for HLA class I molecules, the α chain is encoded by multiple functional genes also found in the HLA locus on chromosome 6. In contrast to HLA class II however, the β_2-microglobin chain is encoded by an invariant gene on chromosome 15. The genes encoding for the α chain are denoted HLA-A, B, and C. Just like HLA class II genes, maternal and paternal HLA class I genes are co-dominantly expressed; which means that for each individual there can be six variants of the HLA class I molecule expressed on all nucleated body cells. However, the levels of expression of the different HLA class I genes differ. For example, 20-50% of individuals depending on the race do not seem express a functional HLA-C protein [36,64].

2.5.3   HLA classification and nomenclature:

In the past, the characterization of HLA variants was performed by serological typing [69,70]. This lead to the classification of HLA molecules into serotypes, where a serotype is defined as the HLA molecule(s) that are detected by an available respective antibody. The serological typing method cannot, however, identify all existing HLA genotypes individually. This is because the available HLA-specific antibodies are smaller in number compared to the ever increasing number of HLA alleles, and are moreover hard to detect. Furthermore, HLA antibodies cannot detect null HLA alleles which are not expressed. As a result of that, each defined HLA serotype usually describes several HLA allelic variants. Furthermore, some of the defined HLA serotypes describe "broad antigens" as they are further sub-categorized into "split antigens" owing to the increase in sensitivity and refinement of Serotyping methods over the years which lead to the identification of new anti-HLA antibodies [71]. As a result of lower specificity, broad antigen serotypes exhibit cross-reactivity with the split antigen serotypes; for example, the HLA-A9 serotype describes a broad antigen whose antibody can detect HLA-A23 and A24 split antigens [71,72].

When HLA DNA typing methods were developed, such as the PCR-based sequence specific oligonucleotide probing (PCR-SSOP), sequence-specific priming (PCR-SSP), and sequencing-based typing (PCR-SBT) methods, a more specific nomenclature system for the characterization of HLA genes and alleles was developed [69,71,72]. Assigning names for new HLA genes and alleles is performed by the World Health Organization's (WHO) Nomenclature Committee for Factors of the HLA System using a specific format [71]. Based on the latest update of the format of the nomenclature system, HLA alleles are to be denoted by an "HLA" prefix, followed by the gene name, followed by 4 fields of digits, and finally a one letter suffix, with all components separated by specific separators which are: a hyphen between the HLA prefix and the gene name, an asterix between the gene name and the digit fields, and finally colons between the digit fields themselves (e.g. HLA-A*02:01:01:02N). The first two digits in the digit fields are used to describe the allele family, which usually corresponds to the serotype of the HLA. The third and fourth digits describe HLA alleles whose sequences harbor non-synonymous mutations in the expressed protein sequence (exons), mainly in the regions coding for the epitope-binding site of the molecule, and are assigned based on the order by which the sequences of the alleles have been elucidated. Synonymous mutations that occur in the exons are described by the fifth and sixth digits of the name. The seventh and eighth digits are used to describe mutations in the introns or the 5' and 3' untranslated regions of the alleles. Finally, the one letter suffix is used to describe the expression status of the allele (e.g. "N" refers to "null allele", while "L" refers to "Low expression level") [73,74]. Depending on the DNA typing method utilized, the resolution of HLA typing can vary between low resolution typing, which characterizes the allele with respect to the allele family only (first digits field), to high resolution typing, which characterizes individual alleles based on the amino acid sequences of the their epitope-binding sites (first two digit fields), to "allelic" resolution, which characterizes the full DNA sequence of the typed allele [73].

In addition to the serological and genetic nomenclature systems, HLA variants can be also classified based on the structural features of the epitopes they bind. X-ray crystallographic analysis of the epitope-binding regions of HLA molecules have shown that their polymorphic residues are located within 'pockets' in the cleft and are

responsible for determining the binding specificity of the HLA molecule by engaging certain anchor residues in their target epitopes. Analysis of the amino-acid sequences of the epitope repertoire of each HLA molecule uncovered a frequent occurrence of certain amino acid residues in these anchor positions; the recurring patterns were termed "motifs" [62,75]. Although each HLA molecule has its own binding specificity and anchor motifs, studies have shown that several HLA molecules within the same class (class I or class II) share similar binding specificities with regard to those motifs. These molecules are said to belong to the same 'Supertype', and the broad anchor motif that describes the anchor motifs of each HLA member of the supertype cluster is referred to as the 'Supermotif' [75,76]. For example, the HLA class I supertype A2, which includes the HLA alleles A*02:01, A*02:02, A*02:03 and others, has a supermotif defined by aliphatic hydrophobic residues (e.g. Alanine, Threonine) at position 2 and the C-terminus of the target epitope.

In contrast to the serological and genetic nomenclature systems, HLA supertype assignments are not standardized. Several research groups have defined HLA supertypes based on the epitope-binding data of individual HLA alleles, which was generated via different computational and empirical methods; consequently, no consensus assignment of individual HLA alleles to a particular supertype exists [75,77]. With regard to HLA class I alleles, the research group of Sidney and Sette [75,76,78] classified the HLA-A and HLA-B loci alleles into 9 major supertypes: four for gene A, namely A1, A2, A3, and A24, and five for gene B, namely B7, B44, B27, B58, and B62. This classification is in general agreement with the results of others performed by the research groups of Tong [79], Lund [80], and Hertz [81], with variations, if present, in the assignment of individual alleles to supertypes and the splitting of supertypes to smaller groups; for example, Lund et al. suggested the splitting of some of the alleles in the A1 supertype into a new A26 supertype [62].

The functional categorization of HLA molecules into supertypes is very beneficial for epitope-driven vaccine design. With regard to HLA class I supertypes, it was found that 6 of the 9 supertypes proposed by the Sidney and Sette research group, namely A1, A2, A3, A24, B7, and B44, account for most of the HLA-A and HLA-B alleles of five

major ethnic groups in the general population (over 85% of HLA-A alleles and 40% of HLA-B alleles in the Caucasian, North American Black, Hispanic, Chinese, and Japanese ethnic groups). Practically, epitopes of those 6 supertypes can afford an average coverage of 99.3% of any population, which means that an epitope-driven vaccine targeting epitopes of all those supertypes at once can potentially be effective in any vaccinated population [78].

2.6   *The iVAX Immunoinformatics toolkit*

iVAX (EpiVax Inc., RI, USA) is a web-based suite of immunoinformatics tools for designing genome-derived T-cell epitope-driven vaccines [82]. The key tool of the iVAX toolkit is called *EpiMatrix*, a frequency matrix-based T-cell epitope mapping tool that has been used and validated for over a decade in *in vitro* and *in vivo* studies [23]. In this section, a description of how *EpiMatrix* and the other tools of the iVAX toolkit that were used in this project work will be discussed.

2.6.1   <u>EpiMatrix:</u>

Similar to other matrix-based T-cell epitope mapping tools, the *EpiMatrix* algorithm is based on a collection of matrices for a set of class I and class II HLA molecules (from here-on referred to also as 'HLA alleles' for simplicity instead of saying 'HLA molecule coded by a given allele'). Within each matrix, a score is given to each of the 20 amino acids at each pocket (position) in the binding cleft of the HLA molecule based on the empirically-determined paucity or preponderance of that amino acid at that position in a large set of natural epitopes of that molecule [23,78]. An example of one of EpiMatrix's frequency matricies is illustrated in figure 12. In a typical analysis aiming to predict 9-mer binder epitopes, an input protein sequence is parsed into 9 amino acid peptide frames, where each frame overlaps the last by 8 amino acids, and each frame is scored for binding to a specific HLA allele additively using frequency scores derived from the matrix corresponding to that allele, giving an allele-specific "raw score". To be able to compare binding affinities of a given peptide across several different alleles, the

allele specific raw score is normalized according to a score distribution derived from a randomly generated large set of peptides to give a Z-score (assessment score).

According to the Z-score scale, an epitope scoring 1.64 or higher is considered to be a potential binder, as peptides scoring 1.64 or higher are usually from amongst the top 5% binders of a randomly analyzed peptide set; hence, such an epitope is considered potentially immunogenic and is a good candidate for further testing. Furthermore, epitopes scoring 2.32 or higher correspond to the top 1% binders of any analyzed peptide set, and hence are the best candidates to be carried forward for further analysis in terms of immunogenicity. Converting allele specific raw scores to Z-scores makes easier the identification of promiscuous epitopes [23,78].

2.6.2   Conservatrix:

When designing vaccines against pathogens with highly mutable genomes such as HCV and the Human Immunodeficiency virus (HIV), it is disadvantageous to include highly mutable epitopes in the vaccine, despite the fact that they may be highly immunogenic, as the pathogen will be able to escape the immune response directed against those epitopes by mutating them. Accordingly, vaccine design must be directed towards including highly conserved and immunogenic epitope targets in the vaccine; such epitopes are usually parts of conserved amino acid sequences that play an important structural or functional role in the pathogen's life cycle and are less likely to be mutated in the course of the pathogen's evolution. To identify those epitopes, the iVAX toolkit employs the "*Conservatrix*" tool, which parses input amino acid sequences into frames of user-defined length, then scans the input dataset, which harbors several sequences obtained from the different pathogen's strains, for matching segments. The tool then calculates the percent conservation of the analyzed segments across the input protein sequences [23].

2.6.3   EpiAssembler:

EpiAssembler is a tool that constructs MHC class II epitopes in an extended form called "Immunogenic Consensus Sequence (ICS)" peptides using highly conserved, promiscuously immunogenic, and overlapping MHC class II epitopes that were scored for binding to MHC class II molecules using *EpiMatrix* and analyzed for conservancy using *Conservatrix* [23]. *EpiAssembler* works by first selecting highly conserved and promiscuously immunogenic (having Z-scores ≥ 1.64 for more than one MHC class II alleles) 9-mer peptides, and then iteratively searching the rest of the *EpiMatrix* results dataset for other highly conserved and promiscuously immunogenic peptides that overlap the N- and C-termini of the initial "core" peptide in order to use those peptides to extent the length of the constructed ICS. In the case where more than one overlapping peptides are identified, the one with the highest conservancy and promiscuous immunogenicity is selected by *EpiAssembler* to extend the ICS. This process of ICS extension is repeated until either no more overlapping peptides are found in the dataset, or until the user-defined ICS length is reached [23,83]. An illustration of how EpiAssembler works is given in figure 13.

Apart from choosing the most conserved and promiscuously immunogenic MHC class II epitopes for vaccine design, which would potentially increase the vaccine's target population coverage and its ability to trigger a protective immune response, *EpiAssembler* has the advantage of enabling the design of a vaccine construct that delivers an 'antigenic payload' which would otherwise be very expensive to develop using conventional methods, as different variants of each antigen will have to be included in the vaccine [23].

2.6.4   JanusMatrix:

As aforementioned, the inclusion of T-cell epitopes in the vaccine construct that bear similarity to autologous epitopes from the human proteome may either result in triggering an autoimmune response against self, or in triggering a suppressive immune response that would decrease the vaccine's efficiency; accordingly, it is advised that such

epitopes not be included in the vaccine [23,60]. Epitopes that are putatively cross-reactive with self can be identified *in silico* using the *JanusMatrix* tool.

JanusMatrix is a tool that identifies T-cell epitopes from the human proteome (genome), protein sequences of the human microbiome, and/or human bacterial and viral pathogen proteomes that are potentially cross-reactive with T-cell epitopes of interest; specifically by identifying T-cell epitopes from those proteomes and protein sequences that bear identical T-cell receptor (TCR)-facing amino acid residues to those of the epitopes of interest [84]. The tool works by dividing putative 9mer T-cell epitopes predicted by the *EpiMatrix* tool into TCR-facing residues (referred to also as the "epitope") and HLA-binding residues (referred to as the "agretope") based on rules defined in literature, and then searching for potentially cross-reactive epitopes across user-selected protein databases from those that have been pre-uploaded to the tool that bear identical TCR-facing residues and are predicted to bind the same HLA molecule as the putative T-cell epitopes of interest, regardless of the existence of discrepancies in the agretope [84]. According to literature, the TCR-facing residues of HLA class I 9mer epitopes are residues number 4, 5, 6, 7, and 8, while those of HLA class II epitopes are 2, 3, 5, 7, and 8 [85]. Figure 14 displays a 3D model of a T-cell epitope in association with both the HLA molecule and the TCR.

3. PROJECT AIM & OBJECTIVES

The aim of this project is to identify, *in silico*, a set of highly immunogenic and conserved T-cell epitopes from within the HCV-4 proteome as putative targets for an epitope-driven vaccine for HCV-4. The objectives of the project are as follows:

1. To use the iVAX immunoinformatics toolkit (EpiVax Inc., RI, USA) to identify highly conserved and immunogenic T-cell epitopes from within 46 non-redundant HCV-4 proteomes, and assess their homology to epitopes derived from the human genome.

2. To retrospectively validate the HLA binding predictions for those epitopes by searching for records of HLA binding assays or T-cell assays performed on those epitopes in previous literature in the context of other HCV genotypes on the Immune Epitope Database (IEDB).

3. To perform a pilot *in vitro* HLA binding assay on selected HLA class II ICS peptides (epitopes), and compare the results with those of the retrospective validations for the same peptides.

4. To assess the conservation of a selected set of the identified epitopes in newly partially re-sequenced HCV genomes derived from the Egyptian Population as an assessment of the HCV subtype coverage that these epitopes would provide if they were to be included in population-specific vaccine targeting the Egyptian population.

4. METHODOLOGY

4.1   HCV-4 Genomes Collection

46 non-redundant HCV genotype 4 full genomic sequences were retrieved from the Los Alamos HCV sequence database (LANL) [86]; their accession numbers were also matched on the Genbank database, these being: DQ418785, DQ418782, DQ418783, DQ418784, DQ418787; DQ418788; DQ418789, DQ516084, DQ988073, DQ988074, DQ988075, DQ988076, DQ988077, DQ988078, DQ988079, GU814265, NC_009825, FJ025854, FJ025855, FJ025856, FJ462435, FJ462436, DQ418786, DQ516083, EU392172, FJ462437, EF589160, EF589161, EU392169, EU392170, EU392174, EU392175, FJ462432, EU392171, EU392173, FJ462438, FJ839870, FJ462433, FJ462441, FJ462440, FJ462431, FJ462434, FJ462439, FJ839869, HQ537008, and HQ537009. The amino acid sequences of the HCV polyprotein genes corresponding to those genomic sequences were downloaded in FASTA format and modified to a format compatible with the iVAX toolkit server, whereby all the single letter ambiguous amino acid codes B, Z, and J, which stand for Asparagine/Aspartic acid, Glutamine/Glutamic acid, and Leucine/Isoleucine, respectively, and the single letter amino acid codes U and O, which stand for the 21st and 22nd identified amino acid residues Selenocysteine and Pyrrolysine, respectively, were replaced with the single letter ambiguous amino acid code X, which declares the amino acid residue as unspecified. After their modification, the 46 protein sequences where uploaded to the iVAX toolkit's website (https://ocs.epivax.com/ivax_ntd/) as a single file bearing the ".PEP" file extension that is required for the sequences to be identified as proteins. More information on the 46 HCV-4 sequences is provided in table 3.

4.2   HLA Class I and Class II Epitopes Identification and Selection

After uploading the sequences onto the iVAX server, the tools of the iVAX-toolkit were used in concession to identify and narrow down the selection of HLA class I

and II epitopes from the HCV-4 sequences to a set that can be putatively included in the final vaccine product.

4.2.1 Conservation Analysis:

The *Conservatrix* tool was used to parse the input sequences into 9-mer peptide frames and analyze the conservation of those frames across the 46 amino acid sequences. The 9-mer peptide frames represent putative HLA class I epitopes and the core regions of putative HLA class II epitopes.

4.2.2 HLA Class I and II Epitope Prediction:

After performing the conservation analysis, the *EpiMatrix* tool was used to predict the binding of the parsed 9-mer frames to the six HLA class I alleles A*01:01, A*02:01, A*03:01, A*24:02, B*07:02, and B*44:03 which belong to the 6 supertypes A1, A2, A3, A24, B7, and B44 respectively, and the eight HLA class II DR alleles DRB1*01:01, DRB1*03:01, DRB1*04:01, DRB1*07:01, DRB1*08:01, DRB1*11:01, DRB1*13:01, and DRB1*15:01. The phenotypic frequencies of those HLA class I and II alleles cover >90% of any human population regardless of its ethnic origin, potentially allowing the vaccine to be effective in any vaccinated population [78,95]. It is important to note here that the DR molecules are referred to by the names of the DRB1 alleles which code for the polymorphic β chains of the molecules.

After obtaining the results of the analysis, only those peptide frames that have Z-scores ≥ 1.64 for binding to a HLA class I allele and are fully conserved in at least 42 out of the 46 sequences (>90% conservancy) were chosen for further analysis. As for peptide frames scored for binding to HLA class II alleles, the *EpiAssembler* tool was used to construct extended Immunogenic Consensus Sequence peptides (ICS) from those frames. Before executing the analysis, *EpiAssembler* was set to only extend core peptides with peptides that overlap with it by at least 4 amino acids at either the N- or C-termini, and to extend the ICS to a maximum length of 21 amino acids. It was also set to construct ICS peptides whose component core 9-mer segments cover, collectively, at least 90% of the

46 uploaded HCV polyprotein sequences in terms of conservation, and to only accept the construction of ICS peptides that have a minimum EpiMatrix Cluster Immunogenicity score of 10. The EpiMatrix Cluster Immunogenicity score is derived from the number of EpiMatrix "hits" (scores ≥ 1.64 per allele) normalized for the length of the ICS peptide; a score of ICS 10 or higher indicates that the ICS possesses a significant immunogenic potential.

4.2.3   Homology to Self Analysis:

In order to avoid a suppressive immune response to the sequences selected for the vaccine, or autoimmune adverse effects due to the induction of immune response to epitopes cross-conserved with self in the final vaccine product, the selected HLA class I epitopes and HLA class II ICS peptides were assessed for homology to the human genome using the *JanusMatrix* tool, where the reviewed human Uniprot protein database that was per-uploaded to the tool server was scanned for cross-reactive epitopes within the human genome that bear identical TCR-facing AA residues and are predicted by *EpiMatrix* to bind to the same alleles for which the analyzed epitopes are predicted to bind to.

4.3   Epitope Mapping on the HCV Genome

The regions of the HCV genome harboring the selected epitopes and their exact locations were determined by aligning the amino acid sequences of the epitopes with the amino acid sequence of the polyprotein gene of the HCV genotype I reference strain H77 (Accession number: NC_004102) using the "Sequence locator" tool of the Los Alamos HCV sequence database (LANL) [86].

4.4   Retrospective Validation of the EpiMatrix Predictions

To investigate whether the selected HLA class I epitopes and constructed HLA class II ICS peptides were tested before *in vitro* in previous literature for HLA binding,

cell-surface presentation, and/or their ability to elicit an immune response, the Immune Epitope Database (IEDB), a large repository of immune epitope data manually curated from peer reviewed literature, patents, and direct submissions from companies and institutions [21], was queried for each of the selected epitopes and ICS peptides. Using the home page search tool (http://www.iedb.org/), the database was queried for available data on experiments for MHC binding and elution (MHC Ligand assays), and T-cell response performed on epitopes/peptides that exactly match the query sequences or harbor those sequences as substrings. For the HLA class II ICS search, peptides that are 70-90% identical to the constructed ICS peptides were also included in the search criteria. Before querying the database, the immune recognition context parameters were restricted to MHC class I or II-restricted matches (depending on the queried epitope), and *Homo sapiens (id:9606)* selected as the host organism for T-cell assay searches.

4.5   HLA Class II in vitro Binding Assay

As a pilot *in vitro* validation of the predictions performed by the iVAX toolkit, 4 MHC class II ICS peptides (table 10) were assessed for binding to 4 different purified, soluble HLA class II DR alleles (molecules), namely DRB1*01:01, DRB1*04:01, DRB1*07:01, and DRB1*15:01 in a competition binding assay against biotinylated standards. The presence/absence and strength of peptide binding to a given HLA molecule are determined via determining the concentration of test peptide required to prevent 50% of standard peptide from binding to the assayed HLA molecule (IC_{50}).

The overall strategy of the experiment involved each ICS peptide undergoing the assay for each HLA allele in triplicates of 6 dilutions of the peptide in four 96-well reaction plates, one plate for each HLA allele. For each reaction plate, positive and negative control reactions were performed in triplicates. The positive control reactions involved the reaction of HLA allele-specific standard peptides with their respective HLA molecules, giving a "maximum fluorescence" reading for the binding of the standard peptide. The negative control reactions involved the addition of only the HLA molecules to the reaction plates without the addition of any peptides.

At first, lyophilized pellets of the test ICS peptides (21st Century Biochemicals, MA, USA) were solubilized in amounts of Dimethyl Sulfoxide (DMSO) (Sigma-Aldrich, MO, USA) pre-determined by the manufacturer to obtain a final concentration of 100 μM of each peptide. After solubilizing the test peptides, 30 mL of reaction Master Mix were prepared by mixing 30 mL of Citrate-phosphate buffer pH 5.4 (prepared by mixing 278 mL of 2M Sodium phosphate and and 222 mL of 1M Citric acid (Sigma-Aldrich, MO, USA)), 300 μL of 1 mM PefaBloc (Roche Applied Science, Penzberg, Germany) and 225mg of Octyl-β-D-Glucopyranaside (OG) (MP Biomedicals, OH, USA). After preparing the master mix, 6 dilutions of the 4 ICS peptides in the master mix were prepared in a tube rack composed of 6-rows x 4-columns, where each column was designated for one of the 4 ICS peptides. 800μL of master mix solution were transferred to all the rows of the tube rack except for the top row, which received 1170μL of solution. To each of the 4 tubes in the top row, 46.8μL of the solubilized ICS peptide solutions was added, one peptide for each tube. After that, 370.4μL of the master mix-ICS peptide mixtures in the top row were serially transferred across the remaining 5 rows, diluting the ICS peptide-master mix solution by a factor of 3.16 across the tubes. The dilution factor was predetermined and pre-optimized by EpiVax Inc., and excess amounts of master mix and peptide solutions were added to the tube rack based on calculations performed according to a project work-plan provided by EpiVax Inc. The final dilutions of each peptide in each of the reaction plates across the top 6 rows were: 100 μmol/mL, 31.6 μmol/mL, 10 μmol/mL, 3.16 μmol/mL, 1 μmol/mL, and 0.316 μmol/mL.

After preparing the ICS dilutions, triplicates of 25μL of each dilution of each peptide were transferred to the top 6 rows of four 96-well reaction plates (Fisherbrand, Loughborough, UK). 25μL of master mix alone were transferred to the first six wells of the bottom row of the four reaction plates where the positive and negative control reactions were carried out. To prepare the HLA molecules required for performing the binding reactions, purified HLA-DRB1 monomers of the molecules of interest (Benaroya Research Institute, WA, USA) were each diluted in 3mL of the prepared master-mix based on calculations performed according to the project work-plan provided by EpiVax Inc. The exact amounts of monomers diluted were: 0.5μL of DRB1*01:01 monomers

from a stock solution with a concentration of 2.5mg/mL, 0.92μL of DRB1*04:01 monomers from a stock solution with a concentration of 1.3mg/mL, 0.8μL of DRB1*07:01 monomers from a stock solution with a concentration of 1.5mg/mL, and 0.67μL of DRB1*15:01 monomers from a stock solution with a concentration of 1.8mg/mL. After mixing the aforementioned amounts with 3mL of master-mix each, 25μL of each diluted monomer solution were directly added to its respective negative control wells of the prepared reaction plates, causing the final reaction volume in those plates to be 50μL.

To prepare the standard biotinylated peptides that will compete with the ICS test peptides for HLA binding, aliquots from stock solutions of the standard peptides (21st Century Biochemicals, MA, USA) were diluted in the prepared HLA monomer-master mix mixtures, where each HLA monomer was mixed with a specific standard peptide. The standards peptides were diluted in such a way that their final concentrations in the wells of the reaction plates became 0.025μmol/mL. Precisely, 6μL from 25μM stock solutions of the YAR, Tetanus Toxin, and MBP protein standard peptides were diluted in the 3mL solutions of the DRB1-04:01, DRB1-07:01, and DRB1-15:01 monomers respectively, and 30μL from a 5μM stock solution of the Influneza A virus Hemaglutinin-derived standard peptide was diluted in the 3mL solution of the DRB1-01:01 monomer. After preparing the biotinylated peptide-HLA mixtures, 25μL of each mixture were transferred to the 96-well reaction plates in all the wells containing the test peptides and in the positive control wells, causing the total reaction volume in all those cells to be 50μL. As aforementioned, each reaction plate is only designated for one kind of HLA-DRB1 allele. After that, the reaction plates were covered and incubated overnight at $37^{o}C$ to allow the binding reactions to take place.

After putting the reaction plates in incubation, four ELISA plates were prepared for capturing the HLA-DR monomers with their bound peptides by coating four 96-well EIA/RIA plates (Corning, NY, USA) with L243 anti-HLA-DRA antibodies (BioXCell, NH, USA). First, a 1000-fold dilution of the L243 antibodies in Borate buffer pH 8.2 (prepared by dissolving 1.19g of Sodium tetraborate decahydrate (Sigma-Aldrich, MO, USA) in 250mL MilliQ water) was prepared from an antibody stock solution of

concentration 10.36mg/mL. Then, 100μL of the prepared antibody solution was transferred to each well of the four ELISA plates, and the plates were covered and incubated overnight at 4°C.

On the following day after incubation, the contents of the ELISA plates were discarded, and the plates were washed three times with Phosphate Buffer Saline PBS-0.05% Tween 20 (prepared by mixing 500mL of 10x Dulbecco's PBS (Life Technologies, CA, USA) with 2.5mL of Tween 20 (Sigma-Aldrich, MO, USA) in 4.5L of MilliQ water) using an automated plate washer. 40μL of 50mM Tris Neutralization buffer (prepared by diluting 1.5mL of 1M Trizma HCL (Sigma-Aldrich, MO, USA) in 28.5mL of MilliQ water) were then added to each well of the ELISA plates. After that, 40μL of HLA monomer-peptide reaction mixtures from each well of the overnight incubated reaction plates were transferred to their corresponding wells in the ELISA plates containing the neutralization buffer. The ELISA plates were then incubated at 37°C for 2.5 hours.

After the 2.5 hour incubation, the contents of the ELISA plates were discarded, and the plates were washed three times with PBS-0.05% Tween 20 using an automated plate washer. A 1000-fold diluted solution of Streptavidin-Europium in Assay buffer was then prepared, where 10μL of Europium-labeled Streptavidin (Perkin-Elmer, MA, USA) were diluted in 10mL of DELFIA Assay buffer (Perkin-Elmer, MA, USA). 100μL of diluted Streptavidin-Europium solution were then transferred to each well of the ELISA plates harboring captured HLA molecules. The plates were then covered and placed in the dark for 1 hour at room temperature. After that, the contents of the plates were discarded, and the plates were washed three times with PBS-0.05% Tween 20 using an automated plate washer, followed by 2 washes with PBS alone to remove any bubbles. After that, 100μL of Delfia Enhancement buffer (Perkin-Elmer, MA, USA) were added to each well of the ELISA plates, and the plates were again covered and placed in the dark for 10 minutes at room temperature.

After that, the plates were inserted into a Victor^{3}V Microtiter Plate Reader (Perkin-Elmer, MA, USA), and the fluorescence of Europium was read under the Europium setting. The resultant raw data file was exported in Excel format and stored for

IC$_{50}$ calculations, which were performed using the SigmaPlot 11.1 software package (Systat Software Inc., IL, USA) via constructing four parameter logistic curves.

4.6 Conservation Analysis in newly partially re-sequenced HCV genomes from the Egyptian Population

In an attempt to assess whether the predicted T-cell epitopes are conserved in the HCV strains currently circulating in the Egyptian population, a pilot conservation analysis has been performed on a group of selected epitopes in partially re-sequenced HCV genomes newly isolated from the Egyptian population; epitopes found to be highly conserved are, as aforementioned, good candidates for population-specific vaccine development. Specifically, the Core, NS3, NS4B, and NS5B regions of the HCV genome -the regions harboring most of the identified epitopes as explained in the results section- were targeted for re-sequencing and conservation analysis. The steps of a protocol outlined in Yao et al. (2005) [96] were followed for the Reverse Transcription Polymerase Chain Reaction amplification (RT-PCR) and nested PCR amplification of the target sequences. The amplification step was followed by Sanger sequencing of the amplified sequences and the *in silico* conservation analysis of the identified epitopes that they harbor/cover.

4.6.1 Primer Design:

At first, oligonucleotide sense and antisense primers were designed *de novo* to be used in the RT-PCR amplification and Sanger sequencing of the target regions. Since the optimal DNA target read length for Sanger sequencing is 700-900 bp long [97], the NS3 (1893 bp long) and NS5B (1776 bp long) regions were accordingly each divided into two frames overlapping by ~140 bp for amplification and sequencing; the Core (573 bp) and NS4B (783 bp) regions, however, were targeted for sequencing in single frames, making the total number of target HCV RNA frames to be amplified and sequenced 6 out of each isolated HCV genome. To design the primers, the 46 HCV-4 sequences retrieved for

epitope prediction (table 3) were aligned by multiple sequence alignment using the Clustal Omega online tool (http://www.ebi.ac.uk/Tools/msa/clustalo/) [98]; the resultant multiple alignment file was then viewed using the ClustalX 2.1 program [99] to allow for the visual identification of the most highly conserved RNA sequences flanking the 6 target regions to be analyzed in order to be used for primer design. Candidate sequences were then analyzed using the freely available online OligoAnalyzer 3.1 tool (http://eu.idtdna.com/analyzer/Applications/OligoAnalyzer/) (Integrated DNA Technologies Inc.) with regards to melting and annealing temperatures (T_m and T_a respectively), Guanosine and Cytosine percentage (GC content), and the formation of hair-pin structures, self-dimers, and heterodimers between sense and antisense primers.

To the best extent possible, all primers were designed to have: (1) a GC content of 40%-60%, which would decrease the likelihood of non-specific priming caused by the strong bonding between G and C nucleotides, (2) a maximum negative Gibbs free energy (ΔG) value of -3 kcal/mol associated with intra-molecular hairpin structure formation and -6 kcal/mol associated with inter-molecular homo and heterodimer structure formation between primers; larger negative values for ΔG associated with these structures may facilitate their formation at the primer annealing temperatures and consequently decrease the availability of these primers for use in the amplification reaction, eventually leading to a lower product yield, and (3) a difference in annealing temperature between sense and anti-sense primers of not more than 5 degrees [100]. Where possible, nested primers were designed in order to carry out nested PCR reactions using the amplicon products of the RT-PCR reactions; this would increase the yield and specificity of the amplification reactions. Following their design, the primers were ordered for synthesis (Bioneer Inc., Daejeon, Republic of Korea). A list of the primer sequences and information about their sizes, locations, melting temperatures, and amplicon sizes are listed in table 4. An example of the alignment of one of the primers with the 46 HCV-4 sequences is illustrated in figure 15.

4.6.2 Sample Collection:

Left-over anonymous serum samples positive for HCV RNA isolated from the Egyptian population were obtained from the sample archives of private diagnostic labs collaborating with our research group "Novel Diagnostics and Therapeutics" for HCV RNA extraction after obtaining the approval of the Internal Review Board of the American University in Cairo.

4.6.3 HCV RNA Extraction:

Viral RNA was extracted using the QIAamp® Viral RNA Mini kit (Qiagen, Limburg, Netherlands) according to the manufacturer's "spin column" extraction protocol. First, 140 µL of each sample were incubated for 10 minutes at room temperature with 560 µL of buffer AVL mixed with 5.6 µL of carrier RNA – AVE buffer solution (1 µg/µL). Buffer AVL contains guanidine thiocyanate for lysis of the viral particles and denaturation of RNases in the serum sample. The carrier RNA binds to the extracted viral RNA, enhancing its subsequent binding to the silica-gel membrane of the spin extraction columns and protecting it from degradation by RNases in the unlikely event that these RNases escape denaturation by buffer AVL. Before mixing with buffer AVL, carrier RNA was first dissolved in buffer AVE, which is RNase-free water containing 0.04% sodium azide as a preservative for microbial growth prevention. Following the 10 minutes incubation, 560 µL of absolute ethanol were added to each sample-AVL-carrier RNA mixture. After that, 630 µL of each mixture were transferred to spin columns fitted inside empty collection tubes. The spin columns were then centrifuged for 1 minute at 8000 rotations per minute (r.p.m.) at room temperature using a 2-16PK centrifuge (Sigma, Osterode, Germany); this step was repeated a second time with the remaining 630 µL of sample mixture. Collection tubes containing the filtrate were discarded, and the spin columns were placed in new collection tubes.

In the following step, 500 µL of AW1 wash buffer were added to each spin column, followed by a 1 minute centrifugation at 8000 r.p.m. at room temperature. The filtrate was then discarded and a new collection tube was added. 500 µL of AW2 wash

buffer were then added to each spin column, followed by a 3 minute centrifugation at 14,000 r.p.m. at room temperature. The wash buffers purify the viral RNA in the sample from other contaminants. After discarding the collection tubes with the filtrate from the last centrifugation, the spin columns were then placed in 1.5 mL microcentrifuge tubes and 60 µL of nuclease-free water were added to the columns. After one minute incubation at room temperature, the spin columns were centrifuged for 1 minute at 8000 rpm at room temperature, and the eluate containing the extracted and purified viral HCV RNA was retained.

4.6.4 <u>Amplification by RT-PCR:</u>

Following HCV RNA extraction, the target Core, NS3, NS4B, and NS5B regions were amplified by RT-PCR using the Access RT-PCR System (Promega, WI, USA) and the set of external primers (not nested) listed in table 4. Six reaction mixtures for each sample, 25 µL each, were prepared for the amplification of the target regions. To prepare the reaction mixtures, the following components were added to 0.5 mL PCR tubes: (1) 5 µL of AMV/*Tfl* Reaction Buffer (5X), (2) 0.5 µL of dideoxy-nucleotide triphosphate (dNTP) mixture (10mM of each nucleotide), (3) 1 µL of Magnesium Sulfate ($MgSO_4$) solution (25mM), (4) 2.5 µL of each forward and reverse primers (10 pmol/µL), (5) 0.5 µL of Avian Myeloblastosis Virus Reverse Transcriptase (AMV RT) (5u/µL), (6) 0.5 µL of thermostable *Thermus flavus* (*Tfl*) DNA polymerase (5u/µL), (7) 10 µL of extracted HCV RNA sample, and (8) nuclease-free water up to 25 µL. Positive control reactions were prepared by adding the same components to the reaction tubes with the exception of adding of 2 µL positive control RNA with carrier (1×10^6 copies) and 3.3 µL of control upstream and downstream primers (50 pmols) provided with the kit instead of the HCV RNA template and the *de novo* designed primers respectively. For negative control reactions, nuclease free water was used instead of the positive control RNA with carrier; the control upstream and downstream primers were also used for the negative reactions.

After preparing the reaction mixtures, the PCR tubes were placed into the Veriti 90-well Thermal Cycler (Applied Biosystems, CA, USA). The PCR cycling profile was set to start with one cycle for first strand (cDNA) synthesis by reverse transcription at

45°C for 45 minutes, followed by one cycle for the denaturation of the AMV RT at 94°C for 2 minutes. After that, the samples were subjected to 45 cycles of PCR amplification which consisted of: (1) a 30 second template denaturation step at 94°C, (2) a 1 minute primer annealing step at the optimized temperature of 45°C, (3) a 2 minute new stand extension step at 68°C. A final extension step at 68°C was performed for 7 minutes to allow the extension of truncated products and enhance the quality of the results.

4.6.5   Nested PCR:

To increase the specificity and yield of the amplification reactions, nested PCR was performed using the internal nested PCR primers from table 4 and amplicons obtained from the RT-PCR step as a template. Nested PCR was performed using either the GoTaq® Colorless Master Mix (Promega, WI, USA) kit or the Maxima Hot Start PCR Master Mix (Thermo Scientific, MA, USA). To prepare the reaction mixtures, the following components were added to 0.5 mL PCR tubes: (1) 25 μL Master mix containing a *Taq* DNA polymerase, (2) 2 μL of each of the forward and reverse primers, (3) 10 μL of the RT-PCR amplicon products as templates, and (4) nuclease free water up to 50 μL.

After preparing the reaction mixtures, the PCR tubes were placed into the Veriti 90-well Thermal Cycler (Applied Biosystems, CA, USA). The PCR cycling profile was set to start with one denaturation step at 95°C for 2 minutes, followed by 35 cycles of: (1) a 30 second denaturation step at 95°C, (2) a 1 minute primer annealing step at optimized annealing temperatures for each primer pair, and (3) a 45 second DNA strand extension step at 72°C. A final extension step for 5 minutes at 72°C was performed before stopping the reaction. The optimized annealing temperatures that gave the best results for each primer pair were: (1) 45°C for the A_F and A_{RN} pair, (2) 47°C for the D_F and D_{RN} pair, and (3) 50°C for the pairs B_{FN} and B_R, C_{FN} and C_{RN}, and E_F and E_{RN}. No optimized annealing temperature was determined for the primer pair F_F and F_R since amplification reactions utilizing this pair did not yield any results by RT-PCR (as will be explained in the results section).

The results of the amplification were analyzed by gel electrophoresis on a 1% agarose gel. The gel was prepared by dissolving 2g of Seakem® LE Agarose (Lonza, Basel, Switzerland) and 6 µL of Ethidium Bromide solution (10mg/mL) (Promega, WI, USA) in 200 mL of 1X Tris-Borate-EDTA buffer (Lonza, Basel, Switzerland) and then allowing the mixture to solidify. After gel solidification, 5 µL of a 100 bp DNA Ladder (Promega, WI, USA) mixed with 2 µL loading dye (provided with the ladder) were loaded into the first well. 5 µL of each of the samples' amplicons were added to the other wells mixed with 2 µL loading dye. The gel was then run at 95V for 1 hour in 1X TBE running buffer. The amplicon bands were then visualized on an ECX-20.M transilluminator (Vilber Lourmat, Marne-la-Vallée, France); images of the gel were taken using a Samsung S2 8.0 megapixel mobile phone camera (Samsung, Seoul, South Korea). The remainder of the sample amplicons was stored at -20°C to be later sequenced.

4.6.6   PCR Purification and Sanger Sequencing:

Purification of the PCR sample amplicons was performed at Sigma Scientific Services Co. (Cairo, Egypt) using the GeneJET™ PCR Purification Kit (Thermo Scientific, MA, USA) according to the manufacturer's instructions; where required, gel extraction of the amplicons was performed using the same purification kit. Sanger sequencing of the purified samples was performed at GATC Biotech AG (Constance, Germany) using an ABI 3730xl DNA Analyzer (Applied Biosystems, CA, USA); sequencing was performed using the forward and reverse primers for each sample.

4.6.7   Conservation Analysis:

After the results of the Sanger sequencing were retrieved, the sequencing chromatograms were viewed using the BioEdit sequence alignment editor (Ibis Biosciences, CA, USA). Base calls made based on broad peaks or regions with strong background "noise" at the 5' and 3' ends of the sequences in the chromatogram were truncated from the nucleotide sequences. The truncated sequences were then analyzed using the Basic Local Alignment Search Tool (BLAST) of the National Center for

Biotechnology Information website (NCBI) (http://blast.ncbi.nlm.nih.gov/Blast.cgi); where the blastn and blastx programs were used in their default settings to confirm the identities of the new sequences. Once confirmed, consensus amino acid sequences of the re-sequence regions were assembled by aligning the forward and reverse sequenced frames of each region using the Clustal X 2.1 software. Following sequence assembly, the conservation of the epitopes across the consensus sequences that cover the locations of those epitopes on the HCV-1a (H77) sequence was assessed using the Epitope Conservancy Analysis tool of the Immune Epitope Database (IEDB) Analysis Resource website (http://tools.immuneepitope.org/tools/conservancy/iedb_input).

5. RESULTS

5.1 HLA Class I Epitope Identification and Validation

5.1.1 iVAX analysis results:

The *Conservatrix* and *EpiMatrix* tools parsed and scored a total of 32,286 9-mer peptide frames from the 46 polyprotein sequences. From those, 151 peptides that are fully conserved in at least 42 out of the 46 sequences (90% conservancy) and that have a Z-score ≥ 1.64 were chosen for further analysis; peptides with Z-scores ≥ 1.64 for a given allele are most likely to be binders of that allele. The *JanusMatrix* analysis revealed that 90 out of the 151 selected epitopes have no cross-reactive hits in the human genome. The remaining 61 epitopes do have cross-reactive epitopes ranging from 1 to 9. Accordingly, the 90 epitopes with no cross-reactive hits were selected as good putative candidates for epitope-driven vaccine design against HCV-4.

Analysis of the locations of the 90 epitopes revealed that 23 epitopes are located in the Core region, 21 in the NS3 region, 14 in the NS4B region, 15 in the NS5B region, 5 in the NS5A region, 4 in the E2 region, 2 in the NS2 region, and 1 in the E1 region. No epitopes were found to be located in the p7 and NS4A regions. The 5 remaining epitopes were found to be located at the junctions between the sequences of non-structural proteins. An example of the results of the binding predictions for 7 of the 90 selected peptides and their percent conservation and locations in the HCV-4 genome are listed in table 5.

5.1.2 Retrospective Validation Results:

The IEDB search revealed that 20 out of the 90 selected 9-mer epitopes have no entries for previously performed tests for MHC binding, cell-surface presentation, nor the ability to elicit a CD8[+] immune response in previous literature; indicating that those epitopes may be novel and have not been previously characterized. As for the remaining 70 epitopes, entries have been found for previously performed tests for HLA binding *in vitro* and/or the ability to elicit a CD8[+] T-cell immune response via *in vitro* T-cell assays. No entries for

MHC elution assays were found for any of the queried epitopes. The 70 epitopes can be divided into three categories based on the IEDB results: (1) a category of 20 epitopes that tested positive in HLA binding and/or CD8[+] T-cell assays as 9-mer peptides (exact match to the identified sequences), (2) a category of 22 epitopes that tested positive in HLA binding and/or CD8[+] T- cell assays as substrings of longer peptides, and (3) a category of 28 epitopes that only tested negative in HLA binding and/or T-cell assays as 9-mer peptides and/or as substrings of longer peptides.

From the first category of 20 epitopes that tested "positive" for HLA binding and/or the ability to trigger a CD8[+] T-cell response in assays where the 9-mer epitope itself has been used as the test peptide or has been empirically identified as the optimal epitope causing the positive response from within a larger tested peptide, 9 epitopes were previously tested for both HLA binding and the ability to trigger a CD8[+] T-cell response, 4 were tested for HLA binding only, and 7 were tested for the ability to trigger a CD8[+] T-cell response only. For all the assays performed on the 20 epitopes, the HLA restricting allele was determined and listed in the epitopes' IEDB entries. The results of the HLA binding assays performed in previous literature on 4 of the epitopes listed in table 5 are listed in table 6 as an example of the returned search results. In table 7, the results of the IEDB search for previously performed T-cell assays also on 4 of the epitopes listed in table 5 are displayed; only the entries where the epitope was tested as a 9-mer in previous literature are listed.

The last column to the right of table 7 describes the methods by which the authors of the listed references demonstrated the HLA allelic restrictions of the test epitopes; the methods in the table are represented as one number and one letter whose meanings are described in table 8. Each number represents a "method" by which the HLA allelic restriction has been demonstrated, and each letter represents the basis upon which the authors have predicted/known the tested epitopes to be restricted by the listed alleles prior to the execution of the assays. The methods of demonstration of HLA restriction can be categorized into three methods: the first method involves the use of different lines of antigen presenting cells or target cells that bear matching and mismatching sets of HLA alleles to the assayed CD8[+] T-cells from individual(s) carrying a certain HLA allele

X (HLA-X+); if the cells are stimulated by APCs or target cells carrying the same X allele but not by ones lacking it, this confirms the restriction of the test epitope by that allele. The second method involves the use of HLA tetramers or APCs expressing only a single HLA allele (e.g. murine cells expressing a human HLA allele) to trigger the T-cell response; if the response is positive, this confirms the restriction of the test epitope by the single allele type present in the assay. The third method involves the use of only APCs with HLA alleles matching those of the assayed T-cells (usually autologous) to trigger a T-cell response; while this method is not sufficient to demonstrate HLA restriction, its repetition in more than one assay by more than one author using different patient cohorts and different testing methods increases the likelihood of the claimed HLA restriction being true. In addition to the methods of demonstration of restriction, authors of the listed assays usually have reasons to predict/know the HLA restriction of the epitope prior to performing the assay; these include *in silico* binding predictions, the presence of the HLA binding motif in the epitope, an *in vitro* HLA binding assay, a previous citation mentioning/demonstrating HLA restriction, or a combination thereof. Where the reason for prediction/knowledge is not listed in table 7, it means that the reason has not been clearly mentioned, or that the assay was performed on pools of peptide epitopes that have not been previously characterized.

For 11 out of the 16 epitopes that tested positive in T-cell assays as 9-mer peptides, the listed HLA restrictions were demonstrated at least once by methods "1" and/or "2" in previous literature, in addition to the demonstration of some of those restrictions by *in vitro* HLA binding assays. As for the 5 remaining epitopes, the listed HLA restrictions were demonstrated only by method "3" in previous literature, which on its own is not sufficient to empirically demonstrate HLA restriction. However, for two of those epitopes the listed HLA restrictions were empirically demonstrated prior to performing the T-cell assays via *in vitro* binding assays; which in itself can validate the *EpiMatrix* binding predictions. As for the 3 other epitopes, the listed HLA restrictions were predicted prior to performing the T-cell assays due to the presence of their binding motifs within those epitopes; accordingly, further empirical demonstration of HLA restriction for those epitopes is required.

When the results of the EpiMatrix binding predictions for the aforementioned 20 epitopes were compared to their empirically demonstrated HLA allelic restrictions, it was observed that the EpiMatrix predictions for 18 out of 20 of those epitopes were retrospectively validated by the empirical results. This is because the HLA alleles that were empirically demonstrated to restrict those epitopes had Z-scores ≥ 1.64, or belonged to the same supertypes of the alleles that scored ≥ 1.64 in the EpiMatrix predictions. Furthermore, the HLA allele A*24:02 which was empirically proven to restrict the epitope ITYSTYGKF had an EpiMatrix Z-score of 1.4, which although is not ≥ 1.64 predicts that the epitope is from amongst the top 10% binders of the allele from within a random set of peptides. Accordingly, it can be said that the EpiMatrix predictions performed in this analysis were retrospectively validated for 19 out of the 20 epitopes that have been positively tested as 9-mers for HLA binding and/or the ability to trigger a CD8$^+$ T-cell response in human subjects. As for the epitope QYLAGLSTL which was empirically demonstrated to bind to the allele A*02:01, the empirical results did not validate the EpiMatrix prediction, where the allele was given a Z-score of 0.95. However, the binding of that epitope to its highest scoring allele, A*24:02 (Z-score = 2.73), was not tested *in vitro*; empirical validation of binding to this allele is still required. These results also take into account the 3 epitopes whose HLA restrictions require further confirmation; however their reported allelic restrictions do validate the EpiMatrix results. The retrospective validation results of 7 of the 20 epitopes are listed in table 9 as an example; these are the same epitopes whose EpiMatrix results are listed in table 5.

Apart from the positive entries found on IEDB for the 20 selected epitopes that were tested as 9-mers, negative results were also found for HLA binding assays and/or CD8$^+$ T-cell assays that contradict with the EpiMatrix predictions. With regard to HLA binding assays, 4 epitopes were shown to be negative binders of HLA class I alleles where the negative results contradict with their EpiMatrix scores; despite the fact that all 4 epitopes possess the binding motifs of the supertypes to which those alleles belong. With regard to CD8$^+$ T-cell assays, entries with "Negative" results for 15 out of the 20 epitopes were found, where the negative result was attributed to the 9-mer itself. For 9 out of those 15 epitopes, the restricting HLA allele associated with the negative response was determined, and the negative results contradicted with the EpiMatrix binding

predictions. However, due to the widely varying responses between patient cohorts and different experimental setups, these negative results cannot be used to rule out those epitopes from further analysis; further standardized testing is required to draw a final conclusion.

In addition to the 20 epitopes that have been tested positive as 9-mers, the IEDB search revealed that 22 out of the 70 chosen epitopes (category 2) had positive entries for HLA binding and/or the ability to trigger a CD8$^+$ T-cell response as substrings of 10-mer and/or multimer peptides (>10 AA). HLA binding assay results for these peptides cannot be used for the validation of the EpiMatrix predictions since they represent the binding affinities of the 10mer and multimer peptides themselves, and accordingly will not be considered. Only the results of T-cell assays will be considered, since multimer peptides undergo processing before presentation on MHC class I molecules. However, since the optimal epitope size for HLA class I molecules is 9 to 11 AA, the results where the peptides were tested as 10-mers or where the optimal/minimal epitope has been identified as a 10-mer are less likely to be attributed to the 9-mer peptide frames harbored within them as substrings, and hence will not be used as evidence for validation of the EpiMatrix results. Of the 22 epitopes mentioned herein, only 9 tested positive for the ability to trigger a CD8$^+$ T-cell response as multimers and had the restricting HLA allele determined. These results warrant support for further investigation via truncation analysis in order to identify the optimal/minimal epitope causing the positive response or maximal stimulation of the CD8$^+$ T-cells. Since in this project only 9-mer binding predictions have been performed, it is not possible to predict which 9-mer, 10-mer, or 11-mer frame is responsible for the optimal positive response from within the multimer peptide. The IEDB search for the remaining 28 of the 70 epitopes (category 3) to which entries were found in IEDB returned only negative results, which, as aforementioned, cannot be used to rule out those epitopes from further analysis.

5.2   HLA Class II Epitope Identification and Validation

5.2.1   iVAX analysis results:

From the 32,286 peptide frames scored by *Conservatrix* and *EpiMatrix* for conservation across the 46 genomes and for binding to the aforementioned DRB1 alleles respectively, the EpiAssembler tool selected 156 9-mer peptide frames to construct 14 extended ICS peptides, which harbor 9-mer frames in their core sequence that are collectively conserved in >90% of the analyzed sequences (Open Reading Frame ORF coverage >90%) and which have a minimum EpiMatrix cluster immunogenicity score of 10. Table 10 lists the results of the EpiAssembler analysis for 5 of the 14 ICS peptides. The EpiMatrix Z-scores assigned for each of the component 9-mer frames of one ICS peptide for binding to the analyzed HLA-DRB1 alleles and their percent conservation calculated by Conservatrix are listed in table 11 as an example. Peptide frames with 4 or more "hits" are referred to as "EpiBars" and are therefore promiscuous binders. All of the 14 constructed ICS peptides harbor at least one EpiBar.

Analysis of the locations of the ICS peptides revealed that 8 lie in the NS4B region, 3 in the NS5B region, 1 in the Core region, 1 in the NS3 region, and 1 at the junction between the NS2 and NS3 protein sequences. *JanusMatrix* analysis showed that all of the 14 ICS peptides possess a certain degree of homology to the human genome as demonstrated by their JanusMatrix cluster scores (table 10). JanusMatrix cluster scores are equal to the EpiMatrix cluster immunogenicity scores minus penalty deductions for each cross-reactive match in the human genome for each of the 9-mer component peptides of the ICS core sequence. The larger the difference between both cluster scores (EpiMatrix and JanusMatrix), the higher the homology to the human genome. Therefore, ICS peptides having a smaller difference between both scores are considered better candidates for inclusion in the final vaccine. Further empirical testing, however, is required to assess the efficacy and safety of the ICS peptides before deciding on which ones to include in the final product.

5.2.2 Retrospective validation results:

The IEDB search revealed that none of the 14 peptides have been tested earlier as exact sequences, however, all 14 have been previously tested for HLA class II binding and/or the ability to trigger a CD4[+] T-cell response as substring peptides or as peptides that are 70-99% similar to the ICS peptides (IEDB BLAST 70%-90% search). Of the 14 peptides, 8 have been previously shown (as substrings and/or similar sequences) to positively bind to different HLA class II alleles via *in vitro* binding assays; the IEDB search results for HLA class II binding assays for 4 of these 8 peptides are listed in table 12. Only the results for test peptides that include at least one 9-mer frame from the core sequence (excluding the 3 AA terminal residues at the N and C termini) of the corresponding ICS peptide within their core sequence are listed. All binding assays performed were purified MHC competition binding assays.

With regard to their ability to trigger a CD4[+] T-cell response, the IEDB search revealed that 13 out of the 14 ICS peptides have been previously tested positive in CD4[+] T-cell assays (as substrings and/or similar sequences). For only 5 out of the 13, the restricting HLA class II allele associated with the positive response has been determined; Table 13 lists the results of the IEDB search for 3 of those 5 ICS peptides; only the entries where the restricting HLA allele has been determined and where the test peptides or the empirically-determined minimal epitopes include at least one 9-mer frame from the core sequence of the ICS peptide within their core sequence are listed. In all of the assays performed on the 5 ICS peptides, the HLA allelic restriction was determined using methods 1 and/or 2 mentioned in table 8.

Upon comparing the IEDB results of the 8 aforementioned ICS peptides with the EpiMatrix predictions, it was found that 39 out of 51 EpiMatrix binding predictions performed for the component 9-mer frames of the 8 peptides collectively have been validated as true positive (TP) or true negative (TN). However, 12 binding predictions had contradicting IEDB results. Further empirical testing must be carried out to confirm or refute the EpiMatrix predictions before rendering these predictions as false positives (FP); especially that, for most of the 12 predictions, the contradicting results were reported in a single corresponding reference. Table 14 lists the results of the retrospective

validation performed for the ICS peptide $NS3_{1246-1265}$(SQGYKVLVLNPSVAATLGFG) as an example. For each IEDB entry for HLA binding or T-cell assays performed on the peptide demonstrated in tables 12 and 13, Table 14 lists whether the HLA restrictions that tested positive or negative for peptide binding confirm or refute the EpiMatrix predictions for each of the eight analyzed HLA class II alleles.

Apart from the 8 ICS peptides that have been tested for HLA binding and the ability to trigger a T-cell response with the restricting allele determined, the remainder 6 peptides were all tested for the ability to trigger a $CD4^+$ T-cell response in previous literature, and 6 of those tested positive, however without the restricting HLA allele being determined. These results need to be tested further in the future, and the restricting HLA class II alleles driving the response need to be determined.

5.3 HLA Class II in vitro binding assay

Table 15 displays the results of the pilot HLA class II binding assay compared to the highest Z-score calculated by EpiMatrix assigned to one of the core 9-mer frames (the frame most likely to bind with high affinity to the binding pockets of the HLA class II molecules). All ICS peptides were found generally to bind with high affinity to the tested HLA class II molecules, validating the EpiMatrix binding predictions, and confirming the results of the retrospective validation. This is true except for the peptide $NS4B_{1769-1786}$(VSGIQYLAGLSTLPGNPA), which failed to bind to the allele DRB1*07:01 in contradiction to the EpiMatrix pedictions. This negative result was also reported from three different references as shown in table 12 [137-139]. This strongly suggests that the EpiMatrix binding prediction for this peptide for binding to the DRB1*07:01 allele is a false positive.

5.4   *Conservation Analysis in newly partially re-sequenced HCV genomes from the Egyptian population*

HCV genomes from 15 serum samples were successfully amplified by RT-PCR; however, not all 6 regions were successfully amplified in all the extracted genomes. Table 16 shows 25 genomic fragments from the 15 sequences that were successfully amplified. The second fragment of NS5B failed to amplify in all serum samples, which suggests that the sequences of one or both primers targeting that region may have not been correctly synthesized. Examples of electrophoresis gel photos showing the bands corresponding to the nested PCR products are shown in figure 16; gel lanes that are not labeled belong to regions from the HCV sample regions that failed to amplify.

Following amplification, only 23 regions from 14 serum samples were successfully sequenced in the forward and/or reverse directions, as shown in table 16. Sample regions K9-NS5B(1) and K10-NS3(2) were not successfully sequenced in neither the forward nor reverse directions because the amount of DNA amplicons in the samples was insufficient for performing the sequencing reactions; this insufficiency was probably due to sample loss by acid hydrolysis upon repeated freeze-thaw cycles and/or during the purification of the PCR amplicons. Sample region K2-NS5B(1) was successfully sequenced in the forward direction only. After examining the resulting sequencing chromatograms of the forward and reverse sequencing reactions performed on the 23 sample regions that were successfully sequenced, it was found that, out of 45 forward and reverse sequences (excluding the reverse sequence of K2-NS5B(1)), 38 sequences were found to be of high quality, with clear chromatograms that have low background noise and distinct sharp peaks for each nucleotide along the sequence (after truncating the low quality nucleotide peaks at the 5' and 3' ends of the sequences). The chromatograms of the remaining 7 sequences were of low quality with a lot of background noise and broad peaks along the lengths of the sequences, and accordingly were not carried forward for further analysis; these 7 sequences, however, had corresponding sequences of the opposite polarity that were of high quality which were used to determine the amino acid sequences of the sample regions. The quality of the successfully sequenced samples and

63

their sizes after truncation are listed in table 17. Examples of good quality and bad quality chromatograms are displayed in figure 17.

The results of the BLAST analysis confirmed the identities of most of the high quality truncated sequences; the highest scoring matches belonged to HCV genotype 4, with alignment scores > 200 and expect (E) values of 0.0. The BLASTX results of the sequence K1-NS3(2)-F showed that the sequence was translated into two reading frames located 10 amino acids apart, with alignment scores > 200 for all aligned results; this could be due to insertion or deletion mutations that caused a frame shift in the analyzed sequence probably inserted during the amplification or sequencing reactions. The BLASTN analysis of the sequences K6-NS3(1)-R, K7- NS3(1)-F, and K7- NS3(1)-R did not return any results with significant similarity, despite the good quality of the chromatograms of the truncated sequences that were analyzed; the sequences of those regions, therefore, were not used for the conservation analysis. Furthermore, BLASTN analysis of the sequences K8-NS5B(1)-R, K10-NS3(2)-F, and K10-NS3(2)-R showed that these sequences had matches with sequences from the human genome, suggesting that samples K8 and K10 were contaminated with human RNA carried over from the HCV RNA extraction step, which at the low annealing temperatures of the PCR reactions may have been non specifically amplified.

Following the BLAST analysis, the high quality forward and reverse sequences of confirmed identities were assembled to derive the consensus sequences of the corresponding regions. If a mismatch was present between the amino acid sequences of the forward and reverse fragments, the amino acid residues used for the construction of the consensus sequence were those matching the corresponding residues of the highest scoring aligned sequences in the BLAST analysis. The locations of the resultant consensus amino acid sequences are listed in table 18.

The results of the conservation analysis using the Epitope Conservancy Analysis tool of the IEDB Analysis Resource website show that, out of 16 HLA class I epitopes in the NS3 region covered by the consensus sequences, 14 were 100% conserved across the sequences that harbored them, while 2 were only 50% conserved; the only HLA class II ICS peptide harbored within that region was also conserved in 100% of the sequences

64

that covered it. With regard to the NS4B region, all 14 HLA class I epitopes that were identified by the *in silico* analysis were 100% conserved across the sequences that harbored them, and the same result was reached for 5 HLA class II ICS peptides harbored within those sequences. As for the Core region, 21 out of the 23 HLA class I epitopes identified *in silico* were found to be 100% conserved across the sequences that harbored them, and only 1 was found to be only 75% conserved; the only ICS class II peptide located in the Core region was not covered by the new sequences. As for the NS5B(1) region, none of the NS5B epitopes identified *in silico* within this region (2 epitopes) were covered by the new sequences; this is because these epitopes are locate in the 5' end of the new sequences which was truncated due to the low quality of the sequencing chromatograms in that region. Most of the identified NS5B epitopes (13 out of 15 HLA class I epitopes and 3 HLA class II ICS peptides) are located within the region NS5B(2) which failed to amplify in all the obtained serum samples.

It is important to note that, as evident from table 17, the number of new sequences analyzed for each region is not the same. For example, the two HLA class I epitopes that were found to be only 50% conserved in the NS3 region were only covered by 2 new NS3 sample region sequences, making them conserved in only 1 of the 2; also, the Core HLA class I epitope that was found to be conserved in 75% of the sequences was covered by only 4 new sequences, making it conserved in 3 out of the 4. Generally, the maximum number of new sequences across which the conservancy of any given epitope was analyzed was 5, and the minimum was 2. For a pilot study, the results of this conservation analysis add to the results of the analysis performed by the *Conservatix* tool of the iVAX toolkit and confirm the high conservancy of the analyzed epitopes. A larger number of new sequences from the Egyptian population are required, however, to draw stronger conclusions about the conservancy of the analyzed epitopes in the HCV RNA genomes currently circulating in the Egyptian population. If significant deviations from the conservation analysis results obtained so far are to be found, this will not only possibly alter the decision of which epitopes are to be included in a population-specific vaccine for Egypt, but will also indicate the evolution of new HCV strains in the population that require investigation. An example of the results of the conservation analysis performed on the new sequences is displayed in table 19.

6. DISCUSSION

In this project, the immunoinformatics-based identification of a group of highly conserved and immunogenic T-cell epitopes from the HCV-4 proteome using the validated iVAX immunoinformatics toolkit was described, as a first step towards the development of an epitope-driven vaccine that targets HCV-4. A retrospective validation of the results was also performed using the repository of HCV immune epitope data on the IEDB; thereby introducing a fast and inexpensive method for the validation of *in silico* epitope prediction. The results of the retrospective analysis of 4 HLA class II constructed ICS peptides were also confirmed by a pilot HLA class II binding assay. Furthermore, a pilot conservation analysis of the predicted epitopes in the proteomes of HCV strains currently circulating in the Egyptian population was also performed. Based on this analysis, a group of 90 HLA class I epitopes and 14 HLA class II ICS peptides were identified as good putative candidates for vaccine design. A flowchart summarizing the epitopes' prediction and selection process and their retrospective validation is presented in figure 18.

Based on the results of the retrospective validation, 20 out of the set of 90 HLA class I epitopes were found to have no previous records of empirical testing via HLA binding assays nor T-cell assays, which warrants support for future research to investigate their roles in the interplay between HCV-4 and the host immune system, and to evaluate their efficacy as vaccine targets. From the remaining 70 HLA class I epitopes, the EpiMatrix prediction results of 19 out of 20 epitopes were validated; where those 20 epitopes were previously tested positive in empirical assays where the test peptides were the 9-mer epitope sequences themselves. The results of most of the EpiMatrix predictions for 8 out 14 HLA class II ICS peptides were also validated by retrospective analysis. These results demonstrate the high accuracy of EpiMatrix as an epitope-mapping tool. In most of the T-cell assays performed on the class I and II epitopes in previous literature, the HCV source genotype of the sequences of the test epitopes and the HCV genotype infecting the human test subjects was either genotype 1 or undetermined; genotypes 2 and 3 were also present as the source or infecting genotypes but to a much lesser extent. This indicates that the tested epitopes are conserved in HCV genotypes other than HCV-4,

meaning that they lie in locations in the viral proteins that affect the structure and function of the protein harboring them and influence the viral life cycle, and hence are difficult to mutate.

Different kinds of HLA binding assays were performed on the epitopes in previous literature, the most frequent of which is the purified MHC binding competition assay, where the binding affinity of the test epitope is measured by determining the concentration of that epitope required to inhibit the binding of 50% of a labeled standard epitope to the same purified MHC allele (Inhibitory Concentration 50 [IC_{50}]) [23]. Cell-bound MHC assays have also been performed; these include: (1) a competition assay, which is similar in concept to the described purified MHC binding competition assay but with the HLA allele bound to an experimental cell line and not purified [113], (2) a dissociation (complex stability) assay, where the binding affinity of the test epitope is measured by determining the time required for it to dissociate from its cell-bound MHC allele [112], and (3) association assays, where the affinity of the test epitope is quantified by its ability to associate to and stabilize MHC molecules on the surface of experimental cell lines compared to controls [105,141-143]. Examples of T-cell assays performed on the epitopes in previous literature include: (1) Enzyme-linked immunosorbent (ELISA) and Enzyme-linked immunosorbent spot (ELISPOT) assays, which detect T-cell responses by measuring cytokine secretion from T-cells (mainly IFN-γ in the IEDB results) in response to stimulation by the test epitopes [23], (2) cytotoxicity assays, which detect the cytotoxic (lytic) response of $CD8^+$ T-cell to target cells displaying the test epitope on their surface by measuring the release of a marker, usually the radioisotope Chromium 51, from the lysed target cells [105], (3) MHC tetramer staining assays, where streptavidin-labeled tetrameric constructs of MHC alleles in association with the test epitope are used to detect and quantify the amount of T-cells in a given blood sample that are specific for that epitope [144], (4) intracellular cytokine staining (ICS) assays, where the T-cell response is measured by detecting cytokines produced intra-cellularly in response to stimulation by the test peptide via intra-cellular cytokine staining followed by flow cytometry for visualization [23], and (5) proliferation assays, where the T-cell response is detected by measuring the proliferation of those T-cells in response to stimulation by the test epitope [23].

When the locations of the identified HLA class I and II epitopes were mapped on the HCV polyprotein sequence, it was found that most epitopes are located within the Core, NS3, NS4B, and NS5B regions. Within the Core protein region, which is the building block of the viral capsid, 23 class I epitopes and one class II ICS peptide were identified. Within the NS3 protein region, 21 class I epitopes and one class II ICS peptide were identified. Within the NS4B protein region, 14 HLA class I epitopes and 8 HLA class II ICS peptides were identified, and finally within the NS5B region, 15 HLA class I epitopes and 3 class II ICS peptides were identified. Due to the important structural and functional roles that these proteins play in the HCV life cycle, it is difficult for the virus to mutate those sequences at low fitness costs, which explains why the biggest proportion of the conserved epitopes identified in this project was located within those regions. Furthermore, a large group of those epitopes was also found to be conserved in other HCV genotypes, as demonstrated by the retrospective validation results performed on IEDB. This does not only demonstrate the importance of those epitopes' sequences in the HCV life cycle, but also means that these epitopes provide excellent targets for a cross-genotype effective vaccine. On the other hand, in the E1 and E2 regions, very few highly conserved epitopes were identified relative to the sizes of those regions, which is explained by the hypervariable nature of those proteins required for the virus to escape recognition from the humoral immune system.

In order to avoid triggering a suppressive or autoimmune response to the final vaccine product, the novel JanusMatrix tool was used to assess the homology of the analyzed epitope set to the human genome; epitopes with several cross-reactive hits with epitopes derived from the human genome are better to be excluded from the final product. In this project, a highly stringent selection of HLA class I epitopes that had 0 cross-reactive hits in the human genome was performed. The same could not be done for the HLA class II ICS peptides, which all had some degree of homology to the human genome; if 9-mer peptide frames within the ICS peptides were to be removed to decrease this degree of homology, the promiscuous coverage of the HLA-DR alleles provided by the peptides, their ability to bind to HLA class II molecules, and their HCV-4 subtype coverage would all be compromised. Here it is important to note that when a large set of empirically characterized T-effector cell epitopes were analyzed by JanusMatrix, it was

found that a T-cell epitope homologous to the human genome may still be able to successfully trigger a T-effector cell response if that degree of homology lies within a certain ratio of cross-reactive hits per number of AA residues in the human genome database [84]. Therefore, future empirical analysis of the kind of responses the HLA class II ICS peptides can trigger is required before deciding on which epitopes to include in the vaccine. While JanusMatrix is efficient in identifying homologous epitopes derived from the human genome, it cannot identify all of the putatively cross-reactive epitopes. This is because, in reality, the TCRs of cross-reactive T-cells can recognize their target epitopes by the conformational availability of their AA side-chain residues and not just their exact sequence identity. However, JanusMatrix can still provide a preliminary prediction of cross-reactivity based on sequence identity that can narrow down the selection of epitopes which would be later tested empirically for the nature of the immune response that they can trigger [84]. Work is currently still in progress to establish a normalized scoring scheme for JanusMatrix (similar to EpiMatrix Z-score) to define a threshold based on which the nature of the analyzed epitopes (whether effector or suppressive) can be determined.

The conservation analysis performed on the newly partially re-sequenced HCV genomes isolated from the Egyptian population showed that almost all the epitopes located within the new sequences were fully conserved across them. This indicates that these epitopes may provide good coverage of the HCV strains currently circulating in the Egyptian population, which makes them good candidates for a vaccine that can be effective in Egypt; however, the number of sequences included in the analysis is relatively small, which decreases the power of the conclusions drawn from the analysis. From the small amount of serum samples that could be obtained to carry out the analysis, not all regions were successfully amplified; which could be attributed to low viral loads in the samples, mismatches between the HCV RNA and the amplification and sequencing primers, and/or the degradation of the HCV RNA with repeated freeze-thaw cycles. Furthermore, HCV RNA is known to have a relatively high GC content, meaning that 2ry RNA structures may have been difficult to denature at the low primer annealing temperatures employed in the reactions. All the above factors may have contributed to the decrease in availability of enough HCV RNA templates to perform the RT-PCR

amplification. This is supported by the fact that, for the sample regions that did amplify, a relatively large volume of extracted HCV RNA was required to carry out the reactions at low optimal annealing temperatures. Better results can be obtained in the future by determining the viral load of the analyzed sequences prior to amplification so as to only use samples with high viral loads, and by using primer sequences that are designed based on only the HCV-4 strains derived from the Egyptian population.

7. CONCLUSION

HCV-4 is the predominant HCV genotype in the Middle East and Africa; in Egypt, the country with the highest HCV infection rates worldwide, more than 90% of the infections are attributed to the same genotype. Despite the direness of the issue, very limited attention has been given to HCV-4 as a vaccine target in HCV vaccine research. In an attempt to fill this gap in research, this project describes the *in silico* identification and selection of a group of highly conserved and immunogenic T-cell epitopes from the HCV-4 proteome as a first step for the development of an epitope-driven vaccine against HCV-4. The iVAX immunoinformatics toolkit was used to identify the epitopes; where its epitope prediction tool, EpiMatrix, was proven to be of the most accurate epitope prediction tools available online. The *in silico* results were retrospectively validated using the repository of empirical HCV immune epitope data on the IEDB; thus demonstrating the accuracy of EpiMatrix as a prediction tool, and the conservation of the identified epitopes across other HCV genotypes. This retrospective method of validation can provide a fast and inexpensive alternative to performing empirical validations from scratch for different strains of previously investigated pathogens, such as HCV, HIV, Influenza, or Coronaviruses. The method can be useful when designing population-specific or personalized vaccines and/or in the case of an outbreak caused by a new strain of a given pathogen. In addition to identifying and retrospectively validating the epitopes, the HCV-4 subtype sequence coverage provided by the identified epitopes for the HCV strains circulating in the Egyptian population has been investigated, as a pilot assessment of the putative efficacy of the vaccine when used in Egypt.

In the future, empirical characterization of the 20 novel HLA class I epitopes identified by the iVAX analysis must be carried out, and further assessment of the efficacy of the retrospectively validated HLA class I and class II T-cell epitopes as vaccine targets in patient populations where HCV-4 prevalence is high needs to be pursued. Pre-clinical trials for previously thoroughly investigated epitopes using small animal models can be directly undertaken. In the coming years, several new drugs for HCV treatment will enter the market, just like the recently approved drug Sofobuvir. These drugs, although they hold promise for increasing the rates of treatment success,

remain very expensive for patient populations in the Middle East and Africa. The development of an effective vaccine targeting those populations will have a strong impact on HCV therapy and prophylaxis.

8. TABLES

Table 1: **HCV Structural & Non-Structural Proteins and their functions.** A summary of the roles that HCV proteins play in the HCV life cycle [12].

Protein	Function
Core	Forms the viral nucleocapsid.
E1 & E2	Mediate viral attachment to cell-surface receptors and fusion of viral and cellular membranes in the viral entry step of the HCV life cycle.
p7	Cation channel (viroporin), essential for assembly and release of infectious HCV particles.
NS2	Constitutes the *cis*-acting auto-protease together with NS3 for cleavage of the NS2-NS3 junction in the precursor polyprotein molecule.
NS3	A multi-functional protein. The N-terminal one-third harbors a serine protease catalytic domain that catalyzes cleavage of the polyprotein molecule at the NS2/NS3, NS3/NS4A, NS4A/NS4B, NS4B/NS5A, and the NS5A/NS5B junctions with the aid of the NS4A co-factor. The C-terminal two-thirds of the protein harbor an RNA helicase/NTPase, which is suggested to unwind extensive secondary RNA structures during RNA replication.
NS4A	Co-factor of NS3 serine protease activity.
NS4B	Induces formation of the membranous web, the site of HCV RNA replication in the host cell
NS5A	Modulates the efficiency of viral RNA replication
NS5B	RNA-dependant-RNA polymerase, catalyzes viral RNA replication

Table 2: HCV Vaccines in Clinical Trials. Examples of HCV vaccines currently undergoing clinical trials, and the results of these trials as reported so far [17].

Vaccine Type, Components, & Investigator Company	Source HCV genotype	Phase (Year)	Tested subjects	Outcome	Notes
Recombinant protein E1/E2 protein + MF59C adjuvant *(Novartis Inc.)*	1a	I (2010)	60 healthy subjects	Safe and elicited antibody and T-cell responses in all subjects.	In pre-clinical trials it elicited antibody and T-cell responses in 21 chimpanzees: 5 developed sterilizing immunity against a homologous challenge with HCV-1a, and 14 managed to clear the challenge virus after infection [57].
Peptide Vaccine IC41: 5 peptides from Core, NS3, and NS4 regions + poly-L-arginine adjuvant *(Intercell AG)*	1 and 2	II (2008)	60 patients, chronic HCV-1 infection, HLA-A2 positive	Safe, induces T-cell response in 67% of patients, and caused decrease in viremia >1 log in 3 subjects.	A new phase II clinical trial for the vaccine has been started in 2011 testing the effects of the vaccine in combination with the new broad spectrum antiviral Nitazoxanide.
DNA Vaccine ChonVac-C: plasmid expressing NS3/4a protein *(Tripep AB)*	N/A	I/IIa (2009)	12 treatment-naive patients infected with HCV-1	Safe, caused >0.5 log decrease in viral load in 4/6 patients who received higher doses	Delivered by electroporation to enhance intramuscular injection immunogenicity.
Vector Vaccine TG4040: Modified vaccinia Ankara (MVA) vector expressing NS3,4, and 5B proteins *(Transgene)*	N/A	I (2009)	15 patients with Chronic HCV	Caused 0.5–1.4 decline in viral load in 6/15 patients.	Phase II trial planned in combination with standard therapy.

Table 3: Analyzed HCV-4 Sequences. A List of the 46 HCV IV full genomes whose corresponding polyprotein amino acid sequences were uploaded to the iVAX server for epitope prediction and analysis.

Accession Number	Genotype/Subtype	RNA Sequence Length	Protein ID	Amino Acid Sequence Length	Sampling Country	Reference
DQ418782	4a	9292	ABD75824	3009	U.S.A.	[87]
DQ418783	4a	9253	ABD75825	3008	U.S.A.	
DQ418784	4a	9298	ABD75826	3008	U.S.A.	
DQ418785	4	9218	ABD75827	2985	U.S.A.	
DQ418786	4d	9273	ABD75828	3007	U.S.A.	
DQ418787	4a	9307	ABD75829	3008	U.S.A.	
DQ418788	4a	9306	ABD75830	3008	U.S.A.	
DQ418789	4a	9295	ABD75831	3008	U.S.A.	
DQ516083	4d	9300	ABF60956	3006	Spain	[88]
DQ516084	4a	9309	ABF60957	3009	Spain	
DQ988073	4a	8957	ABL63005	2899	Egypt	Direct Submission
DQ988074	4a	9054	ABL63006	2932	Egypt	
DQ988075	4a	9057	ABL63007	2933	Egypt	
DQ988076	4a	9051	ABL63008	2931	Egypt	
DQ988077	4a	9043	ABL63009	2928	Egypt	
DQ988078	4a	9050	ABL63010	2932	Egypt	
DQ988079	4a	8965	ABL63011	2902	Egypt	
EF589160	4f	9181	ABU68271	2969	France	[89]
EF589161	4f	9304	ABU68272	3010	France	
EU392169	4f	9309	ACB45490	3009	N/A	[90]
EU392170	4f	9269	ACB45491	3009	N/A	
EU392171	4K	9281	ACB45492	3011	N/A	
EU392172	4d	9299	ACB45493	3006	N/A	
EU392173	4k	9312	ACB45494	3011	N/A	
EU392174	4f	9298	ACB45495	3009	N/A	
EU392175	4f	9297	ACB45496	3009	N/A	
FJ025854	4b	9267	ACL68400	3010	Portugal	[91]
FJ025855	4b	9242	ACL68401	3011	Portugal	
FJ025856	4b	9271	ACL68402	3011	Portugal	
FJ462431	4p	9475	ACS29430	3009	Canada	[20]
FJ462432	4g	9435	ACS29431	3008	Canada	
FJ462433	4m	9425	ACS29432	3006	Canada	
FJ462434	4q	9433	ACS29433	3010	Canada	
FJ462435	4b	9440	ACS29434	3011	Canada	
FJ462436	4c	9468	ACS29435	3022	Canada	
FJ462437	4d	9421	ACS29436	3006	Canada	
FJ462438	4k	9438	ACS29437	3011	Canada	
FJ462439	4r	9440	ACS29438	3011	Canada	
FJ462440	4o	9422	ACS29439	3005	Canada	
FJ462441	4n	9435	ACS29440	3008	Canada	
FJ839869	4t	9488	ACT66294	3007	Canada	
FJ839870	4l	9388	ACT66295	3006	Canada	
GU814265	4a	9497	ADF97233	3008	Egypt	[92]
HQ537008	4v	9174	AEI00316	2973	Cyprus	[93]
HQ537009	4v	9177	AEI00317	2974	Cyprus	
NC_009825	4a	9355	YP_001469632	3008	Egypt	[94]

Table 4: RT-PCR and Sanger Sequencing primers. The polarity of the primers is denoted in subscript letters (F: Forward, R: Reverse). Nested primers are denoted by the subscript letter 'N'; not all nested primers have a nested counterpart of opposite polarity.

Forward Primers				Reverse Primers					Amplicon size	Target Region
Primer	Primer Sequence (5' - 3')	Location	Length (nt)	Tm (°C)	Primer	Primer Sequence (5' - 3')	Location	Length	Tm (°C)		
A_F	TGATAGGGTGCTTGCGAGTGC	294 - 314	21	59.7	A_R	TGCTKGAATTCGGGCAGTCRTT	954 - 975	22	62	679 bp	Core
					A_{RN}	CACGAGAGAAGTGCCAARAGGAAGAT	867 - 892	26	60.9	599 bp	
B_F	YGGCAGGGACACCAAYGA	3485 - 3502	18	56.3	B_R	AGGGCGACCTCCTCTATRTT	4422 - 4441	20	53.7	957 bp	NS3 (1)
B_{FN}	GTACACCAATGTTGACCARGA	3641 - 3661	21	50.9						801 bp	
C_F	ATCATCTGTGACGARTGCCA	4278 - 4297	20	53.3	C_R	GCTGACATACAAGCCATGATGTAYTT	5268 - 5293	26	57.3	1016 bp	NS3 (2)
C_{FN}	GTCCTGGACCAAGCGGARAC	4335 - 4354	20	56.8	C_{RN}	CACTTCCACATGGTGTCCCA	5151 - 5170	20	56	836 bp	
D_F	CAGTTTGACGAAATGGARGARTG	5451 - 5473	23	56.5	D_R	TTTAGCCACGTYTTGAAGTCA	6308 - 6328	21	52.8	878 bp	NS4B
					D_{RN}	GTCTGARAAGGGATGTCACAGT	6195 - 6216	22	50.6	766 bp	
E_F	TGYTGCTCAATGTCATAYTCATGGAC	7596 - 7621	26	59.8	E_R	ACRTAGAGTCTYTCTGTGAGGGC	8364 - 8386	23	54.3	791 bp	NS5B (1)
					E_{RN}	TCGGGCTCCAGGTCACA	8327 - 8343	17	55.1	752 bp	
F_F	ATGAYACCCGCTGCTTTGACTC	8257 - 8278	22	58.1	F_R	GRTTGAGTTCGTGTGGAGAGTATCC	9027 - 9051	25	57.9	793 bp	NS5B (2)

Table 5: Binding prediction results for 7 HLA class I candidate epitopes for HCV-4 vaccine design. The dark blue color code of the EpiMatrix Z-sores indicates that the epitope is from amongst the top 1% binders of that allele in a random set of peptides, the blue color indicates that the epitope is from amongst the top 5% binders, and light blue color code indicates that the epitope is from amongst the top 10% binders. Z-scores of the bottom 90% binder peptides are not highlighted. The 'Hits' column displays the number of HLA alleles each epitope is predicted to bind to with a Z-score ≥ 1.64 as a measure of the epitopes' binding promiscuity.

Epitope Sequence	Percent Conservation	EpiMatrix Z-Scores						Hits	Location
		A*01:01	A*02:01	A*03:01	A*24:02	B*07:02	B*44:03		
DVVCCSMSY	0.98	2.52	-0.14	2.42	0.19	-0.21	0.6	2	NS5A/NS5B
RLLAPITAY	0.98	1.86	1.07	2.64	0.09	0.93	2.02	3	NS2/NS3
FWAKHMWNF	0.96	-0.21	-0.01	-0.54	3.11	1.16	0.97	1	NS4B
QYLAGLSTL	1	-0.5	0.95	-0.69	2.73	0.55	1.13	1	NS4B
DPRRRSRNL	0.98	-1.56	-1.12	-1.46	0.45	2.81	-0.49	1	Core
YLVAYQATV	0.93	0.71	3.2	0.6	0.27	0.56	0.76	1	NS3
ITYSTYGKF	0.93	1.75	0.37	1.24	1.4	1.49	1.27	1	NS3

Table 6: IEDB Results for HLA Class I Binding Assays performed on 4 candidate epitopes for HCV-4 vaccine design. The "Quantitative Measure" column displays the quantitative results of the HLA binding assays, where: (1) the IC_{50} results of competition binding assays are displayed in nanomolar concentrations (nM), and (2) the results for the HLA dissociation assay are displayed in minutes. The "Assay Result" column displays the designated relative binding affinity of the epitopes as determined by the authors based on the quantitative results of the assays where available, or as determined by the IEDB curators.

AA Sequence	Restricting Allele	Assay Description	Quantitative Measure	Assay Result	Ref.
DVVCCSMSY	A26	Purified MHC - Radioactivity Competition	25 nM	Positive	101
RLLAPITAY	A*03:01	Purified MHC - Radioactivity Competition	4 nM	Positive - High	102
YLVAYQATV	A*02:01	Purified MHC - Radioactivity Competition	45 nM	Positive - High	103
			45 nM	Positive - High	104
			20 nM	Positive - High	105
			20 nM	Positive - High	102
			20 nM	Positive - High	106
			20.4 nM	Positive	107
			20.4 nM	Positive - High	108
			5 nM	Positive - High	109
			N/A	Positive	110
			59 nM	Positive - Intermediate	111
		Cell bound MHC - Fluorescence Dissociation	1440 min.	Positive - High	112
		Cell bound MHC - Fluorescence Competition	4.8 nM	Positive - High	113
	A*02:02	Purified MHC - Radioactivity Competition	39.1 nM	Positive - High	108
			4.3 nM	Positive - High	109
			5.2 nM	Positive - High	109
		Cell bound MHC - Fluorescence Competition	1.5 nM	Positive - High	113
	A*02:03	Purified MHC - Radioactivity Competition	15.9 nM	Positive - High	108
			10 nM	Positive - High	109
		Cell bound MHC - Fluorescence Competition	2.1 nM	Positive - High	113
	A*02:05	Purified MHC - Radioactivity Competition	4.3 nM	Positive - High	109
			22.6 nM	Positive - High	109
	A*02:06	Purified MHC - Radioactivity Competition	82.2 nM	Positive - Intermediate	108
			3.7 nM	Positive - High	109
		Cell bound MHC - Fluorescence Competition	15 nM	Positive - High	113

Table 6: Continued.

	A*02:07	Purified MHC - Radioactivity Competition	> 2300 nM	Negative	109
	A*68:02	Purified MHC - Radioactivity Competition	33.3 nM	Positive - High	108
			40 nM	Positive - High	109
		Cell bound MHC - Fluorescence Competition	81 nM	Positive - Intermediate	113
FWAKHMWNF	A*24:02	Cell bound MHC - Fluorescence Association	N/A	Positive	105

Table 7: IEDB Search Results for T-cell Assays performed on 4 candidate epitopes for HCV-4 vaccine design. "N/A" HCV genotype means that the genotype of the source or infecting virus was undetermined. The Method of Restricting Allele Determination column is explained in table 9.

AA Sequence	Restricting HLA Allele	Ref.	T-cell Assay Description	Assay Result	Source HCV Genotype	Patient(s) HCV Genotype	Method of Restricting Allele Determination
DVVCCSMSY	A26	101	ICS (IFN-γ)	Positive	1	1	1C
		114	ICS (IFN-γ)	Positive	1	1	1B
YLVAYQATV	A2	108	^{51}Cr Release Cytotoxicity	Positive	N/A	N/A	3C
		115	Cytotoxicity Assay	Positive	1	N/A	3B
		116	Cytotoxicity Assay	Positive	1	N/A	3B
		117	ELISpot (IFN-γ)	Positive	1	N/A	3B
		118	ELISpot (IFN-γ)	Positive	1	N/A	3A
	A*02:01	102	^{51}Cr Release Cytotoxicity	Positive	N/A	N/A	2C
		111	^{51}Cr Release Cytotoxicity	Negative	1	N/A	3C
		119	ELISpot (IFN-γ)	Negative	1	N/A	3B
		120	ELISpot (IFN-γ)	Negative	1	N/A	3C
FWAKHMWNF	A*24:02	121	MHC Tetramer Staining ELISpot (IFN-γ)	Positive	1	1	2A
		122	MHC Tetramer Staining	Positive	N/A	N/A	2
		105	^{51}Cr Release Cytotoxicity	Positive	1	1	3C
DPRRRSRNL	B7	123	ELISpot (IFN-γ)	Positive	1	1	3B
		124	^{51}Cr Release Cytotoxicity	Positive	N/A	1	1B
		125	ELISpot (IFN-γ)	Positive	1	N/A	3B

Table 7: Continued

	126	^{51}Cr Release Cytotoxicity	Positive	N/A	N/A	3B
	127	ELISpot (IFN-γ)	Positive	1	N/A	3B
	128	ELISpot (IFN-γ)	Negative	1	N/A	3B

Table 8: Restricting HLA allele determination. Methods of restricting HLA allele determination of T-cell assays listed in table 8.

	Method of Restricting Allele Determination
1	T-cell assay performed on CD8[+] T-cells from HLA-X+ individuals utilizing different APCs with matched and mismatched HLA alleles to demonstrate allelic restriction
2	T-cell assay performed on CD8[+] T-cells from HLA-X+ individuals utilizing APC or HLA tetramers presenting the peptide where the is only a single type of HLA present to demonstrate allelic restriction
3	T-cell assay performed on CD8[+] T-cells from HLA-X+ individuals utilizing APCs that are HLA-X+ presenting the test epitope on their surface
	Basis for prediction/knowledge of Restriction
A	*In silico* binding prediction / Contain HLA binding motif
B	Cited reference
C	*In vitro* binding assay

Table 9: Retrospective validation results of 7 predicted HLA class I epitopes. The Z-scores of the supertype alleles whose binding predictions were retrospectively validated are emphasized by an underline and in bold red font.

AA Sequence	EpiMatrix Z-Scores						Empirically demonstrated HLA restrictions (Supertype)	Prediction Validated?
	A*01:01	A*02:01	A*03:01	A*24:02	B*07:02	B*44:03		
DVVCCSMSY	**2.52**	-0.14	2.42	0.19	-0.21	0.6	A26 (A01)	Yes
RLLAPITAY	1.86	1.07	**2.64**	0.09	0.93	2.02	A*03:01 (A03)	Yes
ITYSTYGKF	1.75	0.37	1.24	**1.4**	1.49	1.27	A*24:02 (A24)	Yes (top 10%)
YLVAYQATV	0.71	**3.2**	0.6	0.27	0.56	0.76	A*02:01 (A02) A*02:02 (A02) A*02:03 (A02) A*02:05 (A02) A*02:06 (A02) A*68:02 (A02)	Yes
FWAKHMWNF	-0.21	-0.01	-0.54	**3.11**	1.16	0.97	A*24:02 (A24)	Yes
QYLAGLSTL	-0.5	**0.95**	-0.69	2.73	0.55	1.13	A*02:01 (A02)	No
DPRRRSRNL	-1.56	-1.12	-1.46	0.45	**2.81**	-0.49	B7 (B07)	Yes

Table 10: Five Immunogenic Consensus Sequence peptides constructed by EpiAssembler. The 9-mer frames of the core of the ICS peptides are emphasized in the ICS sequence column. The Open Reading Frame (ORF) coverage column displays the percentage of the HCV-4 sequences covered by each ICS peptide with regard to the conservation of its 9-mer core frames, collectively, across the sequences analyzed.

ICS Sequence	ORF Coverage	EpiMatrix Cluster Score	Janus Matrix Cluster Score	Location	Region
SQG**YKVLVLNPSVAATL**GFG	1	35.24	14.23	1246 - 1265	NS3
VSG**IQYLAGLSTLPG**NPA	1	18.78	4.58	1769 - 1786	NS4B
STQ**QTLLFNILGGWVA**AQI	1	11.32	2.23	1802 - 1820	NS4B
EGA**VQWMNRLIAFASRGN**HVA	1	23.46	19.42	1910 - 1930	NS4B
PTI**WVRMILMTHFFSILQ**XQE	0.98	24.91	19.03	2837 - 2857	NS5B

Table 11: Component 9-mer frames of the ICS peptide NS3$_{1246-1265}$ (SQGYKVLVLNPSVAATLGFG). The table illustrates the *Conservatix* and *EpiMatrix* scores of each overlapping 9-mer frame in the peptide. Z-score color codes follow the same rules stated in table 6. Epibars (9-mer frames that have 4 or more 'hits') are highlighted in yellow.

9-mer Frame	Percent Conservation	EpiMatrix Z-Scores								Hits
		DRB1*01:01	DRB1*03:01	DRB1*04:01	DRB1*07:01	DRB1*08:01	DRB1*11:01	DRB1*13:01	DRB1*15:01	
SQGYKVLVL	0.02	-0.08	-0.37	0.14	0.89	0.26	-0.08	1.1	0.45	0
QGYKVLVLN	0.98	-1.14	-0.83	-1.37	-0.81	0.73	-0.66	-0.37	-1.09	0
GYKVLVLNP	0.98	-0.45	-0.47	0.36	0.15	-0.6	0.24	-0.11	-0.69	0
YKVLVLNPS	0.98	1.81	1.93	2.02	1.43	2.6	3.08	1.35	2.1	6
KVLVLNPSV	1	1.59	0.97	1.31	1.33	0.87	1.61	1.26	1.9	1
VLVLNPSVA	1	1.9	1.39	1.73	1.48	1.31	1.74	1.49	1.58	3
LVLNPSVAA	1	2.15	2.04	2.28	1.76	1.6	1.59	2	1.77	6
VLNPSVAAT	1	0.62	0.02	1.15	0.41	0.35	0.64	0.9	0.19	0
LNPSVAATL	1	1.97	1.83	1.39	1.9	0.82	1.45	1.68	1.38	4
NPSVAATLG	1	0.22	-0.39	0.08	-0.69	-0.18	-0.65	-0.18	-0.06	0
PSVAATLGF	1	-0.51	0.19	0.67	0.24	-0.34	-0.29	0.32	-0.25	0
SVAATLGFG	1	-0.44	0.95	-0.35	-0.68	0.85	0.22	-0.17	1.05	0

Table 12: **IEDB Search Results for HLA Class II Binding Assays performed on 4 ICS peptides.** The "Test Sequence" column lists the peptide sequence actually tested by the authors of the listed references; AA residues identical to the ICS sequence are colored in black, mismatched residues are colored in red and are underlined, AA residues outside the ICS frame are faded in grey. The "IC_{50}" column displays the results of the purified MHC competition binding assays in nanomolar concentrations (nM).

ICS Sequence	Test Sequence	Ref.	Restricting Allele	IC_{50}	Assay Result
SQGYKVLVLNPSVAATLGFG	GYKVLVLNPSVAAT	132	DRB1*01:01	1.4	Positive -High
		133	DRB1*04:01	7.8	Positive -High
			DRB1*04:04	33	Positive -High
			DRB1*04:05	141	Positive -Int
			DRB1*07:01	126	Positive -Int
			DRB1*08:02	21	Positive -High
			DRB1*09:01	266	Positive -Int
			DRB1*11:01	75	Positive -High
			DRB1*13:02	3.5	Positive -High
			DRB1*15:01	39	Positive -High
			DRB1*08:03	1124	Negative
			DRB1*12:01	4604	Negative
			DRB5*01:01	3695	Negative
	GYKVLVLNPSV	134	DRB1*11:01	26	Positive -High
	GYKVLVLNPSVAATL	135	DRB1*04:01	N/A	Positive
	KVLVLNPSVAATLGF	136	DR1	8	Positive -High
			DR4	20	Positive -High
			DR7	7	Positive -High
			DR11	390	Positive
			DR15	210	Positive
			DRB5	447	Positive
			DR3	4183	Negative
			DR13	10000	Negative
			DRB3	40000	Negative
			DRB4	6481	Negative
	AQGGYKVLVLNPSVAATLGFGAYMSKAHGID	136	DR4	5	Positive -High
			DR15	57	Positive -High
			DR11	40	Positive -High
			DR7	15	Positive -High
			DR1	1	Positive -High
			DR13	900	Positive -Low
			DRB5	1225	Negative
			DR3	2500	Negative

			DRB3	>100000	Negative
			DRB4	2449	Negative
	GYKVLVLNPSVAATLGFGAY	137	DRB1*01:01	3.5	Positive -High
			DRB1*04:01	9.7	Positive -High
			DRB1*07:01	23	Positive -High
			DRB1*08:02	80	Positive -High
			DRB1*09:01	20	Positive -High
			DRB1*11:01	240	Positive -Int
			DRB1*13:02	4.1	Positive -High
			DRB1*15:01	42	Positive -High
			DRB5*01:01	8154	Negative
EGAVQWMNRLIAFASRGNHVA	GEGAVQWMNRLIAFASRGNHV	137	DRB1*01:01	3.2	Positive -High
			DRB1*04:01	361	Positive -Int
			DRB1*07:01	221	Positive -Int
			DRB1*08:02	158	Positive -Int
			DRB1*11:01	14	Positive -High
			DRB5*01:01	182	Positive -Int
			DRB1*09:01	6818	Negative
	GPGEGAVQWMNRLIAFASRG	137	DRB1*01:01	8.7	Positive -High
			DRB1*11:01	88	Positive -High
			DRB1*04:01	5328	Negative
			DRB1*07:01	6305	Negative
	GAVQWMNRLIAFASRGNHVS	138	DRB1*01:01	13	Positive -High
			DRB1*04:01	698	Positive -Int
			DRB1*04:04	35.9	Positive -High
			DRB1*15:01	3.03	Positive -High
			DRB1*03:01	92.5	Positive -High
VSGIQYLAGLSTLPGNPA	NFISGIQYLAGLSTLPGNPA	137	DRB1*01:01	10	Positive -High
			DRB1*04:01	84	Positive -High
			DRB1*08:02	70	Positive -High
			DRB1*09:01	441	Positive -Int
			DRB1*11:01	74	Positive -High
			DRB5*01:01	606	Negative
			DRB1*07:01	>10000	Negative
	QYLAGLSTLPGNGN	138	DRB1*01:01	2.9	Positive -High
			DRB1*04:01	3	Positive -High
			DRB1*15:01	99.2	Positive -High
			DRB1*04:04	1.97	Positive -High
			DRB1*03:01	>70000	Negative
	GIQYLAGLSTLPGNPAIASL	135	DRB1*04:01	N/A	Positive
		138	DRB1*04:01	13.2	Positive -High

			DRB5*01:01	357	Positive -Int
			DRB1*15:01	365	Positive -Int
			DRB1*11:01	5.31	Positive -High
			DRB1*04:05	6.28	Positive -High
			DRB1*13:02	214	Positive -Int
			DRB1*01:01	0.946	Positive -High
			DRB1*07:01	2180	Negative
			DRB1*03:01	35000	Negative
		139	DRB1*01:01	1	Positive -High
			DRB1*04:01	13	Positive -High
			DRB1*04:05	6	Positive -High
			DRB1*11:01	5	Positive -High
			DRB1*13:02	214	Positive -Int
			DRB1*15:01	378	Positive -Int
			DRB5*01:01	357	Positive -Int
			DRB1*03:01	>69767	Negative
			DRB1*07:01	2176	Negative
STQQTLLFNILGGWVAAQI	LTTSQTLLFNILGGWVAAQL	138	DRB1*15:01	17.4	Positive -High
			DRB1*07:01	638	Positive -Int
			DRB1*04:05	502	Positive -Int
			DRB1*13:02	101	Positive -Int
			DRB1*01:01	2.5	Positive -High
			DRB1*04:01	2470	Negative
			DRB5*01:01	1950	Negative
			DRB1*11:01	12300	Negative
			DRB1*03:01	3920	Negative

Table 13: IEDB Search Results for CD4[+] T-cell Assays performed on 3 ICS peptides. The "Test/Optimal Sequences" column lists the peptide sequences actually tested by the authors or the optimal sequence within the tested peptides responsible for the maximal T-cell response (denoted by an asterix*) as shown by truncation analysis. Methods of restricting allele determination follow the same rules of table 9.

ICS Sequence	Ref.	Assay	Test/Optimal Sequences	Source Genotype	Subjects Genotype	Restricting Allele	Method of Restricting Allele Determination
SQGYKVLVLNPSVAATLGFG	141	[3]H -thymidine cell proliferation	QGYKVLVLNPSVAATLGFGA	N/A	N/A	DRB1*11:01	1B
		MHC Tetramer staining	GYKVLVLNPSVAAT	N/A	N/A	DRB1*11:01	2B
	135	MHC Tetramer staining	GYKVLVLNPSVAATL	N/A	N/A	DRB1*04:01	2B
	132	[3]H -thymidine cell proliferation	GYKVLVLNPSVAAT	1,2	1, N/A	DRB1*12:01	1C
						DRB1*11:01	
						DRB1*13:02	
						DRB1*04:01	
						DRB1*16:01	
	141	MHC Tetramer staining	GYKVLVLNPSVAATL	N/A	N/A	DR4	2B
	142	[3]H -thymidine cell proliferation	VLVLNPSVAATLGFGAYM	N/A	N/A	DRB1*04:01	2
		Cytokine Release (IL-2)					
VSGIQYLAGLSTLPGNPA	144	[3]H -thymidine cell proliferation	QYLAGLSTLPGN*	N/A	3	DRB1*11:04	1
			MWNFISGIQYLAGLSTLPGN	1	3	DRB1*11:01	
	139	ICS (IFN-γ)	LAGLSTLPGNP*	1	1	DRB1*04:01	1C
						DRB1*11:04	
						DRB1*04:07	
						DRB1*04:04	
	141	MHC Tetramer staining	SGIQYLAGLSTLPGNPAIASL	N/A	N/A	DR4	2B
		Cytotoxicity Assay					
	135	MHC Tetramer staining	SGIQYLAGLSTLPGNPAIASL	N/A	N/A	DRB1*04:01	2C
STQQTLLFNILGGWVAAQI	144	^{3}H–thymidine cell proliferation	LFNILGGWVA*	N/A	1,3	DRB1*01:01	1
			FNILGGWVAAQL	N/A	1,3	DRB1*01:01	

89

Table 14: Retrospective validation results of ICS peptide NS3$_{1246\text{-}1265}$ (SQGYKVLVLNPSVAATLGFG).
True positive (TP) and false positive (FP) assignments validate or refute the *EpiMatrix* binding predictions outlined in table 12. Cells without assignments lack corresponding empirical data for comparison.

Type of Assay	Validation Results							
	DRB1*01:01	DRB1*03:01	DRB1*04:01	DRB1*07:01	DRB1*08:01	DRB1*11:01	DRB1*13:01	DRB1*15:01
HLA Binding Assay	TP	---	TP	TP	TP	TP	TP	TP
	---	---	---	---	---	TP	---	---
	---	---	TP	---	---	---	---	---
	TP	FP	TP	TP	---	TP	---	TP
	TP	---	TP	TP	TP	TP	TP	TP
CD4+ T-cell Assay	---	---	---	---	---	TP	---	---
	---	---	TP	---	---	---	---	---
	---	---	TP	---	---	TP	TP	---
	---	---	TP	---	---	---	---	---
Overall Results	TP	FP	TP	TP	TP	TP	TP	TP

Table 15: **HLA Class II binding assay results.** The binding affinity scale is divided as follows: (1) Strong Binder: IC50 < 10,000 nM, (2) Moderate Binder: 10,000 < IC50 <100,000 nM, (3) Weak binder: IC50 > 100,000 nM.

ICS Sequence	DRB1*01:01		DRB1*04:01		DRB1*07:01		DRB1*15:01	
	Highest Z-Score	IC$_{50}$ (nM)	Highest Z-Score	IC$_{50}$ (nM)	Highest Z-Score	IC$_{50}$ (nM)	Highest Z-Score	IC$_{50}$ (nM)
SQG**YKVLVLNPSVAATL**GFG	2.15	44	2.28	123	1.9	3,010	2.1	711
VSG**IQYLAGLSTLPG**NPA	2.54	148	2.25	694	1.93	Non-binder	2.25	1,239
STQ**QTLLFNILGGWVA**AQI	2.33	215	1.94	721	1.77	122,087	1.6	782
EGA**VQWMNRLIAFASRGN**HVA	2.54	2,726	1.52	20,416	1.55	44,713	2.56	103

Table 16: Amplified and Sequenced HCV Samples from the Egyptian Population. "Amp" means that the region of the corresponding sample was successfully amplified by RT-PCR and Nested PCR. "Seq" means that the region was successfully sequenced in the forward and/or reverse directions.

Sample	Core	NS3 (1)	NS3 (2)	NS4B	NS5B (1)	NS5B (2)
K1	---	---	Amp + Seq	---	---	---
K2	---	Amp + Seq	Amp + Seq	Amp + Seq	Amp + Seq	---
K3	---	---	Amp + Seq	---	---	---
K4	---	Amp + Seq	Amp + Seq	Amp + Seq	---	---
K5	---	---	Amp + Seq	Amp + Seq	---	---
K6	---	Amp + Seq	---	---	---	---
K7	---	Amp + Seq	---	---	---	---
K8	---	---	---	---	Amp + Seq	---
K9	---	---	---	---	Amp Only	---
K10	---	Amp + Seq	Amp Only	---	---	---
K11	Amp + Seq	---	---	---	---	---
K12	Amp + Seq	---	---	Amp + Seq	---	---
K13	Amp + Seq	---	---	Amp + Seq	Amp + Seq	---
K14	Amp + Seq	---	---	---	---	---
K15	---	---	---	---	Amp + Seq	---

Table 17: Egyptian population successfully sequenced sample regions' sizes, chromatogram qualities, and locations on the HCV polyprotein sequence. The letters "F" and "R" in the sample region column label the forward and reverse sequences obtained from the sequencing reactions. The translation reading frames of the sequences were determined using the BLASTX analysis. The locations of the translated amino acid sequences on the reference HCV-1a (H77) AA sequence were determined using the Sequence Locator tool of the HCV Los Alamos database.

Sample	Sample Region	Chromatogram Quality	Length after Truncation (nt)	Reading Frame	Location on HCV polyprotein (AA)
K1	NS3(2)-F	High	741	+1/+3	1354 – 1486 / 1497 - 1599
	NS3(2)-R	Low	---	---	---
K2	NS3(1)-F	High	633	+1	1152 - 1362
	NS3(1)-R	High	616	-1	1111 - 1315
	NS3(2)-F	High	680	+1	1358-1578
	NS3(2)-R	High	738	-2	1352 - 1596
	NS4B-F	Low	---	---	---
	NS4B-R	High	568	-3	1717 - 1904
	NS5B(1)-F	High	651	+3	2447 - 2662
	NS5B(1)-R	---	---	---	---
K3	NS3(2)-F	High	751	+1	1354 - 1603
	NS3(2)-R	High	781	-1	1337 - 1596
K4	NS3(1)-F	High	632	+3	1151 - 1359
	NS3(1)-R	Low	---	---	---
	NS3(2)-F	Low	---	---	---
	NS3(2)-R	High	570	-1	1354 - 1543
	NS4B-F	High	693	+2	1724 - 1952
	NS4B-R	High	659	-2	1717 - 1935
K5	NS3(2)-F	High	766	+1	1349 - 1603
	NS3(2)-R	High	781	-1	1337 - 1596
	NS4B-F	High	678	+2	1730 - 1954
	NS4B-R	High	674	-1	1709 - 1927
K6	NS3(1)-F	Low	---	---	---
	NS3(1)-R	High	443	N/A	N/A
K7	NS3(1)-F	High	464	N/A	N/A
	NS3(1)-R	High	541	N/A	N/A

Table 17: Continued.

K8	**NS5B(1)-F**	**Low**	---	---	---
	NS5B(1)-R	**High**	407	N/A	N/A
K10	**NS3(2)-F**	**High**	478	N/A	N/A
	NS3(2)-R	**High**	487	N/A	N/A
K11	**Core-F**	**Low**	---	---	---
	Core-R	**High**	392	-1	1 - 117
K12	**Core-F**	**High**	503	+3	14 - 180
	Core-R	**High**	503	-2	1 - 155
	NS4B-F	**High**	695	+1	1746 – 1952
	NS4B-R	**High**	704	-3	1710 - 1943
K13	**Core-F**	**High**	504	+1	14 -180
	Core-R	**High**	505	-1	1 – 155
	NS4B-F	**High**	619	+1	1739 - 1944
	NS4B-R	**High**	688	-3	1708 - 1924
	NS5B(1)-F	**High**	650	+3	2450 – 2664
	NS5B(1)-R	**High**	652	-1	2422 – 2637
K14	**Core-F**	**High**	469	+3	25 - 179
	Core-R	**High**	502	-2	1 - 155
K15	**NS5B(1)-F**	**High**	676	+2	2443 – 2665
	NS5B(1)-R	**High**	679	-3	2426 - 2650

Table 18: Consensus assembled HCV sequences and their locations.

Sample	Region	Location (AA)
K1	NS3(2)	1354 – 1486
		1497 - 1599
K2	NS3	1111 - 1596
	NS4B	1717 - 1904
	NS5B(1)	2447 - 2662
K3	NS3(2)	1337 - 1603
K4	NS3	1151 - 1543
	NS4B	1717 - 1952
K5	NS3(2)	1337 - 1603
	NS4B	1709 - 1954
K11	Core	1 - 117
K12	Core	1 - 180
	NS4B	1710 - 1952
K13	Core	1 - 180
	NS4B	1708 - 1944
	NS5B(1)	2422 - 2664
K14	Core	1 - 179
K15	NS5B(1)	2426 - 2665

Table 19: Example of the conservation analysis results across the newly partially re-sequenced HCV genomes from the Egyptian population. The "Sequence Coverage" column displays the number of sequences in which the epitopes were conserved as part of the total number of new sequences that cover their locations.

Epitope Sequence	Location (AA)	Region	Sequence Coverage	% Conservation
HLA Class I Epitopes				
YLVAYQATV	1585 - 1593	NS3	4/4	100%
ITYSTYGKF	1291 - 1299	NS3	1/2	50%
FWAKHMWNF	1760 -1768	NS4B	5/5	100%
RPMDVKFPG	18 - 26	Core	3/4	75%
HLA Class II ICS Peptides				
SQGYKVLVLNPSVAATLGFG	1246 - 1265	NS3	2/2	100%
VSGIQYLAGLSTLPGNPA	1769 - 1786	NS4B	5/5	100%
STQQTLLFNILGGWVAAQI	1802 - 1820	NS4B	5/5	100%
EGAVQWMNRLIAFASRGNHVA	1910 - 1930	NS4B	4/4	100%

9. FIGURES
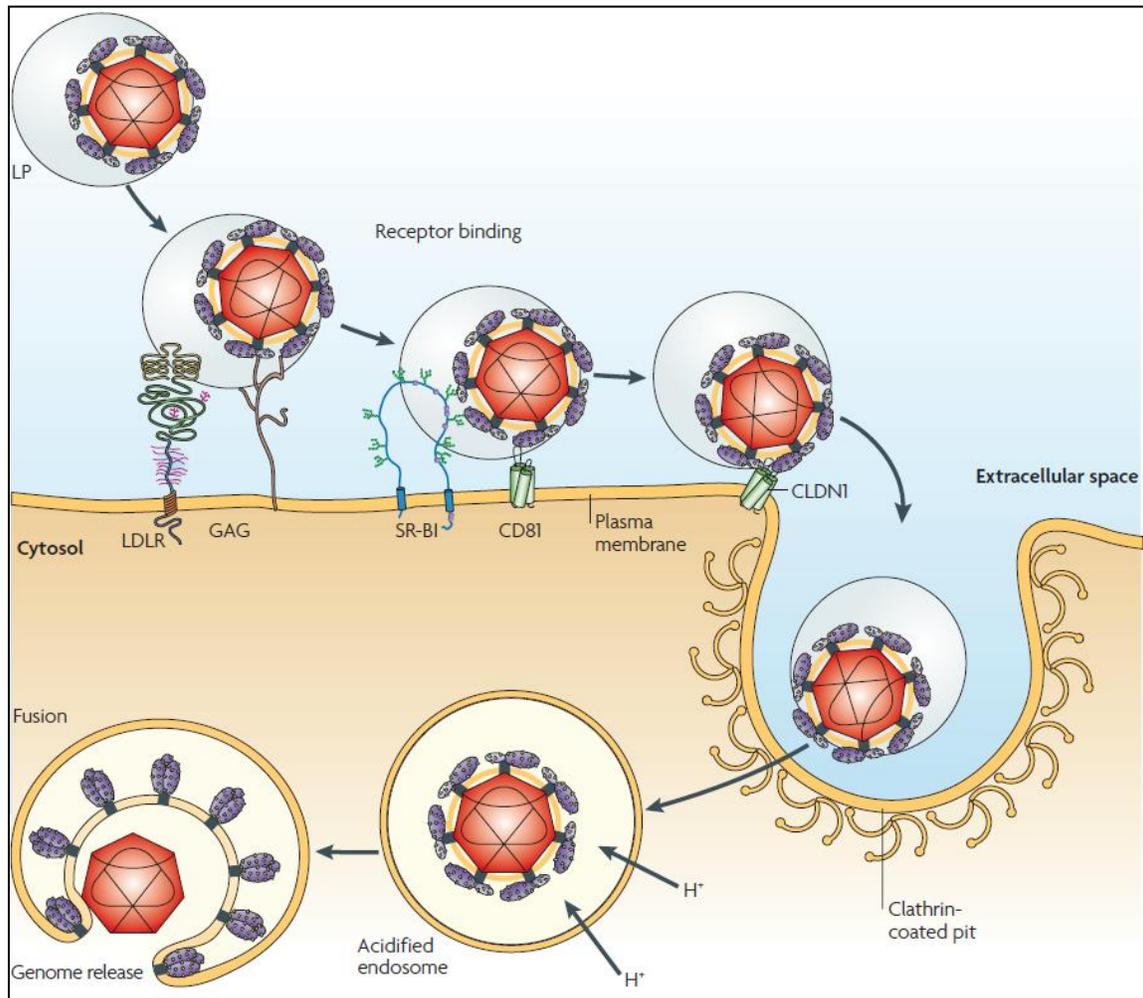
Figure 1: HCV Entry into Hepatocytes. Schematic representation of our current understanding of the viral entry step in the HCV life cycle. *Reprinted by permission from Macmillan Publishers Ltd: Nature Reviews Microbiology [12], copyright 2007.*
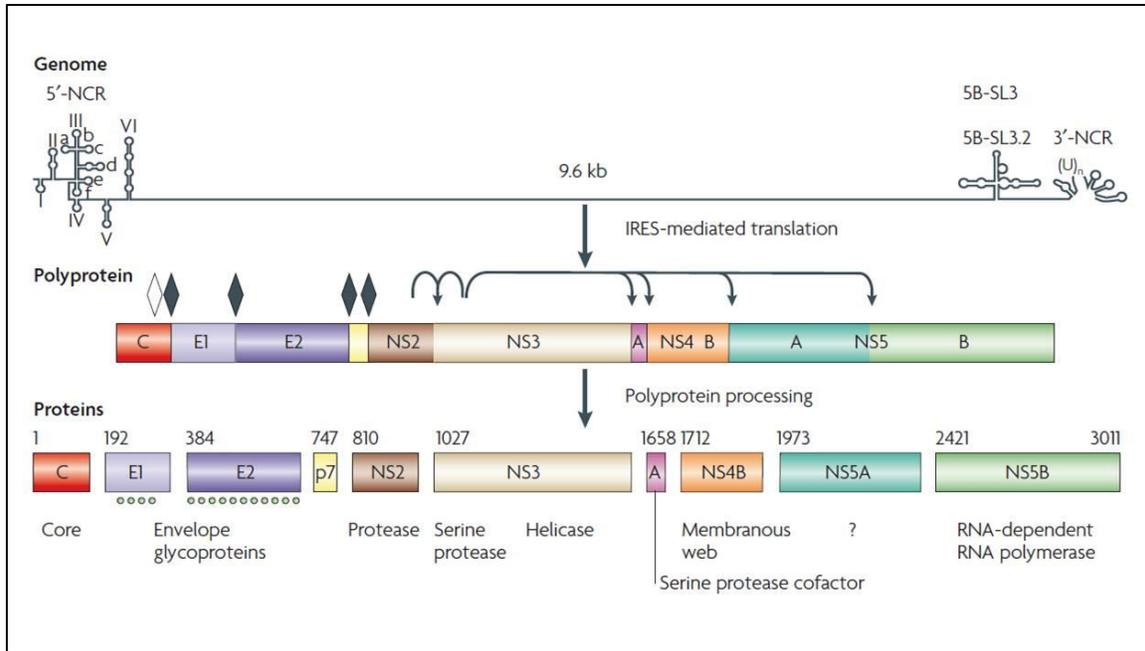
Figure 2: HCV RNA and Polyprotein. Schematic representation of the HCV RNA genome, the precursor polyprotein molecule, and its processing into the structural and non-structural proteins of HCV. Reprinted from [12]. *Reprinted by permission from Macmillan Publishers Ltd: Nature Reviews Microbiology [12], copyright 2007.*
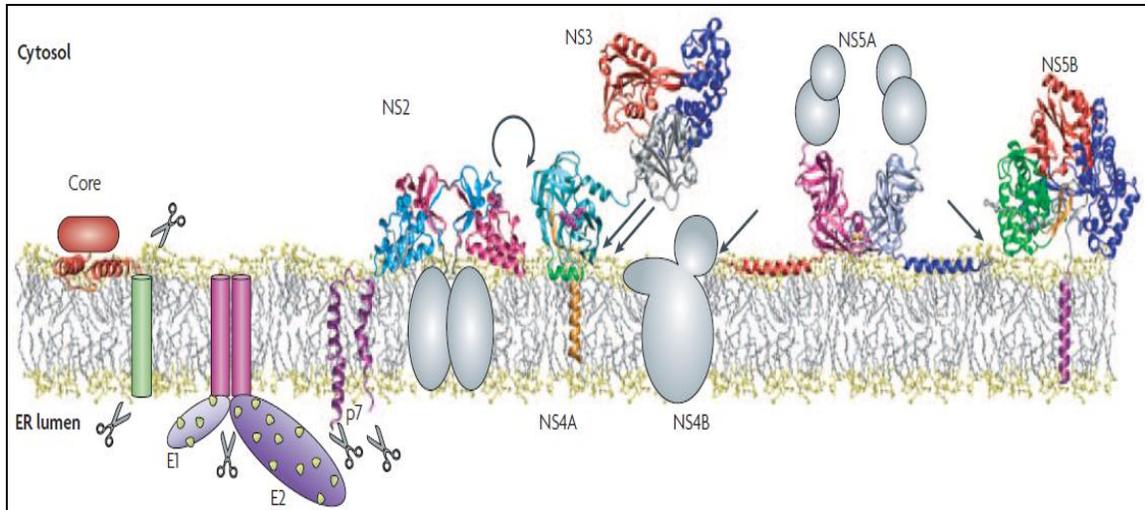
Figure 3: HCV Proteins. Schematic representation of the structures of the HCV proteins and their ER membrane associations. Reprinted from [12]. *Reprinted by permission from Macmillan Publishers Ltd: Nature Reviews Microbiology [12], copyright 2007.*
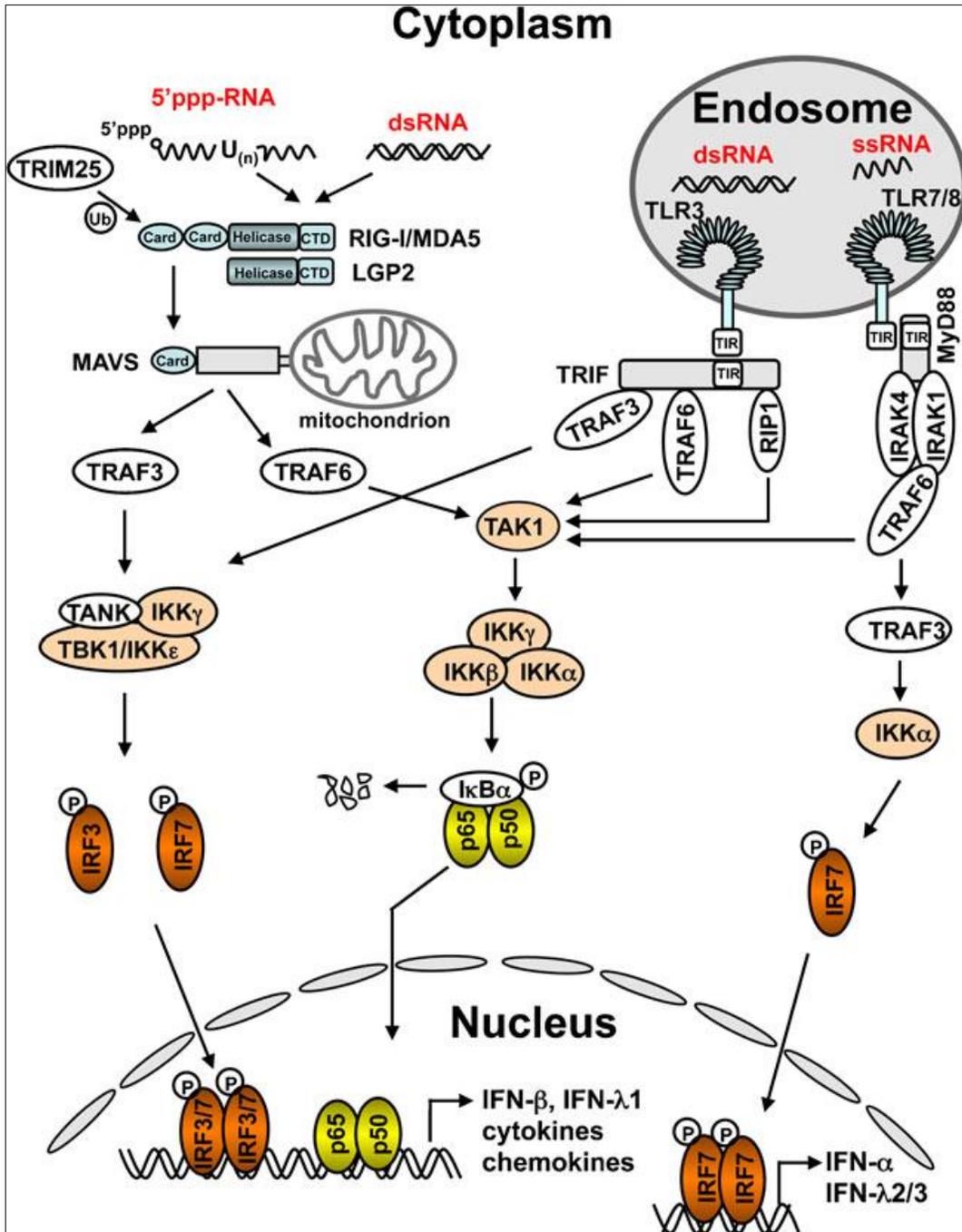
Figure 4: Type I and III IFN production. Detailed schematic depiction of how HCV induces the production of type I and III IFNs by triggering intracellular signaling pathways that lead to their expression. *Reprinted with permission from [29].*
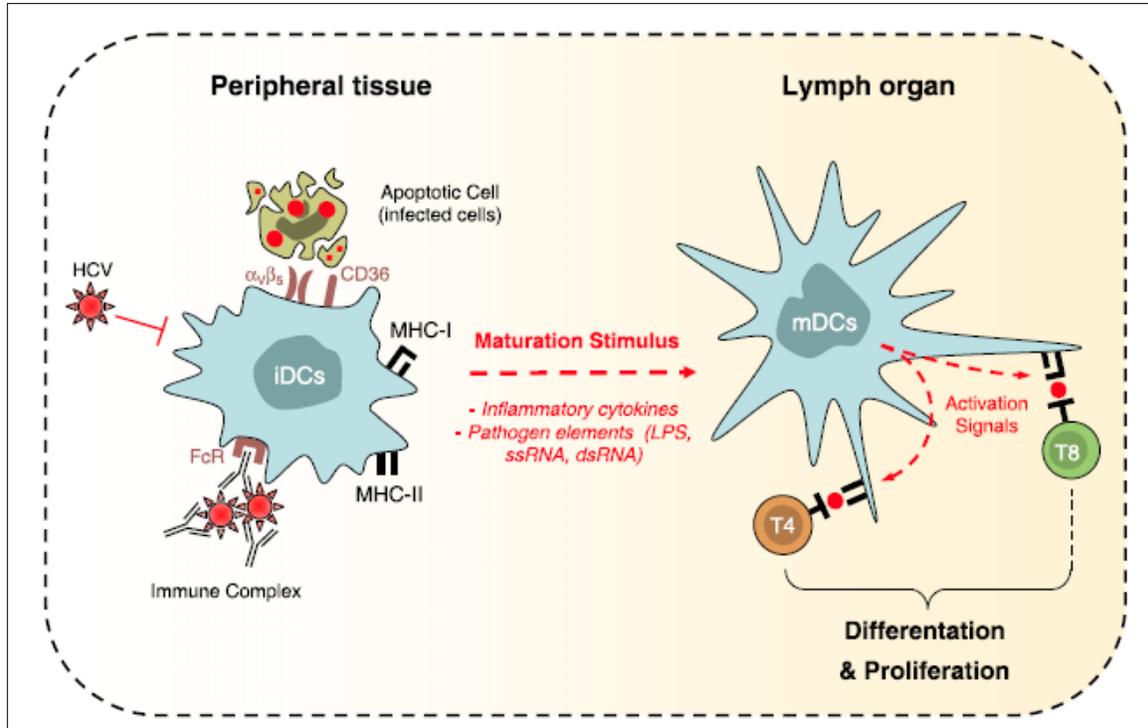
Figure 5: Maturation of Dendritic Cells. Schematic representation of the uptake of HCV antigens by immature dendritic cells (iDCs), and their subsequent maturation into myeloid dendritic cells (mDCs); which in turn activate naïve T-cells in the lymphoid organs. *Reprinted from [40] with permission from Elservier.*
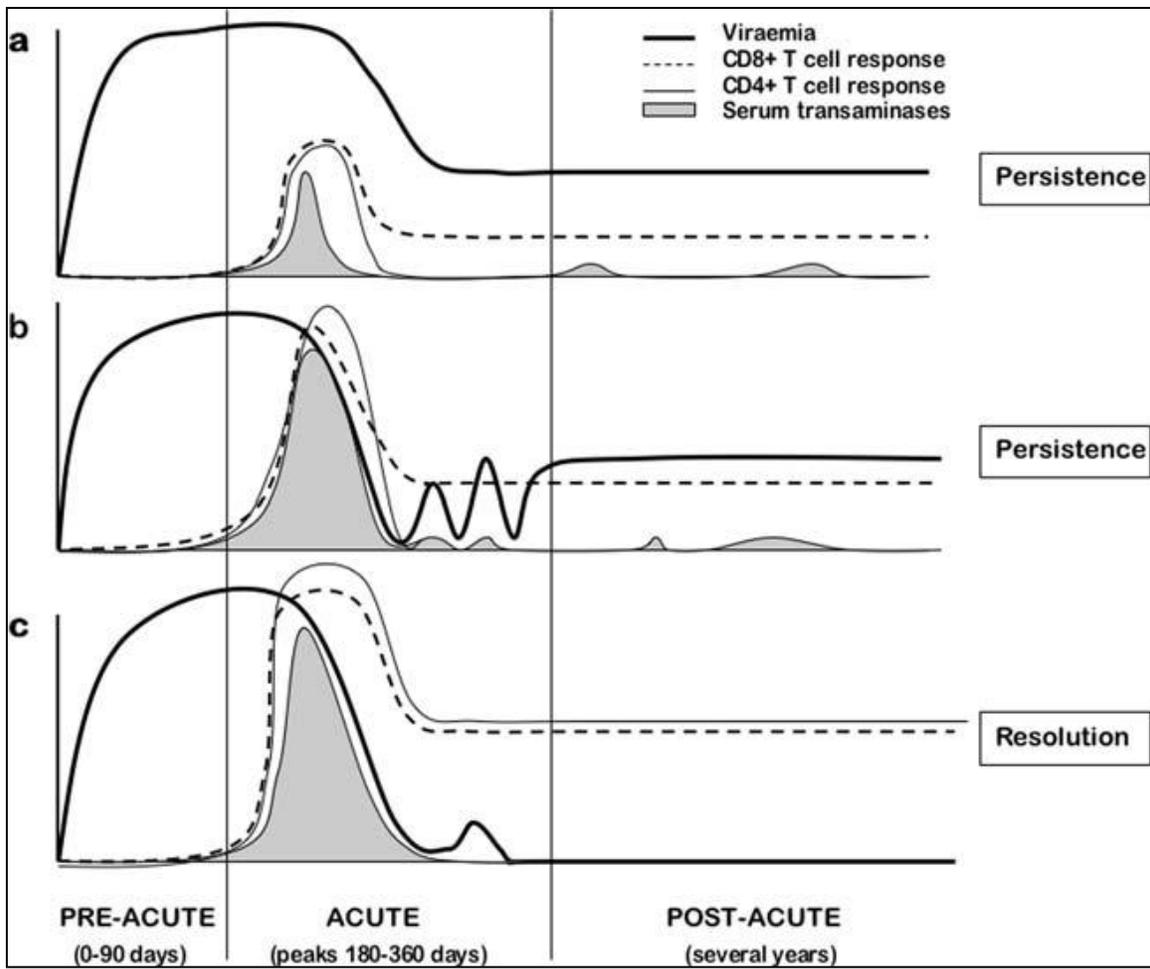
Figure 6: The Evolution of the T-cell response to HCV. Graphical illustration of the changes in viral and T-cell titers across the 3 phases of the T-cell response to HCV; and the possible outcomes of infection: a) weak, short-lasting T-cell response ending in viral persistence, b) initially strong T-cell response that is not sustained, ending in viral persistence, and c) Strong and sustained T-cell response ending in viral clearance. *Reprinted with permission from [43].*
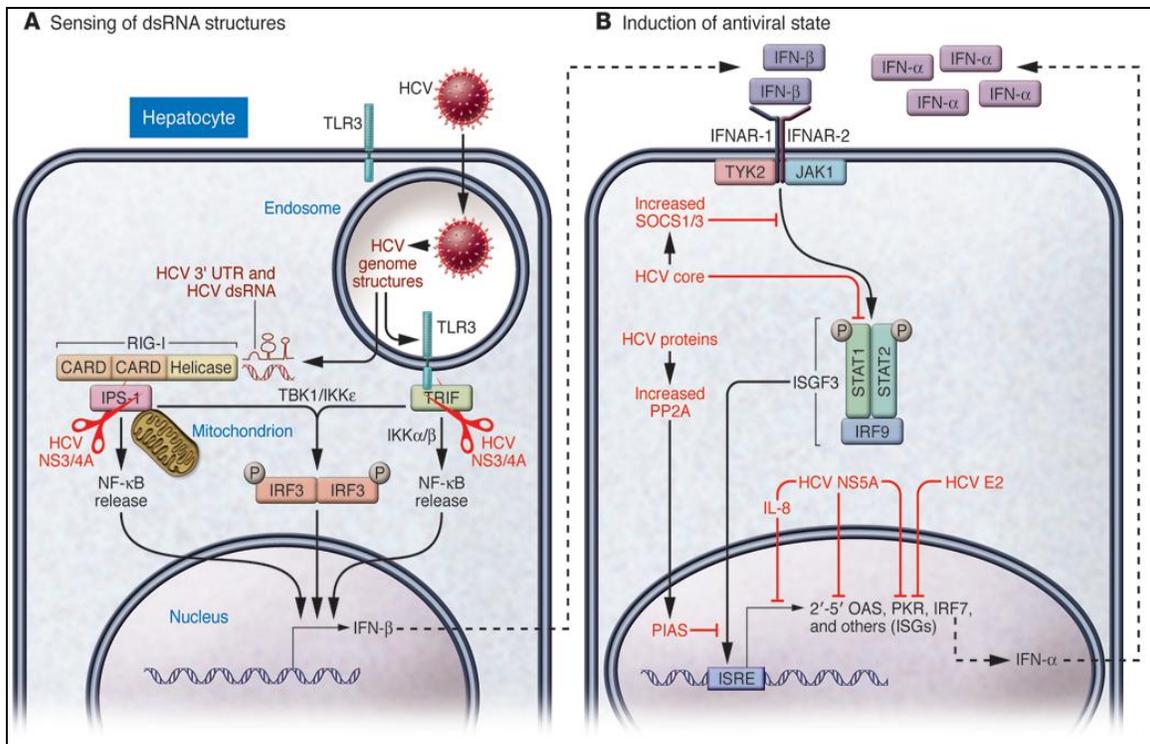
Figure 7: Inhibition of IFN production and ISG protein action. Schematic representation of the interference of different HCV proteins with the expression of type I IFNs and ISG and with the actions of the ISG protein products; the interference of HCV proteins is emphasized in red. *Reprinted with permission from [14].*
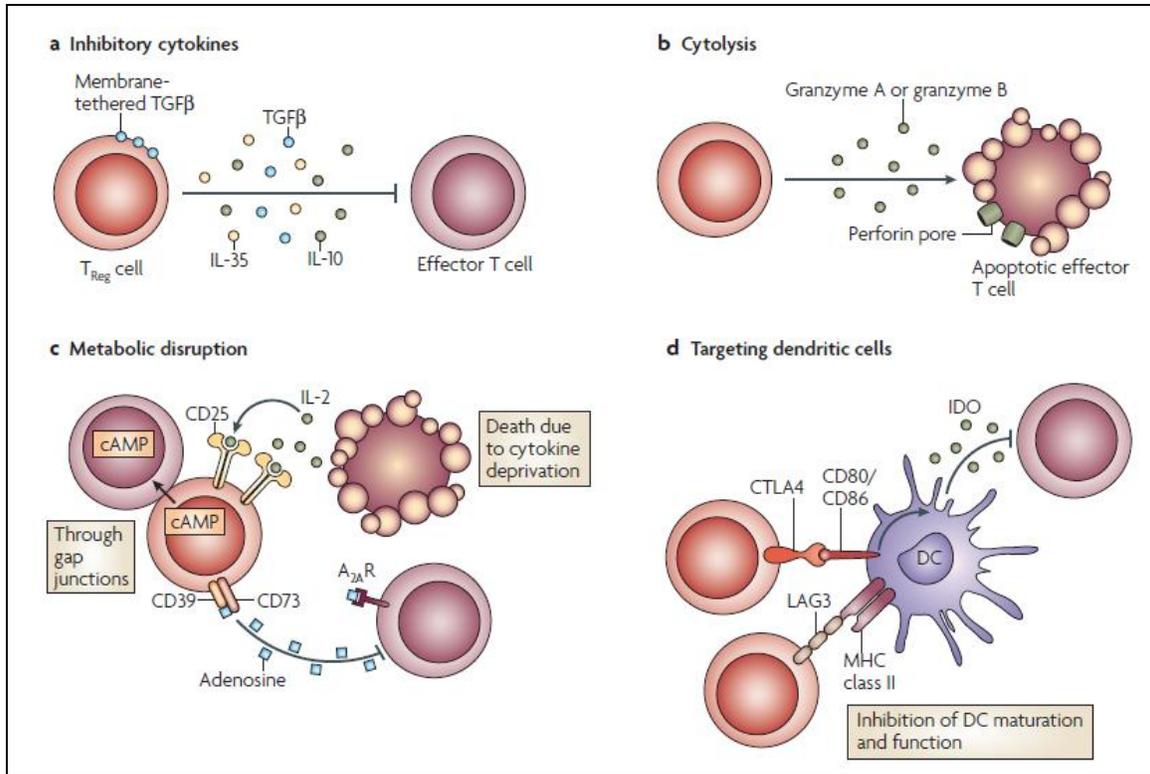
Figure 8: Suppressive actions of T-regulatory lymphocytes. Schematic illustration of the mechanisms by which Tregs suppress the immune response: a) via inhibitory cytokines, b) triggering apoptosis, c) metabolic disruption, and d) inhibition of DC maturation. *Reprinted by permission from Macmillan Publishers Ltd: Nature Reviews Immunology [51], copyright 2008.*
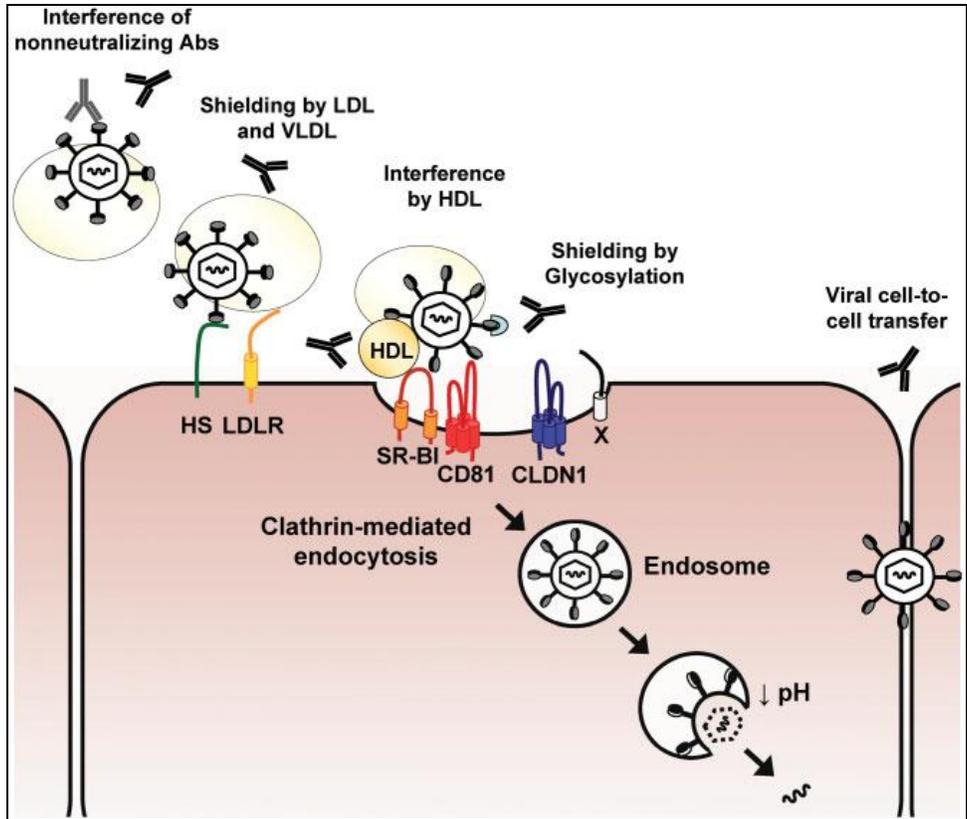
Figure 9: HCV evasion of neutralizing antibodies. Schematic illustration of how HCV evades recognition and capture by neutralizing antibodies. *Reprinted with permission from [15].*
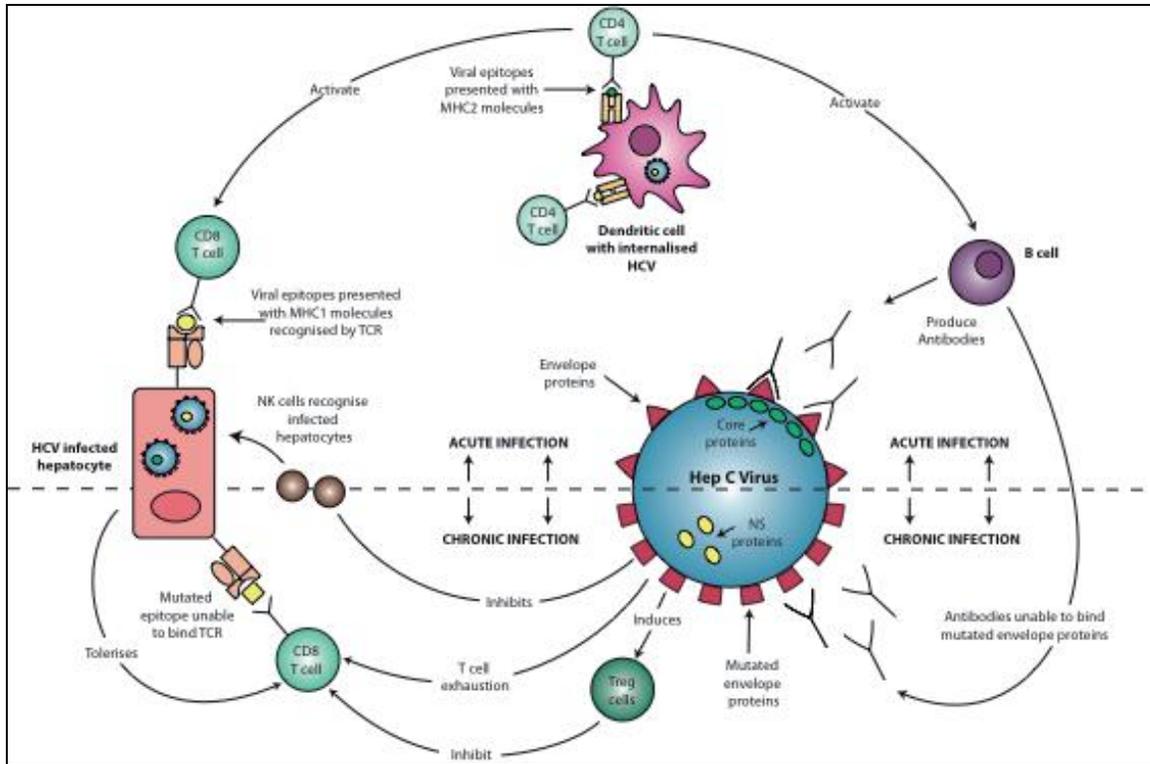
Figure 10: **The interaction between HCV and the host immune system.** A schematic illustration of the immune response to HCV in the acute phase of the infection, and the mechanisms by which HCV evades that response to lead the infection to progress to the chronic phase. *Reprinted with permission from [17].*
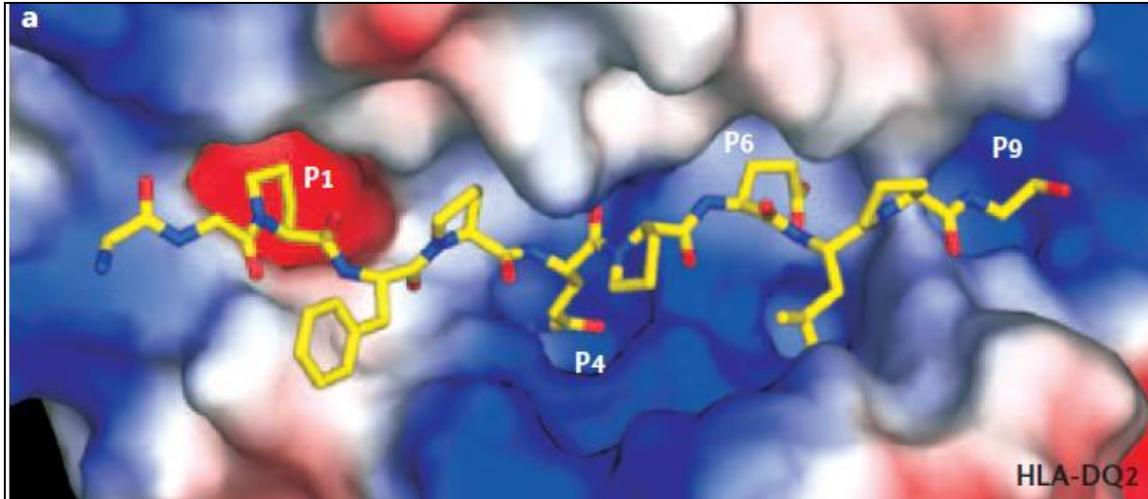
Figure 11: The HLA class II binding groove. A 3-dimentional model of the binding groove of an HLA class II molecule bound to an epitope. The side chains of the amino acid residues P1, 4, 6, and 9 are anchored in the pockets of the binding clefts. *Reprinted by permission from Macmillan Publishers Ltd: Nature Reviews Immunology [67], copyright 2006.*
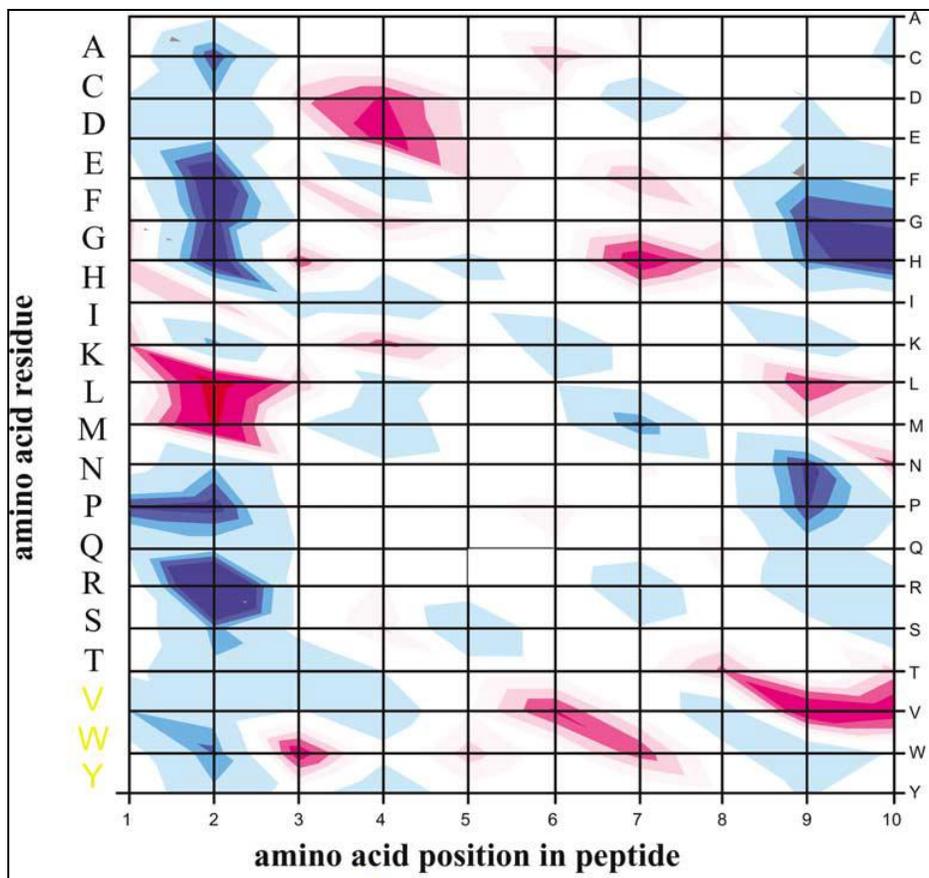
Figure 12: The EpiMatrix frequency matrix for the Class I HLA allele A*02:01. The colors in the grid represents the frequency that a given amino acid residue is observed at a specific amino acid position in a large set of natural HLA-A*02:01 epitopes. The color "Red" indicates the amino acid is frequently present at that position, and therefore promotes epitope binding to the HLA molecule when present. "Blue" indicates the opposite. *Reprinted with permission from [62].*
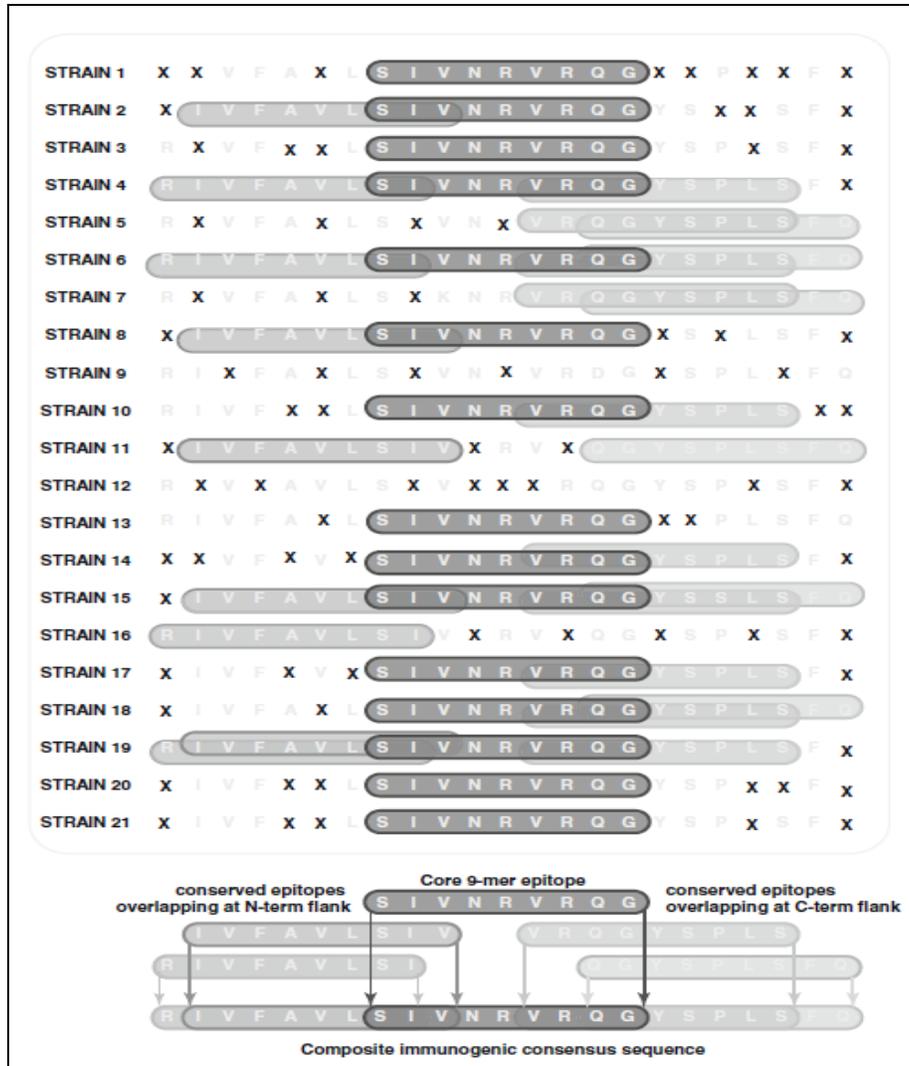
Figure 13: Construction of an Immunogenic Consensus Sequence peptide by EpiAssembler. First, EpiAssembler identifies the most highly conserved epitope across the analyzed sequences and the one that is most promiscuously immunogenic, which here is SIVNRVRQG (an X denotes a mismatch with the conserved peptide), the program then identifies overlapping highly conserved and promiscuously immunogenic peptides, and then assembles those peptides into an extended ICS. *Reprinted with permission from [23]*.
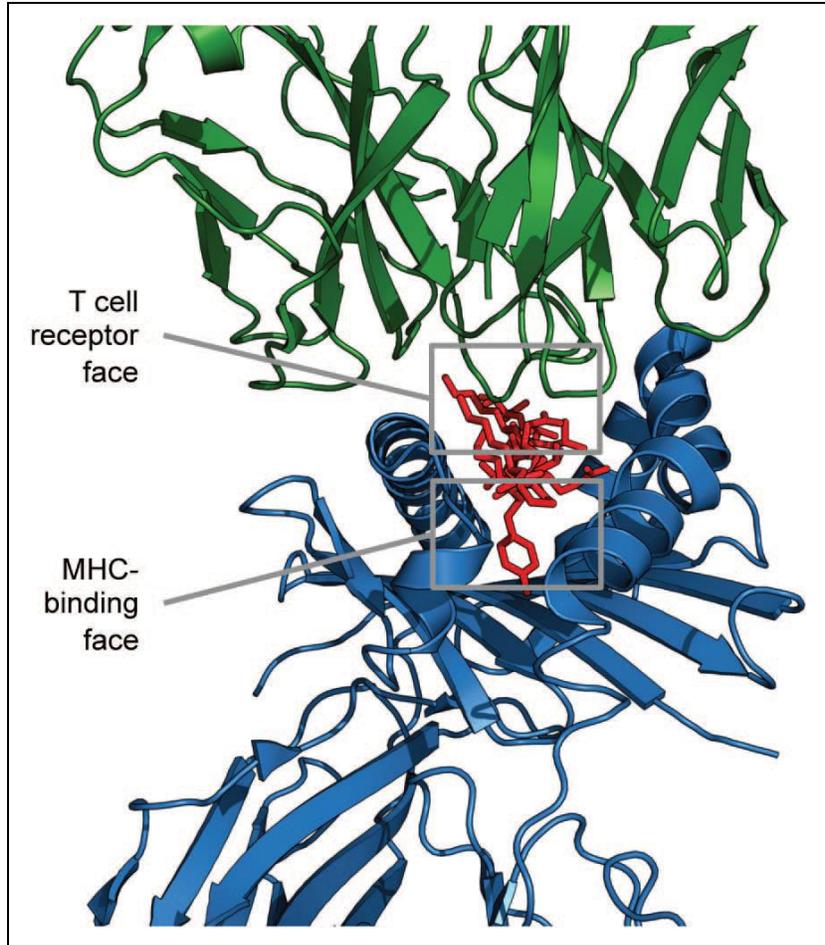
Figure 14: The two faces of the T-cell epitope. A 3D model of a T-cell epitope in association with an HLA molecule and a T-cell receptor, with the 'epitope' residues facing the TCR, and the 'agretope' residues facing the HLA molecule. *Reprinted with permission from [84].*
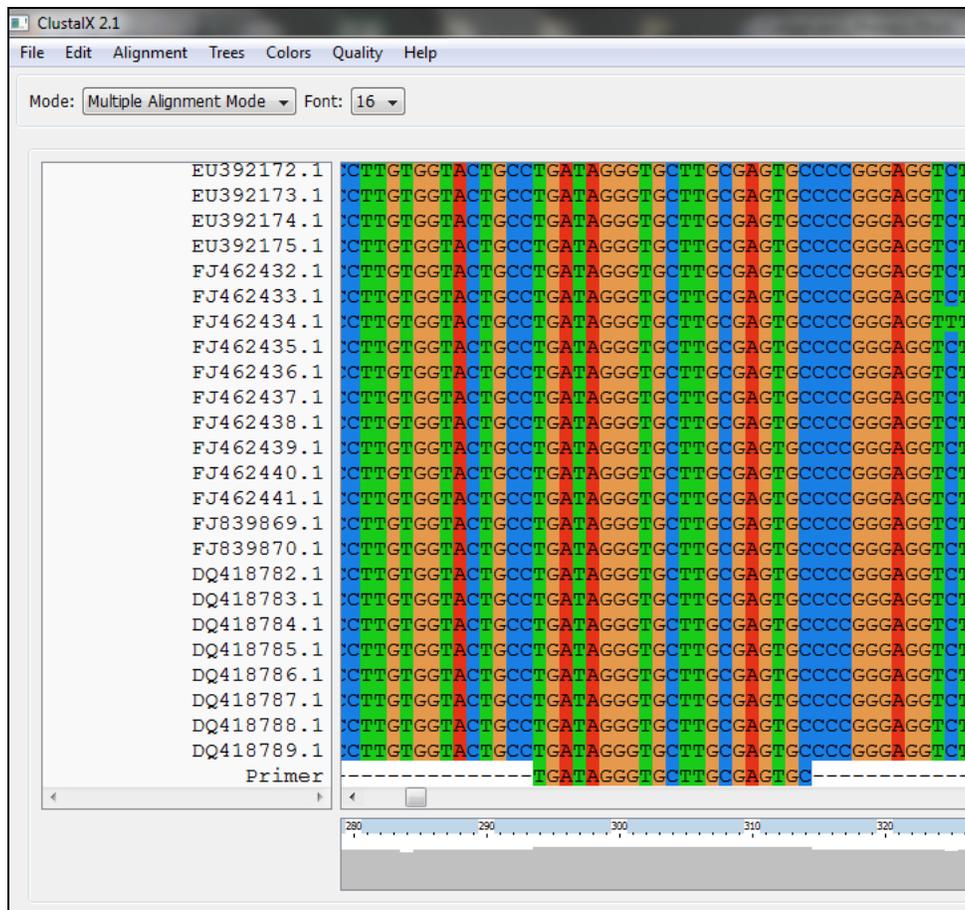
Figure 15: Primer Alignment. An example of the alignment of the primer A_F with the 46 HCV-4 sequences using the ClustalX 2.1 software.
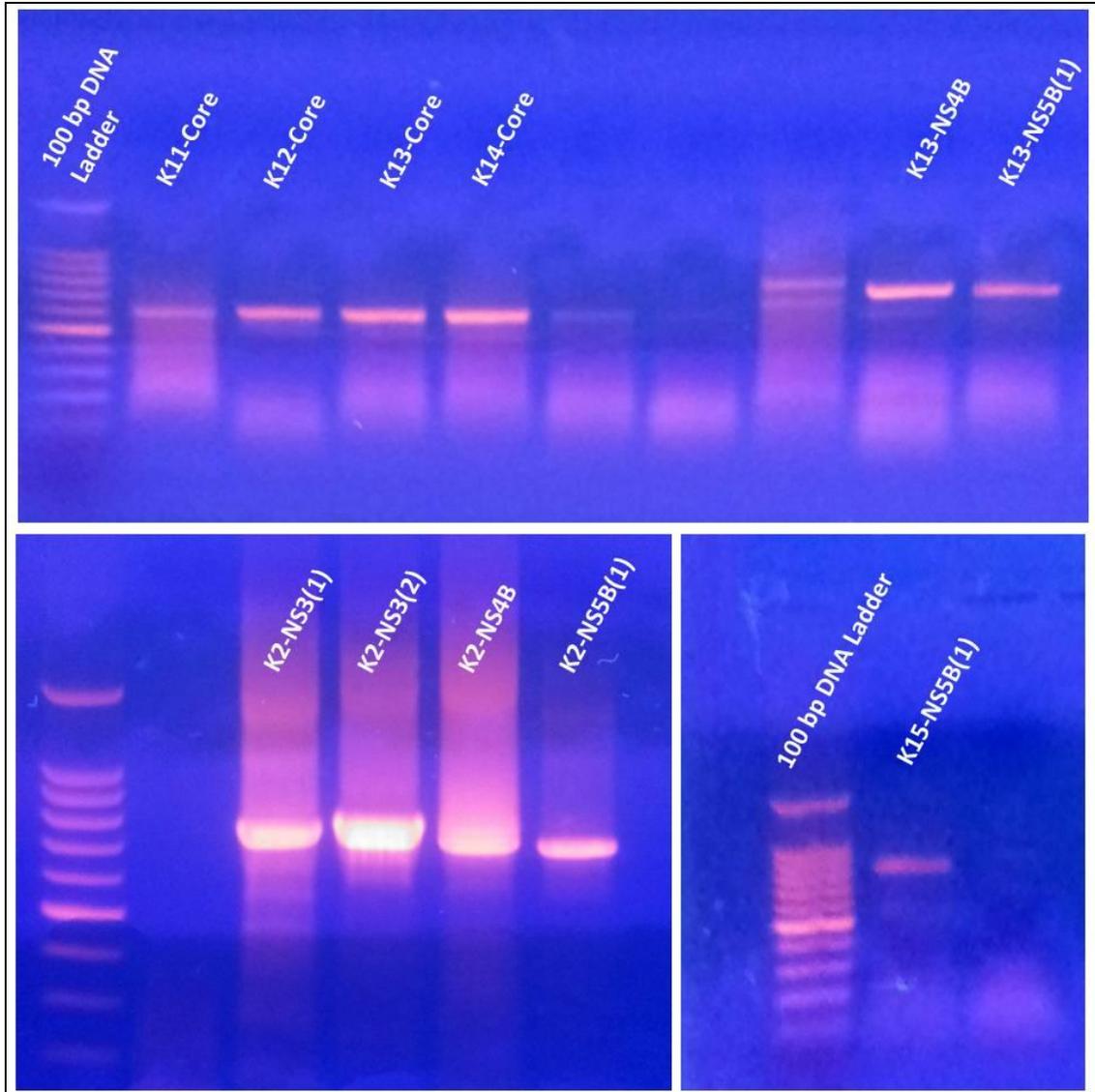
Figure 16: HCV RNA amplification results. Examples of electrophoresis gel photos presenting the results of the nested PCR amplification reactions for 11 sample regions are displayed. The results were analyzed on a 1% agarose gel, with 5 μL of a 100 bp DNA Ladder (Promega, WI, USA) loaded into the first well of each gel, and 5 μL of each of the sample regions amplicons added to the other wells; all mixed with 2 μL loading dye. The gel was then run at 95V for 1 hour in 1X TBE running buffer, and the amplicon bands were visualized on an ECX-20.M transilluminator (Vilber Lourmat, Marne-la-Vallée, France). Images of the gels were taken using a Samsung S2 8.0 megapixel mobile phone camera (Samsung, Seoul, South Korea).

Figure 17: Sequencing Chromatograms Qualities. An example of a high quality sequencing chromatogram is illustrated at the top (part of the sequencing chromatogram of sample region K2-NS3(1)-F) with sharp peaks and low background noise. An example of a low quality chromatogram is illustrated at the bottom (part of the sequencing chromatogram of sample region K2-NS4B-F) with high background noise.
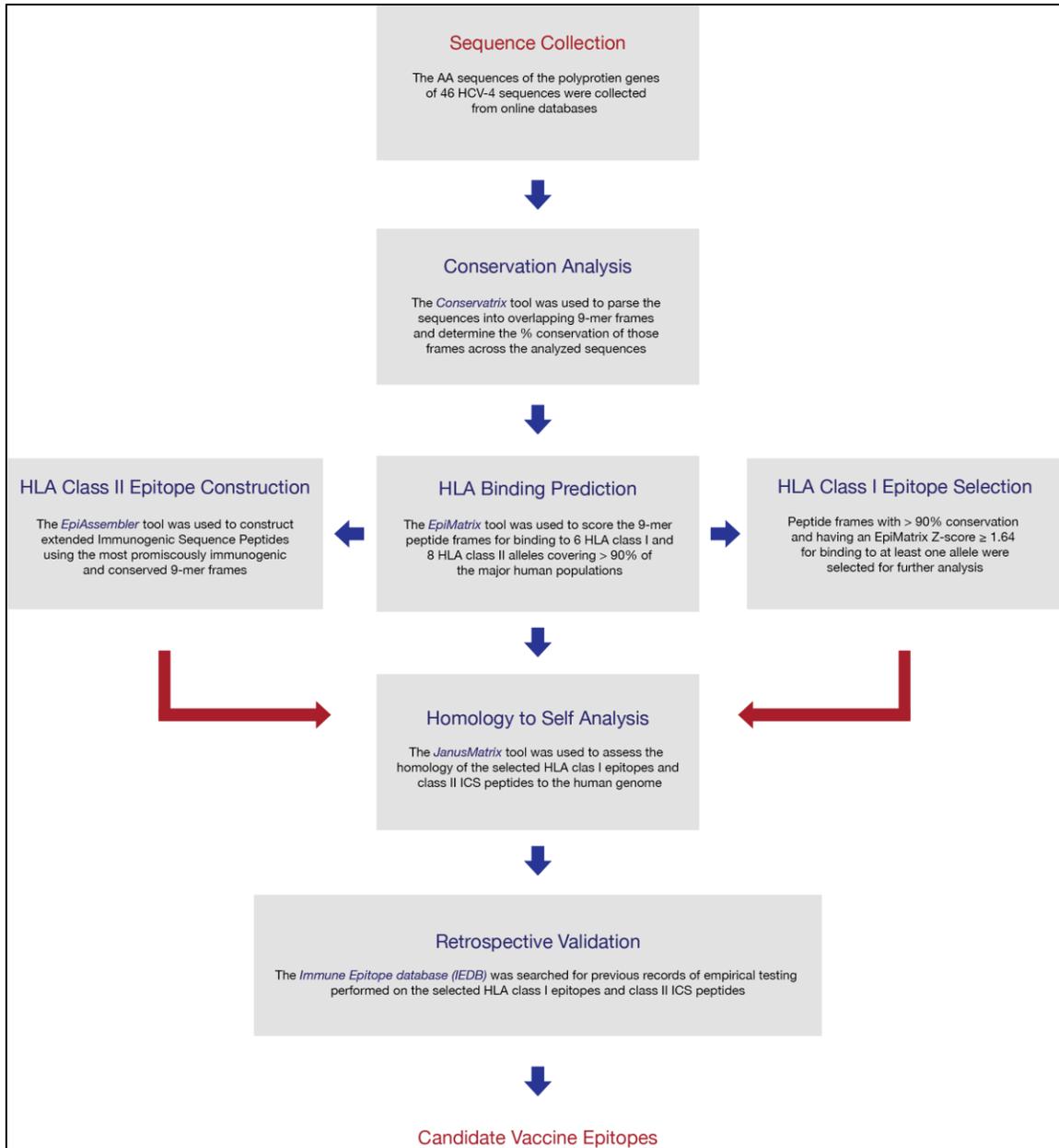
Figure 18: Flowchart summarizing the performed *in silico* T-cell epitope prediction and retrospective validation process.

10. REFERENCES

1. World Health Organization. Hepatitis C fact sheet No164. 2011. http://www.who.int/mediacentre/factsheets/ fs164/en/. Accessed 27 December 2013.

2. Simmonds P, Bukh J, Combet C, Deléage G, Enomoto N, Feinstone S, et al. Consensus proposals for a unified system of nomenclature of hepatitis C virus genotypes. Hepatology. 2005; 42:962-73.

3. Miller FD, Abu-Raddad LJ. Evidence of intense ongoing endemic transmission of hepatitis C virus in Egypt. Proc Natl Acad Sci U S A. 2010;107:14757-62.

4. Yahia M. Global health: a uniquely Egyptian epidemic. Nature. 2011; 474:S12-3.

5. Sievert W, Altraif I, Razavi HA, Abdo A, Ahmed EA, Alomair A, et al. A systematic review of hepatitis C virus epidemiology in Asia, Australia and Egypt. Liver Int. 2011; 31:61-80.

6. Khattab MA, Ferenci P, Hadziyannis SJ, Colombo M, Manns MP, Almasio PL, et al. Management of hepatitis C virus genotype 4: Recommendations of An International Expert Panel. J Hepatol. 2011; 54:1250-62.

7. Pybus OG, Drummond AJ, Nakano T, Robertson BH, Rambaut A. The epidemiology and iatrogenic transmission of hepatitis C virus in Egypt: a Bayesian coalescent approach. Mol Biol Evol. 2003; 20:381-7.

8. Shepard CW, Finelli L, Alter MJ. Global epidemiology of hepatitis C virus infection. Lancet Infect Dis. 2005; 5:558-67.

9. Centers for Disease Control and Prevention. Hepatitis C information for health professionals. 2011. http://www.cdc.gov/hepatitis/HCV/index.htm. Accessed 28 Dec 2013.

10. Ghany MG, Strader DB, Thomas DL, Seeff LB, American Association for the Study of Liver Diseases. Diagnosis, management, and treatment of hepatitis C: an update. Hepatology. 2009;49:1335–74.

11. Sharma SD. Hepatitis C, virus: molecular biology and current therapeutic options. Indian J Med Res. 2010;131:17–34.

12. Moradpour D, Penin F, Rice CM. Replication of hepatitis C virus. Nat Rev Microbiol. 2007; 5:453-63.

13. Ishii S, Koziel MJ. Immune responses during acute and chronic infection with hepatitis C virus. Clin Immunol. 2008; 128:133-47.

14. Rehermann B. Hepatitis C virus versus innate and adaptive immune responses: A tale of coevolution and coexistence. J Clin Invest. 2009; 119:1745-54.

15. Zeisel MB, Cosset FL, Baumert TF. Host neutralizing responses and pathogenesis of hepatitis C virus infection. Hepatology. 2008; 48:299-307.

16. Ghany MG, Nelson DR, Strader DB, Thomas DL, Seeff LB; American Association for Study of Liver Diseases. An Update on Treatment of Genotype 1 Chronic

Hepatitis C Virus Infection: 2011 Practice Guideline by the American Association for the Study of Liver Diseases. Hepatology. 2011; 54:1433-44.

17. Halliday J, Klenerman P, Barnes E. Vaccination for hepatitis C virus: Closing in on an evasive target. Expert Rev Vaccines. 2011; 10:659-72.

18. Heller T, Saito S, Auerbach J, Williams T, Moreen TR, Jazwinski A, et al. An in vitro model of hepatitis C virion production. Proc Natl Acad Sci U S A. 2005; 102:2579-83.

19. Yu CI, Chiang BL. A new insight into hepatitis C vaccine development. J Biomed Biotechnol. 2010; 2010:548280.

20. Li C, Lu L, Wu X, Wang C, Bennett P, Lu T, Murphy D. Complete genomic sequences for hepatitis C virus subtypes 4b, 4c, 4d, 4g, 4k, 4l, 4m, 4n, 4o, 4p, 4q, 4r and 4t. J Gen Virol. 2009; 90:1820-6.

21. Peters B, Sidney J, Bourne P, Bui HH, Buus S, Doh G, et al. The immune epitope database and analysis resource: from vision to blueprint. PLoS Biol. 2005; 3:e91.

22. Kim Y, Vaughan K, Greenbaum J, Peters B, Law M, Sette A. A Meta-Analysis of the Existing Knowledge of Immunoreactivity against Hepatitis C Virus (HCV). PLoS One. 2012; 7:e38028.

23. De Groot, AS. Epitope-based immunome-derived vaccines: A strategy for improved design and safety. In: Falus A. Clinical Applications of Immunomics. NY, USA: Springer US. 2009:39-69.

24. Burlone ME, Budkowska A. Hepatitis C virus cell entry: role of lipoproteins and cellular receptors. J Gen Virol. 2009; 90:1055-70.

25. Chevaliez S, Pawlotsky J. HCV Genome and Life Cycle. In: Tan SL. Hepatitis C Viruses: Genomes and Molecular Biology. Norfolk, UK: Horizon Bioscience. 2006: 5-47.

26. Lindenbach BD, Rice CM. Unravelling hepatitis C virus replication from genome to function. Nature. 2005; 436:933-8.

27. Borden EC, Sen GC, Uze G, Silverman RH, Ransohoff RM, Foster GR, Stark GR. Interferons at age 50: past, current and future impact on biomedicine. Nat Rev Drug Discov. 2007; 6:975-90.

28. Uzé G, Monneron D. IL-28 and IL-29: Newcomers to the interferon family. Biochimie. 2007; 89:729-34.

29. Li K, Lemon SM. Innate immune responses in hepatitis C virus infection. Semin Immunopathol. 2013; 35:53-72.

30. Lemon, S.M. Induction and evasion of innate antiviral responses by hepatitis C virus. J Biol Chem. 2010; 285:22741-7.

31. Abbas AK, Lichtman AH. Effector Mechanisms of Immune Responses. In: Abbas AK, Lichtman AH. Cellular and Molecular Immunology. 5th ed. PA, USA: Saunders. 2005:241-366.

32. Mak TW, Saunders ME. Bridging Innate and Adaptive Immunity: NK, γδT, and NKT Cells. In: Mak TW, Saunders ME. The Immune Response: Basic and Clinical Principles. MA, USA: Elservier Academic Press. 2006:517-549.

33. Boehm U, Klamp T, Groot M, Howard JC. Cellular responses to interferon-gamma. Annu Rev Immunol. 1997; 15:749-95.

34. Zeromski J, Mozer-Lisewska I, Kaczmarek M, Kowala-Piaskowska A, Sikora J. NK cells prevalence, subsets and function in viral hepatitis C. Arch Immunol Ther Exp (Warsz). 2011; 59:449-55.

35. Bozzano F, Marras F, Biassoni R, Maria AD. Natural killer cells in hepatitis C virus infection. Expert Rev Clin Immunol. 2012; 8:775-88.

36. Abbas AK, Lichtman AH. Recognition of Antigens. In: Abbas AK, Lichtman AH. Cellular and Molecular Immunology. 5th ed. PA, USA: Saunders. 2005:41-125.

37. Abbas AK, Lichtman AH. Introduction to Immunology. In: Abbas AK, Lichtman AH. Cellular and Molecular Immunology. 5th ed. PA, USA: Saunders. 2005:3-39.

38. Mak TW, Saunders ME. Antigen Processing and Presentation. In: Mak TW, Saunders ME. The Immune Response: Basic and Clinical Principles. MA, USA: Elservier Academic Press. 2006:279-309.

39. Losikoff PT, Self AA, Gregory SH. Dendritic cells, regulatory T cells and the pathogenesis of chronic hepatitis C. Virulence. 2012; 3:610-620.

40. Albert ML, Decalf J, Pol S. Plasmacytoid dendritic cells move down on the list of suspects: In search of the immune pathogenesis of chronic hepatitis C. J Hepatol. 2008; 49:1069-78.

41. Brenndörfer ED, Sällberg M. Hepatitis C Virus-Mediated Modulation of Cellular Immunity. Arch Immunol Ther Exp (Warsz). 2012; 60:315-29.

42. Abbas AK, Lichtman AH. Maturation, Activation, and Regulation of Lymphocytes. In: Abbas AK, Lichtman AH. Cellular and Molecular Immunology. 5th ed. PA, USA: Saunders. 2005:127-240.

43. Bertolino P, McCaughan GW, Bowen DG. Immunological Parameters Influencing Adaptive Immune Responses to the Hepatitis C Virus. In: Jirillo E. Hepatitis C Virus Disease: Immunobiology and Clinical Applications. NY, USA: Springer New York. 2008:39-70.

44. Petrovic D, Dempsey E, Doherty DG, Kelleher D, Long A. Hepatitis C virus--T-cell responses and viral escape mutations. Eur J Immunol. 2012; 42:17-26.

45. Wang Y, Keck ZY, Foung SK. Neutralizing antibody response to hepatitis C virus. Viruses. 2011; 3:2127-45.

46. Nattermann J, Nischalke HD, Hofmeister V, Ahlenstiel G, Zimmermann H, Leifeld L, Weiss EH, Sauerbruch T, Spengler U. The HLA-A2 Restricted T Cell Epitope HCV Core35–44 Stabilizes HLA-E Expression and Inhibits Cytolysis Mediated by Natural Killer Cells. Am J Pathol. 2005; 166:443-53.

47. Nattermann J, Nischalke HD, Hofmeister V, Ahlenstiel G, Zimmermann H, Leifeld L, Weiss EH, Sauerbruch T, Spengler U. Upregulation of Major Histocompatibility Complex Class I on Liver Cells by Hepatitis C Virus Core Protein via p53 and TAP1 Impairs Natural Killer Cell Cytotoxicity. J Virol. 2003; 77:8299-309.

48. Ward S, Lauer G, Isba R, Walker B, Klenerman P. Cellular immune responses against hepatitis C virus: the evidence base 2002. Clin Exp Immunol. 2002; 128:195-203.

49. Wherry EJ. T cell exhaustion. Nat Immunol. 2011; 12:492-9.

50. Zeisel MB, Fafi-Kremer S, Robinet E, Habersetzer F, Baumert TF, Stoll-Keller F. Adaptive immunity to hepatitis C virus. Viruses. 2009; 1:276-97.

51. Vignali DA, Collison LW, Workman CJ. How regulatory T cells work. Nat Rev Immunol. 2008; 8:523-32.

52. Corthay, A. How do Regulatory T Cells Work? Scand J Immunol. 2009; 70:326-36.

53. Cusick MF, Schiller JJ, Gill JC, Eckels DD. Hepatitis C virus induces regulatory T cells by naturally occurring viral variants to suppress T cell responses. Clin Dev Immunol. 2011; 2011:806061.

54. Di Lorenzo C, Angus AG, Patel AH., Hepatitis C virus evasion mechanisms from neutralizing antibodies. Viruses. 2011; 3:2280-300.

55. Helle F, Duverlie G, Dubuisson J. The hepatitis C virus glycan shield and evasion of the humoral immune response. Viruses. 2011; 3:1909-32.

56. Bartosch B, Dubuisson J, Cosset FL. Infectious hepatitis C virus pseudo-particles containing functional E1-E2 envelope protein complexes. J Exp Med. 2003; 197:633-42.

57. Major ME. Prophylactic and Therapeutic Vaccination against Hepatitis C Virus (HCV): Developments and Future Perspectives. Viruses. 2009; 1:144-65.

58. Vivona S, Gardy JL, Ramachandran S, Brinkman FS, Raghava GP, Flower DR, Filippini F. Computer-aided biotechnology: from immuno-informatics to reverse vaccinology. Trends Biotechnol. 2008; 26:190-200.

59. He Y, Rappuoli R, De Groot AS, Chen RT. Emerging vaccine informatics. J Biomed Biotechnol. 2010; 2010:218590.

60. De Groot AS. Immunomics: discovering new targets for vaccines and therapeutics. Drug Discov Today. 2006; 11:203-9.

61. Sirskyj D, Diaz-Mitoma F, Golshani A, Kumar A, Azizi A. Innovative bioinformatic approaches for developing peptide-based vaccines against hypervariable viruses. Immunol Cell Biol. 2011; 89:81-9.

62. Martin W, Sbai H, De Groot AS. Bioinformatics tools for identifying class I-restricted epitopes. Methods. 2003 Mar;29(3):289-98.

63. Sturniolo T, Bono E, Ding J, et al. Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. Nat Biotechnol. 1999; 17:555-61.

64. Mak TW, Saunders ME. MHC: The Major Histocompatibility Complex. In: Mak TW, Saunders ME. The Immune Response: Basic and Clinical Principles. MA, USA: Elservier Academic Press. 2006:247-277.

65. Wang P, Sidney J, Dow C, Mothé B, Sette A, Peters B. A Systematic Assessment of MHC Class II Peptide Binding Predictions and Evaluation of a Consensus Approach. PLoS Comput Biol. 2008; 4:e1000048.

66. Godkin AJ, Smith KJ, Willis A, Tejada-Simon MV, Zhang J, Elliott T, Hill AV. Naturally processed HLA class II peptides reveal highly conserved immunogenic flanking region sequence preferences that reflect antigen processing rather than peptide-MHC interactions. J Immunol. 2001; 166:6720-7.

67. Jones EY, Fugger L, Strominger JL, Siebold C. MHC class II proteins and disease: a structural perspective. Nat Rev Immunol. 2006; 6:271-82.

68. Sidney J, Peters B, Frahm N, Brander C, Sette A. HLA class I supertypes: a revised and updated classification. BMC Immunol. 2008; 9:1.

69. Erlich H. HLA DNA typing: past, present, and future. Tissue Antigens. 2012; 80:1-11.

70. Dunn PP. Human leucocyte antigen typing: techniques and technology, a critical appraisal. Int J Immunogenet. 2011; 38:463-73.

71. HLA Nomeclature. Anthony Nolan Research Institute. 2011. http://hla.alleles.org/antigens/index.html. Accessed 27 Dec 2013.

72. Society for Biomedical Diabetes Research. Genotype, Serotype, and Supertype classification.2009. http://www.socbdr.org/rds/authors/unit_tables_conversions_and_genetic_dictionaries/e5220/. Accessed 27 Dec 2013.

73. Nunes E, Heslop H, Fernandez-Vina M, Taves C, Wagenknecht DR, Eisenbrey AB, et al. Definitions of histocompatibility typing terms. Blood. 2011; 118:e180-3.

74. Marsh SG, Albert ED, Bodmer WF, Bontrop RE, Dupont B, Erlich HA, et al. Nomenclature for factors of the HLA system, 2010. Tissue Antigens. 2010; 75:291-455.

75. Sette A, Sidney J. HLA supertypes and supermotifs: a functional perspective on HLA polymorphism. Curr Opin Immunol. 1998; 10:478-82.

76. Sidney J, Peters B, Frahm N, Brander C, Sette A. HLA class I supertypes: a revised and updated classification. BMC Immunol. 2008; 9:1.

77. Kuniholm MH, Anastos K, Kovacs A, Gao X, Marti D, Sette A, et al. Relation of HLA class I and II supertypes with spontaneous clearance of hepatitis C virus. Genes Immun. 2013; 14:330-5.

78. Sette A, Sidney J. Nine major HLA class I supertypes account for the vast preponderance of HLA-A and -B polymorphism. Immunogenetics. 1999; 50:201-12.

79. Tong JC, Tan TW, Ranganathan S. In silico grouping of peptide/HLA class I complexes using structural interaction characteristics. Bioinformatics. 2007; 23:177-83.

80. Lund O, Nielsen M, Kesmir C, et al. Definition of supertypes for HLA molecules using clustering of specificity matrices. Immunogenetics. 2004; 55:797-810.

81. Hertz T, Yanover C. Identifying HLA supertypes by learning distance functions. Bioinformatics. 2007; 23:e148-55.

82. EpiVax Inc. iVAX Web-based Vaccine Design – Immunoinformatics Tools. 2010. http://www.epivax.com/vaccine-design-redesign/ivax-web-based-vaccine-design/. Accessed 27 Dec 2013.

83. De Groot AS, Bishop EA, Khan B, Lally M, Marcon L, Franco J, Mayer KH, Carpenter CC, Martin W. Engineering immunogenic consensus T helper epitopes for a cross-clade HIV vaccine. Methods. 2004; 34:476-87.

84. Moise L, Gutierrez AH, Bailey-Kellogg C, Terry F, Leng Q, Abdel Hady KM, Verberkmoes NC, et al. The two-faced T cell epitope: Examining the host-microbe interface with JanusMatrix. Hum Vaccin Immunother. Hum Vaccin Immunother. 2013; 9:1577-86.

85. Rudolph MG, Stanfield RL, Wilson IA. How TCRs Bind MHCs, Peptides, and Coreceptors. Annu Rev Immunol. 2006; 24:419-66.

86. Kuiken C, Yusim K, Boykin L, Richardson R. The Los Alamos hepatitis C sequence database. Bioinformatics. 2005; 21:379-84.

87. Timm J, Neukamm M, Kuntzen T, Kim AY, Chung RT, Brander C, Lauer GM, Walker BD, Allen TM. Characterization of full-length hepatitis C virus genotype 4 sequences. J Viral Hepat. 2007; 14:330-7.

88. Franco S, Tural C, Clotet B, Martínez MA. Complete nucleotide sequence of genotype 4 hepatitis C viruses isolated from patients co-infected with human immunodeficiency virus type 1. Virus Res. 2007; 123:161-9.

89. Hmaied F, Legrand-Abravanel F, Nicot F, Garrigues N, Chapuy-Regaud S, Dubois M, Njouom R, Izopet J, Pasquier C. Full-length genome sequences of hepatitis C virus subtype 4f. J Gen Virol. 2007; 88:2985-90.

90. Kuntzen T, Berical A, Ndjomou J, Bennett P, Schneidewind A, Lennon N, et al. A set of reference sequences for the hepatitis C genotypes 4d, 4f, and 4k covering the full open reading frame. J Med Virol. 2008; 80:1370-8.

91. Koletzki D, Dumont S, Vermeiren H, Peixe P, Nina J, Camacho RJ, Stuyver LJ. Full genome sequence of three isolates of hepatitis C virus subtype 4b from Portugal. Arch Virol. 2009; 154:127-32.

92. Gottwein JM, Scheel TK, Callendret B, Li YP, Eccleston HB, Engle RE, et al. Novel infectious cDNA clones of hepatitis C virus genotype 3a (strain S52) and 4a (strain ED43): genetic analyses and in vivo pathogenesis studies. J Virol. 2010; 84:5277-93.

93. Demetriou VL, Kostrikis LG. Near-full genome characterization of unclassified hepatitis C virus strains relating to genotypes 1 and 4. J Med Virol. 2011 Dec;83(12):2119-27.

94. Chamberlain RW, Adams N, Saeed AA, Simmonds P, Elliott RM. Complete nucleotide sequence of a type 4 hepatitis C virus variant, the predominant genotype in the Middle East. J Gen Virol. 1997; 78:1341-7.

95. Southwood S, Sidney J, Kondo A, del Guercio MF, Appella E, Hoffman S, Kubo RT, Chesnut RW, Grey HM, Sette A. Several common HLA-DR types share largely overlapping peptide binding repertoires. J Immunol. 1998; 160:3363-73.

96. Yao E, Tavis JE; Virahep-C Study Group. A general method for nested RT-PCR amplification and sequencing the complete HCV genotype 1 open reading frame. Virol J. 2005; 2:88.

97. Schadt EE, Turner S, Kasarskis A. A window into third-generation sequencing. Hum Mol Genet. 2010; 19:R227-40.

98. Sievers F, Higgins DG. Clustal omega, accurate alignment of very large numbers of sequences. Methods Mol Biol. 2014; 1079:105-16.

99. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. Bioinformatics. 2007; 23:2947-8.

100. Premier Biosoft. PCR Primer Design Guidelines. 2013. http://www.premierbiosoft.com/tech_notes/PCR_Primer_Design.html. Accessed 27 Dec 2013.

101. Neumann-Haefelin C, Killinger T, Timm J, Southwood S, McKinney D, Blum HE, Thimme R. Absence of viral escape within a frequently recognized HLA-A26-restricted CD8+ T-cell epitope targeting the functionally constrained hepatitis C virus NS5A/5B cleavage site. J Gen Virol. 2007; 88:1986-91.

102. Wentworth PA, Sette A, Celis E, Sidney J, Southwood S, Crimi C, et al. Identification of A2-restricted hepatitis C virus-specific cytotoxic T lymphocyte epitopes from conserved regions of the viral genome. Int Immunol. 1996; 8:651-9.

103. Grey HM, Ruppert J, Vitiello A, Sidney J, Kast WM, Kubo RT, Sette A. Class I MHC-peptide interactions: structural requirements and functional implications. Cancer Surv.; 22:37-49.

104. Ruppert J, Sidney J, Celis E, Kubo RT, Grey HM, Sette A. Sette Prominent role of secondary anchor residues in peptide binding to HLA-A2.1 molecules. Cell. 1993; 74:929-37.

105. Mashiba T, Udaka K, Hirachi Y, Hiasa Y, Miyakawa T, Satta Y, Osoda T, Kataoka S, Kohara M, Onji M. Identification of CTL epitopes in hepatitis C virus by a genome-wide computational scanning and a rational design of peptide vaccine. Immunogenetics. Immunogenetics. 2007; 59:197-209.

106. Bertoni R, Sette A, Sidney J, Guidotti LG, Shapiro M, Purcell R, Chisari FV. Human class I supertypes and CTL repertoires extend to chimpanzees. J Immunol. 1998; 161:4447-55.

107. Alexander J, Del Guercio MF, Fikes JD, Chesnut RW, Chisari FV, Chang KM, Appella E, Sette A. Recognition of a novel naturally processed, A2 restricted, HCV-NS4 epitope triggers IFN-gamma release in absence of detectable cytopathicity. Hum Immunol. 1998; 59:776-82.

108. Scognamiglio P, Accapezzato D, Casciaro MA, Cacciani A, Artini M, Bruno G, et al. Presence of effector CD8+ T cells in hepatitis C virus-exposed healthy seronegative donors. J Immunol. 1999; 162:6681-9.

109. Sidney J, Southwood S, Mann DL, Fernandez-Vina MA, Newman MJ, Sette A. Majority of peptides binding HLA-A*0201 with high affinity crossreact with other A2-supertype molecules. Hum Immunol. 2001; 62:1200-16.

110. Himoudi N, Abraham JD, Fournillier A, Lone YC, Joubert A, Op De Beeck A, et al. Comparative vaccine studies in HLA-A2.1-transgenic mice reveal a clustered organization of epitopes presented in hepatitis C virus natural infection. J Virol. 2002; 76:12735-46.

111. Battegay M, Fikes J, Di Bisceglie AM, Wentworth PA, Sette A, Celis E, et al. Patients with chronic hepatitis C have circulating cytotoxic T cells which recognize hepatitis C virus-encoded peptides binding to HLA-A2.1 molecules. J Virol. 1995; 69:2462-70.

112. Ohno S, Moriya O, Yoshimoto T, Hayashi H, Akatsuka T, Matsui M. Immunogenic variation between multiple HLA-A*0201-restricted, Hepatitis C Virus-derived epitopes for cytotoxic T lymphocytes. Viral Immunol. 2006; 19:458-67.

113. Ishizuka J, Grebe K, Shenderov E, Peters B, Chen Q, Peng Y, et al. Quantitating T cell cross-reactivity for unrelated peptide antigens. J Immunol. 2009; 183:4337-45.

114. Neumann-Haefelin C, Timm J, Spangenberg HC, Wischniowski N, Nazarova N, Kersting N, Roggendorf M, Allen TM, Blum HE, Thimme R. Virological and immunological determinants of intrahepatic virus-specific CD8+ T-cell failure in chronic hepatitis C virus infection. Hepatology. 2008; 47:1824-36.

115. Tsai SL, Sheen IS, Chien RN, Chu CM, Huang HC, Chuang YL, et al. Activation of Th1 immunity is a common immune mechanism for the successful treatment of hepatitis B and C: tetramer assay and therapeutic implications. J Biomed Sci. 2003; 10:120-35.

116. Tsai SL, Lee TH, Chien RN, Liao SK, Lin CL, Kuo GC, Liaw YF. A method to increase tetramer staining efficiency of CD8+ T cells with MHC-peptide complexes: therapeutic applications in monitoring cytotoxic T lymphocyte activity during hepatitis B and C treatment. J Immunol Methods. 2004; 285:71-87.

117. Thammanichanond D, Moneer S, Yotnda P, Aitken C, Earnest-Silveira L, Jackson D, Hellard M, McCluskey J, Torresi J, Bharadwaj M. Fiber-modified recombinant

adenoviral constructs encoding hepatitis C virus proteins induce potent HCV-specific T cell response. Clin Immunol. 2008; 128:329-39.

118. Guo Z, Zhang H, Rao H, Jiang D, Cong X, Feng B, Wang J, Wei L, Chen H. DCs pulsed with novel HLA-A2-restricted CTL epitopes against hepatitis C virus induced a broadly reactive anti-HCV-specific T lymphocyte response. PLoS One. 2012;7:e38390.

119. Anthony DD, Valdez H, Post AB, Carlson NL, Heeger PS, Lehmann PV. Comprehensive determinant mapping of the hepatitis C-specific CD8 cell repertoire reveals unpredicted immune hierarchy. Clin Immunol. 2002; 103:264-76.

120. Fitzmaurice K, Petrovic D, Ramamurthy N, Simmons R, Merani S, Gaudieri S, et al. Molecular footprints reveal the impact of the protective HLA-A*03 allele in hepatitis C virus infection. Gut. 2011; 60:1563-71.

121. Nakamoto Y, Kaneko S, Takizawa H, Kikumoto Y, Takano M, Himeda Y, Kobayashi K. Analysis of the CD8-positive T cell response in Japanese patients with chronic hepatitis C using HLA-A*2402 peptide tetramers. J Med Virol. 2003; 70:51-61.

122. Kaji K, Nakamoto Y, Kaneko S. Analysis of hepatitis C virus-specific CD8+ T-cells with HLA-A*24 tetramers during phlebotomy and interferon therapy for chronic hepatitis C. Oncol Rep. 2007; 18:993-8.

123. Gruener NH, Jung MC, Ulsenheimer A, Gerlach JT, Zachoval R, Diepolder HM, Baretton G, Schauer R, Pape GR, Schirren CA. Analysis of a successful HCV-specific CD8+ T cell response in patients with recurrent HCV-infection after orthotopic liver transplantation. Liver Transpl. 2004; 10:1487-96.

124. Christie JM, Chapel H, Chapman RW, Rosenberg WM. Immune selection and genetic sequence variation in core and envelope regions of hepatitis C virus. Hepatology. 1999; 30:1037-44.

125. Lauer GM, Barnes E, Lucas M, Timm J, Ouchi K, Kim AY, Day CL, et al. High resolution analysis of cellular immune responses in resolved and persistent hepatitis C virus infection. Gastroenterology. 2004; 127:924-36.

126. Cucchiarini M, Kammer AR, Grabscheid B, Diepolder HM, Gerlach TJ, Grüner N, Santantonio T, Reichen J, Pape GR, Cerny A. Vigorous peripheral blood cytotoxic T cell response during the acute phase of hepatitis C virus infection. Cell Immunol. 2000; 203:111-23.

127. Grüner NH, Gerlach TJ, Jung MC, Diepolder HM, Schirren CA, Schraut WW, et al. Association of hepatitis C virus-specific CD8+ T cells with viral clearance in acute hepatitis C. J Infect Dis. 2000; 181:1528-36.

128. Lauer GM, Ouchi K, Chung RT, Nguyen TN, Day CL, Purkis DR, Reiser M, Kim AY, Lucas M, Klenerman P, Walker BD. Comprehensive analysis of CD8(+)-T-cell responses against hepatitis C virus reveals multiple unpredicted specificities. J Virol. 2002; 76:6104-13.

129. Diepolder HM, Gerlach JT, Zachoval R, Hoffmann RM, Jung MC, Wierenga EA, et al. Immunodominant CD4+ T-cell epitope within nonstructural protein 3 in acute hepatitis C virus infection. J Virol. 1997; 71:6011-9.

130. Pape GR, Gerlach TJ, Diepolder HM, Grüner N, Jung M, Santantonio T. Role of the specific T-cell response for clearance and control of hepatitis C virus. J Viral Hepat. 1999; 6:36-40.

131. Shoukry NH, Sidney J, Sette A, Walker CM. Conserved hierarchy of helper T cell responses in a chimpanzee during primary and secondary hepatitis C virus infections. J Immunol. 2004; 172:483-92.

132. Day CL, Seth NP, Lucas M, Appel H, Gauthier L, Lauer GM, et al. Ex vivo analysis of human memory CD4 T cells specific for hepatitis C virus using MHC class II tetramers. J Clin Invest. 2003; 112:831-42.

133. Castelli FA, Leleu M, Pouvelle-Moratille S, Farci S, Zarour HM, Andrieu M, Auriault C, Ménez A, Georges B, Maillere B. Differential capacity of T cell priming in naive donors of promiscuous CD4+ T cell epitopes of HCV NS3 and Core proteins. Eur J Immunol. 2007; 37:1513-23.

134. Lamonaca V, Missale G, Urbani S, Pilli M, Boni C, Mori C, et al. Conserved hepatitis C virus sequences are highly immunogenic for CD4(+) T cells: implications for vaccine development. Hepatology. 1999; 30:1088-98.

135. Schulze Zur Wiesch J, Sidney J, Walker B, Sette A. Direct Submission. 2006.

136. Schulze zur Wiesch J, Lauer GM, Day CL, Kim AY, Ouchi K, Duncan JE, et al. Walker Broad repertoire of the CD4+ Th cell response in spontaneously controlled hepatitis C virus infection includes dominant and highly promiscuous epitopes. J Immunol. 2005; 175:3603-13.

137. Harcourt GC, Lucas M, Sheridan I, Barnes E, Phillips R, Klenerman P. Longitudinal mapping of protective CD4+ T cell responses against HCV: analysis of fluctuating dominant and subdominant HLA-DR11 restricted epitopes. J Viral Hepat. 2004; 11:324-31.

138. Ebinuma H, Nakamoto N, Li Y, Price DA, Gostick E, Levine BL, Tobias J, Kwok WW, Chang KM. Identification and in vitro expansion of functional antigen-specific CD25+ FoxP3+ regulatory T cells in hepatitis C virus infection. J Virol. 2008; 82:5043-53.

139. Eckels DD, Bian T, Gill JC, Sønderstrup G. Epitopes of the NS3 protein of hepatitis C virus: recognition in HLA-DR4 transgenic mice. Immunol Cell Biol. 2002; 80:106-12.

140. Gerlach JT, Ulsenheimer A, Grüner NH, Jung MC, Schraut W, Schirren CA, et al. Minimal T-cell-stimulatory sequences and spectrum of HLA restriction of immunodominant CD4+ T-cell epitopes within hepatitis C virus NS3 and NS4 proteins. J Virol. 2005; 79:12425-33.

141. Vertuani S, Bazzaro M, Gualandi G, Micheletti F, Marastoni M, Fortini C, et al. Effect of interferon-alpha therapy on epitope-specific cytotoxic T lymphocyte responses in hepatitis C virus-infected individuals. Eur J Immunol. 2002; 32:144-54.

142. Sarobe P, Huarte E, Lasarte JJ, López-Díaz de Cerio A, García N, Borrás-Cuesta F, Prieto J. Characterization of an immunologically conserved epitope from hepatitis C virus E2 glycoprotein recognized by HLA-A2 restricted cytotoxic T lymphocytes. J Hepatol. 2001; 34:321-9.

143. Schweitzer S, Schneiders AM, Langhans B, Kraas W, Jung G, Vidalin O, Inchauspe G, Sauerbruch T, Spengler U. Flow cytometric analysis of peptide binding to major histocampatibility complex class I for hepatitis C virus core T-cell epitopes. Cytometry. 2000; 41:271-8.

144. Accapezzato D, Francavilla V, Paroli M, Casciaro M, Chircu LV, Cividini A, Abrignani S, Mondelli MU, Barnaba V. Hepatic expansion of a virus-specific regulatory CD8(+) T cell population in chronic hepatitis C virus infection. J Clin Invest. 2004; 113:963-72.