

Aus dem Institut für Virologie  
der Heinrich-Heine-Universität Düsseldorf  
Direktor: Univ.-Prof. Dr. med. Jörg Timm

Analyse der Hepatitis C Virusevolution anhand des Hepatitis C-Virus-Ausbruchs  
durch kontaminiertes Immunglobulin im Jahre 1979 (Anti-D-Ausbruch)

Dissertation

zur Erlangung des Grades eines Doktors der Medizin  
der Medizinischen Fakultät der Heinrich-Heine-Universität Düsseldorf

vorgelegt von  
Elias Johannes Kortenbruck  
2025

Als Inauguraldissertation gedruckt mit Genehmigung der  
Medizinischen Fakultät der Heinrich-Heine-Universität Düsseldorf

Gez.:

Dekan: Univ.-Prof. Dr. med. Nikolaj Klöcker

Erstgutachter: Univ.-Prof. Dr. med. Jörg Timm

Zweitgutachter: Univ.-Prof. Dr. med. Hans Bock

Meiner Familie

## Zusammenfassung

Das Hepatitis C-Virus (HCV) bleibt auch im Zeitalter der direkten antiviralen Agenzien ein globales gesundheitspolitisches Problem, unter anderem weil die ausgeprägte genetische Variabilität des Hepatitis C-Virus bis heute die Entwicklung einer wirksamen Impfung gegen das Virus erschwert. Des Weiteren ermöglicht sie die Selektion von Immunevasionsvarianten und trägt damit zur Chronifizierung der Infektion mit der Folge von schwerwiegenden gesundheitlichen Auswirkungen auf die betroffenen Patienten bei. Um die Anpassung des Hepatitis C-Virus an seinen Wirt und die zugrundeliegenden evolutionären Mechanismen zu untersuchen, war das erste Ziel dieser Arbeit die Erstellung einer Sequenzdatenbank mit Proben aus einem großen HCV-Ausbruch in den Jahren 1978/79 („Anti-D-Ausbruch“). In diesem Ausbruch erkrankten rund 3000 Personen an einer Hepatitis C durch eine Exposition mit HCV-infizierten Immunglobulin-Präparaten zur Rhesusprophylaxe. Dadurch, dass diese alle auf eine einzige HCV-positive Blutspenderin zurückgehen, mit deren Blut fünf rhesusnegative Spender zur Gewinnung der Immunglobuline immunisiert wurden, kam es zu einem *Single-Source*-Ausbruch aus genetisch sehr eng verwandten Hepatitis C-Viren. Wir konnten ein PCR-Protokoll bestehend aus RNA-Extraktion, cDNA-Synthese und *Nested* PCR etablieren, mit der sich das HCV-Genom von *Core* bis NS2/3 in drei Abschnitten zuverlässig amplifizieren und mittels Sanger-Methode sequenzieren ließ. Aufgrund der schlechten RNA-Qualität aus frühen Proben direkt nach Beginn des Ausbruchs war eine Sequenzierung dieser Proben nicht möglich. Bei später gesammelten Proben aus den 90er-Jahren war dieser Effekt weniger stark ausgeprägt. Insgesamt konnten aus Verlaufsproben aus den 2000er Jahren 218 Sequenzen von 138 Patientinnen in einer Sequenzdatenbank zusammengestellt werden, darunter 138 Sequenzen von *Core* bis NS2. Durch diese sowie vorhandene Sequenzen der Nicht-Struktur-Proteine der entsprechenden Proben konnten 74 komplette HCV-Genome generiert werden. In phylogenetischen Analysen zeigte sich die klare Abgrenzung der HCV-Genome aus der Anti-D-Kohorte von HCV-Referenzsequenzen. Auch die Auftrennung der Isolate in drei *Clades*, aufgrund von drei in der Infektionsquelle vorliegenden Varianten, konnte eindeutig nachgewiesen werden. Des Weiteren wurde die für den *Entry* in die Hepatozyten vermutlich wichtige Hypervariable Region 1 und die dortige Anti-D-spezifische Insertion quantitativ und mittels *Principal Component Analysis* analysiert. Ein signifikanter Zusammenhang zwischen den HLA-Typen und den inserierten Aminosäuren konnte nicht gefunden werden. Um den Einfluss von CD8<sup>+</sup>-T-Zellen auf die Virusevolution in der Anti-D-Kohorte zu untersuchen, wurde der HLA-Typ von 138 Patientinnen mittels SeqFeatR mit der Virussequenz korreliert; zehn Positionen mit einer statistischen Assoziation zwischen HLA-Typ und Sequenz konnten ermittelt werden. Mit dem MHC I *Binding Prediction Tool* der *Immune Epitope Database* (IEDB) konnten diese Positionen neun potenziellen T-Zell-Epitopen zugewiesen werden. Zwei wurden bereits in früheren Arbeiten beschrieben, drei weitere wurden im Zusammenhang mit anderen HLA-Typen beschrieben. Die restlichen vier potenziellen Epitope könnten neue Epitope darstellen und müssten in weiteren Arbeiten *in vitro* bestätigt werden.

Zusammenfassend konnten in dieser Studie weitere vollständige HCV-Genome aus der Anti-D-Kohorte entschlüsselt werden. Die phylogenetischen Analysen bestätigten frühere Studien und bei der Analyse der HVR konnte die Anti-D-typische Insertion weiter untersucht werden. Durch statistische Analysen konnten potenzielle T-Zell Epitope identifiziert werden.

## Abstract

Even in the age of direct antiviral agents, the hepatitis C virus (HCV) remains a global health problem, partly because the pronounced genetic variability of the hepatitis C virus still makes it difficult to develop an effective vaccine against the virus. Furthermore, it enables the selection of immune evasion variants, which contributes to the chronicity of the infection with serious health consequences for the affected patients. In order to investigate the adaptation of the hepatitis C virus to its host and the underlying evolutionary mechanisms, the first aim of this work was to create a sequence database with samples from a large HCV outbreak in 1978/79 (“anti-D-outbreak”). In this outbreak, around 3000 people contracted hepatitis C through exposure to HCV-infected immunoglobulin preparations for rhesus prophylaxis. The fact that these preparations all originated from a single HCV-positive blood donor, whose blood was used to immunize five rhesus-negative donors to obtain the immunoglobulins, resulted in a single-source outbreak of genetically very closely related hepatitis C viruses. In the first phase of the project, we were able to establish a PCR protocol consisting of automated RNA extraction, cDNA synthesis/reverse transcription and subsequent nested PCR, with which the HCV genome from Core to NS2/3 could be reliably amplified in three sections and then sequenced using the Sanger method. Due to the poor RNA quality of early samples taken directly after the outbreak began, sequencing these samples was not possible. This effect was less pronounced in samples collected later in the 1990s. Using follow-up samples from the 2000s, a total of 218 sequences from 138 patients were compiled in a sequence database, including 138 sequences from Core to NS2. Using these as well as corresponding sequences of the non-structural proteins from previous work, 74 complete HCV genomes were generated. Phylogenetic analyses showed the clear separation of the HCV genomes belonging to the anti-D cohort from HCV reference sequences. The separation of the isolates into three clades, based on three variants present in the source of infection, was also clearly demonstrated. Furthermore, the hypervariable region 1, which is presumably important for entry into the hepatocytes, and the anti-D-specific insertion there were analyzed quantitatively and by means of principal component analysis. A significant correlation between the HLA types and the inserted amino acids could not be found. To investigate the influence of CD8<sup>+</sup> T cells on viral evolution in the anti-D cohort, the HLA type of 138 patients with a known HLA type was correlated with the viral sequence using SeqFeatR. Ten positions with a statistical association between HLA type and sequence were identified. Using the MHC I Binding Prediction Tool of the Immune Epitope Database (IEDB), these positions could be assigned to nine potential T cell epitopes. Two have already been described in previous work, three others have been described, but in connection with other HLA types. The remaining four potential epitopes could represent new epitopes and would need to be confirmed *in vitro* in further work.

In summary, further complete HCV genomes from the anti-D cohort could be deciphered in this study. The phylogenetic analyses confirmed previous studies and the analysis of the HVR allowed further investigation of the anti-D-typical insertion. Statistical analyses identified known as well as potential T-cell epitopes.

## Abkürzungsverzeichnis

°C	Grad Celsius
aa	Aminosäuren (engl. <i>amino acids</i> )
AD	Anti-D
APC	Antigenpräsentierende Zelle (engl. <i>antigen-presenting cell</i> )
apoB, apoC, apoE	Apolipoprotein B, C, E
bp	Basenpaare (engl. <i>base pairs</i> )
bzw	beziehungsweise
CD8, CD81	CD8-, CD81-Oberflächenprotein (engl. <i>cluster of differentiation</i> )
cDNA	Komplementäre Desoxyribonukleinsäure (engl. <i>complementary desoxyribonucleic acid</i> )
CDH1	Cadherin-1
CE	Cholesterinester
CHKA	Cholinkinase alpha
CLDN1	Claudin-1
CYBA	Cytochrom b-245 alpha/light chain
DAA	Direkt antiviral wirkende Arzneimittel (engl. <i>direct antiviral agent</i> )
DC	Dendritische Zelle
DDR	Deutsche Demokratische Republik
DNA	Desoxyribonukleinsäure (engl. <i>desoxyribonucleic acid</i> )
dNTP(s)	Desoxynukleosidtriphosphat(e)
DTT	Dithiothreitol
E1, E2	Virale Hüllproteine <i>Envelope</i> 1 und 2
EDTA	Ethylendiamintetraacetat
EGF(-R)	Epidermaler Wachstumsfaktor (-Rezeptor) (engl. <i>epidermal growth factor</i> )
F	Leserichtung vorwärts (engl. <i>forward</i> )
FDR	Falscherkennungsrate (engl. <i>false discovery rate</i> )

<i>H+</i>	Proton bzw. Wasserstoffion (engl. <i>hydrogen</i> )
<i>HCV</i>	Hepatitis C-Virus
<i>HDL</i>	<i>High density lipoprotein</i>
<i>HLA_</i>	<i>Human Leukocyte Antigen</i>
<i>HSPG</i>	Heparansulfat-Proteoglykane
<i>HVR</i>	Hypervariable Region
<i>IC50</i>	Mittlere inhibitorische Konzentration
<i>ID</i>	<i>Identifier</i>
<i>IEDB</i>	<i>Immune Epitope Database</i>
<i>INFg</i>	Interferon gamma
<i>IL-2</i>	Interleukin 2
<i>IRES</i>	Interne ribosomale Eintrittsstelle (engl. <i>internal ribosomal entry site</i> )
<i>IU</i>	Internationale Einheit (engl. <i>international unit</i> )
<i>kb</i>	Kilobasen (1000 Basenpaare)
<i>KCl</i>	Kaliumchlorid
<i>l (μl, ml)</i>	Liter (Mikroliter, Milliliter)
<i>LDL(-R)</i>	<i>Low density lipoprotein (-Rezeptor)</i>
<i>LVP</i>	<i>Lipo Viro Particles</i>
<i>M (nM, μM, mM)</i>	Molar (mol/l) (Nano-, Mikro-, Millimolar)
<i>MgCl<sub>2</sub></i>	Magnesiumchlorid
<i>MMLV</i>	Moloney Murines Leukämievirus
<i>MSA</i>	<i>Multiple Sequence Alignment</i>
<i>NADPH</i>	Protonierte Form von Nicotinamidadenindinukleotidphosphat
<i>ng</i>	Nanogramm
<i>NK</i>	Natürliche Killerzelle
<i>NS_</i>	Nicht-Strukturprotein (2, 3, 4A, 4B, 5A, 5B)
<i>nt</i>	Nukleotide (engl. <i>nucleotides</i> )
<i>NTR</i>	Nicht-translatierte Region
<i>MHC_</i>	<i>Major Histocompatibility Complex</i>
<i>OCN</i>	Occludin
<i>p7</i>	Viroporin p7

<i>PC1, 2</i>	<i>Principal Component 1, 2</i>
<i>PCA</i>	Hauptkomponentenanalyse (engl. <i>principal component analysis</i> )
<i>PCR</i>	Polymerasekettenreaktion (engl. <i>polymerase chain reaction</i> )
<i>pDC</i>	Plasmazytoide dendritische Zelle
<i>pmol</i>	Pikomol
<i>PWID</i>	i.v.-Drogennutzer (engl. <i>people who inject drugs</i> )
<i>R</i>	Leserichtung rückwärts (engl. <i>reverse</i> )
<i>RAC1</i>	<i>Ras-related C3 botulinum toxin substrate 1</i>
<i>RdRp</i>	RNA-abhängige RNA-Polymerase (engl. <i>RNA-dependent RNA-Polymerase</i> )
<i>RNA</i>	Ribonukleinsäure (engl. <i>ribonucleic acid</i> )
<i>rpm</i>	Drehzahl bzw. Umdrehungen pro Minute (engl. <i>revolutions per minute</i> )
<i>RT</i>	Reverse Transkription
<i>SD</i>	Standardabweichung (engl. <i>standard deviation</i> )
<i>SMAD6</i>	<i>Mothers against decapentaplegic homolog 6</i>
<i>SR-BI/SCARB1</i>	HDL-Rezeptor Scavenger-Rezeptor B1
<i>SVR</i>	<i>Sustained Viral Response</i>
<i>TAG</i>	<i>Tri(acyl)glycerid</i>
<i>Tris-HCl</i>	Tris(hydroxymethyl)aminomethan
<i>U</i>	Einheit (engl. <i>unit</i> )
<i>UTR</i>	<i>Untranslated Region</i>
<i>UV</i>	Ultraviolett
<i>VL</i>	Viruslast
<i>VLDL</i>	<i>Very low density lipoprotein</i>
<i>WHO</i>	<i>World Health Organisation</i>

# Inhaltsverzeichnis

Zusammenfassung .....	I
Abstract .....	II
Abkürzungsverzeichnis .....	III
Inhaltsverzeichnis .....	VI
Abbildungsverzeichnis .....	VIII
Tabellenverzeichnis .....	IX
1 Einleitung .....	1
1.1 Hepatitis C-Virus (HCV) .....	1
1.2 HCV Transmission.....	2
1.3 Anti-D-Ausbruch von 1978/79 .....	3
1.4 HCV Struktur.....	5
1.5 Entry .....	6
1.6 Immunsystem und Human-Leukocyte-Antigen (HLA) .....	8
1.7 Ziele der Arbeit.....	10
2 Material und Methoden .....	11
2.1 Material.....	11
2.1.1 Geräte.....	11
2.1.2 Verbrauchsmaterialien.....	11
2.1.3 Chemikalien und Enzyme .....	12
2.1.4 Kits.....	13
2.1.5 Oligonukleotide .....	13
2.1.6 Software und Online-Tools.....	13
2.2 Proben.....	14
2.3 Ethikvotum .....	14
2.4 Methoden.....	14
2.4.1 Probenaufbereitung/RNA-Extraktion.....	14
2.4.2 Ermittlung der Viruslast.....	15
2.4.3 Reverse Transkription.....	15
2.4.4 Nested PCR .....	16
2.4.5 Kapillarelektrophorese.....	18
2.4.6 DNA-Aufreinigung .....	18
2.4.7 Konzentrationsmessung.....	19

2.4.8	Sequenzierung.....	20
2.5	Datenverarbeitung und statistische Auswertung .....	20
2.5.1	Sequenzanalyse in „Geneious“ .....	20
2.5.2	SPSS .....	20
2.5.3	Prism.....	21
2.5.4	SeqFeatR.....	21
2.5.5	IEDB .....	21
2.5.6	FigTree.....	22
2.5.7	Hauptkomponentenanalyse (PCA) mittels R.....	22
3	Ergebnisse.....	23
3.1	Etablierung der PCR .....	23
3.2	Amplifikation früher Proben .....	26
3.3	Amplifikation später Proben .....	29
3.4	Analyse der Hypervariablen Region (HVR).....	33
3.5	SeqFeatR-Analyse.....	38
4	Diskussion .....	44
4.1	Schlussfolgerungen .....	44
4.2	Amplifikation und Analyse der „späten“ Proben .....	46
4.3	Aminosäure-Positionen unter Selektionsdruck durch HLA Klasse-I-Moleküle .....	49
4.4	Fazit und Ausblick.....	53
5	Literatur- und Quellenverzeichnis .....	54
6	Anhang.....	60
	Danksagung .....	66

## Abbildungsverzeichnis

Abbildung 1: Weltweite Verteilung der HCV-Genotypen.....	1
Abbildung 2: Abschnitte des HCV-Genoms .....	5
Abbildung 3: Entry von HCV in die Hepatozyten .....	7
Abbildung 4: Interaktion zwischen infizierter Zelle und Immunsystem .....	9
Abbildung 5: Konzept der Nested PCR .....	16
Abbildung 6: Amplifizierte Abschnitte des HCV-Genoms und verwendete Primer .....	24
Abbildung 7: Verbesserte PCR-Effizienz mit dem AntiD-917-F-Primer .....	25
Abbildung 8: Kapillarelektrophorese der Core-PCRs früher Proben .....	27
Abbildung 9: Prozentsatz hoch qualitativer Sequenzanteile nach Abschnitt.....	31
Abbildung 10: Phylogenetischer Baum der sequenzierten Anti-D-Proben sowie einer Auswahl an HCV-Genotyp 1b-Referenzsequenzen .....	32
Abbildung 11: Schematische Darstellung des Übergangs E1/E2 mit beispielhaften HVR- Insertionen.....	33
Abbildung 12: Häufigkeit von Insertionen in den Clades und in der gesamten Kohorte .....	34
Abbildung 13: PCA der HVR .....	36
Abbildung 14: Sequenzlogo der inserierten Aminosäuren und Aminosäuren-Frequenz nach Clades.....	37
Abbildung 15: HLA-Klasse-I assoziierte Polymorphismen in HCV-Proteinen .....	39

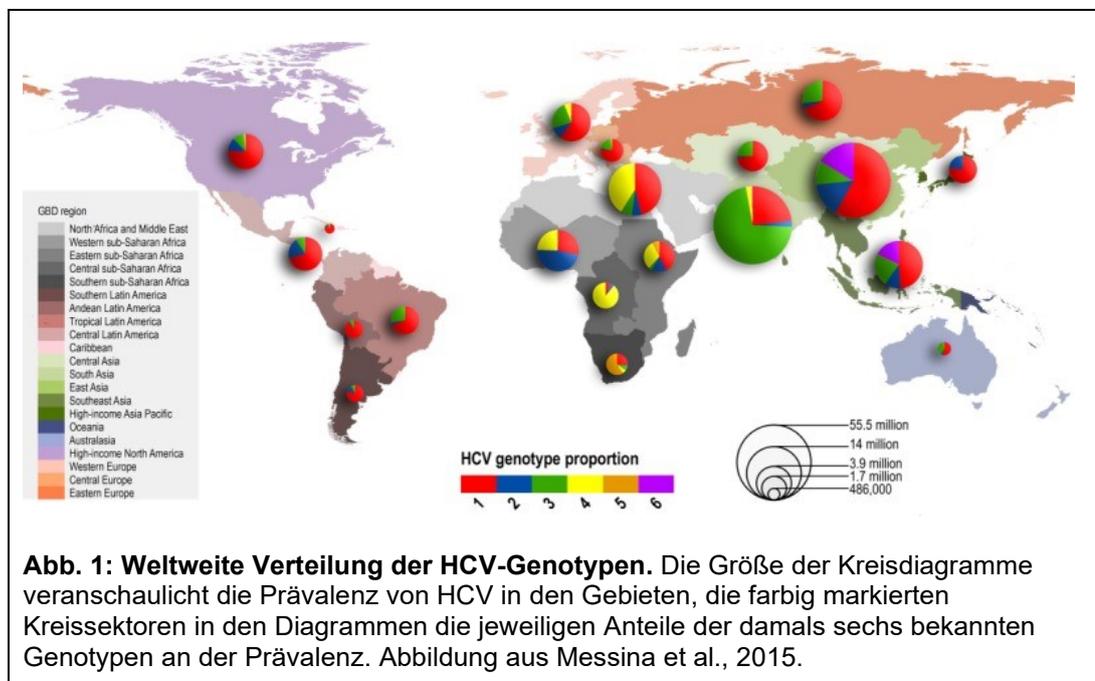
## **Tabellenverzeichnis**

Tabelle 1: Verwendete PCR-Primer .....	13
Tabelle 2: Phasen der verwendeten PCR („NS5B_GT1+4“) .....	17
Tabelle 3: Mittels COBAS ermittelte Viruslasten der frühen Proben .....	28
Tabelle 4: Mittels IEDB-Tool vorhergesagte Epitope im Bereich der zehn mit SeqFeatR ermittelten Mutationen mit den besten p-Werten und einem q-Wert von unter 0,21 .....	41
Tabellen 5 und 6: Mutationen der Aminosäuren-Positionen 470 bis 478 unter HLA-B*15-positiven und HLA-B*15-negativen Patientinnen und Virussequenzen einer Auswahl von HLA-B*15-positiven Patientinnen .....	42

# 1 Einleitung

## 1.1 Hepatitis C-Virus (HCV)

Das Hepatitis C-Virus (HCV) ist ein Virus aus der Familie der *Flaviviridae*, das in acht Genotypen unterteilt wird (Guntipalli et al., 2021). In Deutschland sind die Genotypen 1 und 3 am häufigsten. In Nord- und Südamerika, Europa, dem Norden Afrikas und Zentralasien ist Genotyp 1, in Westafrika Genotyp 2, in Teilen Skandinaviens sowie Süd- und Ostasien Genotyp 3, in Zentralafrika und dem Nahen Osten Genotyp 4 am häufigsten. Genotyp 5 und 6 sind nur in vergleichsweise kleinen Gebieten, in Südafrika und Laos, der jeweils dominante Genotyp (siehe Abb. 1) (Messina et al., 2015).



Unabhängig vom Genotyp können Hepatitis C-Viren im Menschen akute Hepatitiden verursachen, die in der Mehrheit als chronische Infektion persistieren. Nur ungefähr 30 % (15 – 45 %) der Patienten überwinden die meist asymptomatische und nicht lebensbedrohliche Akutinfektion ohne Behandlung. Bei den restlichen 70 % (55 – 85 %) der Patienten hingegen geht diese in eine chronische HCV-Infektion über (WHO, 2023), welche mit einem signifikant erhöhten Risiko für Leberzirrhose (15 – 30 % nach 20 Jahren; WHO 2023) und die Entwicklung eines hepatozellulären Karzinoms (ca. 5 % pro Jahr bei Vorliegen einer HCV-induzierten Leberzirrhose; Sarrazin et al., 2018) einhergeht.

Mit den seit 2013/2014 verfügbaren direkt antiviral wirkenden Medikamenten (DAA) erreicht man in über ca. 95 % der Anwendungen eine Heilung (*sustained viral response*, SVR) (Falade-Nwulia et al., 2017; WHO, 2023).

Obwohl die Kosten der Therapie weiterhin hoch sind und in einigen Ländern ihre Verfügbarkeit damit begrenzt ist, wurden in den letzten Jahren erhebliche Fortschritte bei der medizinischen Versorgung HCV-Infizierter gemacht: Während 2015 nur fünf Prozent der HCV-Infektionen diagnostiziert und davon wiederum nur fünf Prozent einer Therapie zugeführt wurden (Cox, 2015), waren es Ende 2019 schon 21 Prozent und davon 62 Prozent (WHO, 2021). Dies sind aber nur 13 Prozent der zu behandelnden Infizierten (WHO, 2021), sodass chronische Hepatitiden durch HCV weiterhin ein globales Problem bleiben: Weltweit sind ungefähr 58 Millionen Menschen an einer chronischen HCV-Infektion erkrankt (WHO, 2023) und jedes Jahr versterben schätzungsweise 300.000 Menschen (230.000 – 580.000), hauptsächlich an ihren schwerwiegenden Spätfolgen wie hepatozellulären Karzinomen (WHO, 2023).

In Deutschland betrug die Inzidenz von Hepatitis C im Jahr 2020 5,5 Erstdiagnosen/100.000 Einwohnern (7,7 bei Männern, 3,2 bei Frauen) (Meyer et al., 2021), während die Prävalenz von Anti-HCV-Antikörpern in der Bevölkerung zwischen 0,3 % (Poethko-Müller et al., 2013) und 0,95 % (Wolffram et al., 2015) lag. Die wahre Prävalenz liegt aber vermutlich höher, da in den Studien Bevölkerungsgruppen mit einer höheren Prävalenz, wie zum Beispiel Insassen von Justizvollzugsanstalten (Prävalenz 14,3 %) (Schulte et al., 2009) und PWIDs (*people who inject drugs*, Prävalenz 66 %) (Enkelmann et al., 2020) unterrepräsentiert oder nicht inkludiert waren (RKI, 2016).

## 1.2 HCV Transmission

Die Übertragung des Hepatitis C-Virus erfolgt im Großteil der Fälle auf parenteralem Weg durch Kontakt mit kontaminiertem Blut. Die Infektiosität des Hepatitis C-Virus ist dabei zwischen der des Hepatitis B-Virus (HBV) und des Humanen Immundefizienz-Virus (HIV) anzusiedeln; bei einer Nadelstichverletzung beträgt das Übertragungsrisiko ungefähr 1,8 % und liegt somit zwischen dem Risiko bei HBV (6 – 30 %) und HIV (0,3 %) (Beltrami et al., 2000).

In der heutigen Zeit sind in entwickelten Ländern vor allem die oben genannten Risikogruppen infektionsgefährdet; von den 23,9 % der Infektionen, bei denen der wahrscheinlichste Übertragungsweg bekannt ist, erfolgten 76,2 % im Rahmen von injizierendem Drogenkonsum (RKI, 2016).

Bei diagnostischen und chirurgischen Eingriffen wird HCV in Deutschland im Gegensatz zu weniger entwickelten Ländern nur noch sehr selten übertragen, und auch das Übertragungsrisiko über infizierte Blutprodukte ist minimal. Seit der Einführung der diagnostischen Testung der Blutprodukte auf Anti-HCV-Antikörper (April 1991) und HCV-RNA (1999) wurden bei jährlich vier bis fünf Millionen durchgeführten Transfusionen (RKI, 2016) bisher nur drei Übertragungen von HCV beschrieben; eine im Jahr 2004 (Kretzschmar et al., 2007), sowie zwei im Jahr 2019 durch die Blutprodukte eines Anti-HCV-negativen, aber HCV-RNA-positiven Spenders (Himmelsbach et al., 2023). Die Wahrscheinlichkeit einer HCV-Infektion durch eine frische, im neuntägigen diagnostischen Fenster noch nicht nachweisbare Infektion des Spenders wird rechnerisch mit 1:3 Millionen angegeben (an der Heiden et al., 2015).

### **1.3 Anti-D-Ausbruch von 1978/79**

Ein besonderer Fall der iatrogenen HCV-Transmission ereignete sich 1978/79 in der ehemaligen DDR, also noch bevor HCV 1988/89 erstmalig beschrieben wurde. Aufgrund eines befürchteten Lieferengpasses von Anti-D-Immunglobulinen wurde die Rhesusprophylaxe nach der Schwangerschaft in diesem Zeitraum mit Präparaten aus HCV-kontaminierten Chargen durchgeführt.

Als Infektionsquelle wird eine Antigenspenderin vermutet, deren Erythrozyten im Februar 1978 trotz einer seit Dezember 1977 bekannten aktiven Hepatitis mit Appetitlosigkeit, Völlegefühl, Oberbauchschmerzen und einer tastbaren, druckdolenten Leber fünf Plasmaspendern verabreicht wurden. Diese fünf Plasmaspender erkrankten im April 1978 ebenfalls an einer Virushepatitis. Im Mai 1978 wurde diese bei den zwei Spendern, deren Plasma im März 1978 an das Bezirksinstitut für Blutspende- und Transfusionswesen in Halle (Saale) geschickt worden war, mit einer Leberbiopsie gesichert. Die dort aus dem gespendeten Plasma produzierten Anti-D-Immunglobulin-Chargen 6 und 7 wurden gesperrt und auf das damals bekannte Hepatitis B-Virus sowie ihre „Hepatitis-sicherheit“ nach Fraktionierung“ getestet.

Im Januar 1979 wurden trotz der Sperrung erste Hepatitiden nach Gabe von Anti-D-Immunglobulinen bekannt. Es kam zutage, dass das Plasma der zwei potenziell infektiösen Spender in den Chargen 8 bis 14 und Material aus den Chargen 6 und 7 in der Charge 15 verwendet worden war. Die Chargen 8 bis 15 wurden gesperrt, zu diesem Zeitpunkt waren aber bereits ca. 4000 Ampullen der gesperrten Chargen verwendet worden. Nachdem im Rahmen der Verwendung der Chargen 16 und 17 auch Erkrankungen auffielen, stellte sich heraus, dass die für die Produktion der Chargen 16 bis 23 verwendete Waschflüssigkeit aus Arbeitsschritten der jeweils zuvor hergestellten Charge stammte. Diese mussten also ebenfalls gesperrt werden.

Insgesamt wird davon ausgegangen, dass 6773 Ampullen der potenziell kontaminierten Chargen verwendet wurden und bis Februar 1981 2916 Personen an einer Hepatitis C erkrankten, davon mindestens 133 Kontaktpersonen wie Partner oder Säuglinge (Steger et al., 2017).

Es wird in *Follow-Up*-Studien davon ausgegangen, dass 93 % der exponierten jungen Frauen durch diese iatrogene HCV-Infektion im Verlauf eine akute Hepatitis C entwickelten, 46 % wurden auch 25 Jahre später noch positiv auf HCV-RNA im Blut getestet (Wiese M et al., 2005).

Dass die fast 3000 Fälle umfassende Infektionskette auf eine einzelne Infektionsquelle zurückzuführen ist, stellt eine fast einmalige Besonderheit dieses Ausbruchs dar; es ist nur ein weiterer Ausbruch dieser Art bekannt. Dieser ereignete sich um das Jahr 1977 in Irland (Expert Group on the Blood Transfusion Service Board & Ireland Dept. of Health, 1995).

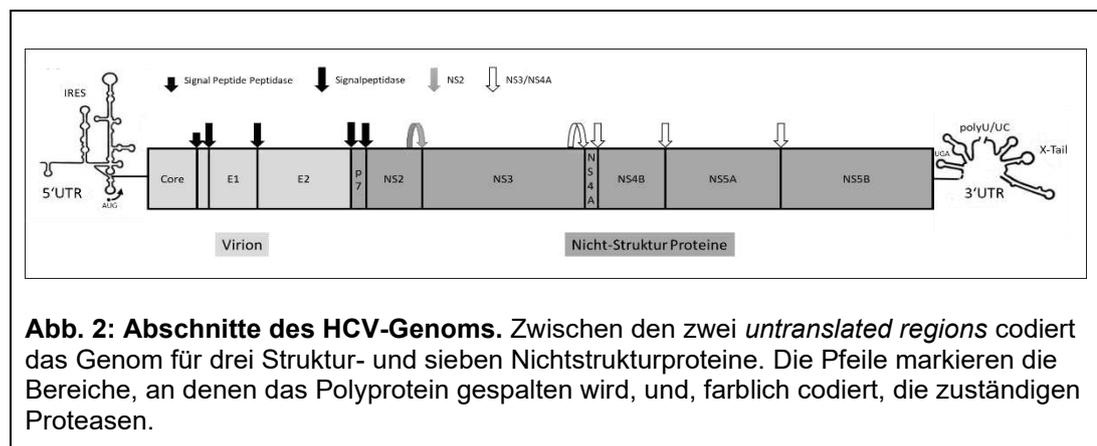
Eine weitere Besonderheit ist, dass in der ursprünglichen Infektionsquelle drei Varianten des Hepatitis C-Virus vorlagen. Dies ermöglicht es, die HCV-Isolate dieses Ausbruchs jeweils einem von drei *Clades*, also einer Gruppe mit einem gemeinsamen Ursprung, zuzuordnen, da jeweils nur eine der drei Varianten in den chronisch infizierten Patientinnen selektiert wurde.

Mit Hilfe der später eingeführten Genom-Sequenzierung und Genotypisierung konnten die Varianten des Anti-D-Ausbruchs dem HCV-Genotyp 1b zugeordnet werden.

## 1.4 HCV Struktur

Das Genom des einsträngigen, positiv geladenen RNA-Virus hat eine Länge von 9646 Nukleotiden.

Diese kodieren zwischen der 5'-*untranslated region* und der 3'-*untranslated region* für ein Polyprotein, das in drei Strukturproteine (*Core*, *Envelope 1*, *Envelope 2*) und sieben Nicht-Strukturproteine gespalten wird. Die Nicht-Strukturproteine sind ein Viroporin (p7), eine Cystein-Protease (NS2), eine Serin-Protease/RNA-Helicase (NS3), ein Protease-Cofaktor (NS4A, bildet einen Komplex mit NS3), das Protein NS4B (ermöglicht vermutlich eine Membranbindung und -veränderung), ein Phosphoprotein (NS5A, für die Replikation bedeutsam) und die RNA-abhängige RNA-Replacase/-Polymerase (RdRp, NS5B) (siehe Abb. 2) (Penin et al., 2004; Bartenschlager et al., 2013).



Bis heute ist noch keine effektive Prävention einer HCV-Infektion in Form einer Impfung gegen das Hepatitis C-Virus verfügbar. Dies liegt auch an der für das Hepatitis C-Virus charakteristischen hohen Vermehrungs- sowie Mutationsrate durch dessen fehleranfällige RNA-abhängige RNA-Polymerase. Diese führen zu einer stark ausgeprägten genetischen Variabilität und spiegeln sich in mittlerweile acht Genotypen und über 57 Subtypen (Guntipalli et al., 2021) wider. Außerdem existiert HCV im infizierten Wirt als sogenannte Quasispezies. Damit wird ein „Schwarm“ einer Vielzahl von verschiedenen und genetisch eng verwandten Varianten innerhalb eines einzelnen Wirts bezeichnet. Die virale Quasispezies ist Grundlage für die rasche Selektion von neuen Varianten und damit verbundene Fähigkeit von HCV, sich an veränderte Replikationsbedingungen anzupassen. Während innerhalb eines Genotyps 65,7 – 68,9 % und innerhalb eines Subtyps 76,9 – 80,1 % der Nukleotide

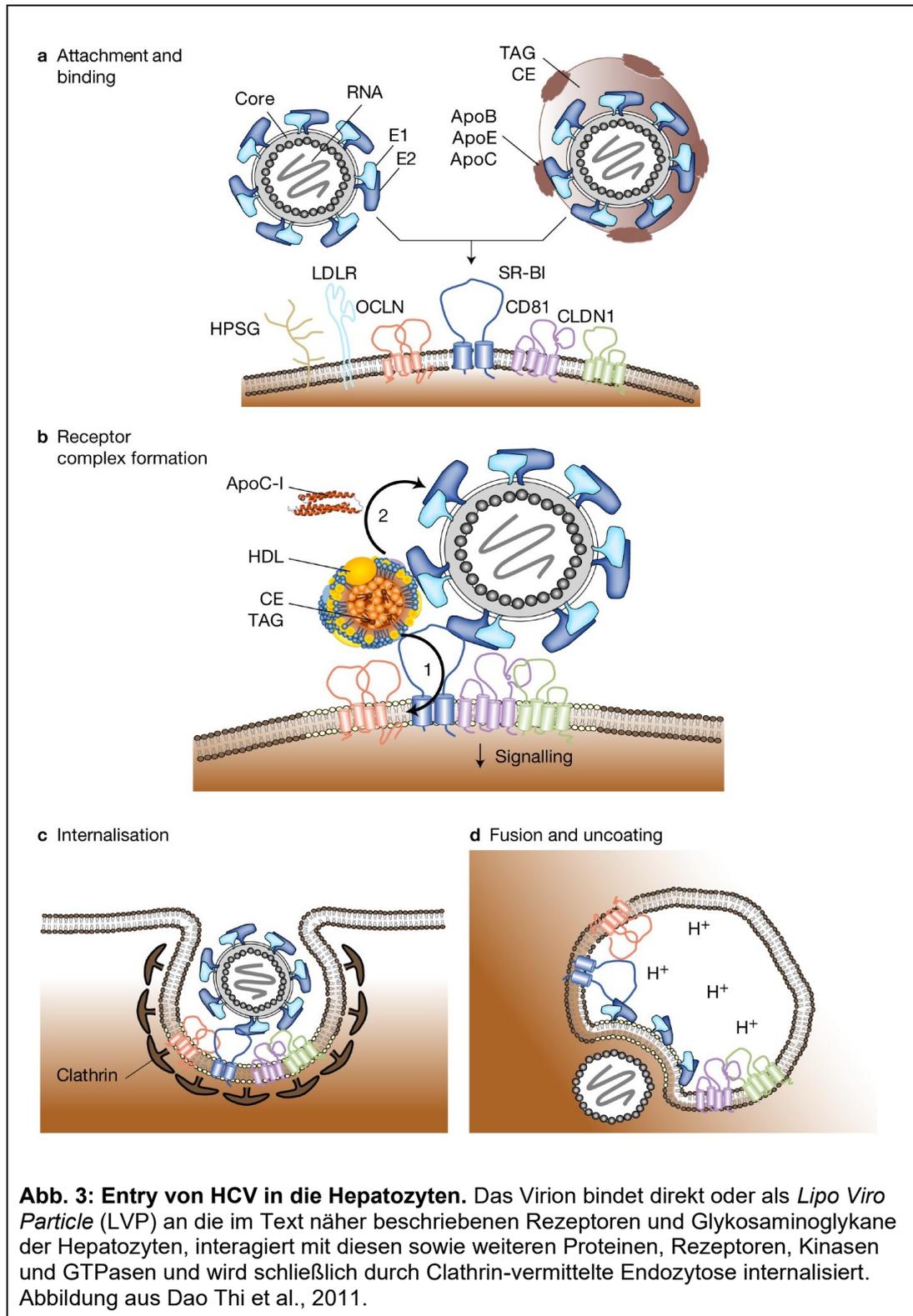
gleich sind, sind die Varianten innerhalb einer Quasispezies zu 90,8 – 99 % gleich aufgebaut (Zein, 2000).

Diese genetische Variabilität geht eng einher mit der Fähigkeit, dem Immunsystem des menschlichen Wirts zu entkommen, indem Evasionsvarianten selektiert werden, die eine Chronifizierung der Infektion begünstigen. Ein wichtiges Ziel der Immunantwort sind dabei virale Hüllproteine, die für den Eintritt von HCV in die Hepatozyten notwendig sind.

## 1.5 Entry

Der Eintritt des Hepatitis C-Virus in die Hepatozyten des Wirts (*Entry*) ist stark abhängig vom Fettstoffwechsel; HCV hat in der Form von *Lipo Viro Particles* (LVP), also verbunden mit Phospholipiden, neutralen Fetten und Apolipoproteinen wie *Very low density* Lipoproteinen (VLDL), die höchste Infektiosität (Ogden & Tang, 2015). HCV bzw. LVP binden von den blutgefüllten Leber-Sinusoiden aus zunächst direkt mit dem Strukturprotein *Envelope 2* oder über Apolipoproteine an Glykosaminoglykane wie Heparansulfat-Proteoglykane (HSPG) und Lipoprotein-Rezeptoren wie den LDL-Rezeptor (LDL-R) in der basolateralen Membran der Hepatozyten. Dort interagiert HCV bzw. das LVP über das Oberflächenprotein E2 und dessen N-terminale 27 Aminosäuren (Hypervariable Region 1, HVR1) mit dem *Scavenger*-Rezeptor B1 (SR-BI/SCARB1) (Scarselli et al., 2002) und mit dem Tetraspanin CD81. Das Virion diffundiert zu den *Tight junctions* und interagiert wiederum direkt mit einem E1-E2-Dimer oder über CD81 mit den *Tight Junction*-Proteinen Claudin-1 (CLDN1) und Occludin (OCLN). Diese Interaktion wird durch den *epidermal growth factor*-Rezeptor EGFR sowie E-Cadherin (CDH1), die Rho-GTPase RAC1, eine Cholin-Kinase (CHKA), das Protein SMAD6 und eine Untereinheit der NADPH-Oxidase (CYBA) verstärkt. Nach diesen Interaktionen wird das Virion über Clathrin-vermittelte Endozytose in die Hepatozyten aufgenommen (siehe Abb. 3) (Dao Thi et al., 2011).

Die Bindung an die Hepatozyten mit dem darauffolgenden *Entry*-Prozess bildet eines der wichtigsten Ziele der adaptiven Immunantwort (Dustin et al., 2014). Vor allem gegen das oben genannte Oberflächenprotein E2 und dessen N-terminale 27 Aminosäuren (Hypervariable Region 1, HVR1) werden neutralisierende Antikörper gebildet.



Durch die hohe Sequenzvariabilität kann sich die HVR aber an das Immunsystem anpassen und dient damit als eine Art immunologischer Täuschkörper. Durch ihre hohe Immunogenität werden viele Antikörper gegen die HVR gebildet, während umliegende, stärker konservierte Epitope „kaschiert“ werden. Dank der geringen

funktionellen Einschränkungen der HVR kann das Virus den gebildeten, oft Isolat-spezifischen HVR-bindenden Antikörpern schnell entkommen, indem Varianten aus der Quasispezies selektiert werden, die *Escape*-Mutationen tragen (Dustin et al., 2014).

Eine Besonderheit der Anti-D-Kohorte ist eine Insertion in dieser für *Entry* und *Immune-Escape* wichtigen Hypervariablen Region 1 des Oberflächenproteins E2.

Diese kommt nur in seltenen Fällen in anderen Isolaten des Genotyps 1b vor (Casino et al., 1999; Gerotto et al., 2001; Torres-Puente et al., 2007), in den Proben des Anti-D-Ausbruchs allerdings sehr häufig. Die bekannten Insertionen sind unterschiedlich lang und es ist nicht bekannt, ob sie eine funktionelle Bedeutung haben. Das vergleichsweise häufige Auftreten dieser Insertionen könnte rein zufällig sein oder in Zusammenhang mit Antikörpern des Wirts, *Entry*-Faktoren oder anderen *Host*- oder Virus-spezifischen Faktoren stehen.

Entkommen die Hepatitis C-Viren der humoralen Abwehr und gelangen durch die oben genannten *Entry*-Mechanismen in die Hepatozyten, übernehmen andere Teile des Immunsystems die Aufgabe der Viruselimination.

## **1.6 Immunsystem und Human-Leukocyte-Antigen (HLA)**

Für die Immunantwort gegen HCV spielen, wie bei anderen Viren und intrazellulären Erregern auch, neben neutralisierenden Antikörpern vor allem die zytotoxischen CD8<sup>+</sup>-T-Zellen eine große Rolle.

Damit diese T-Zellen körperfremde Antigene als solche erkennen, müssen diese ihnen durch HLA-Proteine (*Human-Leukocyte-Antigen*-Proteine) der Klasse I präsentiert werden (sog. HLA-Restriktion, siehe auch Abb. 4).

Während HLA-Proteine der Klasse II nur auf phagozytierenden Immunzellen vorkommen, werden die besagten HLA-Proteine der Klasse I auf allen kernhaltigen Zellen des menschlichen Körpers exprimiert. Sie werden weiter unterteilt in HLA A, B und C und sind im humanen Haupthistokompatibilitätskomplex (engl. MHC für *Major histocompatibility complex*) auf dem kurzen Arm des Chromosoms 6 codiert.



Molekülabschnitte eines Antigens, die wie oben beschrieben von einem HLA-Protein an den T-Zell-Rezeptor der T-Zelle präsentiert werden, werden auch Epitope oder antigene Determinante genannt. Die von der HLA-Klasse I gebundenen Epitope sind meist acht bis elf Aminosäuren lang und dienen als Ausgangsstruktur für die Bildung von spezifischen T-Zellen mit passenden T-Zell-Rezeptoren im Zuge der adaptiven Immunantwort.

Viren, die genau innerhalb dieser Epitope Mutationen des Virus-Genoms besitzen und so dieser Immunantwort entgehen, haben einen evolutionären Vorteil gegenüber anderen Viren und werden im Verlauf selektiert. Die Mutationen, die dies ermöglichen, werden auch *Escape*-Mutationen genannt.

## 1.7 Ziele der Arbeit

Ziel der Arbeit ist es, anhand von frühen Proben des Anti-D-Ausbruchs die Anpassung des Hepatitis C-Virus an seinen Wirt und die zugrundeliegenden evolutionären Mechanismen zu untersuchen. Hierzu sollen die kompletten Genome der in den Anti-D-Proben enthaltenen Hepatitis C-Viren sequenziert und in einer Sequenzdatenbank zusammengeführt werden. Mit diesen Daten wird es möglich sein, Insertionen und andere Mutationen im HCV-Genom zu identifizieren und funktionell zu charakterisieren, sowie vorliegende Quasispezies zu rekonstruieren.

Neben der allgemeinen Analyse der Kohorte und der neu generierten viralen Sequenzen soll ein besonderer Fokus auf die Hypervariable Region 1 im viralen Hüllprotein E2, die dortige für die Anti-D-Kohorte charakteristische Insertion sowie mögliche Konsequenzen für den Entry in die Hepatozyten gelegt werden.

Des Weiteren soll der Umfang der Anti-D-Sequenzdatenbank genutzt werden, um mögliche Epitope des HCV-Genoms für bestimmte HLA-Typen zu finden.

## 2 Material und Methoden

### 2.1 Material

#### 2.1.1 Geräte

- -80 °C-Gefrierschrank: New Brunswick Ultra-Low Temperature Freezer, Eppendorf AG, Hamburg, Deutschland
- 4 °C-Kühlschrank: Robert Bosch GmbH, Stuttgart, Deutschland
- Automatisierte Kapillarelektrophorese: QIAxcel Advanced System, Qiagen N.V., Hilden, Deutschland
- Automatisiertes Nukleinsäure-Extraktionssystem: EZ1 Advanced XL mit EZ1 Advanced Virus Card v2.0, Qiagen N.V., Hilden, Deutschland
- Heizblöcke: Thermostat Plus, Eppendorf AG, Hamburg, Deutschland
- Sicherheitswerkbank: PCR-Chamber, Bära GmbH & Co. KG, Leichlingen, Deutschland
- Spektrophotometer: Nanodrop 2000, Thermo Fisher Scientific Inc., Waltham, USA
- Thermocycler: Mastercycler nexus, Eppendorf AG, Hamburg, Deutschland; Biometra T3000, Analytik Jena AG, Überlingen, Deutschland
- Vollautomatisiertes System zur Viruslastbestimmung: cobas 6800, F. Hoffmann-La Roche AG, Basel, Schweiz
- Vortex-Schüttler: VV3, VWR International, Radnor, USA
- Zentrifugen: Centrifuge 5415D und Centrifuge 5417C, Eppendorf AG, Hamburg, Deutschland; Color Sprout Mini-Zentrifuge, Biozym Scientific GmbH, Wien, Österreich

#### 2.1.2 Verbrauchsmaterialien

- Desinfektionsmittel: Softa-man acute, B. Braun Melsungen AG, Melsungen, Deutschland; Sterilium, Bode Chemie, Hamburg, Deutschland
- Gestell für Reaktionsgefäße: neoLab Top-Rack, neoLab Migge GmbH, Heidelberg, Deutschland

- Pipetten: Eppendorf Research plus Pipette, Eppendorf AG, Hamburg, Deutschland
- Präzisionstücher: Kimtech Science Präzisionstücher, Kimberly-Clark Corporation, Dallas, USA; Tapira Plus, GVS-Grossverbraucherspezialisten eG, Friedewald, Deutschland
- Reagiergefäße: SafeSeal Reagiergefäß 1,5 ml und 2 ml, Sarstedt AG, Nümbrecht, Deutschland; QX 0,2 ml 12-Tube Strips, Qiagen N.V., Hilden, Deutschland; PCR Tube Strips 0,2 ml, Eppendorf AG, Hamburg, Deutschland; Polypropylen Röhren 15 ml Cellstar, Greiner Bio-One GmbH, Kremsmünster, Österreich

### **2.1.3 Chemikalien und Enzyme**

- 0,1 M DTT, Invitrogen AG, Waltham, USA
- 5X First-Strand Buffer, Invitrogen AG, Waltham, USA
- 5X Colorless/Green GoTaq® Reaction Buffer, Promega Corporation, Madison, USA
- dNTP Mix (10 mM each), Invitrogen AG, Waltham, USA
- Ethanol, Merck KGaA, Darmstadt, Deutschland
- ExoSAP-IT™ PCR Product Cleanup Reagent, Thermo Fisher Scientific Inc., Waltham, USA
- GoTaq® DNA Polymerase (M300), Promega Corporation, Madison, USA
- Nuclease-Free Water, Invitrogen AG, Waltham, USA
- QX DNA Dilution Buffer 15 ml, Qiagen N.V., Hilden, Deutschland
- QX Size Marker 250 bp – 4 kb (50 µl) v2.0, Qiagen N.V., Hilden, Deutschland
- RNase Inhibitor, Murine, New England BioLabs Inc., Ipswich, USA
- SuperScript™ III Reverse Transcriptase (200 U/µl), Invitrogen AG, Waltham, USA
- Tris-borat-EDTA-Buffer, Sigma-Aldrich, Merck KGaA, Darmstadt, Deutschland

### 2.1.4 Kits

- EZ1 Virus Mini Kit v2.0, Qiagen N.V., Hilden, Deutschland
- QIAamp Viral RNA Kit, Qiagen N.V., Hilden, Deutschland
- QIAquick® PCR Purification Kit, Qiagen N.V., Hilden, Deutschland
- QIAxcel DNA High Resolution/Screening Kit, Qiagen N.V., Hilden, Deutschland

### 2.1.5 Oligonukleotide

Die Oligonukleotide (siehe Tab. 1) wurden von Eurofins Scientific SE, Luxemburg, bezogen, in Nuklease-freiem Wasser auf eine Konzentration von 100 µM eingestellt und dann auf eine 10 µM Gebrauchslösung verdünnt.

Verwendung	Primer	Leserichtung	Nukleotid-Position in Referenzsequenz	Länge in Nukleotiden	Sequenz in 5'-3'-Richtung, Mismatch(es) hervorgehoben
cDNA-Synthese	Oligo dA	Reverse	9419 - 9468	20+	AAAAAAAAAAAAAAAAAAAA
Core-PCR I	p417s	Forward	416 - 435	20	GGYGGYGGNCAGATCGTTGG
	p874A	Reverse	851 - 873	23	ARGAAGATAGARAARGAGCAACC
Core-PCR II	p439s	Forward	438 - 459	22	GAGTWTACBTGYTGCCGCGCAG
	p1As	Reverse	731 - 750	20	ATRTACCCCATGAGRTCGGC
UTR-PCR I	HCV1b_1a-F	Forward	35 - 51	17	ATCACTCCCCGTGAGG
	HCV1b_2a-R	Reverse	1797 - 1814	18	GAGGYGCGTAGTGCCAGC
UTR-PCR II	HCV1b_1a-F_int	Forward	70 - 88	19	GAAAGCGTCTAGCCATGGC
	HCV1b_1e-R	Reverse	1592 - 1608	17	TTGATGTGCCAGCTGCC
E1-PCR I	HCV1b_1e-F	Forward	1015 - 1033	20	CAYGTCGAYTTGCTCGTTGG
	HCV1b_2d-R	Reverse	2949 - 2966	18	GGAGGATGATGGCATCGC
	AntiD-917-F	Forward	917 - 936	20	GAAGTGCACAACGTGTCCGG
	HCV1b_3a-R	Reverse	3068 - 3086	19	GAGCGGYACRAAGTACGG
E1-PCR II	HCV1b_2a-F	Forward	1299 - 1318	20	GGGATATGATGATGAAGTGG
	HCV1b_2c-R	Reverse	2488 - 2507	19	GCTATCAGCAGCATCATCC
	HCV1b_2d-R	Reverse	2949 - 2966	18	GGAGGATGATGGCATCGC
NS2-PCR I	HCV1b_2e-F	Forward	2649 - 2668	20	TGTTCTTCTGTGCGCTGG
	HCV1b_3e-R	Reverse	4188 - 4205	18	CGCCCGTGGTGTGGTCC
NS2-PCR II	HCV1b_3a-F	Forward	2780 - 2799	20	ATGGCTGCATCGTGCGGAGG
	HCV1b_3d-R	Reverse	4015 - 4032	18	GGAGCGTGTAGATGGGCC

**Tabelle 1: Verwendete PCR-Primer.**

### 2.1.6 Software und Online-Tools

- FigTree, <http://tree.bio.ed.ac.uk/software/figtree/>
- Geneious® 10.2.2, Biomatters Ltd., Auckland, Neuseeland
- GraphPad Prism, GraphPad Software Inc., San Diego, USA
- IEDB, <https://www.iedb.org/>
- Microsoft Excel, PowerPoint und Word, Microsoft Corporation, Redmond, USA

- SeqFeatR, <https://seqfeatr.zmb.uni-due.de>
- SPSS Statistics, IBM Corporation, Armonk, USA

## 2.2 Proben

Bei den verwendeten Blutproben handelt es sich um diagnostische Rückstellproben der Anti-D-Kohorte, die im Rahmen der Behandlung der Patientinnen entnommen, in dem Institut für Virologie der Charité Berlin archiviert und dann 2013 in die Biobank des Kompetenznetz Hepatitis Essen überführt wurden.

Die zur Datenanalyse verwendeten Proben wurden 138 Patientinnen und Patienten entnommen, die zum Zeitpunkt der Infektion im Durchschnitt ungefähr 25 (Mittelwert) bzw. 24 (Median) Jahre alt waren.

## 2.3 Ethikvotum

Die Ethikkommission der Heinrich-Heine-Universität erteilte dieser retrospektiven, anonymisierten Datenanalyse mit der Studiennummer 2018-25-RetroDEuA ein positives Ethikvotum.

## 2.4 Methoden

### 2.4.1 Probenaufbereitung/RNA-Extraktion

Zum größten Teil erfolgt die Nukleinsäure-Reinigung aus Blutproben mit dem EZ1 *Advanced* XL-System von Qiagen. Dieses führt mithilfe der auf der EZ1 *Advanced* XL Card gespeicherten Protokolle automatisiert die Arbeitsschritte durch, die nötig sind, um HCV-RNA aus den Blutproben zu extrahieren. Nach dem Einlesen der Reagenzien- und Probanden werden die Proben mittels Proteinkinasen lysiert, die enthaltenen Nukleinsäuren an Magnet-Partikel gebunden, gewaschen und eluiert. Aus 200 µl Blutplasma werden durch diese Schritte Nukleinsäuren in einem Volumen von 60 µl gewonnen, welches sofort weiterverwendet oder bei -20 °C gelagert wird. Zuletzt erstellt das System eine Reportdatei und kann mittels UV-Strahlung dekontaminiert werden.

In einzelnen Fällen wird die HCV-RNA manuell mittels QIAamp *Viral* RNA Kit von Qiagen aus den Proben extrahiert. Die Proben werden zuerst lysiert, um RNasen zu inaktivieren und intakte virale RNA zu isolieren.

Nach dem Anpassen der Pufferkonditionen werden die Proben in die QIAamp Mini-*Spin* Gefäße gegeben, in welchen die RNA durch Zentrifugation an die Silica-basierte Membran bindet. Daraufhin werden Kontaminanten durch zwei Waschschrte mit zwei verschiedenen Pufferlösungen entfernt und zuletzt die virale RNA in einem RNase-freien Puffer eluiert.

#### **2.4.2 Ermittlung der Viruslast**

Das vollautomatisierte PCR-System cobas 6800 von F. Hoffmann-La Roche AG (Basel, Schweiz) ermöglicht eine quantitative Detektion von HCV-RNA aus EDTA-Plasma oder Serum. Die Ermittlung der Viruslasten erfolgt durch eine automatisierte Probenvorbereitung (Extraktion und Aufreinigung), eine Amplifikation basierend auf Real-Time-PCR-Technologie und anschließende Detektion.

#### **2.4.3 Reverse Transkription**

Um aus der extrahierten viralen RNA cDNA zu synthetisieren, wird ein etabliertes zweischrittiges Protokoll verwendet.

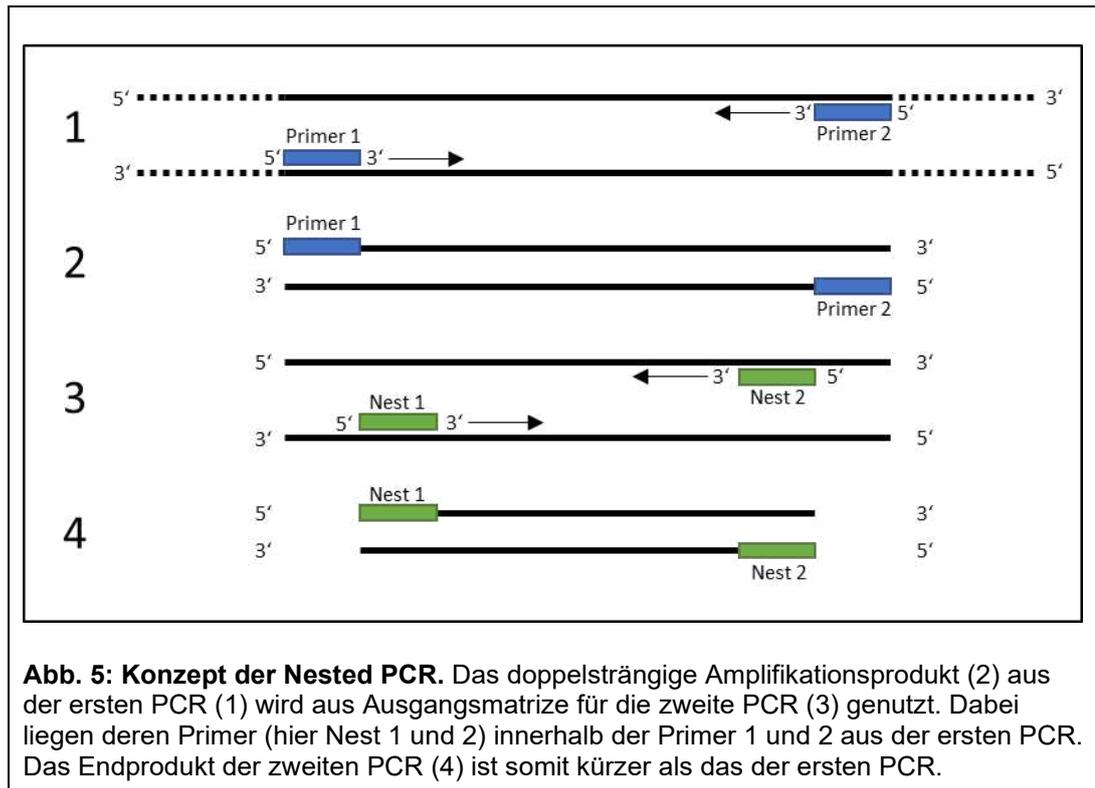
Im ersten Schritt werden pro Probe 1 µl Nuklease-freies Wasser, 1 µl dNTP-Mix, 1µl des Primers Oligo d(A) (50 pmol/µl) sowie 10 µl RNA-Eluat gemischt und im Thermocycler für fünf Minuten auf 65 °C erhitzt.

Nach der Abkühlung auf 25 °C wird im zweiten Schritt 7 µl RT-Reaktionsansatz dazugegeben. Dieser besteht aus je 1 µl *SuperScript*<sup>TM</sup> III Reverse Transcriptase, einer gentechnologisch hergestellten Moloney *Murine Leukemia Virus* (MMLV)-Reverse Transkriptase mit reduzierter RNase H-Aktivität, verbesserter Halbwertszeit und verbesserter thermischer Stabilität, sowie 4 µl *5X First-Strand Buffer* (Tris-HCl, KCl, MgCl<sub>2</sub>), 1 µl 0,1 M DTT und 1 µl eines murinen RNase Inhibitors.

Die Reverse Transkription erfolgt in einem *Thermocycler* während einer zehnminütigen Phase bei 25 °C, einer sechzigminütigen Phase bei 42 °C und zwei jeweils dreißigminütigen Phasen bei 50 °C und 55 °C. Um die RNA und die Reverse Transkriptase zu denaturieren, wird der Ansatz zuletzt für 15 Minuten auf 75 °C erhitzt, dann auf 4 °C abgekühlt. Sofern die cDNA nicht sofort weiterverwendet wird, wird sie bei -20 °C gelagert.

### 2.4.4 Nested PCR

Um die gewonnene cDNA zu amplifizieren, werden *Nested* PCRs (verschachtelte PCRs) durchgeführt. Bei der *Nested* PCR werden Anteile des Produkts einer ersten PCR als Ausgangsmatrix für eine zweite PCR verwendet. Da bei der zweiten PCR zwei neue, innerhalb des im ersten Durchgang amplifizierten Genabschnitts liegende Primer benutzt werden, entstehen als Endprodukt kleinere DNA-Fragmente als im ersten PCR-Durchgang (siehe Abb. 5).



**Abb. 5: Konzept der Nested PCR.** Das doppelsträngige Amplifikationsprodukt (2) aus der ersten PCR (1) wird als Ausgangsmatrix für die zweite PCR (3) genutzt. Dabei liegen deren Primer (hier Nest 1 und 2) innerhalb der Primer 1 und 2 aus der ersten PCR. Das Endprodukt der zweiten PCR (4) ist somit kürzer als das der ersten PCR.

Ein Vorteil der *Nested* PCR im Vergleich zur „herkömmlichen“ PCR ist eine höhere Sensitivität, da die Verwendung von zwei Paaren von Oligonukleotiden eine höhere Anzahl an durchführbaren Zyklen ermöglicht. Außerdem ist durch die Verwendung von zwei Primer-Sets im Bereich derselben Zielsequenz die Spezifität der *Nested* PCR höher, da nur das PCR-Produkt aus der ersten PCR als Matrize der zweiten Reaktion infrage kommt (Green & Sambrook, 2019). Dadurch sollen weniger unspezifische Banden entstehen; anstatt unspezifischer Amplikons wird nur noch die gewünschte Zielsequenz amplifiziert.

Der Nachteil des Verfahrens liegt darin, dass diese gewünschte Sequenz bekannt sein muss, um passende Primer auszuwählen (Green & Sambrook, 2019). Außerdem ist die Kontaminationsgefahr durch Verschleppung („*carry over*“) durch die zusätzlichen Arbeitsschritte erhöht (Carr et al., 2010).

Für den Ansatz der PCR I werden 5 µl der durch Reverse Transkription gewonnenen cDNA, 10 µl farbloser 5X GoTaq *Reaction Buffer*, 1 µl dNTP-Mix (jeweils 2,5 mM), 2,5 µl *Forward*- und 2,5 µl *Reverse*-Primer (Konzentration 0,1–1 µM) sowie 0,25 µl GoTaq-Polymerase (5 U/µl) in 28,75 µl Nuklease-freiem Wasser gemischt.

Ein Anteil von 3 µl des gewonnenen PCR I-Produkts dient als Ausgangsmatrix für die PCR II. Wie in dem ersten Durchgang werden 1 µl dNTP-Mix (jeweils 2,5 mM) sowie 0,25 µl GoTaq-Polymerase (5 U/µl) verwendet. Abweichend vom Protokoll der PCR I enthält der Ansatz der PCR II zwei neue Primer, jeweils in der gleichen Menge und Konzentration wie zuvor, zur besseren Unterscheidung der Proben 10 µl grünen statt farblosen 5X GoTaq *Reaction Buffer* und 30,75 µl Nuklease-freies Wasser.

Sowohl die PCR I als auch die PCR II erfolgen in einem *Thermocycler* unter den in Tabelle 2 dargestellten Bedingungen.

Zyklen	Phase	Temperatur	Dauer
1x	Aufwärmen	94 °C	1 Minute
35x	Denaturierung	94 °C	1 Minute
	Annealing	55 °C	30 Sekunden
	Elongation	72 °C	2 Minuten
1x	Beendigung der Elongation	72 °C	10 Minuten
1x	Lagerung	10 °C	Unendlich

**Tabelle 2: Phasen der verwendeten PCR („NS5B\_GT1+4“).** Die Temperatur während der *Annealing*-Phase ist abhängig von den verwendeten Primern, die Dauer der Elongationsphase von der Größe der amplifizierten Abschnitte. Im Verlauf wurde diese auf 2:30 Minuten pro Zyklus erhöht.

Eine sensitive *Nested* PCR zur Überprüfung von Proben ist die *Core*-PCR, die bereits mit Proben mit nur 100 IU/ml HCV funktioniert. Zur Amplifikation des *Core*-Fragments mit einer Länge von 312 Basenpaaren wird der zuvor beschriebene PCR-Ansatz verwendet, aber unter Verwendung der Primer p417s und p874A (PCR I) sowie p439s und p1As (PCR II).

Während die Temperaturen der Phasen dieselben sind, sind die Zeiten der Zyklen im Programm „Geno2“ anders als bei den zuvor beschriebenen Strukturprotein-PCRs. Nach einer zweiminütigen Aufwärmphase erfolgt pro Zyklus eine 30 Sekunden lange Denaturierungsphase, eine ebenfalls 30 Sekunden lange *Annealing*-Phase und eine einminütige Elongationsphase.

#### **2.4.5 Kapillarelektrophorese**

Der Erfolg der Amplifikation der Proben wird mithilfe des QIAxcel *Advanced System* für automatisierte Kapillarelektrophoresen überprüft.

Von den zu analysierenden Proben werden jeweils 10 µl in die *Tubes* der QX 0,2 ml 12-*Tube-Strips* gegeben; die Negativprobe sowie nicht genutzte *Tubes* werden mit der gleichen Menge QX DNA *Dilution Buffer* aufgefüllt.

Der oder die *Tube*-Streifen wird bzw. werden vorsichtig in die entsprechende Halterung gestellt und die Öffnung des QIAxcel *Advanced System* geschlossen. In der QIAxcel *ScreenGel Software* wird das Prozessprofil ausgewählt und im nächsten Schritt die benötigte, zu den eingesetzten *Tube*-Streifen passende Anzahl an Reihen sowie, falls gewünscht, Marker-Einstellungen festgelegt. Optional kann in dem darauffolgenden Schritt noch die *ID* der einzelnen Proben eingegeben werden. Nach einem letzten Check-Bildschirm kann das System mit der Kapillarelektrophorese beginnen.

Die einzelnen Proben laufen unterschiedlich schnell in individuellen Kapillaren einer vorgefertigten Gelkartusche in Richtung einer positiv geladenen Anode. Ein Detektor identifiziert sie und misst ihr Fluoreszenzsignal, welches dann von einem Photoelektronenvervielfacher-Detektor (*Photomultiplier Detector*) in elektronische Daten konvertiert und an die QIAxcel *ScreenGel Software* weitergegeben wird.

Dank der hohen Sensitivität dieses Systems lassen sich DNA-Fragmente mit einer höheren Genauigkeit als bei herkömmlichen Gelelektrophoresen nachweisen.

#### **2.4.6 DNA-Aufreinigung**

Die Aufreinigung von erfolgreich amplifizierter DNA wird fast ausschließlich mithilfe des QIAQuick PCR *Purification Kit* von Qiagen durchgeführt.

Dieses enthält die Materialien, um PCR-Produkte mit einer Größe von über 100 Basenpaaren in einem Silicamembran-basierten Verfahren aufzureinigen.

In einem ersten Bindungsschritt wird die Filtersäule in ein Reagiergefäß eingesetzt und mit 450 µl Bindungspuffer PB sowie den 45 µl PCR-Produkt befüllt. Durch eine 30 Sekunden lange Zentrifugation bei 13.000 *rpm* (Umdrehungen pro Minute) bindet die DNA an den Filter in der Säule.

Nachdem der Durchfluss verworfen wurde, wird im zweiten Schritt 750 µl Waschpuffer PE auf die Filtersäule gegeben, diese für eine Minute stehen gelassen und danach erneut 30 Sekunden lang zentrifugiert. Der Durchfluss wird wieder verworfen und die Filtersäule für 30 Sekunden zentrifugiert, um verbliebene Ethanolreste zu entfernen.

Im letzten Schritt wird die DNA eluiert. Dazu wird die Filtersäule in ein neues, beschriftetes Reagiergefäß eingesetzt, mit 30 µl Elutionspuffer EB befüllt und eine weitere Minute stehen gelassen. Durch eine letzte, 30 Sekunden lange Zentrifugation wird die aufgereinigte DNA in einem Elutionsvolumen von 30 – 50 µl in dem Reagiergefäß gesammelt.

Das aufgereinigte PCR-Produkt kann dann gemessen und für die Sequenzierung vorbereitet oder bei -20 °C gelagert werden.

#### **2.4.7 Konzentrationsmessung**

Die Konzentration der aufgereinigten DNA wird mit dem Vollspektrum-UV-Spektrophotometer NanoDrop 2000 von Thermo Scientific™ und der dazugehörigen *Software* gemessen. So ist es möglich, innerhalb von unter fünf Sekunden pro Probe DNA-Mengen ab 2 ng/µl in einem Volumen von 0,5 – 2 µl zu quantifizieren.

Dazu wird eine entsprechende Menge (standardmäßig 1 µl) des zu messenden, aufgereinigten PCR-Produkts auf den unteren Messsockel des Spektrophotometers pipettiert und der Probenarm des Geräts vorsichtig abgesenkt. Die Oberflächenspannung der Probe hält diese zwischen der sendenden optischen Faser im Probenarm und der empfangenden optischen Faser im Messsockel.

Die Messung wird am PC in der entsprechenden *Software* gestartet und erfolgt mittels einer Xenon-Lampe als Lichtquelle und einem Spektrometer mit linearem CCD-*Array* als Sensor, der das durch die Probe tretende Licht analysiert.

Die DNA-Konzentration der Probe wird in der Software angezeigt und auf dem Reagiergefäß mit Ethanol-festem Stift notiert. Nach jeder Probenmessung werden beide Messsockel mit einem trockenen Präzisionstuch vorsichtig abgewischt, damit

kein Anteil der Probe im Gerät verbleibt. Um valide Werte zu erhalten und Probleme auszuschließen, sollte vor der ersten Messung oder alle 30 Minuten eine Messung mit einem „*Blanking*“ Puffer durchgeführt werden.

#### **2.4.8 Sequenzierung**

Zur Sequenzierung nach Sanger mittels Kettenabbruch- bzw. Didesoxymethode werden die aufgereinigten PCR-Produkte in einem vorgegebenen Ansatz in mit zugehörigen Etiketten versehenen *SafeSeal* Reagiergefäßen an Eurofins Genomics GmbH (Ebersberg, Deutschland) oder LGC Genomics GmbH (Berlin, Deutschland) geschickt.

Für die Sequenzierung bei Eurofins besteht der Ansatz aus demjenigen Volumen an PCR-Produkt, das 300 ng DNA enthält, 2 µl Sequenzierungsprimer (Konzentration 10 µM) sowie der Menge an Nuklease-freiem Wasser, die benötigt wird, um das Gesamtvolumen auf 17 µl zu ergänzen. Pro Probe wird dieser Ansatz zweimal hergestellt, einmal mit dem *Forward*-Primer und einmal mit dem *Reverse*-Primer.

Der bei ausreichend hoher Nukleinsäure-Konzentration bevorzugt verwendete Ansatz für die Sequenzierung bei LGC besteht aus 560 ng DNA in 12 µl Volumen sowie 2 µl Sequenzierungsprimer.

## **2.5 Datenverarbeitung und statistische Auswertung**

### **2.5.1 Sequenzanalyse in „Geneious“**

Die von Eurofins und LGC übermittelten Sequenzdateien werden in die *Software* „Geneious® 10.2.2“ (<https://www.geneious.com>) von Biomatters Ltd. (Auckland, Neuseeland) eingepflegt. Mit dieser lassen sich die *Forward*- und *Reverse*-Sequenz jeder Probe und deren Chromatogramme in einem *Alignment* darstellen, auf Fehler überprüfen und zu einer Konsensussequenz zusammenführen.

Die überprüften Sequenzen werden dann in einer Datenbank gesammelt und in den *Alignments* sowie phylogenetischen Bäumen weiter analysiert.

### **2.5.2 SPSS**

SPSS Statistics ist eine *Software* der IBM Corporation (Armonk, USA) für die Durchführung, Darstellung und Weiterverarbeitung statistischer Analysen.

### 2.5.3 Prism

GraphPad Prism von GraphPad Software Inc. (San Diego, USA) ist eine *Software* zur Organisation, Analyse und grafischen Darstellung von Daten, die in dieser Arbeit vor allem zur Erstellung von Abbildungen und Graphen verwendet wurde.

### 2.5.4 SeqFeatR

SeqFeatR wird genutzt, um Sequenzmuster zu identifizieren, die signifikant häufiger bei bestimmten HLA-Merkmalen des Wirts auftreten und mögliche virale *Immunescape*-Mutationen darstellen können (Budeus et al., 2016).

Als *Input* in SeqFeatR dient ein *Multiple Sequence Alignment* (MSA) der zu analysierenden Sequenzen im FASTA-Format. Wichtig ist, dass die Sequenzen im *Alignment* und somit in der FASTA-Datei so benannt sind, dass die Zahlen der HLA-Merkmale immer die gleichen Positionen in dem Namen einnehmen. In der Maske der SeqFeatR-*Website* wird dann angegeben, an welchen Positionen im Sequenznamen die jeweiligen HLA-Typen stehen. Außerdem kann man die Werte des grafischen Outputs festlegen, die Rechenmethode auswählen (Exakter Test nach Fisher) und weitere Einstellungen vornehmen, zum Beispiel ob eine p-Wert-Korrektur mittels *False discovery rate* (FDR) durchgeführt werden soll.

Die *Software* erstellt dann intern für jede Aminosäuren-Position im *Alignment* eine Häufigkeitstabelle in Bezug auf das Vorliegen oder Fehlen jeweils einer bestimmten Aminosäure an der Position und eines gewissen HLA-Typs und führt einen Exakten Test nach Fisher durch, um zu ermitteln, ob eine Häufung der Aminosäure bei dem HLA-Typ signifikant ist. Des Weiteren errechnet SeqFeatR eine *Odds Ratio*, die die Stärke der Assoziation quantifiziert.

Zusätzlich werden die Ergebnisse grafisch in Form von Manhattan *Plots* ausgegeben. In diesen sind die Aminosäure-Positionen der Sequenz auf der X-Achse und die dazugehörigen p-Werte der Aminosäuren auf der Y-Achse dargestellt.

### 2.5.5 IEDB

Die IEDB, also *Immune Epitope Database and Analysis Resource*, ist eine Online-Datenbank, die unter anderem ein MHC I-*Binding Prediction Tool* anbietet. Dieses erlaubt, möglicherweise an MHC I bindende Epitope innerhalb einer Sequenz zu ermitteln (Vita et al., 2019).

Hierzu werden die Sequenzen im FASTA-Format hochgeladen und die von IEDB empfohlene Prädiktionmethode benutzt (zur Zeit der Benutzung „NetMHCpan 4.0“). Außerdem stellt die Website eine umfassende Datenbank an Epitopen zur Verfügung, mit deren Hilfe man nachschlagen kann, ob bestimmte Epitope in Publikationen erwähnt und in *Assays* untersucht worden sind.

### **2.5.6 FigTree**

FigTree ist ein Online-Programm, welches ermöglicht, phylogenetische Bäume aus Sequenz-*Alignments* zu generieren, diese zu bearbeiten und zu speichern.

### **2.5.7 Hauptkomponentenanalyse (PCA) mittels R**

Die Hauptkomponentenanalyse ist ein Verfahren aus der multivariaten Statistik, mit dem umfangreiche Datensätze vereinfacht und übersichtlich dargestellt werden können. Dazu werden miteinander korrelierende Variablen in möglichst wenigen Hauptkomponenten zusammengefasst. Diese können dann bildlich in einem *Plot* dargestellt werden.

In dieser Arbeit wurde die Hauptkomponentenanalyse benutzt, um mögliche Zusammenhänge zwischen der Hypervariablen Region und den exprimierten Aminosäuren aufzudecken.

Dazu wurde die Funktion „prcomp“ des Programms R verwendet.

### 3 Ergebnisse

Um anhand von frühen Proben des Anti-D-Ausbruches die Anpassung des Hepatitis C-Virus an seinen Wirt und die zugrunde liegenden evolutionären Mechanismen zu untersuchen, sollten die Genome der in uns zur Verfügung stehenden frühen sowie späten Anti-D-Proben enthaltenen Hepatitis C-Viren sequenziert werden.

Zu diesem Zweck wurden Anti-D-Proben von 1979 bis zu den 2000er-Jahren zusammengetragen, sortiert sowie katalogisiert und ein PCR-Protokoll etabliert.

Die sequenzierten Genome sollten dann in einer Anti-D-Sequenzdatenbank zusammengeführt, in Form von *Alignments* und phylogenetischen Bäumen dargestellt und longitudinal analysiert werden.

Dies sollte ermöglichen, Insertionen und andere Mutationen im HCV-Genom zu identifizieren, möglichst funktionell zu charakterisieren und vorliegende Quasispezies zu rekonstruieren.

Ein besonderes Augenmerk lag dabei auf den Strukturproteinen und der Hypervariablen Region 1 (HVR1) im E2-Protein, um dort die für die Anti-D-Kohorte charakteristischen Insertionen, sowie Mutationen in Abhängigkeit vom HLA-Typ zu untersuchen.

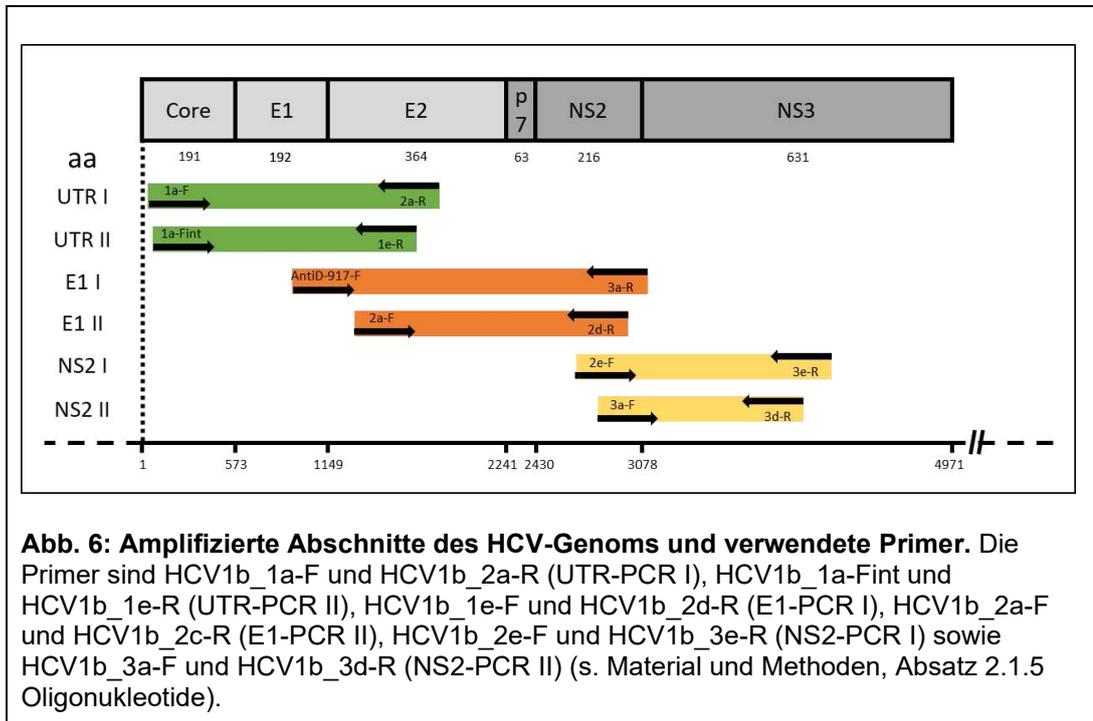
#### 3.1 Etablierung der PCR

Um die Proben aus der Anti-D-Kohorte zu amplifizieren und sequenzieren, wurde ein Reverse Transkriptions- und PCR-Protokoll etabliert und im Verlauf mehrerer Experimente optimiert.

Als Protokoll für die Reverse Transkription diente das im Abschnitt 2.4.3 zuvor beschriebene, in der Arbeitsgruppe bereits etablierte Protokoll.

Wie in Abbildung 6 dargestellt, sollte das HCV-Genom von der 5'-*untranslated region* (UTR) bis über das Nichtstrukturprotein NS2 hinaus mithilfe jeweils zweischrittiger *Nested* PCRs (Protokoll s. Material und Methoden, Absatz 2.4.4 *Nested* PCR) in drei Abschnitten amplifiziert werden. Diese reichen von der UTR- bzw. *Core*-Region bis E2 („UTR-PCR“), von E1 bis NS2/3 („E1-PCR“) und von NS2

bis NS3 („NS2-PCR“) und decken somit überlappend das HCV-Genom von der UTR bis in NS3 hinein ab.



Die UTR-PCR mit den Primern HCV1b\_1a-F und HCV1b\_2a-R (PCR I) sowie HCV1b\_1a-Fint und HCV1b\_1e-R (PCR II) war in dieser ursprünglichen Form zur Amplifikation des ersten Abschnitts geeignet.

Auch mithilfe der NS2-PCR mit den Primern HCV1b\_2e-F und HCV1b\_3e-R (PCR I) sowie HCV1b\_3a-F und HCV1b\_3d-R (PCR II) gelang es, den gewünschten dritten Abschnitt zuverlässig zu amplifizieren.

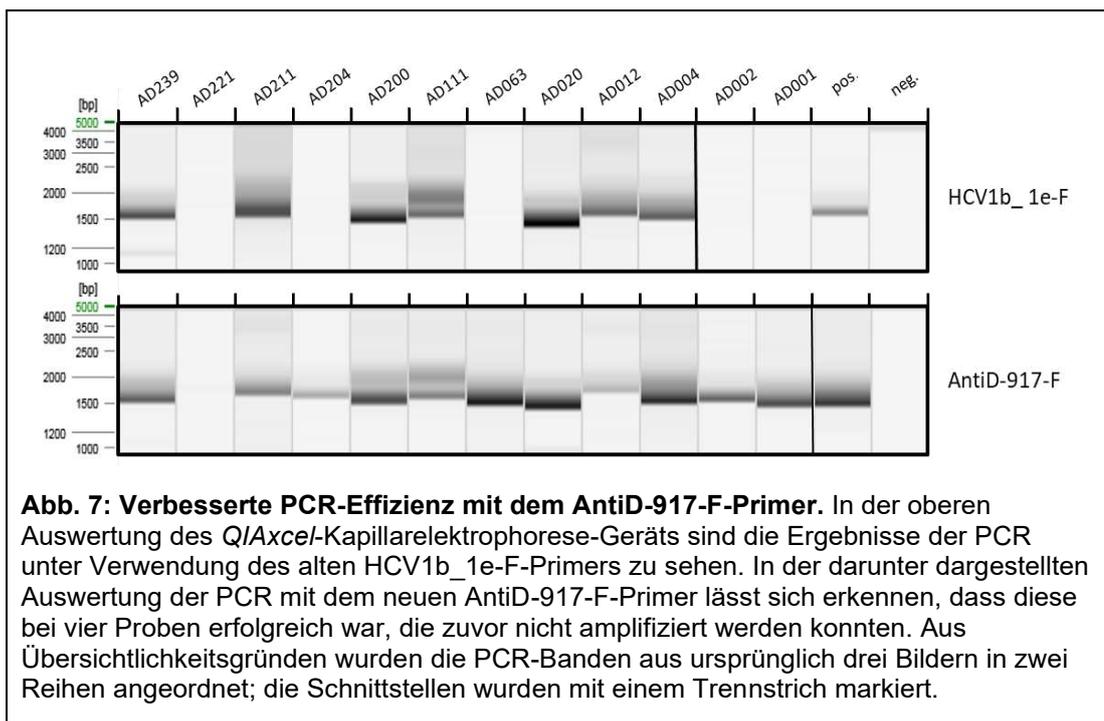
Für die E1-PCR wurden die Primer HCV1b\_1e-F und HCV1b\_3a-R (PCR I, Länge des Fragments: 2028bp) sowie HCV1b\_2a-F und HCV1b\_2d-R (PCR II, Länge des Fragments: 1728bp) neu ausgewählt.

Um zu analysieren, ob der Grund für die dennoch weiterhin geringe Amplifikationseffizienz der E1-PCR-Primer HCV1b\_1e-F und HCV1b\_3a-R sowie HCV1b\_2a-F und HCV1b\_2d-R die schlechte Bindung an das Genom sind, wurden die Primer-Bindungsstellen in den bereits generierten 5'UTR-E1- und NS2-NS3-Sequenzen in der Software „Geneious“ auf Variabilität überprüft. Während die Bindungsstelle für den Primer HCV1b\_3a-R keinerlei Varianz aufwies, zeigten sich bei der Bindungsstelle der Primer HCV1b\_2a-F und HCV1b\_2d-R in einigen Isolaten

unterschiedliche Nukleotide. Da diese sehr selten waren und überwiegend in der Mitte der Primer auftraten, beeinträchtigten sie die Bindung der Primer wahrscheinlich nicht oder nur sehr wenig. Bei der Analyse der Bindungsstelle des Primers HCV1b\_1e-F zeigte sich jedoch eine hohe Varianz, was wahrscheinlich die geringe Amplifikationseffizienz erklärt.

Die Basen 917 bis 936 der Referenzsequenz 1b.DE.HCV-AD78P1.AJ132997, lokalisiert kurz nach dem Anfang der E1-Region, wurden als in der Kohorte wenig variable Region identifiziert. Die entsprechende Sequenz GAAGTGCACAACGTGTCCGG wurde als neuer Primer AntiD-917-F synthetisiert und überprüft.

Mit dem ursprünglichen Primer 1b\_1e-F konnten sieben von zwölf Isolaten amplifiziert werden, mit dem neuen AntiD-917-F-Primer elf von zwölf Isolaten (siehe Abb. 7).



Da das neue E1-PCR I-Produkt unter Verwendung des AntiD-917-F-Primers 2170 statt 2028 Basen lang war, wurde die Polymerisationsphase des PCR I-Programms von zwei Minuten auf zweieinhalb Minuten pro Zyklus erhöht. Um ein einfacheres Handling des Protokolls zu ermöglichen, wurden alle drei Abschnitte von nun an sowohl in PCR I als auch in PCR II zweieinhalb Minuten pro Zyklus polymerisiert.

### 3.2 Amplifikation früher Proben

Nachdem die PCR erfolgreich etabliert worden war, konnte mit der Amplifikation früher Proben begonnen werden.

Dazu wurden zunächst Seren aus den Jahren zwischen 1978/79 und 2009 aus der Biobank des Kompetenznetz Hepatitis Essen abgeholt. Die Proben waren aus Kapazitätsgründen nur bei -20 °C gelagert. Nach dem Transport nach Düsseldorf in mit Trockeneis gekühlten Transportboxen wurden sie bei -80 °C gelagert.

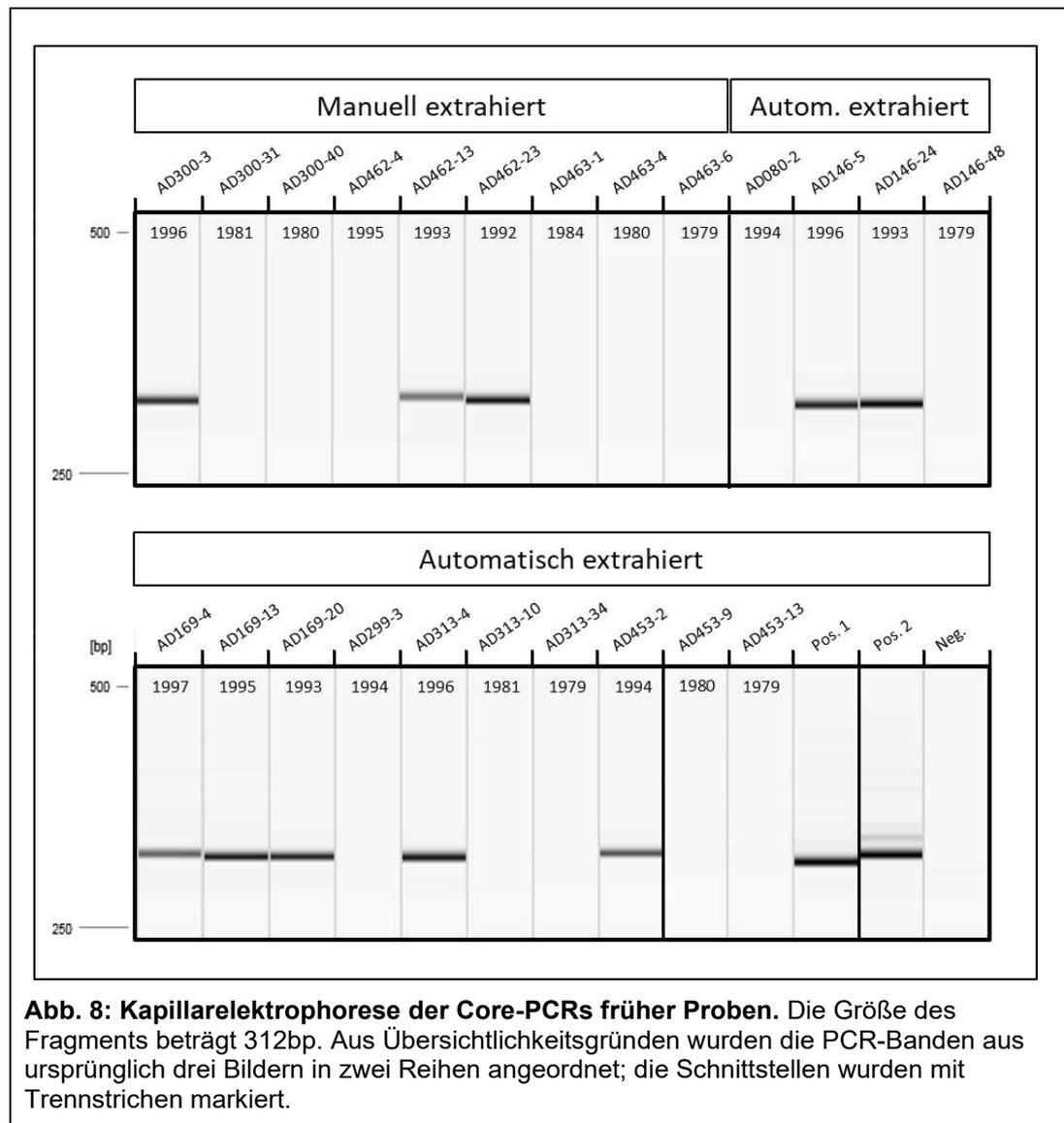
In den folgenden Wochen wurden im -80 °C-Kühlraum aus 15 Kisten 115 Proben-Boxen von rund 477 Patientinnen und Patienten katalogisiert und, soweit möglich, nach Anti-D-Nummern (AD \_\_\_ - \_\_) sortiert.

Da viele der Proben in alten Gefäßen wie handbeschrifteten Glasampullen oder Filmdosen, in oft undichtem oder beschädigtem Zustand, lose in kleinen Kisten gelagert waren, wurden möglichst viele der Proben in moderne Gefäße umgefüllt. Um ein häufiges Auftauen zu vermeiden, wurden die zum Umfüllen aufgetauten Proben im selben Arbeitsschritt einer RNA-Extraktion und/oder Viruslast-Bestimmung nach vorheriger Behandlung zugeführt.

Initial wurden zunächst 216 Proben von neun Patientinnen in neue *Cryo*-Gefäße umgefüllt, beschriftet und bei -80 °C gelagert.

Nach dieser anfänglichen Sortierung und Katalogisierung wurde mit der Durchführung der UTR-, E1- und NS2-PCRs begonnen. Dabei stellte sich heraus, dass es nicht möglich war, die drei Strukturprotein-Abschnitte von Proben aus den Jahren 1979 bis 1996 zu amplifizieren (Daten nicht gezeigt). Aus diesem Grund wurde die RNA-Qualität von neun Proben mittels der sehr sensitiven *Core*-PCR (s. Material und Methoden, Abschnitt 2.4.4 *Nested* PCR) überprüft. Von diesen neun *Core*-PCRs waren nur die von drei Proben aus den Jahren 1992, 1993 und 1996 erfolgreich.

Auch wenn in der Literatur beschrieben ist, dass eine automatische RNA-Extraktion ebenso gute Ergebnisse wie die manuelle RNA-Extraktion liefert (Knepp et al., 2003), wurde, um dies zu überprüfen, die RNA aus den neun Seren nochmal mittels *QiAmp Viral RNA Kit* manuell extrahiert und wie zuvor cDNA synthetisiert. Aus der gewonnenen cDNA konnten aber erneut nur die drei gleichen Proben aus den Jahren 1992, 1993 und 1996 amplifiziert werden (Daten nicht gezeigt).



Die automatische RNA-Extraktion, cDNA-Synthese und *Core*-PCR wurde mit 14 weiteren Proben von sechs Patientinnen durchgeführt. Insgesamt konnten in diesem Schritt von neun manuell extrahierten Proben (AD 300, 462, 463) und 14 automatisch extrahierten Proben (AD 080, 146, 169, 299, 313, 453) mit der sensitiven *Core*-PCR nur zehn Proben aus den Jahren 1992 – 1997 amplifiziert werden, während 13 Proben aus den Jahren 1979 – 1995 nicht amplifizierbar waren (siehe Abb. 8).

Um festzustellen, ob das Alter der Probe ausschlaggebend für die geringe Amplifikationseffizienz war, wurden zusätzlich zu den in Abbildung 8 dargestellten Proben weitere Proben untersucht. Insgesamt wurde RNA aus 49 Proben aus den Jahren 1979 – 1997 von 19 Patientinnen, von denen longitudinale Proben vorhanden waren, extrahiert, cDNA synthetisiert und diese mittels *Core*-PCR untersucht.

Von den 20 Proben aus den Jahren 1979 – 1989 war keine Probe mittels *Core*-PCR amplifizierbar. Von den 29 Proben aus den Jahren 1992 – 1997 konnten hingegen 23 (79,3 %) mittels *Core*-PCR amplifiziert werden.

Um einen möglichen Grund für die negativen *Core*-PCRs abzuklären, wurde die Viruslast (VL) einiger Proben ermittelt. Bei sieben der Proben aus den Jahren 1979, 1993, 1994, 1995, 1995 und 1997 war das Serum trotz vorheriger Behandlung mit EDTA und Zentrifugation zu stark geronnen für eine Messung. Es standen somit noch 16 Proben von acht Patientinnen zur Analyse zur Verfügung.

Identifier	Entnahmedatum	Core-PCR erfolgreich?	Viruslast in IU/ml
AD453-13	31.10.1979	Nein	112
AD313-34	01.11.1979	Nein	Sample clot
AD463-6	06.11.1979	Nein	Target not detected
AD453-9	08.01.1980	Nein	< 140
AD463-4	14.04.1980	Nein	Target not detected
AD300-40	26.06.1980	Nein	Target not detected
AD300-31	05.01.1981	Nein	420
AD313-10	08.09.1981	Nein	217
AD463-1	31.01.1984	Nein	Target not detected
AD462-23	11.05.1992	Ja	< 210
AD462-13	03.05.1993	Ja	Sample clot
AD146-24	27.08.1993	Ja	52.346
AD169-20	03.12.1993	Ja	18.004
AD080-2	05.01.1994	Nein	Sample clot
AD299-3	01.06.1994	Nein	Target not detected
AD453-2	16.09.1994	Ja	1.932
AD462-4	10.04.1995	Nein	Sample clot
AD169-13	06.11.1995	Ja	Sample clot
AD300-3	07.03.1996	Ja	1.456
AD146-5	07.05.1996	Ja	179.508
AD313-4	17.07.1996	Ja	273.070
AD169-4	06.05.1997	Ja	Sample clot

**Tabelle 3: Mittels COBAS ermittelte Viruslasten der frühen Proben.** Die Proben wurden nach dem Entnahmedatum in aufsteigender Reihenfolge sortiert. Dargestellt ist des Weiteren das Ergebnis der *Core*-PCR.

In den acht verwertbaren Proben aus den Jahren 1979 – 1984 war eine Viruslast in vier Fällen nicht mehr nachweisbar (unter 15 IU/ml), in vier anderen Fällen sehr gering (unter 500 IU/ml) (siehe Tab. 3). Keine dieser acht Proben hatte sich zuvor mittels *Core*-PCR amplifizieren lassen. Bei den acht verwertbaren Proben aus den Jahren 1992 – 1996 ließen sich in je zwei Fällen eine Viruslast über 1.000 IU/ml, 10.000 IU/ml oder sogar 100.000 IU/ml feststellen, aber auch aus diesem Zeitraum gibt es jeweils eine Probe mit sehr geringer (AD 462-23 aus 1992, VL < 210 IU/ml)

oder nicht nachweisbarer Viruslast (Probe AD 299-3 aus 1994) (siehe Tab. 3). Letztere Probe sowie zwei nicht messbare Proben aus den Jahren 1994 und 1995 hatten sich zuvor ebenfalls nicht mittels *Core*-PCR amplifizieren lassen.

Die nicht nachweisbaren oder sehr niedrigen Viruslasten in den stichprobenartig untersuchten Proben aus den Jahren 1979 – 1984 sowie 1992 und 1994 gehen vermutlich auf eine Degradierung der Proben durch unsachgemäße Lagerung bei zu hohen Temperaturen und in teils ungeeigneten, undichten oder sogar beschädigten Gefäßen zurück. Die älteren Proben aus den Jahren 1979 – 1984 sind deutlich stärker von dieser Entwicklung betroffen als die Proben aus den 90er-Jahren, da sie höchstwahrscheinlich länger und häufiger schädlichen Einflüssen wie erhöhten Temperaturen oder häufigem Auftauen ausgesetzt waren.

Diese Umstände hatten zur Folge, dass keine der Proben aus den Jahren 1979 – 1984 in der sensitiven *Core*-PCR und den etablierten drei Strukturprotein-PCRs amplifizierbar war. Da die Amplifikation mittels Strukturprotein-PCRs auch bei den 23 *Core*-positiven Proben aus den 90er-Jahren nur bei vier der 69 Abschnitte möglich war (ca. 5,8 %; AD 168-3 (08/1997, NS2, VL unbekannt), AD 305-3 (04/1995, NS2, VL unbekannt), AD 313-4 (07/1996, E1 und NS2, VL 273.070 IU/ml)), sahen wir davon ab, weitere Versuche mit Probenmaterial aus den „frühen Proben“ durchzuführen.

### **3.3 Amplifikation später Proben**

Da die frühen Proben aus der Zeit des Ausbruchs und den 90er-Jahren nicht verlässlich mit der UTR-, E1- und NS2-PCR amplifizierbar waren, konzentrierten wir uns stattdessen auf die Amplifikation und Sequenzierung weiterer Anti-D-Proben aus den 2000er-Jahren, um die Anti-D-Sequenzdatenbank auszubauen und diese zu analysieren.

Insgesamt stand in Düsseldorf laut Aufzeichnungen positives Probenmaterial aus den 2000er-Jahren von 113 Patientinnen zur Verfügung, von mindestens acht Patientinnen war kein Material mehr vorrätig (AD 049, 062, 070, 072, 093, 194, 212, 371). Zehn der Proben (AD 343, 346, 347, 353, 366, 399, 408, 413, 421, 431) standen erst nach Beendigung dieser Arbeit zur Verfügung. Zwei weitere, AD 385 und AD 429, wurden nur zu Beginn für die PCR-Etablierung verwendet, hier lagen die Sequenzen bereits aus vorherigen Arbeiten vor (Ruhl et al., 2011; Lipskoch et al., 2014).

Insgesamt wurde aus 39 Proben (35 aus dem Jahr 2009, zwei aus 2010, zwei aus 2011), von denen RNA-Eluate bereits vorhanden waren, cDNA synthetisiert. Im Verlauf wurde RNA aus Serum-Proben von elf dieser Patientinnen sowie 66 weiteren (zwei aus dem Jahr 2006, 13 aus 2008, 62 aus 2009) extrahiert und aus dieser cDNA synthetisiert.

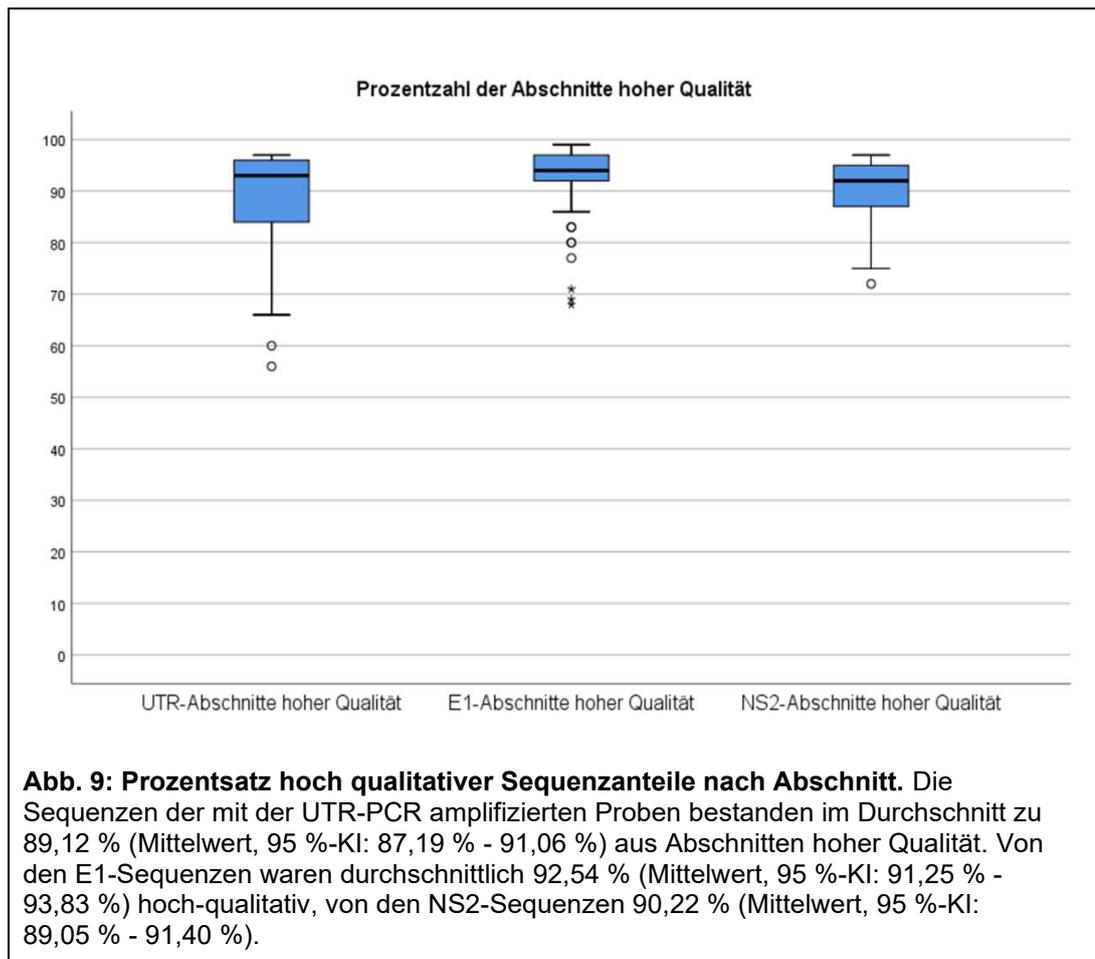
Die cDNA aus 101 Proben – 25 aus vorhandenen RNA-Eluaten, 76 aus extrahierter RNA – wurde in den Strukturprotein-PCRs weiterverwendet.

Die Anti-D-Proben aus den 2000er-Jahren waren mit den etablierten Strukturprotein-PCRs effizient amplifizierbar; die UTR-PCR war in 97 der 101 Proben (96 % (SD: 1,96 %)) positiv, die E1- und NS2-PCR in 98 der 101 Proben (97 % (SD: 1,71 %)). Drei der Proben (AD 103, 265 und 339) waren in allen drei Abschnitten negativ. Die Probe AD 024 war in sechs Versuchen nicht mit der UTR-PCR amplifizierbar, die anderen beiden Strukturprotein-PCRs waren hingegen erfolgreich.

Mittels Sanger-Sequenzierung konnte die amplifizierte DNA zuverlässig sequenziert werden. Von den 25 PCR-positiven Abschnitten aus RNA-Eluaten waren nur ein UTR-Abschnitt und ein E1-Abschnitt (je 4 %) nicht zufriedenstellend zu sequenzieren (beide von AD 111). Von den aus extrahierter RNA gewonnenen PCR-Produkten ließen sich 68 von 72 (UTR, 94,4 %; AD 175, 235, 249, 258 negativ), 68 von 73 (E1, 93,2 %; AD 249, 258, 288, 298 negativ) und 73 von 73 (NS2, 100 %) zufriedenstellend sequenzieren.

Die generierten Sequenzen wurden im Programm „Geneious“ zusammengetragen und bearbeitet. Im Durchschnitt wurden in „Geneious“ 89,12 % jedes UTR-Abschnitts, 92,54 % jedes E1-Abschnitts und 90,22 % jedes NS2-Abschnitts als von sehr hoher Qualität eingestuft (siehe Abb. 9). Die Qualität wird vom *base calling* Programm festgelegt; je höher die Qualität, desto niedriger ist die Wahrscheinlichkeit eines Fehlers, also einer falsch zugeordneten Base, an einer Position.

Da die Segmente von schlechterer Qualität meist Primer-fern an den Enden der generierten Abschnittsequenzen lagen, konnte durch das Abgleichen der vier Chromatogramme der jeweiligen Forward- und Reverse-Sequenzen in den überlappenden Abschnitten dennoch mit hoher Verlässlichkeit eine Konsensussequenz jeder Probe erstellt werden.



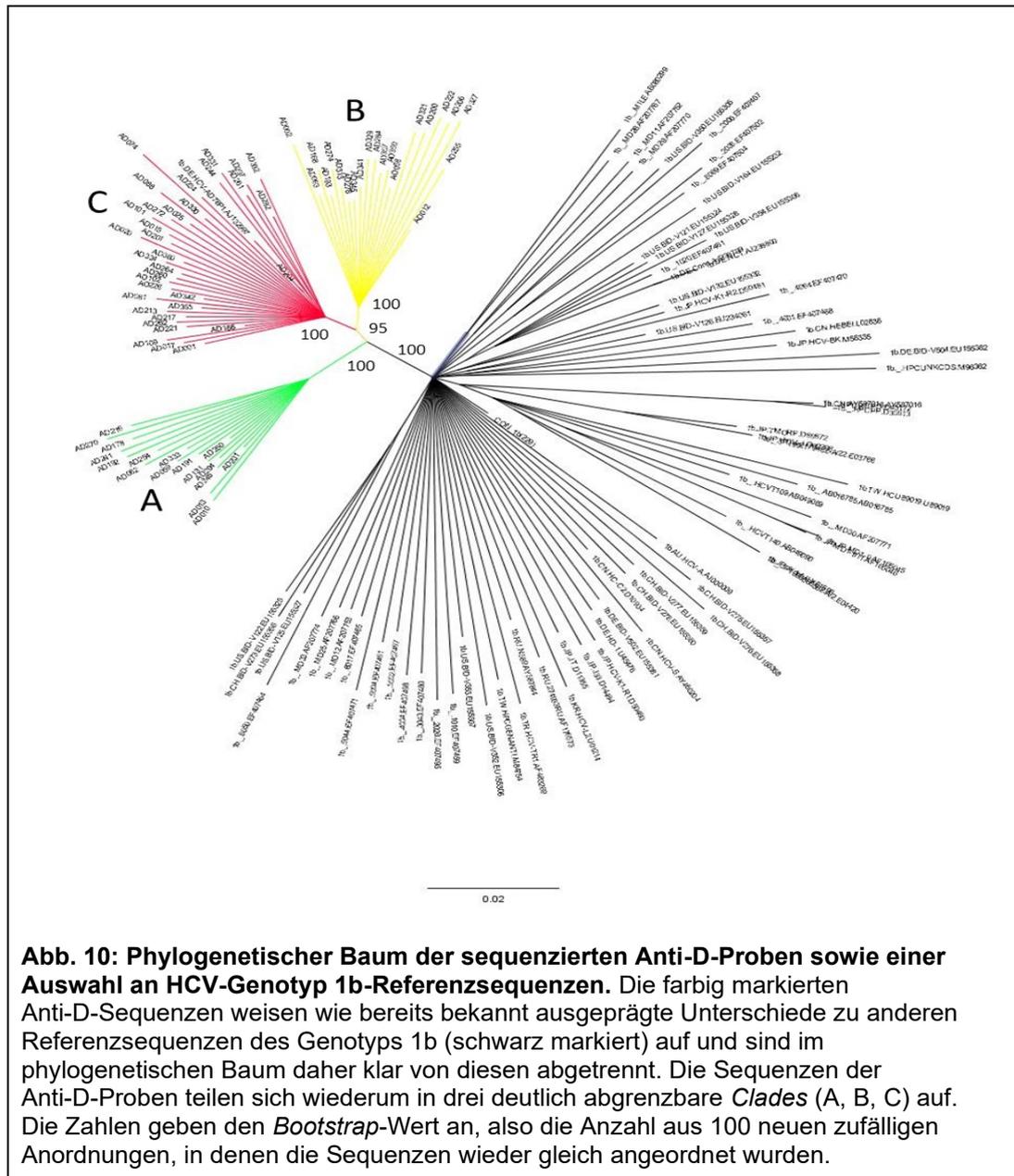
Insgesamt konnten so 89 Anti-D-Sequenzen von UTR bis NS2 sowie acht unvollständige Sequenzen generiert werden.

Diese wurden zusammengetragen mit bereits vorhandenen, unvollständigen, meist NS3 bis NS5B oder auch *Core* bis NS2 umfassenden Anti-D-Sequenzen aus vorherigen Arbeiten (Ruhl et al., 2011; Lipskoch et al., 2014).

So konnte eine Anti-D-Sequenzdatenbank mit insgesamt 218 Sequenzen von 138 Patientinnen erstellt werden, darunter 74 komplette HCV-Genome.

In der Literatur ist beschrieben, dass sich die Genom-Sequenzen der Anti-D-Kohorte deutlich von den Sequenzen anderer Hepatitis C-Viren des Genotyps 1b unterscheiden und sich wiederum in drei voneinander abgrenzbare *Clades* unterteilen lassen (Ruhl et al., 2011).

Um dies zu untersuchen, wurden die 74 kompletten Genome mit 84 Referenzisolaten *aligned* und mit der *Neighbor-Joining*-Methode unter Verwendung des Tamura-Nei-Distanzmodells ein phylogenetischer Baum errechnet (siehe Abb. 10).



Mit dem phylogenetischen Baum konnte bestätigt werden, dass sich die Sequenzen der Anti-D-Kohorte deutlich von Referenzsequenzen des Genotyps 1b unterscheiden. Des Weiteren sieht man die charakteristische Aufteilung der Anti-D-Proben auf die drei *Clades* A, B und C. Diese Aufteilung ist zurückzuführen auf drei Varianten in dem Blut der Infektionsquelle des Ausbruchs, welche sich dann unterschiedlich stark in den hergestellten Anti-D-Chargen und den damit behandelten Patientinnen durchgesetzt haben.

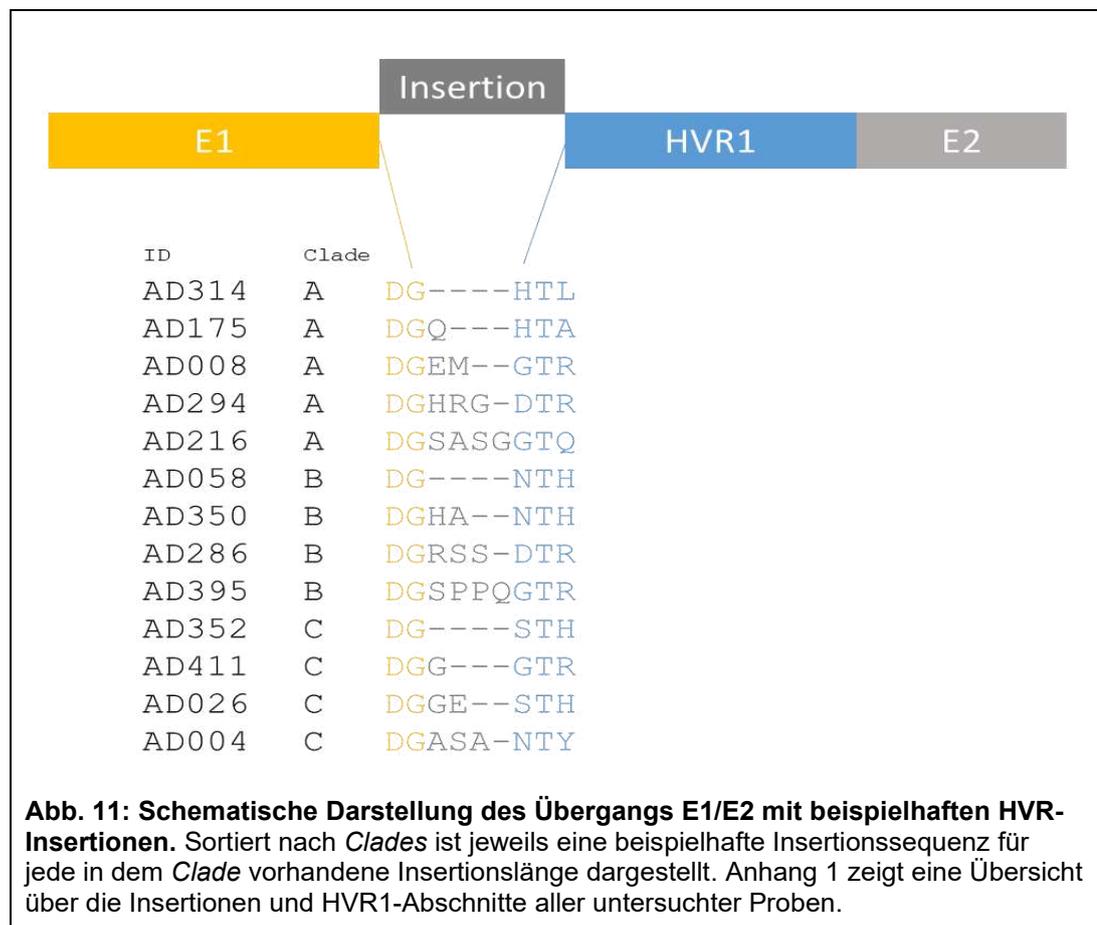
In der Abbildung sind an jeder Aufzweigung des Baums *Bootstrap*-Werte dargestellt. Im *Bootstrapping*-Verfahren, einer statistischen Methode mittels *Resampling*,

beschreiben diese Werte in diesem Fall die Anzahl von hundert zufälligen Neuankordnungen, in denen die Proben wieder wie in der Abbildung dargestellt angeordnet wurden. Die hohen *Bootstrap*-Werte zeigen, dass sowohl die Abgrenzung der Anti-D-Proben von den Referenzsequenzen als auch die der *Clades* untereinander sehr verlässlich ist. Die Anti-D-Sequenzen werden in 100 aus 100 Fällen erneut abgegrenzt von den Referenzsequenzen des Genotyps 1b dargestellt.

Eine der Besonderheiten der Anti-D-Kohorte, die für die klare Abgrenzung von anderen Genotyp-1b-Sequenzen mitverantwortlich sein dürfte, ist die für die Kohorte charakteristische Insertion zwischen E1 von E2 direkt vor der Hypervariablen Region 1 (HVR1).

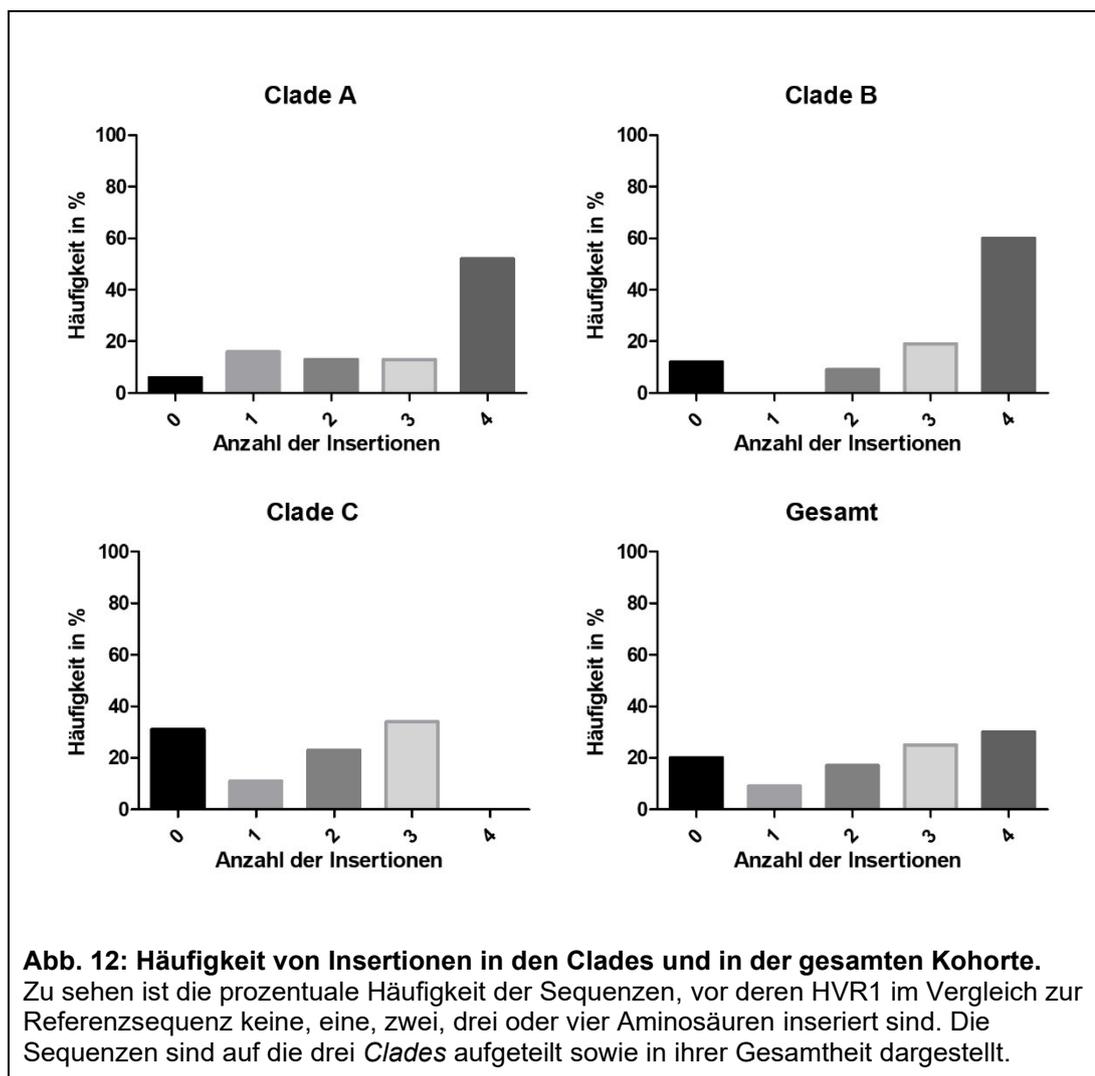
### 3.4 Analyse der Hypervariablen Region (HVR)

Um diese Insertion und die Hypervariable Region näher zu charakterisieren, wurden die 138 Anti-D-Sequenzen in einem *Alignment* angeordnet und nach Anzahl der Insertionen geordnet betrachtet, sowohl auf *Clades* aufgeteilt als auch in ihrer Gesamtheit.



Von den 138 untersuchten Sequenzen entfielen 31 Proben auf *Clade A* (22,5 %), 43 Proben auf *Clade B* (31,2 %) und 64 Proben auf *Clade C* (46,4 %).

In allen *Clades* gab es auch Proben ohne Insertion; in der gesamten Kohorte enthielten 20 Prozent keine Insertion. Unter den Proben mit Insertion stieg mit zunehmender Länge der Insertionen auch deren Häufigkeit. In neun Prozent der Proben lag eine Insertion einer Aminosäure vor, in 17 Prozent der Fälle waren zwei Aminosäuren inseriert, in 25 Prozent drei Aminosäuren und in 30 Prozent der Proben hatte die Insertion eine Länge von vier Aminosäuren (siehe Abb. 12).



Im Rahmen der auf *Clades* aufgeteilten Betrachtung fallen Unterschiede bezüglich der Verteilung der Häufigkeiten auf.

In *Clade A* sind fehlende Insertionen seltener als in der gesamten Kohorte (6 %) und Insertionen mit einer Länge von einer, zwei oder drei Aminosäuren ähnlich

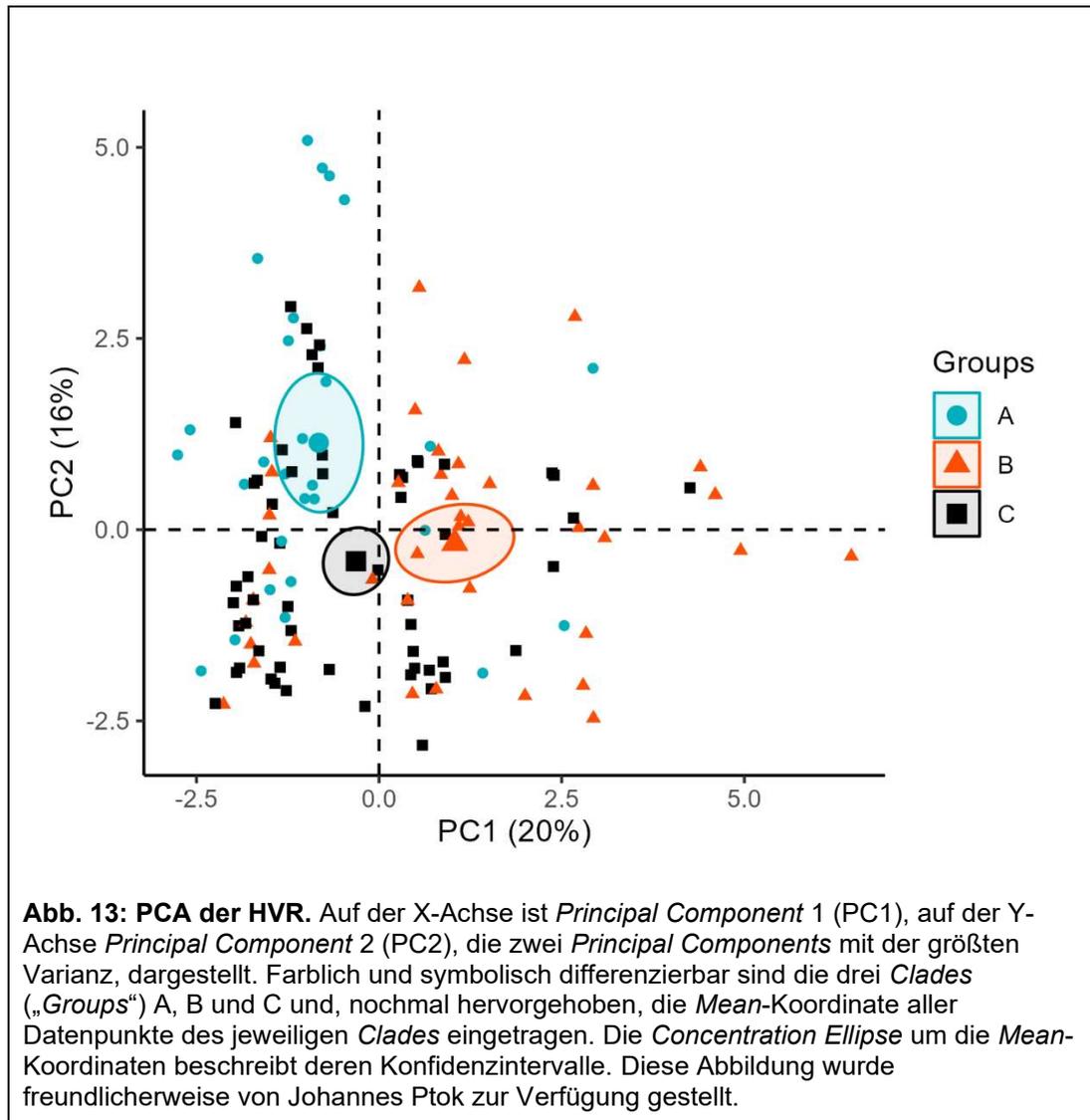
häufig (16 %, 13 % und 13 %). Überdurchschnittlich häufig (52 %) treten hingegen Insertionen mit einer Länge von vier Aminosäuren auf.

Letzteres gilt auch für *Clade B*; hier sind 60 Prozent aller Insertionen vier Aminosäuren lang. Auffällig ist, dass keine der Proben eine Insertion einer einzigen Aminosäure vorweist. Die Insertion fehlt in nur zwölf Prozent der Proben aus *Clade B*, Insertionen mit einer Länge von zwei und drei Aminosäuren liegen in neun und 19 Prozent der Proben vor.

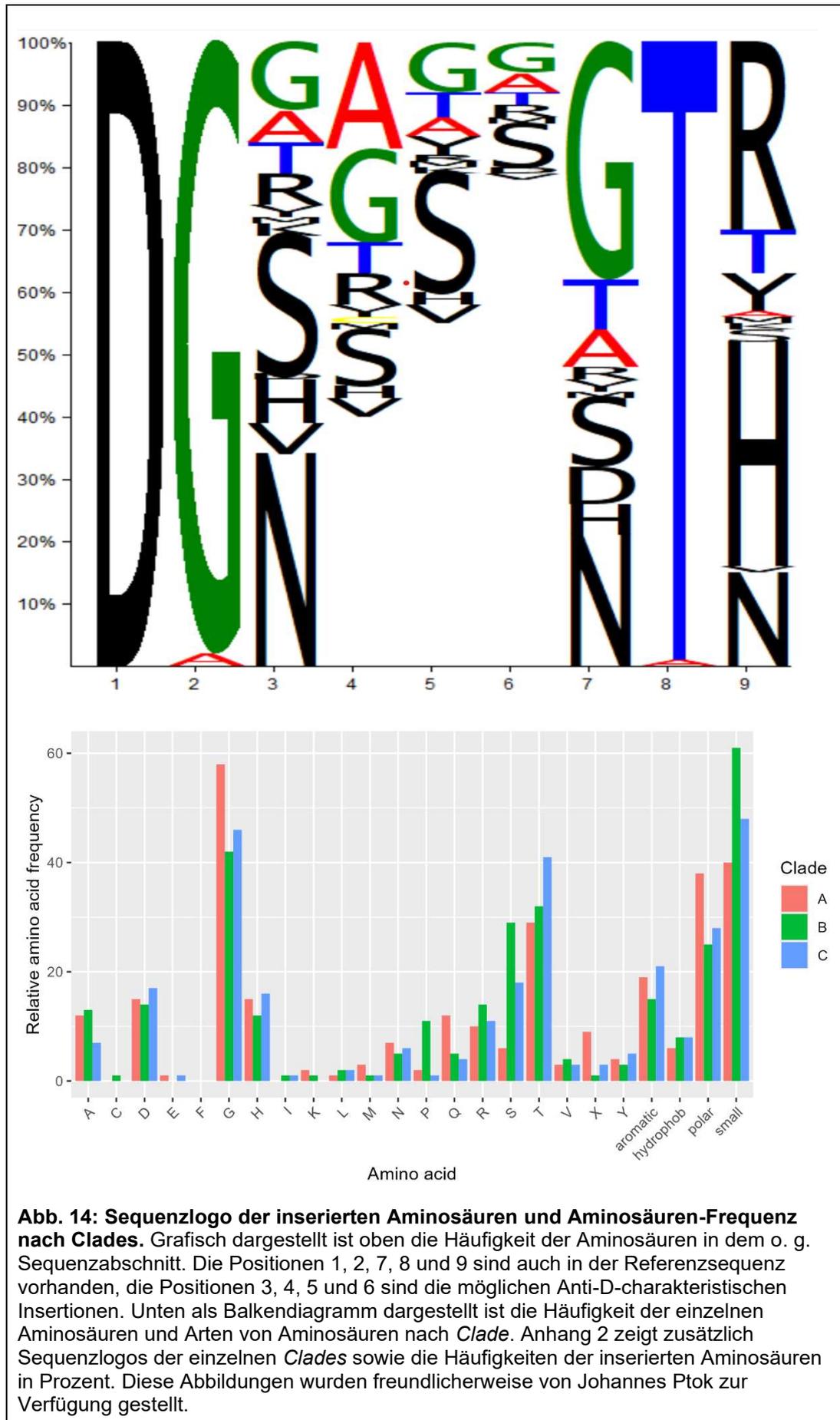
Die zu *Clade C* zugehörigen Sequenzen weichen von denen der anderen beiden *Clades* auffallend ab. So kommt die in *Clade A* und *B* am häufigsten vorliegende Insertionslänge von vier Aminosäuren in *Clade C* gar nicht vor. Überdurchschnittlich häufig (31 %) fehlt die Insertion ganz. Die restlichen Insertionslängen von einer, zwei und drei Aminosäuren steigen dagegen wie in der Gesamtkohorte relativ gleichmäßig an (11 %, 23 %, 34 %).

Um zu untersuchen, ob es über die Unterschiede im Hinblick auf die Insertionslänge und deren Häufigkeit hinaus weitergehende Unterschiede zwischen den Insertionen in den einzelnen *Clades* gibt, wurde eine Hauptkomponentenanalyse (*Principal Component Analysis*, PCA) durchgeführt.

Die Häufigkeit des Vorkommens der einzelnen Aminosäuren wurde genutzt, um eine PCA zu errechnen mit deren zusammengefasster Expression. Die Proben-IDs wurden je nach *Clade* unterschiedlich gefärbt. Proben von *Clade A* und *C* schienen sich im Hinblick auf die Expression von Aminosäuren zu ähneln, während *Clade B* sich von den anderen beiden *Clades* etwas zu unterscheiden scheint (siehe Abb. 13).



Betrachtet man die Häufigkeit der Aminosäuren an den Positionen der Insertion (wie in Abbildung 14 dargestellt) getrennt nach *Clades*, scheinen die Sequenzen aus *Clade A* dort durchaus eine abweichende Sequenz aufzuweisen (siehe Abbildungen im Anhang 2). So enthält die Insertion in *Clade A* an jeder Position am häufigsten die Aminosäure Glycin, während dies in *Clade C* nur an der ersten und zweiten Stelle der Insertion der Fall ist und in *Clade B* an jeder Stelle eine andere Aminosäure die häufigste ist.



**Abb. 14: Sequenzlogo der inserierten Aminosäuren und Aminosäuren-Frequenz nach Clades.** Grafisch dargestellt ist oben die Häufigkeit der Aminosäuren in dem o. g. Sequenzabschnitt. Die Positionen 1, 2, 7, 8 und 9 sind auch in der Referenzsequenz vorhanden, die Positionen 3, 4, 5 und 6 sind die möglichen Anti-D-charakteristischen Insertionen. Unten als Balkendiagramm dargestellt ist die Häufigkeit der einzelnen Aminosäuren und Arten von Aminosäuren nach *Clade*. Anhang 2 zeigt zusätzlich Sequenzlogos der einzelnen *Clades* sowie die Häufigkeiten der inserierten Aminosäuren in Prozent. Diese Abbildungen wurden freundlicherweise von Johannes Ptok zur Verfügung gestellt.

Des Weiteren wurde untersucht, ob die Proben der drei *Clades* unterschiedlich häufig z. B. polare oder aromatische Aminosäuren in ihrer Hypervariablen Region 1 exprimieren. Dies konnte nicht bestätigt werden; die Sequenzen aus *Clade A* scheinen mehr polare und weniger kleine Aminosäuren zu enthalten als diejenigen aus *Clades B* und *C*, aber dieser Effekt war nicht signifikant.

Signifikante Zusammenhänge zwischen den inserierten Aminosäuren und den HLA-Typen der Patientinnen konnten im Rahmen der PCA ebenfalls nicht gefunden werden.

Weitere Untersuchungen im Hinblick auf virusspezifische Faktoren, die möglicherweise mit den Insertionen assoziiert sein könnten, oder wirtsspezifische Faktoren wie die *Entry*-Rezeptoren konnten aufgrund von fehlenden Informationen nicht durchgeführt werden.

### 3.5 SeqFeatR-Analyse

Um zu untersuchen, ob gewisse Mutationen im HCV-Genom bei bestimmten HLA-Typen des Wirts gehäuft vorkommen, wurden die Sequenzen der Anti-D-Datenbank mit dem *Online-Tool* „SeqFeatR“ (Budeus et al., 2016) analysiert. Wenn sich in einem viralen Genom Mutationen in normalerweise an T-Zellen präsentierten Abschnitten (T-Zell-Epitopen) entwickeln, um der Erkennung durch besagte Zellen zu entgehen, spricht man von *Escape*-Mutationen. Treten diese *Escape*-Mutationen als Antwort auf Selektionsdruck durch T-Zellen abhängig vom HLA-Typ häufiger auf, spricht man von einem *HLA-Footprint*. Im Umkehrschluss können *HLA-Footprints* als Hinweis auf T-Zell-Epitope dienen.

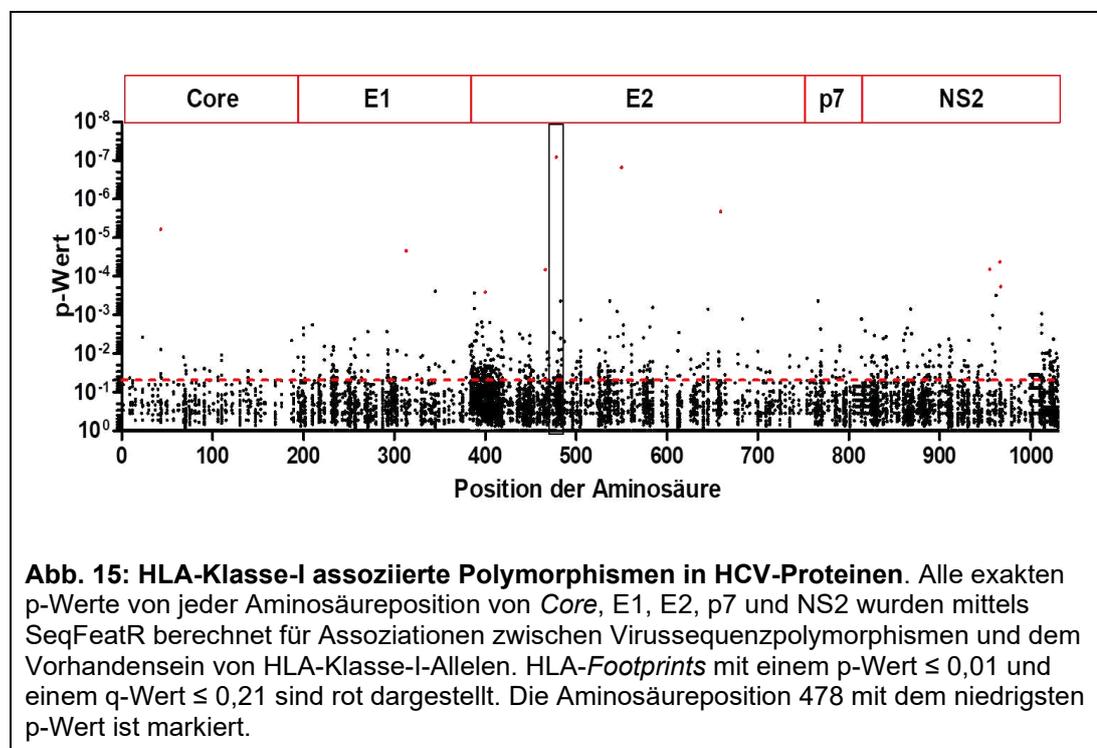
Zu diesem Zweck wurden die neu generierten und weitere bereits vorhandene Sequenzen in einem *Alignment* zusammengetragen und die Sequenzen auf die Abschnitte zwischen *Core* und NS2 gekürzt. Außerdem wurde der HVR1-Bereich der Sequenzen aufgrund seiner zu hohen Variabilität entfernt.

Sequenzen von Patientinnen, deren HLA-Typ nicht bekannt war (AD 013, AD 101, AD 186, AD 397, AD 418, AD 429, AD 430 und AD 445) wurden von diesem *Alignment* ausgeschlossen. Der HLA-Typ (bzw. die vier HLA-A- und HLA-B-Allele) war zuvor im Rahmen einer anderen Arbeit (Ziegler et al., 2016) für unter anderem diese 134 Patientinnen bestimmt worden. Von vier weiteren Patientinnen (AD 063, AD 241, AD 254, AD 261) waren drei der vier Merkmale (HLA-A1, HLA-

A2, HLA-B1, HLA-B2) bekannt, diese wurden zunächst in die Analyse miteingeschlossen. Zwei weitere Sequenzen, von denen drei HLA-Merkmale bekannt waren (AD 247 und AD 264), wurden wegen ihrer Sequenzqualität von der SeqFeatR-Analyse ausgeschlossen.

Somit standen 138 Sequenzen von *Core* bis einschließlich NS2 zur Verfügung, davon 134 Proben mit vollständig bekanntem HLA-Typ des jeweiligen Wirts. Die Sequenzen wurden im FASTA-Format im SeqFeatR-*Tool* hochgeladen zur Identifizierung von statistischen Assoziationen zwischen viralen Sequenzpolymorphismen und HLA-Typ.

Das SeqFeatR-*Tool* ermittelte 631 Aminosäure-Positionen mit einer Häufung an Mutationen in Bezug auf den jeweiligen HLA-Typ, an denen dementsprechend ein möglicher Hinweis auf Selektionsdruck besteht. Von den 631 Aminosäure-Positionen wiesen neun Positionen einen q-Wert für die Falscherkennungsrate (*False Discovery Rate*, FDR) von 0,2, sowie eine Position einen q-Wert von 0,208 auf. In wissenschaftlichen Studien hat es sich etabliert, einen q-Wert von 0,2 als *Cut off* zu wählen. Die oben genannten zehn Aminosäure-Positionen wurden in weitere Untersuchungen eingeschlossen.



In dem dargestellten Manhattan *Plot* (Abbildung 15) entsprechen die rot hervorgehobenen Punkte den Aminosäure-Positionen im HCV-Genom zwischen *Core* und NS2 und den zugehörigen p-Werten der zehn Mutationen.

Um zu untersuchen, ob an diesen Positionen des viralen Genoms, an denen ein erhöhter Selektionsdruck besteht, möglicherweise Epitope vorliegen, also Molekülabschnitte eines Antigens, die Ziele der adaptiven Immunantwort durch CD8<sup>+</sup>-T-Zellen darstellen, wurde die *Immune Epitope Database and Analysis Resource* (IEDB) verwendet. Diese Datenbank ist eine öffentlich zugängliche Ressource, in der in wissenschaftlichen Veröffentlichungen beschriebene Epitope zusammengetragen und katalogisiert werden (Vita et al., 2019). Die Datenbank umfasste im Januar 2024 über 1,6 Millionen Epitope.

Auf der Website der IEDB ist außerdem ein *MHC I Binding Prediction Tool* verfügbar. Dieses Werkzeug der Datenbank ermöglicht es, *in silico* statt experimentell Epitope innerhalb von eingegebenen Sequenzen vorherzusagen.

Das dazu verwendete Prädiktionsmodell *NetMHCpan 4.0* erstellt quantitative Prädiktionen der Affinität (IC<sub>50</sub> Werte) von Peptid-HLA-Interaktionen (Nielsen et al., 2007) mithilfe von *artificial neural network* Modellen (Nielsen et al., 2003).

Um potenzielle Epitope innerhalb der generierten Sequenzen vorherzusagen, wurden diese inklusive des zugehörigen HLA-Typs der Patientinnen und Patienten im FASTA-Format in das *MHC I Binding Prediction Tool* der IEDB hochgeladen. Als Prädiktionsmethode wurde die aktuelle *NetMHCpan*-Version (4.0) verwendet, als MHC-Ursprung „menschlich“ ausgewählt. Das Ergebnis der Suche wurde unter „*Output*“ in Form einer Tabelle angefordert, in der die potenziellen Epitope nach vorhergesagter Bindungsstärke (IC<sub>50</sub> in nM) sortiert aufgeführt wurden.

Für die zehn oben genannten Aminosäure-Positionen konnten so *in silico* die in Tabelle 4 dargestellten wahrscheinlichsten Epitope, die die jeweilige Position enthalten, prognostiziert werden.

HLA	n	AS-Pos. (Protein)	p-Wert	q-Wert	AS mit kleinstem p-Wert	Predicted Epitope (* = potenziell neues Epitop)
B15	23	478 (E2)	8,23E-08	<0,001	Y	* AQGWGPITY
B57	14	550 (E2)	1,53E-07	<0,001	Q	NTRPPQGNWFGCTW
B40	16	659 (E2)	2,13E-06	<0,01	E	LEDRDRSEL
B8	28	966 (NS2)	4,27E-05	0,10	A	* HAGLRDLAV
B7	27	43 (Core)	6,18E-06	0,01	R	LPRRGPR LGV
A25	4	313 (E1)	2,22E-05	0,02	L	* HVTGHRMAW
B44	41	955 (NS2)	6,66E-05	0,15	D	YDHLTPLRDW
B57	14	466 (E2)	6,72E-05	0,05	I	MASCRPIDKFAQGW
B8	28	967 (NS2)	0,0002	0,21	S	* HAGLRDLAV
A25	4	400 (E2)	0,0003	0,20	V	* TTSGFTSLF

**Tabelle 4: Mittels IEDB-Tool vorhergesagte Epitope im Bereich der zehn mit SeqFeatR ermittelten Mutationen mit den besten p-Werten und einem q-Wert von unter 0,21.** Die vier mit Stern markierten Epitope (eines davon mit zwei mutierten Positionen) sind bisher noch nicht als Epitope des jeweiligen HLA-Typs beschrieben worden.

Zwei der potenziellen Epitope, das B\*40-Epitop LEDRDRSEL und das B\*57-Epitop NTRPPQGNWFGCTW, wurden in dieser oder, im Falle des B\*57-Epitops, abgekürzter Form bereits als Epitope des jeweiligen HLA-Typs beschrieben.

Vier der potenziellen Epitope – das B\*15-Epitop AQGWGPITY, das B\*8-Epitop HAGLRDLAV (welches an der zweiten und dritten Position signifikante Häufungen an Mutationen aufweist), das A\*25-Epitop HVTGHRMAW und das A\*25-Epitop TTSGFTSLF – sind in der IEDB-Database hingegen bisher noch nicht in dieser Form als CD8+-T-Zell-Epitope beschrieben und könnten somit unter Umständen neue Epitope darstellen.

Von den noch nicht als Epitope des jeweiligen HLA-Typs beschriebenen Sequenzen wurde das potenzielle B\*15-Epitop AQGWGPITY im Verlauf noch genauer analysiert (siehe Tabellen 5 und 6). Hierzu standen 136 Sequenzen zur Verfügung, da zwei Sequenzen von Patientinnen mit unvollständigem HLA-B-Status ausgeschlossen wurden.

	470	478	Anzahl der Patienten
	A	Q	
	Q	G	
	W	G	
	P	I	
	T	Y	
HLA-B*15-positiv (n=23)	A	Q	5
	-	D	1
	-	P	1
	-	H	9
	D	-	1
	T	-	2
	S	-	2
	P	-	1
	-	X	1
HLA-B*15-negativ (n=113)	A	Q	89
	-	V	1
	-	H	2
	D	-	1
	D	-	3
	V	-	3
	T	-	1
	-	K	1
	-	L	1
	-	S	7
	-	S	1
	-	V	1
	-	L	1
	-	X	1

Identifizier	HLA A1	HLA A2	HLA B1	HLA B2	Sequenz 470-478	A	Q	G	W	G	P	I	T	Y
AD012	11	24	15	40	AXXWGPITY	-	X	X	-	-	-	-	-	-
AD096	3	32	15	51	AQGWGPITD	-	-	-	-	-	-	-	-	D
AD102	2	0	15	44	AQGWGPITH	-	-	-	-	-	-	-	-	H
AD178	1	2	15	35	AQGWGPITH	-	-	-	-	-	-	-	-	H
AD194	1	24	15	35	AQGWGPITH	-	-	-	-	-	-	-	-	H
AD213	2	0	7	15	PQGWGPITY	P	-	-	-	-	-	-	-	-
AD221	2	3	15	51	AQGWGPITH	-	-	-	-	-	-	-	-	H
AD234	2	24	15	50	TQGWGPITY	T	-	-	-	-	-	-	-	-
AD267	3	24	14	15	SQGWGPITY	S	-	-	-	-	-	-	-	-
AD286	2	0	13	15	AQGWGPITH	-	-	-	-	-	-	-	-	H
AD298	1	2	15	57	SQGWGPITY	S	-	-	-	-	-	-	-	-
AD327	1	3	8	15	AQGWGPITH	-	-	-	-	-	-	-	-	H
AD329	2	0	15	44	AQGWGPITH	-	-	-	-	-	-	-	-	H
AD331	3	0	14	15	TQGWGPITY	T	-	-	-	-	-	-	-	-
AD385	2	0	13	15	DQGWGPITH	D	-	-	-	-	-	-	-	H
AD386	3	29	15	44	AQGWGPITP	-	-	-	-	-	-	-	-	P
AD389	2	0	15	44	AQGWGPITH	-	-	-	-	-	-	-	-	H
AD399	24	33	14	15	AQGWGPITH	-	-	-	-	-	-	-	-	H

**Tabellen 5 und 6: Mutationen der Aminosäuren-Positionen 470 bis 478 unter HLA-B\*15-positiven und HLA-B\*15-negativen Patientinnen und Virussequenzen einer Auswahl von HLA-B\*15-positiven Patientinnen.** Diese wiesen signifikant häufiger Mutationen an der AS-Position 478 nach Beginn des Core-Abschnitts auf. Ebenfalls abgebildet mit zugehörigem Identifier und HLA-Typ sind die sechs Sequenzen mit unveränderter Position 478, aber einer anderen Mutation. Nicht gesondert in Tabelle 6 abgebildet sind die der Konsensussequenz AQGWGPITY entsprechenden Sequenzen (davon fünf HLA-B\*15-positive) und die HLA-B\*15-negativen Patientinnen mit Mutationen. Bei Patientinnen mit HLA-A2 „0“ ist der HLA-A-Typ homozygot.

Die Aminosäure Tyrosin an Position 478 nach Beginn der *Core*-Region im vorhergesagten Epitop AQQWGPITY ist signifikant häufiger bei Patientinnen mit dem HLA-Typ B\*15 verändert. Während von den 113 Proben aus HLA B\*15-negativen Patientinnen nur vier Virussequenzen an der Position eine andere Aminosäure als Tyrosin aufwiesen (ein Y478N, drei Y478H; entspricht ca. 3,5 %), wiesen zwölf Virussequenzen aus den 23 Proben aus HLA B\*15-positiven Patientinnen eine andere Aminosäure auf (zehn Y478H, ein Y478D, ein Y478P; entspricht ca. 52,2 %).

Die anderen elf Sequenzen aus Proben aus HLA B\*15-positiven Patientinnen entsprachen in fünf Fällen der Konsensussequenz (ca. 21,7 %), in sechs Fällen lag eine andere Mutation in dem potenziellen Epitop vor (ca. 26,1 %).

Von den restlichen 109 Sequenzen aus Proben aus HLA B\*15-negativen Patientinnen entsprachen 89 der Konsensussequenz (ca. 78,7 %), 20 trugen andere Mutationen in dem betrachteten Abschnitt (ca. 17,7 %).

Der HLA-assoziierte Polymorphismus an dieser Aminosäuren-Position deutet darauf hin, dass hier ein *Footprint* des HLA-Typs B\*15 vorliegt, also dieser Bereich vom HLA B\*15-Molekül präsentiert und von T-Zell-Rezeptoren erkannt wird.

## 4 Diskussion

### 4.1 Schlussfolgerungen

Wie am Anfang dieser Arbeit beschrieben, stellt das Hepatitis C-Virus auch knapp zehn Jahre nach der Einführung der hochwirksamen direkt antiviral wirkenden Arzneimittel (DAA) ein globales gesundheitspolitisches Problem dar, weil Reinfektionen möglich sind und die hohe genetische Variabilität durch die fehleranfällige RNA-Polymerase des Virus in vielen Fällen ein *Immune Escape* sowie eine Chronifizierung der Infektion mit schwerwiegenden Spätfolgen für den Wirt ermöglicht. Ebendiese hohe genetische Variabilität erschwert außerdem bis heute die Entwicklung einer wirksamen Impfung gegen HCV, die die effektivste Lösung der o.g. Probleme darstellen würde. Das Ziel dieser Arbeit war die Erstellung einer Sequenzdatenbank mit Proben aus dem Anti-D-Ausbruch von 1978/79, um mit Hilfe dieser Datenbank die Anpassung des Hepatitis C-Virus an seinen Wirt und die zugrunde liegenden evolutionären Mechanismen näher zu untersuchen. Außerdem sollte die sogenannte Hypervariable Region 1 zwischen den *Envelope*-Proteinen 1 und 2 als das Hauptziel von Antikörpern näher charakterisiert werden, da ein besseres Verständnis, wie die Neutralisation der Hepatitis C-Viren durch Antikörper funktioniert, notwendig ist. Zuletzt sollte der Einfluss von CD8<sup>+</sup>-T-Zellen auf die Virusevolution in der Anti-D-Kohorte untersucht werden, um HLA-*Footprints* und darüber mögliche T-Zell-Epitope als Ziele der zellulären Immunantwort auf HCV zu ermitteln.

Die Anti-D-Kohorte ist für solche Analysen besonders gut geeignet, da sie auf einen *Single-Source-Outbreak* zurückgeht, sich alle Isolate aus der Kohorte also zu einer einzigen Infektionsquelle zurückverfolgen lassen.

Wir etablierten ein *Nested* PCR-Protokoll, mit dem sich das HCV-Genom von *Core* bis NS2/3 in drei Abschnitten zuverlässig amplifizieren ließ. Es wurde ein neuer Primer AntiD-917-F synthetisiert, mithilfe dessen die PCR I des mittleren Abschnitts deutlich bessere Ergebnisse erbrachte. Auf eine zeitsparende enzymatische Aufreinigung der PCR-Produkte wurde aufgrund von schlechteren Sequenzierungsergebnissen verzichtet.

Nachdem die Anti-D-Proben aus den Jahren 1978/79 bis 2009 abgeholt, sortiert und ein Teil der Proben in adäquate Behältnisse umgefüllt wurde, zeigte sich, dass sich die frühen Proben nicht mit den etablierten PCR-Protokollen vervielfältigen ließen.

Um die Probenqualität zu überprüfen, wurde versucht, 49 Proben von 19 Patientinnen mit einer möglichst sensitiven PCR, der *Core*-PCR, zu amplifizieren. Dies war nur in 23 Fällen und ausschließlich bei Proben aus den 90er-Jahren möglich. Um zu überprüfen, ob eine manuelle RNA-Extraktion unter Umständen bessere Ergebnisse erzielen kann, wurde aus neun der Patientenseren RNA manuell extrahiert. Aus der aus manuell extrahierter RNA synthetisierten cDNA ließen sich ebenso wenige Proben mittels *Core*-PCR vervielfältigen wie zuvor, woraufhin von 23 der 49 Proben die Viruslast bestimmt wurde. Dies war trotz vorheriger Behandlung der Proben wegen zu starker Degradation und Verklumpung nur in 16 Fällen möglich. Die für eine Viruslast-Bestimmung geeigneten Proben aus den Jahren zwischen 1979 und 1984 wiesen in der Hälfte der Fälle Viruslasten von maximal 420 IU/ml auf, in der anderen Hälfte der Proben waren gar keine Viren mehr nachweisbar. Während unter den untersuchten Proben aus den Jahren zwischen 1992 und 1997 einige höhere Viruslasten aufwiesen, fanden sich auch Proben mit Viruslasten, die unter 210 IU/ml lagen oder nicht mehr nachweisbar waren.

Es wurde versucht, 18 der 23 *Core*-positiven Proben mittels der etablierten Strukturprotein-PCRs zu vervielfältigen. Dies war nur bei drei Proben möglich, alle drei Strukturprotein-PCRs waren bei keiner der Proben erfolgreich.

Die niedrigen oder nicht mehr nachweisbaren Viruslasten der frühen Proben sind vermutlich den häufigeren Tau-Frost-Zyklen sowie der teils unsachgemäßen Lagerung bei 4 °C oder -20 °C anstatt bei -80 °C geschuldet. Da die Proben aus den 90er-Jahren diesen äußeren Einflüssen seltener und kürzer ausgesetzt waren, sind in den Proben aus diesen Jahren höhere Viruslasten nachweisbar als in den Proben aus den Jahren des Ausbruchs und die *Core*-PCR vereinzelt erfolgreich.

Die im ersten Schritt etablierte *Nested*-PCR in drei Abschnitten scheint aber für den Großteil dieser Proben nicht sensitiv genug zu sein – möglicherweise können in der Zukunft mit neuen Aufreinigungsverfahren und Sequenziermethoden auch bei diesen Proben noch Sequenzen gewonnen werden.

## 4.2 Amplifikation und Analyse der „späten“ Proben

Die „späten“ Proben aus *Follow-Up*-Studien in den 2000er-Jahren ließen sich aufgrund der besseren Lagerungsbedingungen und besseren RNA-Qualität in den meisten Fällen mit den etablierten *Nested* PCRs amplifizieren.

Zusammen mit 78 Sequenzen von NS3 bis NS5B (Ruhl et al., 2011) und 91 Sequenzen von *Core* bis NS2 (Lipskoch et al., 2014) aus vorherigen Arbeiten konnten 218 Sequenzen von 138 Patientinnen zusammengestellt werden, darunter 74 komplette Genome. In weiteren Arbeiten könnten zusätzliche Proben aufgereinigt oder bereits aufgereinigte Proben, deren Abschnitte von NS3 bis NS5 noch nicht vorliegen, mit den etablierten PCR-Protokollen für die Abschnitte NS3 bis NS5 vervielfältigt und sequenziert werden, um so die Anzahl der verfügbaren kompletten Genome zu erhöhen.

Aus den 74 kompletten Genomen konnte ein phylogenetischer Baum erstellt werden, der die klare Distinktion der Anti-D-Kohorte von HCV-Referenzsequenzen sowie der drei *Clades* untereinander verdeutlicht. Dies steht im Einklang mit den Ergebnissen früherer Publikationen (Ruhl et al., 2011), in welchen die drei Varianten, aus denen sich die drei unterschiedlichen *Clades* entwickelten, in den Proben der ersten Blutspenderin, also der Infektionsquelle des Ausbruchs, nachgewiesen werden konnten. In der besagten Arbeit wurde mithilfe von 78 Sequenzen der Abschnitte von NS3 bis NS5B festgestellt, dass sich jeweils nur eine der drei Varianten aus der Infektionsquelle in den Patientinnen und Patienten des Ausbruchs durchgesetzt hat und sich ihre Proben somit klar einem der drei *Clades* zuordnen lassen. In dieser Arbeit, in der 74 komplette Genome zur Erstellung des phylogenetischen Baums verwendet wurden, konnte bestätigt werden, dass sich die Anti-D-Sequenzen als Sequenzen eines *Single-Source-Outbreaks* klar von den HCV-Referenzsequenzen abgrenzen und sich in drei *Clades* aufteilen lassen.

In vorherigen Arbeiten (Lipskoch et al., 2014) wurde außerdem anhand von 91 Anti-D-Sequenzen von *Core* bis NS2 untersucht, ob die Selektion einer der drei Virusvarianten in jeder Patientin abhängig von Polymorphismen der für den *Entry* wichtigen Rezeptoren des Wirts sein könnte. Dabei konnte keine relevante Korrelation zwischen den Rezeptor-Polymorphismen des Patienten und der dominierenden Virusvariante hergestellt werden.

In dieser Arbeit wurden die für den *Entry* vermutlich wichtige Hypervariable Region 1 und die in der Anti-D-Kohorte besonders häufige Insertion zwischen E1 und der HVR1 näher betrachtet.

In einer Arbeit, in der 25 Genotyp 1a- sowie 48 Genotyp 1b-Proben von spanischen Patienten auf Insertionen an der besagten Stelle untersucht wurden, wiesen eine Genotyp 1a-Probe (4 %) sowie fünf Genotyp 1b-Proben (10,4 %) dort Insertionen auf (Torres-Puente et al., 2007). Zuvor waren nur vereinzelte Insertionen an dieser Stelle beschrieben (Abe et al., 1992; Kato et al., 1992; Aizaki et al., 1996), so auch zunächst in der Anti-D-Kohorte (Höhne et al., 1994) und interessanterweise auch in den Proben einer Patientin, die sekundär über eine Bluttransfusion von einer der Patientinnen der irischen Anti-D-Kohorte von 1977 infiziert wurde (Casino et al., 1999). In Gerotto et al., 2001, wurden fünf Proben mit Insertionen einer Aminosäure (und eine mit einer Deletion) in einer Kohorte von 21 Genotyp 1b-positiven, an Typ 2-Kryoglobulinämie erkrankten Patienten nachgewiesen und ein Zusammenhang mit der Erkrankung vermutet, da in der Kontrollgruppe von 20 Patienten keine Insertion (oder Deletion) auftrat. Die Autorinnen und ihre Co-Autoren vermuten eine Stimulation von B-Zellen und dadurch deren Proliferation durch die exponierte Lage der veränderten Stelle des E2-Proteins und deren Nähe zur CD81-Bindungsstelle.

Im Rahmen der Literaturrecherche fanden wir über Insertionen in Isolaten anderer Genotypen ebenfalls nur vereinzelte Berichte, so trat eine Insertion zum Beispiel in einem von zehn Patienten bzw. in 30 von 834 Klonen in einer Genotyp 3a-Studie auf (Moreau et al., 2008a).

Sowohl in Studien zu Genotyp 1b (Gerotto et al., 2001) als auch zu anderen Genotypen (Moreau et al., 2008b) werden die Insertionen vor der HVR 1 ein „seltenes Ereignis“ genannt. Anders ist dies in der von uns untersuchten Anti-D-Kohorte, in der 80 % aller Isolate an der Stelle eine Insertion vorwiesen. Eine Erklärung für diese hohe Frequenz könnte sein, dass Viren mit Insertion bereits in der Infektionsquelle vorhanden waren und dadurch viele Patientinnen mit insertionshaltigen Viren infiziert wurden. Es konnte leider nicht entschlüsselt werden, warum diese Insertion in dem Großteil der Patientinnen erhalten blieb und in anderen wieder verschwand.

Bei der quantitativen Untersuchung der charakteristischen Insertion vor der Hypervariablen Region 1 zeigte sich, dass die Anzahl an inserierten Aminosäuren von

*Clade* zu *Clade* unterschiedlich war, wobei sich *Clade C* von den anderen beiden *Clades* im Hinblick auf die Häufigkeit der jeweiligen Längen der Insertion noch einmal stärker zu unterscheiden scheint. Der Grund für diese Unterschiede ist jedoch weiterhin unklar.

Um zu untersuchen, ob zwischen den *Clades* Unterschiede in der Aminosäuren-Komposition der Hypervariablen Region 1 und der dortigen Anti-D-spezifischen Insertion bestehen, oder ein Zusammenhang mit dem HLA-Typ des Wirts besteht, wurde eine Hauptkomponenten-Analyse mit den Daten durchgeführt. In der durchgeführten *Principal Component Analysis* zeigten sich keine deutlichen Trends, die Proben aus *Clade A* und *C* scheinen sich aber in Bezug auf die an der Stelle exprimierten Aminosäuren ähnlicher zu sein als die Proben aus *Clade B*. Wenn man die Häufigkeit der Aminosäuren an den Positionen der Insertion getrennt nach *Clades* betrachtet, enthält die Insertion in *Clade A* an jeder Position am häufigsten die Aminosäure Glycin, während dies in *Clade C* nur an der ersten und zweiten, nicht aber der dritten, Stelle der Insertion der Fall ist. In *Clade B* ist an jeder Stelle der Insertion eine andere Aminosäure die häufigste.

Es ergaben sich keine signifikanten Auffälligkeiten im Hinblick auf die Häufigkeit der Expression von polaren oder aromatischen Aminosäuren, die Sequenzen aus *Clade A* scheinen aber mehr polare und weniger kleine Aminosäuren zu enthalten als diejenigen aus *Clades B* und *C*.

Signifikante Zusammenhänge zwischen der Hypervariablen Region 1 der sequenzierten Proben und den HLA-Typen der jeweiligen Patientinnen und Patienten konnten nicht hergestellt werden. Gegebenenfalls wären diese bei einer größeren Anzahl an verfügbaren Genomen und zugehörigen HLA-Typen der Patientinnen und Patienten feststellbar gewesen. In folgenden Arbeiten könnten die Sequenzen und insbesondere die Hypervariable Region 1 auf weitere mögliche Korrelationen zwischen dieser und Virus-spezifischen Faktoren wie der umliegenden Sequenz sowie der Viruslast oder wirtsspezifischen Faktoren wie exprimierten *Entry*-Rezeptoren untersucht werden.

### 4.3 Aminosäure-Positionen unter Selektionsdruck durch HLA Klasse-I-Moleküle

Um den Einfluss von CD8<sup>+</sup>-T-Zellen auf die Virusevolution in der Anti-D-Kohorte zu untersuchen, wurden 138 Virussequenzen von *Core* bis NS2, von deren *Hosts* die HLA-Typen aus vorherigen Arbeiten bekannt waren (Ziegler et al., 2013) mittels SeqFeatR auf einen möglichen Zusammenhang zwischen dem HLA-Typ und der Sequenz untersucht. Potenzielle Epitope waren zuvor in der Arbeitsgruppe bereits in NS3 bis NS5B (Ruhl et al., 2011) und vereinzelt auch in den Strukturproteinen (z. B. das B\*57-Epitop NTRPPLGNWF in E2; Lauer et al., 2004; Kuntzen et al., 2007; Neumann-Haefelin et al., 2008) identifiziert worden.

In dieser Arbeit sollten virale Polymorphismen, die unter HLA Klasse-I-Selektionsdruck stehen, systematisch in den Strukturproteinen gesucht werden. Dabei wurden mit der SeqFeatR-Analyse 631 Aminosäuren-Positionen gefunden, die häufiger bei bestimmten HLA-Typen mutiert waren. Nach einer Korrektur mittels *False Discovery Rate* verblieben zehn Aminosäure-Positionen, die signifikant häufig Mutationen bei bestimmten HLA-Typen des *Hosts* aufwiesen. Diese Positionen wurden mittels IEDB und MHC I *Binding Prediction Tool* weiter untersucht, um potenziell an diesen Sequenzabschnitten liegende Epitope zu ermitteln. Zwei der vom Tool vorhergesagten Epitope wurden bereits in früheren Arbeiten beschrieben, drei weitere wurden in der gleichen oder ähnlichen Form beschrieben, aber im Zusammenhang mit anderen HLA-Typen. Die restlichen vier möglichen Epitope sind in der Datenbank, sofern überhaupt enthalten, nur in stark abgewandelter Form und ohne assoziierten HLA-Typ zu finden, könnten also neue Epitope darstellen.

Das B\*40-Epitop LEDRDRSEL (Epitop ID 35407) wurde in genau dieser Form bereits in einigen Veröffentlichungen beschrieben, zuerst als potenzielles B\*60-Epitop (Wong et al., 2001 und Lauer et al., 2002). Das Epitop war in weiteren Arbeiten in je einem HLA-B\*40:01- und HLA-B\*40:02-MHC *Ligand-Assay* positiv (Burke et al., 2012) und in drei von sieben IFN $\gamma$ -*Release*- sowie einem von zwei *Cytotoxicity*-T-Zell-*Assays* positiv.

Ein im Vergleich zum vorhergesagten B\*57-Epitop NTRPPQGNWFGCTW kürzeres und um die mutierte Aminosäure abweichendes Epitop NTRPPLGNW (Epitop ID 142246) wurde bereits als mögliches B\*57-Epitop mit Austausch der Aminosäure Leucin durch Alanin, Arginin, Methionin oder, wie im vorhergesagten Epitop,

Glutamin beschrieben. In der Publikation (Kim AY et al., 2011) wurde in einer amerikanischen HCV-positiven Kohorte festgestellt, dass das Vorliegen von HLA-B\*57 mit einer besseren spontanen *Clearance* von HCV assoziiert ist. Da dieser protektive Effekt in der Kohorte eines *Single-Source-Outbreaks* in Irland nicht besteht (McKiernan et al., 2004), wurden die Sequenzen der Kohorten verglichen und festgestellt, dass in der Kohorte dieser Arbeit bei HLA-B\*57-Positiven häufiger Sequenzvariationen, die mit der Immunogenität interferieren, vorliegen. Das um eine Aminosäure längere Epitop NTRPPLGNWF (Epitop ID 46243) wurde in drei früheren Publikationen als potenzielles B\*57-Epitop beschrieben (Lauer et al., 2004; Kuntzen et al., 2007; Neumann-Haefelin et al., 2008).

Diese zwei vorbeschriebenen Epitope eines bestimmten HLA-Typs konnten also in gleicher oder ähnlicher Form in dieser Arbeit bestätigt werden. Drei der anderen potenziellen Epitope, das B\*07-Epitop LPRRGPR LGV, das B\*44-Epitop YDHLTPLRDW und das B\*57-Epitop MASCRPIDKFAQGW wurden in ähnlicher Form als Epitope vorbeschrieben, allerdings nur im Zusammenhang mit anderen HLA-Typen.

Das vorhergesagte B\*07-Epitop LPRRGPR LGV (Epitop ID 38683) wurde bisher laut IEDB-Datenbank im Hinblick auf eine Restriktion durch HLA-A\*02 untersucht und war in dem zugehörigen MHC *Ligand-Assay* negativ (Schweitzer et al., 2000). In der Publikation hatte ein Peptid von AS-Position 35 bis 44, also ??LPRRGPR L zuvor im *Flow Cytometric Assay* eine Bindung an die meisten HLA-A\*02-positiven Zellen gezeigt. Das an diesen Positionen liegende Epitop YLLPRRGPR L (Epitop ID 74798) wird in 54 Publikationen erwähnt (u.a. als HLA-A\*02-Epitop in Neumann-Haefelin et al., 2008), war in 21 von 22 HLA-A\*02 MHC *Ligand-Assays* positiv und in 43 von 54 IFNg-*Release*- sowie 32 von 37 *Cytotoxicity-T-Zell-Assays* positiv. Des Weiteren fielen je ein MHC *Ligand-Assay* mit HLA-A\*24-, 26-, 31-, 33-Molekülen sowie eines mit HLA-E-Molekül positiv, zwei mit HLA-A\*68:02 und eines mit HLA-A\*11:01 negativ aus. Auf eine Restriktion durch HLA-B\*07 scheinen LPRRGPR LGV beziehungsweise YLLPRRGPR L nicht untersucht worden zu sein.

Das vorhergesagte B\*44-Epitop YDHLTPLRDW wird nicht in exakt dieser Form in der IEDB-Datenbank aufgeführt. Das zwei Aminosäuren früher beginnende Epitop TYVYDHLTPL (Epitop ID 67503) wurde zuvor auf eine Restriktion durch

HLA-B\*40 untersucht und fiel in vier von fünf T-Zell-*Assays* positiv aus (Hakamada et al., 2004).

Das exakte potenzielle B\*57-Epitop MASCRPIDKFAQGW wird in der Datenbank ebenfalls nicht geführt. Die früher beginnende Sequenz ASGCSERMASCRPID (Epitop ID 4414) war jedoch in einem von zwei HLA-A\*02 MHC *Ligand-Assays* sowie einem IL-2 *Release* T-Zell-*Assay* positiv ausgefallen (Sarobe et al., 2001 und 2006).

Die vier weiteren mögliche Epitope finden sich ebenfalls nur in deutlich abgewandelter Form in der Datenbank wieder, wurden aber bisher mit keinem bestimmten HLA-Typ assoziiert.

Die vom vorhergesagten A\*25-Epitop HVTGHRMAW an der Position der mutierten Aminosäure abweichenden und früher (CSIYPGHITGHRMAW (Epitop ID 7016), Cox et al., 2005) oder später (ITGHRMAWMMMMNWS (Epitop ID 28911), Meunier et al., 2008) in dem viralen Genom beginnenden Epitope werden ohne assoziierten HLA-Typ aufgeführt. Das zuerst genannte Epitop war in zwei von zwei IFN $\gamma$ -T-Zell-*Assays* positiv, letzteres in keinem von sechs IFN $\gamma$ -T-Zell-*Assays*.

Das dem potenziellen B\*15-Epitop AQQWGPITY ähnliche, längere Epitop QFDQGWGPITYNESH (Epitop ID 50748, Sarobe et al., 2001 und 2006) war im IL-2 *Release* T-Zell-*Assay* positiv ausgefallen, zeigte im HLA-A\*02 MHC *Ligand-Assay* aber keine Bindung.

Die vor dem vorhergesagten B\*08-Epitop HAGLRDLAV beginnenden Epitope HLTPLQDWAHAGLRD (Epitop ID 2153379) beziehungsweise HLTPLRDWAHAGLRD (Epitop ID 2153380) sowie das das mögliche A\*25-Epitop TTSGFTSLF enthaltende Epitop STHVTGGAAGRTTSGFTSLFTPGPSQK (Epitop ID 61657) wurden nur in B-Zell-*Assays* untersucht.

Die potenziellen neuen Epitope könnten noch in *Assays* untersucht werden, um sie als Epitope *in vitro* zu bestätigen oder zu widerlegen. Des Weiteren könnten die Virussequenzen in anderen HCV-positiven Kohorten auf die Epitope hin untersucht werden, um zu bestätigen, dass diese auch außerhalb der Anti-D-Kohorte durch Selektionsdruck von CD8<sup>+</sup>-T-Zellen entstehen und somit Isolate-übergreifend relevant sind.

Wie auch in anderen Arbeiten kann ein *Bias* zugunsten von häufigen HLA-Typen vorliegen. Dass die Mutationen auffallend häufig HLA-B-assoziiert sind, ist zuvor schon in anderen Arbeiten mit Proben aus dem Anti-D-Ausbruch aufgefallen (Ruhl et al., 2011); auch in dieser Arbeit waren unter den neun mutierten Aminosäure-Positionen mit den besten p-Werten eine HLA-B\*07-, eine B\*08-, eine B\*15-, eine B\*40-, eine B\*44- und zwei B\*57-assoziiert, demgegenüber stehen nur zwei HLA-A\*25-assoziierte Positionen. Die vergleichsweise häufige HLA-B-Restriktion könnte darauf hindeuten, dass HLA-B-restringierte CD8<sup>+</sup>-T-Zellen einen höheren Selektionsdruck auf HCV ausüben als HLA-A-restringierte, was im Zusammenhang mit HIV bereits vorbeschrieben wurde (Kiepiela et al., 2004).

Im Vergleich zu den Nicht-Strukturproteinen scheint es in den Strukturproteinen etwas weniger häufig Epitope zu geben, so konnten in vorherigen Arbeiten mit Proben der Anti-D-Kohorte allein in der aus 631 Aminosäuren bestehenden NS3-Region acht (Ruhl et al., 2011) sowie mindestens zwei weitere Epitope beschrieben werden (Kuntzen et al., 2008), in der aus 747 Aminosäuren bestehenden Strukturregion sind nur sieben der neun in dieser Arbeit beschriebenen Epitope lokalisiert (eins in *Core*, eins in E1, fünf in E2).

Dies kann darauf hindeuten, dass die Strukturproteine unter einem geringeren Selektionsdruck durch CD8<sup>+</sup>-T-Zellen stehen als die Nicht-Strukturproteine oder die Mutationen in den Strukturproteinen von Hepatitis C-Viren weniger gut toleriert werden, weil die Fitnesskosten von potenziellen *Escape*-Mutationen hier zu groß sind.

Damit könnten T-Zell-Impfungen, die auf Basis von Antigenen/Epitopen in den Strukturproteinen entwickelt werden, einen Vorteil gegenüber solchen, die auf Basis von Nicht-Strukturprotein-Sequenzabschnitten entwickelt werden, bieten. Die letzte vielversprechende Impfung, die in einem Phase-1- und Phase-2-Trial getestet wurde und in diesem zwar HCV-spezifische T-Zell-Antworten produzieren und die HCV-RNA-Level der Erkrankten reduzieren, aber keine chronischen HCV-Infektionen verhindern konnte, waren zwei rekombinante Vakzine, die Nicht-Strukturproteine enkodierten (Page et al., 2021).

In zukünftigen Arbeiten könnten vielversprechende Epitope in den Strukturproteinen *in vitro* sowie später *in vivo* getestet werden. Es sind außerdem weitere Analysen bezüglich der Fitnesskosten von Mutationen in den Strukturproteinen notwendig.

## 4.4 Fazit und Ausblick

Zusammenfassend konnten in dieser Arbeit die Sequenzen der Strukturproteine von weiteren 138 Proben von Patientinnen und Patienten der Anti-D-Kohorte generiert und in einer Anti-D-Sequenzdatenbank zusammengeführt werden. Die phylogenetische Analyse bestätigte die vorbeschriebene klare Abgrenzung der Anti-D-Sequenzen von Genotyp 1b-Referenzsequenzen sowie die der drei *Clades* untereinander. Die Analyse der Hypervariablen Region 1 zeigte komplexe Selektionsmuster, die bis dato noch nicht entschlüsselt werden konnten. Es konnten 631 Aminosäurepositionen unter Selektionsdruck ermittelt werden, von denen zehn Positionen statistisch signifikant waren. An diesen Positionen konnten mit Hilfe eines *Tools* der *Immune Epitope Database* neun mögliche Epitope vorhergesagt werden, von denen vier potenziell neue T-Zell-Epitope darstellen.

Sollten in der Zukunft neue Methoden verfügbar sein, die eine Gewinnung von verwertbarem Material aus den degradierten Proben ermöglichen, könnten die Sequenzen longitudinal analysiert werden. Bis dahin können in weiteren Arbeiten die potenziellen Epitope bestätigt oder widerlegt und in anderen Isolaten gesucht werden. Des Weiteren könnte die Sequenzdatenbank ausgebaut werden, um tiefgreifende Analysen der Hypervariablen Region und die Entdeckung von weiteren potenziellen Epitopen zu ermöglichen.

## 5 Literatur- und Quellenverzeichnis

- Abdel-Hakeem, M. S., & Shoukry, N. H. (2014). Protective immunity against hepatitis C: many shades of gray. *Front Immunol*, 5, 274. doi:10.3389/fimmu.2014.00274
- Abe, K., Inchauspe, G., & Fujisawa, K. (1992). Genomic characterization and mutation rate of hepatitis C virus isolated from a patient who contracted hepatitis during an epidemic of non-A, non-B hepatitis in Japan. *J Gen Virol*, 73 (Pt 10), 2725-2729. doi:10.1099/0022-1317-73-10-2725
- Aizaki, H., Saito, A., Kusakawa, I., Ashiwara, Y., Nagamori, S., Toda, G., . . . Miyamura, T. (1996). Mother-to-child transmission of a hepatitis C virus variant with an insertional mutation in its hypervariable region. *J Hepatol*, 25(5), 608-613. doi:10.1016/s0168-8278(96)80227-3
- an der Heiden, M., Ritter, S., Hamouda, O., & Offergeld, R. (2015). Estimating the residual risk for HIV, HCV and HBV in different types of platelet concentrates in Germany. *Vox Sang*, 108(2), 123-130. doi:10.1111/vox.12204
- Bartenschlager, R., Lohmann, V., & Penin, F. (2013). The molecular and structural basis of advanced antiviral therapy for hepatitis C virus infection. *Nat Rev Microbiol*, 11(7), 482-496. doi:10.1038/nrmicro3046
- Beltrami, E. M., Williams, I. T., Shapiro, C. N., & Chamberland, M. E. (2000). Risk and Management of Blood-Borne Infections in Health Care Workers. *Clinical Microbiology Reviews*, 13(3), 385-407. doi:10.1128/cmr.13.3.385
- Budeus, B., Timm, J., & Hoffmann, D. (2016). SeqFeatR for the Discovery of Feature-Sequence Associations. *PLoS One*, 11(1), e0146409. doi:10.1371/journal.pone.0146409
- Burke, K. P., Munshaw, S., Osburn, W. O., Levine, J., Liu, L., Sidney, J., . . . Cox, A. L. (2012). Immunogenicity and cross-reactivity of a representative ancestral sequence in hepatitis C virus infection. *J Immunol*, 188(10), 5177-5188. doi:10.4049/jimmunol.1103008
- Carr, J., Williams, D. G., Hayden, R. T. (2010). Molecular Detection of Multiple Respiratory Viruses. *Molecular Diagnostics*, 289-300. <https://doi.org/10.1016/B978-0-12-369428-7.00024-0>
- Casino, C., McAllister, J., Davidson, F., Power, J., Lawlor, E., Yap, P. L., . . . Smith, D. B. (1999). Variation of hepatitis C virus following serial transmission: multiple mechanisms of diversification of the hypervariable region and evidence for convergent genome evolution. *J Gen Virol*, 80 ( Pt 3), 717-725. doi:10.1099/0022-1317-80-3-717
- Cox, A. L. (2015). Global control of hepatitis C virus. *Science*, 349(6250), 790-791. doi:10.1126/science.aad1302
- Cox, A. L., Mosbrugger, T., Lauer, G. M., Pardoll, D., Thomas, D. L., & Ray, S. C. (2005). Comprehensive analyses of CD8+ T cell responses during longitudinal study of acute human hepatitis C. *Hepatology*, 42(1), 104-112. doi:10.1002/hep.20749

- Dao Thi, V. L., Dreux, M., & Cosset, F. L. (2011). Scavenger receptor class B type I and the hypervariable region-1 of hepatitis C virus in cell entry and neutralisation. *Expert Rev Mol Med*, 13, e13. doi:10.1017/S1462399411001785
- Dustin, L. B., Cashman, S. B., & Laidlaw, S. M. (2014). Immune control and failure in HCV infection--tipping the balance. *J Leukoc Biol*, 96(4), 535-548. doi:10.1189/jlb.4RI0214-126R
- Enkelmann, J., Gassowski, M., Nielsen, S., Wenz, B., Ross, S., Marcus, U., . . . group, D. S. (2020). High prevalence of hepatitis C virus infection and low level of awareness among people who recently started injecting drugs in a cross-sectional study in Germany, 2011-2014: missed opportunities for hepatitis C testing. *Harm Reduct J*, 17(1), 7. doi:10.1186/s12954-019-0338-y
- Expert Group on the Blood Transfusion Service Board & Ireland. Dept. of Health. (1995). *Report of the Expert Group on the Blood Transfusion Service Board*. Stationery Office.
- Falade-Nwulia, O., Suarez-Cuervo, C., Nelson, D. R., Fried, M. W., Segal, J. B., & Sulkowski, M. S. (2017). Oral Direct-Acting Agent Therapy for Hepatitis C Virus Infection: A Systematic Review. *Ann Intern Med*, 166(9), 637-648. doi:10.7326/M16-2575
- Gerotto, M., Dal Pero, F., Loffreda, S., Bianchi, F. B., Alberti, A., & Lenzi, M. (2001). A 385 insertion in the hypervariable region 1 of hepatitis C virus E2 envelope protein is found in some patients with mixed cryoglobulinemia type 2. *Blood*, 98(9), 2657-2663. doi:10.1182/blood.v98.9.2657
- Green, M. R., & Sambrook, J. (2019). Nested Polymerase Chain Reaction (PCR). *Cold Spring Harb Protoc*, 2019(2). doi:10.1101/pdb.prot095182
- Guntipalli, P., Pakala, R., Kumari Gara, S., Ahmed, F., Bhatnagar, A., Endaya Coronel, M. K., . . . Shahini, E. (2021). Worldwide prevalence, genotype distribution and management of hepatitis C. *Acta Gastroenterol Belg*, 84(4), 637-656. doi:10.51821/84.4.015
- Hakamada, T., Funatsuki, K., Morita, H., Ugajin, T., Nakamura, I., Ishiko, H., . . . Imawari, M. (2004). Identification of novel hepatitis C virus-specific cytotoxic T lymphocyte epitopes by ELISpot assay using peptides with human leukocyte antigen-A\*2402-binding motifs. *J Gen Virol*, 85(Pt 6), 1521-1531. doi:10.1099/vir.0.79801-0
- Himmelsbach, K., Mueller, S., Kress, J., Fiedler, S. A., Miskey, C., Ivics, Z., . . . Chudy, M. (2023). Second hepatitis C virus transmission by blood components since introduction of mandatory NAT screening in Germany. *Transfusion*, 63(2), 339-347. doi:10.1111/trf.17224
- Hohne, M., Schreier, E., & Roggendorf, M. (1994). Sequence variability in the encoding region of hepatitis C virus isolated from patients infected during a single source outbreak. *Arch Virol*, 137(1-2), 25-34. doi:10.1007/BF01311170
- Immune Epitope Database & Tools (2024, January 21). *Summary Metrics*. Immune Epitope Database & Tools. Retrieved January 30, 2024, from <https://www.iedb.org/>

- Jurtz, V., Paul, S., Andreatta, M., Marcatili, P., Peters, B., & Nielsen, M. (2017). NetMHCpan-4.0: Improved Peptide-MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data. *J Immunol*, 199(9), 3360-3368. doi:10.4049/jimmunol.1700893
- Kato, N., Ootsuyama, Y., Tanaka, T., Nakagawa, M., Nakazawa, T., Muraiso, K., . . . Shimotohno, K. (1992). Marked sequence diversity in the putative envelope proteins of hepatitis C viruses. *Virus Res*, 22(2), 107-123. doi:10.1016/0168-1702(92)90038-b
- Kiepiela, P., Leslie, A. J., Honeyborne, I., Ramduth, D., Thobakgale, C., Chetty, S., . . . Goulder, P. J. (2004). Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA. *Nature*, 432(7018), 769-775. doi:10.1038/nature03113
- Kim, A. Y., Kuntzen, T., Timm, J., Nolan, B. E., Baca, M. A., Reyor, L. L., . . . Lauer, G. M. (2011). Spontaneous control of HCV is associated with expression of HLA-B 57 and preservation of targeted epitopes. *Gastroenterology*, 140(2), 686-696 e681. doi:10.1053/j.gastro.2010.09.042
- Knepp, J. H., Geahr, M. A., Forman, M. S., & Valsamakis, A. (2003). Comparison of automated and manual nucleic acid extraction methods for detection of enterovirus RNA. *J Clin Microbiol*, 41(8), 3532-3536. doi:10.1128/JCM.41.8.3532-3536.2003
- Kretzschmar, E., Chudy, M., Nubling, C. M., Ross, R. S., Kruse, F., & Trobisch, H. (2007). First case of hepatitis C virus transmission by a red blood cell concentrate after introduction of nucleic acid amplification technique screening in Germany: a comparative study with various assays. *Vox Sang*, 92(4), 297-301. doi:10.1111/j.1423-0410.2007.00903.x
- Kuniholm, M. H., Kovacs, A., Gao, X., Xue, X., Marti, D., Thio, C. L., . . . Carrington, M. (2010). Specific human leukocyte antigen class I and II alleles associated with hepatitis C virus viremia. *Hepatology*, 51(5), 1514-1522. doi:10.1002/hep.23515
- Kuntzen, T., Timm, J., Berical, A., Lennon, N., Berlin, A. M., Young, S. K., . . . Allen, T. M. (2008). Naturally occurring dominant resistance mutations to hepatitis C virus protease and polymerase inhibitors in treatment-naïve patients. *Hepatology*, 48(6), 1769-1778. doi:10.1002/hep.22549
- Kuntzen, T., Timm, J., Berical, A., Lewis-Ximenez, L. L., Jones, A., Nolan, B., . . . Allen, T. M. (2007). Viral sequence evolution in acute hepatitis C virus infection. *J Virol*, 81(21), 11658-11668. doi:10.1128/JVI.00995-07
- Lauer, G. M., Barnes, E., Lucas, M., Timm, J., Ouchi, K., Kim, A. Y., . . . Klenerman, P. (2004). High resolution analysis of cellular immune responses in resolved and persistent hepatitis C virus infection. *Gastroenterology*, 127(3), 924-936. doi:10.1053/j.gastro.2004.06.015
- Lauer, G. M., Ouchi, K., Chung, R. T., Nguyen, T. N., Day, C. L., Purkis, D. R., Reiser, M., Kim, A. Y., Lucas, M., Klenerman, P., & Walker, B. D. (2002). Comprehensive analysis of CD8(+)-T-cell responses against hepatitis C virus reveals multiple unpredicted specificities. *J Virol*, 76(12), 6104-6113. doi:10.1128/jvi.76.12.6104-6113.2002

- Lipskoch, M., Wiese, M., Timm, J., Roggendorf, M., & Viazov, S. (2014). Genetic heterogeneity of hepatitis C virus cell entry receptors seems to have no influence on selection of virus variants. *Virol J*, 11, 50. doi:10.1186/1743-422X-11-50
- McKiernan, S. M., Hagan, R., Curry, M., McDonald, G. S., Kelly, A., Nolan, N., . . . Kelleher, D. (2004). Distinct MHC class I and II alleles are associated with hepatitis C viral clearance, originating from a single source. *Hepatology*, 40(1), 108-114. doi:10.1002/hep.20261
- Messina, J. P., Humphreys, I., Flaxman, A., Brown, A., Cooke, G. S., Pybus, O. G., & Barnes, E. (2015). Global distribution and prevalence of hepatitis C virus genotypes. *Hepatology*, 61(1), 77-87. doi:10.1002/hep.27259
- Meunier, J. C., Russell, R. S., Goossens, V., Priem, S., Walter, H., Depla, E., . . . Purcell, R. H. (2008). Isolation and characterization of broadly neutralizing human monoclonal antibodies to the e1 glycoprotein of hepatitis C virus. *J Virol*, 82(2), 966-973. doi:10.1128/JVI.01872-07
- Meyer, E., Steffen, G., Krings, A., Ullrich, A., Kollan, C., Dudareva, S., & Zimmermann, R. (2021). Zur Situation bei wichtigen Infektionskrankheiten in Deutschland – Virushepatitis C im Jahr 2020. *Epid Bull*, 28:3 -19. doi:10.25646/8790
- Moreau, I., Levis, J., Crosbie, O., Kenny-Walsh, E., & Fanning, L. J. (2008). Correlation between pre-treatment quasispecies complexity and treatment outcome in chronic HCV genotype 3a. *Virol J*, 5, 78. doi:10.1186/1743-422X-5-78
- Moreau, I., O'Sullivan, H., Murray, C., Levis, J., Crosbie, O., Kenny-Walsh, E., & Fanning, L. J. (2008). Separation of Hepatitis C genotype 4a into IgG-depleted and IgG-enriched fractions reveals a unique quasispecies profile. *Virol J*, 5, 103. doi:10.1186/1743-422X-5-103
- Neumann-Haefelin, C., Timm, J., Spangenberg, H. C., Wischniowski, N., Nazarova, N., Kersting, N., . . . Thimme, R. (2008). Virological and immunological determinants of intrahepatic virus-specific CD8<sup>+</sup> T-cell failure in chronic hepatitis C virus infection. *Hepatology*, 47(6), 1824-1836. doi:10.1002/hep.22242
- Nielsen, M., & Andreatta, M. (2017). NNAlign: a platform to construct and evaluate artificial neural network models of receptor-ligand interactions. *Nucleic Acids Res*, 45(W1), W344-W349. doi:10.1093/nar/gkx276
- Nielsen, M., Lundegaard, C., Blicher, T., Lamberth, K., Harndahl, M., Justesen, S., . . . Buus, S. (2007). NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and -B locus protein of known sequence. *PLoS One*, 2(8), e796. doi:10.1371/journal.pone.0000796
- Nielsen, M., Lundegaard, C., Worning, P., Lauemoller, S. L., Lamberth, K., Buus, S., . . . Lund, O. (2003). Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. *Protein Sci*, 12(5), 1007-1017. doi:10.1110/ps.0239403
- Ogden, S. C., & Tang, H. (2015). The missing pieces of the HCV entry puzzle. *Future Virol*, 10(4), 415-428. doi:10.2217/FVL.15.12

- Page, K., Melia, M. T., Veenhuis, R. T., Winter, M., Rousseau, K. E., Massaccesi, G., . . . Cox, A. L. (2021). Randomized Trial of a Vaccine Regimen to Prevent Chronic HCV Infection. *N Engl J Med*, 384(6), 541-549. doi:10.1056/NEJMoa2023345
- Palmer, B. A., Moreau, I., Levis, J., Harty, C., Crosbie, O., Kenny-Walsh, E., & Fanning, L. J. (2012). Insertion and recombination events at hypervariable region 1 over 9.6 years of hepatitis C virus chronic infection. *J Gen Virol*, 93(Pt 12), 2614-2624. doi:10.1099/vir.0.045344-0
- Penin, F., Dubuisson, J., Rey, F. A., Moradpour, D., & Pawlotsky, J. M. (2004). Structural biology of hepatitis C virus. *Hepatology*, 39(1), 5-19. doi:10.1002/hep.20032
- Poethko-Muller, C., Zimmermann, R., Hamouda, O., Faber, M., Stark, K., Ross, R. S., & Thamm, M. (2013). Die Seroepidemiologie der Hepatitis A, B und C in Deutschland: Ergebnisse der Studie zur Gesundheit Erwachsener in Deutschland (DEGS1). *Bundesgesundheitsbl.*, 56(5-6), 707-715. doi:10.1007/s00103-013-1673-x
- Robert-Koch-Institut (2016). GBE-Themenheft Hepatitis C. Gesundheitsberichterstattung des Bundes. Gemeinsam getragen von RKI und Destatis. doi:10.17886/RKI-GBE-2016-016
- Ruhl, M., Knuschke, T., Schewior, K., Glavinic, L., Neumann-Haefelin, C., Chang, D. I., . . . East German, H. C. V. S. G. (2011). CD8+ T-cell response promotes evolution of hepatitis C virus nonstructural proteins. *Gastroenterology*, 140(7), 2064-2073. doi:10.1053/j.gastro.2011.02.060
- Sarobe, P., Huarte, E., Lasarte, J. J., Lopez-Diaz de Cerio, A., Garcia, N., Borrascueta, F., & Prieto, J. (2001). Characterization of an immunologically conserved epitope from hepatitis C virus E2 glycoprotein recognized by HLA-A2 restricted cytotoxic T lymphocytes. *J Hepatol*, 34(2), 321-329. doi:10.1016/s0168-8278(00)00018-0
- Sarobe, P., Lasarte, J. J., Garcia, N., Civeira, M. P., Borrascueta, F., & Prieto, J. (2006). Characterization of T-cell responses against immunodominant epitopes from hepatitis C virus E2 and NS4a proteins. *J Viral Hepat*, 13(1), 47-55. doi:10.1111/j.1365-2893.2005.00653.x
- Sarrazin, C., Zimmermann, T., Berg, T., Neumann, U. P., Schirmacher, P., Schmidt, H., . . . Robert, K.-I. (2018). S3-Leitlinie "Prophylaxe, Diagnostik und Therapie der Hepatitis-C-Virus (HCV) -Infektion". *Z Gastroenterol*, 56(7), 756-838. doi:10.1055/a-0599-1320
- Scarselli, E., Ansuini, H., Cerino, R., Roccasecca, R. M., Acali, S., Filocamo, G., . . . Vitelli, A. (2002). The human scavenger receptor class B type I is a novel candidate receptor for the hepatitis C virus. *EMBO J*, 21(19), 5017-5025. doi:10.1093/emboj/cdf529
- Schulte, B., Stover, H., Thane, K., Schreiter, C., Gansefort, D., & Reimer, J. (2009). Substitution treatment and HCV/HIV-infection in a sample of 31 German prisons for sentenced inmates. *Int J Prison Health*, 5(1), 39-44. doi:10.1080/17449200802692144

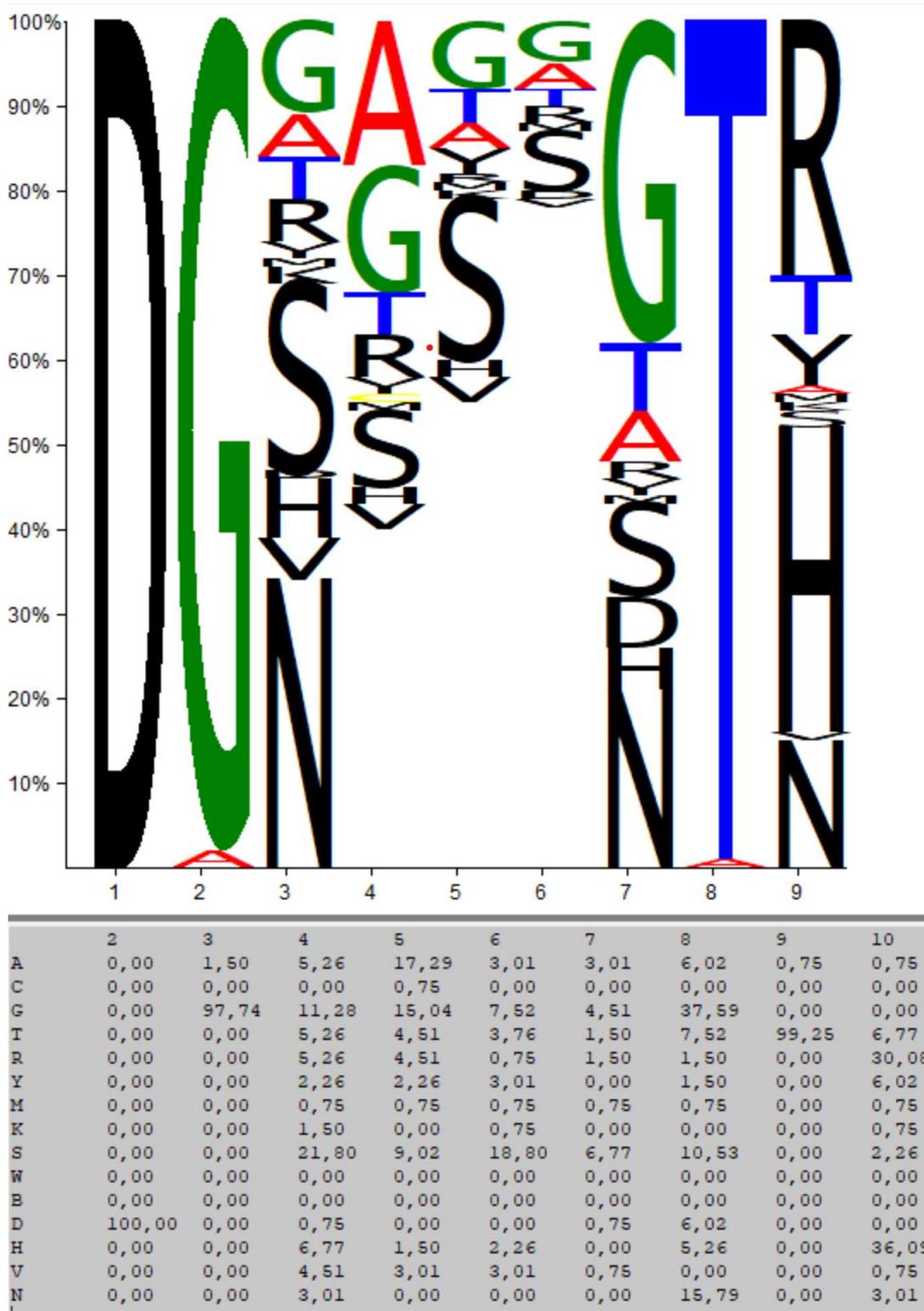
- Schweitzer, S., Schneiders, A. M., Langhans, B., Kraas, W., Jung, G., Vidalin, O., . . . Spengler, U. (2000). Flow cytometric analysis of peptide binding to major histocompatibility complex class I for hepatitis C virus core T-cell epitopes. *Cytometry*, 41(4), 271-278. doi:10.1002/1097-0320(20001201)41:4<271::aid-cyto5>3.0.co;2-m
- Steger, F., Wiethoff, C., & Schochow, M. (2017). *Vertuschter Skandal – Die kontaminierte Anti-D-Prophylaxe in der DDR 1978/1979 und ihre Folgen*. Mitteldeutscher Verlag.
- Torres-Puente, M., Cuevas, J. M., Jimenez-Hernandez, N., Bracho, M. A., Garcia-Robles, I., Carnicer, F., . . . Gonzalez-Candelas, F. (2007). Contribution of insertions and deletions to the variability of hepatitis C virus populations. *J Gen Virol*, 88(Pt 8), 2198-2203. doi:10.1099/vir.0.82855-0
- Vita, R., Mahajan, S., Overton, J. A., Dhanda, S. K., Martini, S., Cantrell, J. R., . . . Peters, B. (2019). The Immune Epitope Database (IEDB): 2018 update. *Nucleic Acids Res*, 47(D1), D339-D343. doi:10.1093/nar/gky1006
- Wiese, M., Grungreiff, K., Guthoff, W., Lafrenz, M., Oesen, U., Porst, H., & East German Hepatitis, C. S. G. (2005). Outcome in a hepatitis C (genotype 1b) single source outbreak in Germany--a 25-year multicenter study. *J Hepatol*, 43(4), 590-598. doi:10.1016/j.jhep.2005.04.007
- Wolffram, I., Petroff, D., Batz, O., Jedrysiak, K., Kramer, J., Tenckhoff, H., . . . German Check-Up 35+ Study, G. (2015). Prevalence of elevated ALT values, HBsAg, and anti-HCV in the primary care setting and evaluation of guideline defined hepatitis risk scenarios. *J Hepatol*, 62(6), 1256-1264. doi:10.1016/j.jhep.2015.01.011
- Wong, D. K., Dudley, D. D., Dohrenwend, P. B., Lauer, G. M., Chung, R. T., Thomas, D. L., & Walker, B. D. (2001). Detection of diverse hepatitis C virus (HCV)-specific cytotoxic T lymphocytes in peripheral blood of infected persons by screening for responses to all translated proteins of HCV. *J Virol*, 75(3), 1229-1235. doi:10.1128/JVI.75.3.1229-1235.2001
- World Health Organization (2021). Global progress report on HIV, viral hepatitis and sexually transmitted infections, 2021. Accountability for the global health sector strategies 2016–2021: actions for impact. Geneva: World Health Organization; 2021. Licence: CC BY-NC-SA 3.0 IGO.
- World Health Organization (2023, July 18). *Hepatitis C*. Retrieved October 20, 2023, from <https://www.who.int/news-room/fact-sheets/detail/hepatitis-c>
- Zein, N. N. (2000). Clinical significance of hepatitis C virus genotypes. *Clin Microbiol Rev*, 13(2), 223-235. doi:10.1128/CMR.13.2.223
- Ziegler, S., Ruhl, M., Tenckhoff, H., Wiese, M., Heinemann, F. M., Horn, P. A., . . . East-German, H. C. V. S. G. (2013). Susceptibility to chronic hepatitis C virus infection is influenced by sequence differences in immunodominant CD8+ T cell epitopes. *J Hepatol*, 58(1), 24-30. doi:10.1016/j.jhep.2012.08.016

# 6 Anhang

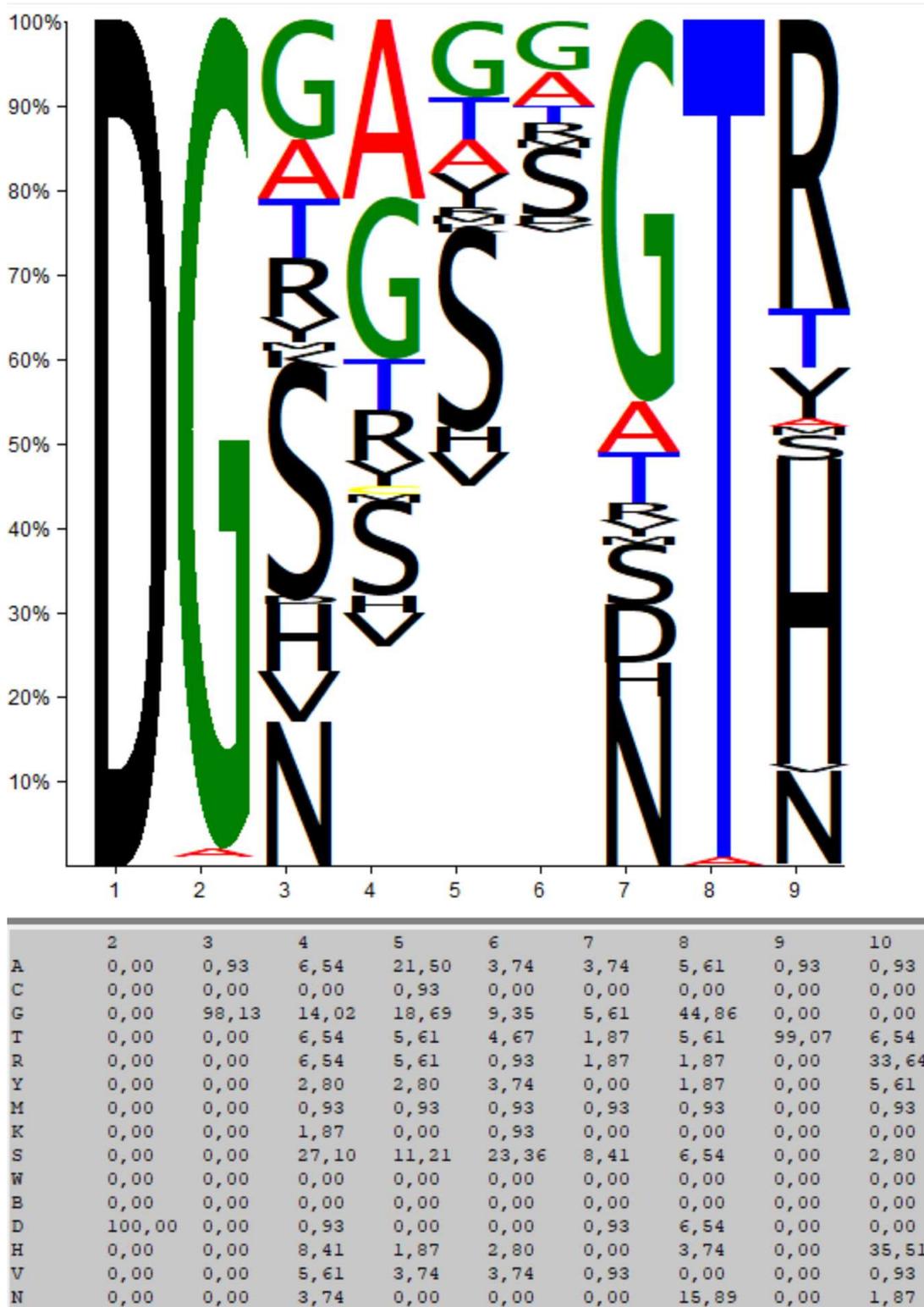
## Anhang 1: Insertionen und HVR1-Abschnitte aller untersuchten Proben

Consensus Identity	1	10	20	30	40	50	59
1. AD001x_A02:24...	WAKVLLVLM	LLFAGVDG	---	MTNTIGG	QAQITXVGLAMFMXG	SP	SQIIXQXNTSSG
2. AD054x_A01:25...	WAKVLLVLM	LLFAGVDG	---	GTTHVGG	AARSTFRLLS	FTLGG	SQKIQLVNTNG
3. AD058x_A24:36...	WAKVLLVLM	LLFAGVDG	---	NTHVGG	AARSTFRLLS	FTLGG	SQKIQLVNTNG
4. AD088A_A24:6...	WAKVLLVLM	LLFAGVDG	---	QTRTAGG	TAHHTHTLASF	QPG	SQRILQINTNG
5. AD093x_A11:26...	WAKVLLVLM	LLFAGVDG	---	ATTTIGG	SQAASHTSGFMGL	FTS	SQKIQLVNTNG
6. AD102x_A02:00...	WAKVLLVLM	LLFAGVDG	---	STYVGG	AARARGAGG	MSF	SPG
7. AD111x_A11:24...	WAKVLLVLM	LLFAGVDG	---	NHTLIGG	QAQATGFWAL	ARG	SP
8. AD177B_A03:26...	WAKVLLVLM	LLFAGVDG	---	HTRITGG	TAAFTLSGL	TR	FLG
9. AD217x_A01:26...	WAKVLLVLM	LLFAGVDG	---	HTXVGG	QAQATGFWAL	ARG	SP
10. AD222x_A03:0...	WAKVLLVLM	LLFAGVDG	---	DTRTVGG	AARARGAGG	MSF	SPG
11. AD235x_A01:0...	WAKVLLVLM	LLFAGVDG	---	TRVSGG	SAAHATS	MF	AGL
12. AD260x_A03:2...	WAKVLLVLM	LLFAGVDG	---	ATTTVGG	QAQATGFWAL	ARG	SP
13. AD264x_A02:4...	WAKVLLVLM	LLFAGVDG	---	HTVGG	QAQATGFWAL	ARG	SP
14. AD314B_A03:2...	WAKVLLVLM	LLFAGVDG	---	HTLVGG	AQASQ	TS	SL
15. AD321x_A01:2...	WAKVLLVLM	LLFAGVDG	---	TTKVGG	QAQAS	STK	GM
16. AD334B_A01:2...	WAKVLLVLM	LLFAGVDG	---	GTTHVGG	SQAQATGFWAL	ARG	SP
17. AD352A_A02:...	WAKVLLVLM	LLFAGVDG	---	GTTHVGG	SQAQATGFWAL	ARG	SP
18. AD376x_A03:0...	WAKVLLVLM	LLFAGVDG	---	NTYTTGG	AQGRVTHS	F	ASL
19. AD380x_A02:3...	WAKVLLVLM	LLFAGVDG	---	STHTVGG	SQAQATGFWAL	ARG	SP
20. AD381x_A02:2...	WAKVLLVLM	LLFAGVDG	---	STHTVGG	AQASQ	TS	SL
21. AD383B_A01:0...	WAKVLLVLM	LLFAGVDG	---	STHTVGG	AQASQ	TS	SL
22. AD385A_A02:0...	WAKVLLVLM	LLFAGVDG	---	STHTVGG	TQAF	AARGL	TS
23. AD386x_A03:2...	WAKVLLVLM	LLFAGVDG	---	STOTTGG	SQAQATGFWAL	ARG	SP
24. AD389x_A02:0...	WAKVLLVLM	LLFAGVDG	---	STHTVGG	SQAQATGFWAL	ARG	SP
25. AD392A_A2:3...	WAKVLLVLM	LLFAGVDG	---	STHTVGG	AQASQ	TS	SL
26. AD407B_A02:0...	WAKVLLVLM	LLFAGVDG	---	TTNTVGG	AQAQAS	S	FT
27. AD015A_A68:0...	WAKVLLVLM	LLFAGVDG	---	RTYVGG	SQAQATGFWAL	ARG	SP
28. AD017x_A01:0...	WAKVLLVLM	LLFAGVDG	---	RTYVGG	SQAQATGFWAL	ARG	SP
29. AD025x_A03:0...	WAKVLLVLM	LLFAGVDG	---	ATRTVGG	AQAQATGFWAL	ARG	SP
30. AD155B_A03:0...	WAKVLLVLM	LLFAGVDG	---	NTRTVGG	TQAHTNRF	MGL	SP
31. AD175x_A02:2...	WAKVLLVLM	LLFAGVDG	---	HTVGG	AQAQATGFWAL	ARG	SP
32. AD191A_A01:0...	WAKVLLVLM	LLFAGVDG	---	QTRTAGG	TAHHTHTLASF	QPG	SQRILQINTNG
33. AD250x_A01:3...	WAKVLLVLM	LLFAGVDG	---	NTYTTGG	AQASQ	TS	SL
34. AD272x_A02:2...	WAKVLLVLM	LLFAGVDG	---	HTVGG	AQAQATGFWAL	ARG	SP
35. AD331A_A03:0...	WAKVLLVLM	LLFAGVDG	---	YTTVGG	AQAQATGFWAL	ARG	SP
36. AD369B_A03:3...	WAKVLLVLM	LLFAGVDG	---	GRTVGG	AQASQ	TS	SL
37. AD394x_A26:3...	WAKVLLVLM	LLFAGVDG	---	GRTVGG	AQASQ	TS	SL
38. AD411x_A02:3...	WAKVLLVLM	LLFAGVDG	---	GRTVGG	AQASQ	TS	SL
39. AD003A_A01:1...	WAKVLLVLM	LLFAGVDG	---	YTTVGG	AQAQATGFWAL	ARG	SP
40. AD008A_A02:...	WAKVLLVLM	LLFAGVDG	---	GRTVGG	SRAAQL	SL	FA
41. AD010x_A24:6...	WAKVLLVLM	LLFAGVDG	---	GRTVGG	SRAAQL	SL	FA
42. AD026A_A01:...	WAKVLLVLM	LLFAGVDG	---	STHTVGG	SQAQATGFWAL	ARG	SP
43. AD042B_A02:...	WAKVLLVLM	LLFAGVDG	---	HTVGG	AQAQATGFWAL	ARG	SP
44. AD046x_A24:0...	WAKVLLVLM	LLFAGVDG	---	HTVGG	AQAQATGFWAL	ARG	SP
45. AD074x_A01:3...	WAKVLLVLM	LLFAGVDG	---	GTMTVGG	TQASRAAN	F	TS
46. AD106A_A02:...	WAKVLLVLM	LLFAGVDG	---	ATTTVGG	SQAQATGFWAL	ARG	SP
47. AD109x_A01:0...	WAKVLLVLM	LLFAGVDG	---	GTMTVGG	TQASRAAN	F	TS
48. AD213x_A02:0...	WAKVLLVLM	LLFAGVDG	---	GRTVGG	TQASRAAN	F	TS
49. AD221x_A02:0...	WAKVLLVLM	LLFAGVDG	---	GRTVGG	TQASRAAN	F	TS
50. AD237x_A03:2...	WAKVLLVLM	LLFAGVDG	---	STHTVGG	SQAQATGFWAL	ARG	SP
51. AD291A_A04:2...	WAKVLLVLM	LLFAGVDG	---	STHTVGG	SQAQATGFWAL	ARG	SP
52. AD298x_A01:0...	WAKVLLVLM	LLFAGVDG	---	TTTVMGG	SQAQATGFWAL	ARG	SP
53. AD305x_A24:6...	WAKVLLVLM	LLFAGVDG	---	NTRVGG	AQAQATGFWAL	ARG	SP
54. AD309B_A02:3...	WAKVLLVLM	LLFAGVDG	---	DTRTVGG	AARARGAGG	MSF	SPG
55. AD321x_A01:2...	WAKVLLVLM	LLFAGVDG	---	DTRTVGG	AARARGAGG	MSF	SPG
56. AD330A_A24:3...	WAKVLLVLM	LLFAGVDG	---	STHTVGG	AQAQATGFWAL	ARG	SP
57. AD336B_A03:3...	WAKVLLVLM	LLFAGVDG	---	NTRVGG	AQAQATGFWAL	ARG	SP
58. AD350B_A02:3...	WAKVLLVLM	LLFAGVDG	---	GRTVGG	AQAQATGFWAL	ARG	SP
59. AD357B_A03:2...	WAKVLLVLM	LLFAGVDG	---	GRTVGG	AQAQATGFWAL	ARG	SP
60. AD377x_A02:2...	WAKVLLVLM	LLFAGVDG	---	GRTVGG	AQAQATGFWAL	ARG	SP
61. AD390B_A02:0...	WAKVLLVLM	LLFAGVDG	---	NTRVGG	AQAQATGFWAL	ARG	SP
62. AD399x_A24:3...	WAKVLLVLM	LLFAGVDG	---	GRTVGG	AQAQATGFWAL	ARG	SP
63. AD002x_A02:2...	WAKVLLVLM	LLFAGVDG	---	GRTVGG	AQAQATGFWAL	ARG	SP
64. AD004B_A02:2...	WAKVLLVLM	LLFAGVDG	---	NTYTTGG	SQAQATGFWAL	ARG	SP
65. AD007B_A02:2...	WAKVLLVLM	LLFAGVDG	---	NTRVGG	AQAQATGFWAL	ARG	SP
66. AD009B_A01:2...	WAKVLLVLM	LLFAGVDG	---	GRTVGG	AQAQATGFWAL	ARG	SP
67. AD020B_A29:3...	WAKVLLVLM	LLFAGVDG	---	GTSTVGG	AQAQATGFWAL	ARG	SP
68. AD021A_A02:0...	WAKVLLVLM	LLFAGVDG	---	TTHTVGG	SQAQATGFWAL	ARG	SP
69. AD069B_A26:6...	WAKVLLVLM	LLFAGVDG	---	NTYVGG	AQAQATGFWAL	ARG	SP
70. AD081A_A01:1...	WAKVLLVLM	LLFAGVDG	---	NTRVGG	AQAQATGFWAL	ARG	SP
71. AD108A_A01:0...	WAKVLLVLM	LLFAGVDG	---	DTRTVGG	AARARGAGG	MSF	SPG
72. AD119B_A03:0...	WAKVLLVLM	LLFAGVDG	---	NHTVGG	AQAQATGFWAL	ARG	SP
73. AD123B_A02:2...	WAKVLLVLM	LLFAGVDG	---	GTMTVGG	TQASRAAN	F	TS
74. AD201x_A01:0...	WAKVLLVLM	LLFAGVDG	---	STTVMGG	AQAQATGFWAL	ARG	SP
75. AD226x_A01:0...	WAKVLLVLM	LLFAGVDG	---	NHTVGG	AQAQATGFWAL	ARG	SP
76. AD234x_A02:2...	WAKVLLVLM	LLFAGVDG	---	TTHTVGG	SQAQATGFWAL	ARG	SP
77. AD239x_A01:2...	WAKVLLVLM	LLFAGVDG	---	NTRVGG	AQAQATGFWAL	ARG	SP
78. AD244x_A02:0...	WAKVLLVLM	LLFAGVDG	---	DHTVGG	QTSRMT	QGF	TL
79. AD259x_A01:2...	WAKVLLVLM	LLFAGVDG	---	DTVMGG	KAARHHS	R	LAG
80. AD262x_A02:6...	WAKVLLVLM	LLFAGVDG	---	GRTVGG	AQAQATGFWAL	ARG	SP
81. AD267B_A03:2...	WAKVLLVLM	LLFAGVDG	---	GTHTVGG	AQAQATGFWAL	ARG	SP
82. AD270x_A02:0...	WAKVLLVLM	LLFAGVDG	---	GTHTVGG	AQAQATGFWAL	ARG	SP
83. AD275x_A02:2...	WAKVLLVLM	LLFAGVDG	---	STHTVGG	AQAQATGFWAL	ARG	SP
84. AD282A_A01:0...	WAKVLLVLM	LLFAGVDG	---	STHTVGG	AQAQATGFWAL	ARG	SP
85. AD286B_A02:0...	WAKVLLVLM	LLFAGVDG	---	DTRTVGG	AARARGAGG	MSF	SPG
86. AD294x_A02:2...	WAKVLLVLM	LLFAGVDG	---	DTRTVGG	AARARGAGG	MSF	SPG
87. AD311B_A01:0...	WAKVLLVLM	LLFAGVDG	---	GTHTVGG	AQAQATGFWAL	ARG	SP
88. AD332A_A01:2...	WAKVLLVLM	LLFAGVDG	---	STHTVGG	AQAQATGFWAL	ARG	SP
89. AD337B_A02:3...	WAKVLLVLM	LLFAGVDG	---	STRTVGG	AQAQATGFWAL	ARG	SP
90. AD342B_A03:1...	WAKVLLVLM	LLFAGVDG	---	XTHTVGG	SQAQATGFWAL	ARG	SP
91. AD358x_A01:0...	WAKVLLVLM	LLFAGVDG	---	ATHTVGG	SQAQATGFWAL	ARG	SP
92. AD361x_A02:2...	WAKVLLVLM	LLFAGVDG	---	GTHTVGG	AQAQATGFWAL	ARG	SP
93. AD365x_A02:2...	WAKVLLVLM	LLFAGVDG	---	GTVMGG	AQAQATGFWAL	ARG	SP
94. AD404B_A02:0...	WAKVLLVLM	LLFAGVDG	---	GTOTTGG	SQAQATGFWAL	ARG	SP
95. AD406B_A24:2...	WAKVLLVLM	LLFAGVDG	---	GTTRVGG	AQAQATGFWAL	ARG	SP
96. AD412x_A02:6...	WAKVLLVLM	LLFAGVDG	---	GTVMGG	AQAQATGFWAL	ARG	SP
97. AD006A_A03:2...	WAKVLLVLM	LLFAGVDG	---	SHTVGG	AQAQATGFWAL	ARG	SP
98. AD012x_A11:2...	WAKVLLVLM	LLFAGVDG	---	PTVGG	AQAQATGFWAL	ARG	SP
99. AD019A_A01:3...	WAKVLLVLM	LLFAGVDG	---	PTVGG	AQAQATGFWAL	ARG	SP
100. AD022A_A02:...	WAKVLLVLM	LLFAGVDG	---	HTVGG	AQAQATGFWAL	ARG	SP
101. AD052B_A02:...	WAKVLLVLM	LLFAGVDG	---	HTVGG	AQAQATGFWAL	ARG	SP
102. AD059x_A02:...	WAKVLLVLM	LLFAGVDG	---	HTVGG	AQAQATGFWAL	ARG	SP
103. AD063x_A02:...	WAKVLLVLM	LLFAGVDG	---	HTVGG	AQAQATGFWAL	ARG	SP
104. AD082x_A24:...	WAKVLLVLM	LLFAGVDG	---	HTVGG	AQAQATGFWAL	ARG	SP
105. AD096x_A03:...	WAKVLLVLM	LLFAGVDG	---	HTVGG	AQAQATGFWAL	ARG	SP
106. AD118x_A01:...	WAKVLLVLM	LLFAGVDG	---	HTVGG	AQAQATGFWAL	ARG	SP
107. AD131A_A02:...	WAKVLLVLM	LLFAGVDG	---	HTVGG	AQAQATGFWAL	ARG	SP
108. AD141A_A01:...	WAKVLLVLM	LLFAGVDG	---	HTVGG	AQAQATGFWAL	ARG	SP
109. AD168x_A02:...	WAKVLLVLM	LLFAGVDG	---	HTVGG	AQAQATGFWAL	ARG	SP
110. AD169x_A02:...	WAKVLLVLM	LLFAGVDG	---	HTVGG	AQAQATGFWAL	ARG	SP
111. AD178x_A01:...	WAKVLLVLM	LLFAGVDG	---	HTVGG	AQAQATGFWAL	ARG	SP
112. AD192x_A03:...	WAKVLLVLM	LLFAGVDG	---	HTVGG	AQAQATGFWAL	ARG	SP
113. AD193x_A03:...	WAKVLLVLM	LLFAGVDG	---	HTVGG	AQAQATGFWAL	ARG	SP
114. AD194x_A01:...	WAKVLLVLM	LLFAGVDG	---	HTVGG	AQAQATGFWAL	ARG	SP
115. AD200x_A02:...	WAKVLLVLM	LLFAGVDG	---	HTVGG	AQAQATGFWAL	ARG	SP
116. AD206x_A03:...	WAKVLLVLM	LLFAGVDG	---	HTVGG	AQAQATGFWAL	ARG	SP
117. AD211x_A03:...	WAKVLLVLM	LLFAGVDG	---	HTVGG	AQAQATGFWAL	ARG	SP
118. AD216x_A02:...	WAKVLLVLM	LLFAGVDG	---	HTVGG	AQAQATGFWAL	ARG	SP
119. AD231x_A02:...	WAKVLLVLM	LLFAGVDG	---	HTVGG	AQAQATGFWAL	ARG	SP
120. AD241A_A02:...	WAKVLLVLM	LLFAGVDG	---	HTVGG	AQAQATGFWAL	ARG	SP
121. AD245x_A25:...	WAKVLLVLM	LLFAGVDG	---	HTVGG	AQAQATGFWAL	ARG	SP
122. AD254x_A02:...	WAKVLLVLM	LLFAGVDG	---	HTVGG	AQAQATGFWAL	ARG	SP
123. AD255A_A01:...	WAKVLLVLM	LLFAGVDG	---	HTVGG	AQAQATGFWAL	ARG	SP
124. AD274x_A01:...	WAKVLLVLM	LLFAGVDG	---	HTVGG	AQAQATGFWAL	ARG	SP
125. AD284x_A24:...	WAKVLLVLM	LLFAGVDG	---	HTVGG	AQAQATGFWAL	ARG	SP
126. AD307B_A24:...	WAKVLLVLM	LLFAGVDG	---	HTVGG	AQAQATGFWAL	ARG	SP
127. AD313A_A01:...	WAKVLLVLM	LLFAGVDG	---	HTVGG	AQAQATGFWAL	ARG	SP
128. AD327x_A01:...	WAKVLLVLM	LLFAGVDG	---	HTVGG	AQAQATGFWAL	ARG	SP
129. AD329x_A02:...	WAKVLLVLM	LLFAGVDG	---	HTVGG	AQAQATGFWAL	ARG	SP
130. AD333x_A01:...	WAKVLLVLM	LLFAGVDG	---	HTVGG	AQAQATGFWAL	ARG	SP
131. AD341B_A02:...	WAKVLLVLM	LLFAGVDG	---	HTVGG	AQAQATGFWAL	ARG	SP
132. AD345B_A24:...	WAKVLLVLM	LLFAGVDG	---	HTVGG	AQAQATGFWAL	ARG	SP
133. AD351x_A01:...	WAKVLLVLM	LLFAGVDG	---	HTVGG	AQAQATGFWAL	ARG	SP
134. AD356B_A01:...	WAKVLLVLM	LLFAGVDG	---	HTVGG	AQAQATGFWAL	ARG	SP
135. AD364A_A01:...	WAKVLLVLM	LLFAGVDG	---	HTVGG	AQAQATGFWAL	ARG	SP
136. AD370x_A03:...	WAKVLLVLM	LLFAGVDG	---	HTVGG	AQAQATGFWAL	ARG	SP
137. AD379B_A01:...	WAKVLLVLM	LLFAGVDG	---	HTVGG	AQAQATGFWAL	ARG	SP
138. AD395x_A03:...	WAKVLLVLM	LLFAGVDG	---	HTVGG	AQAQATGFWAL	ARG	SP

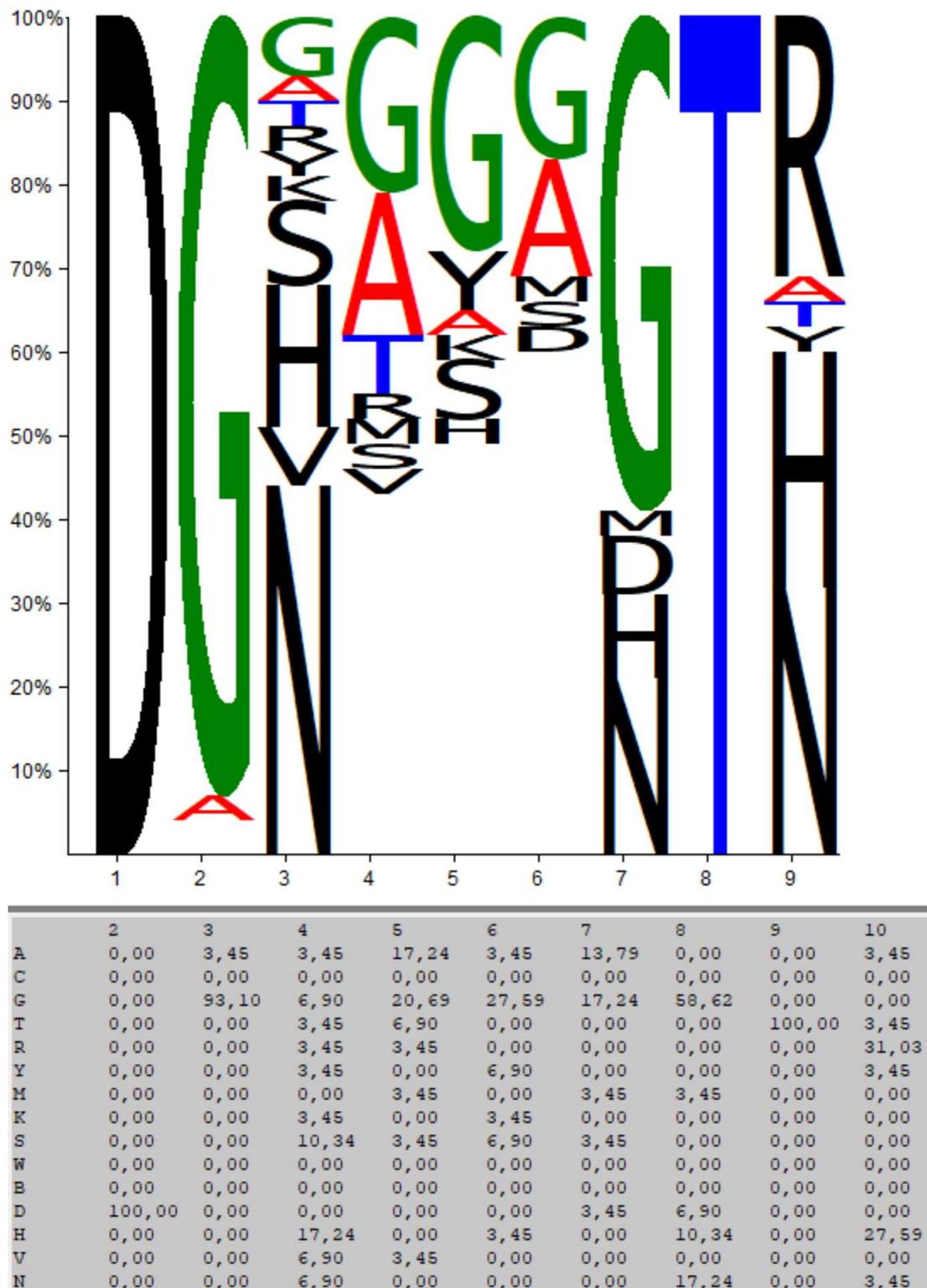
Anhang 2: Sequenzlogos der Insertionen und Häufigkeiten in Prozent

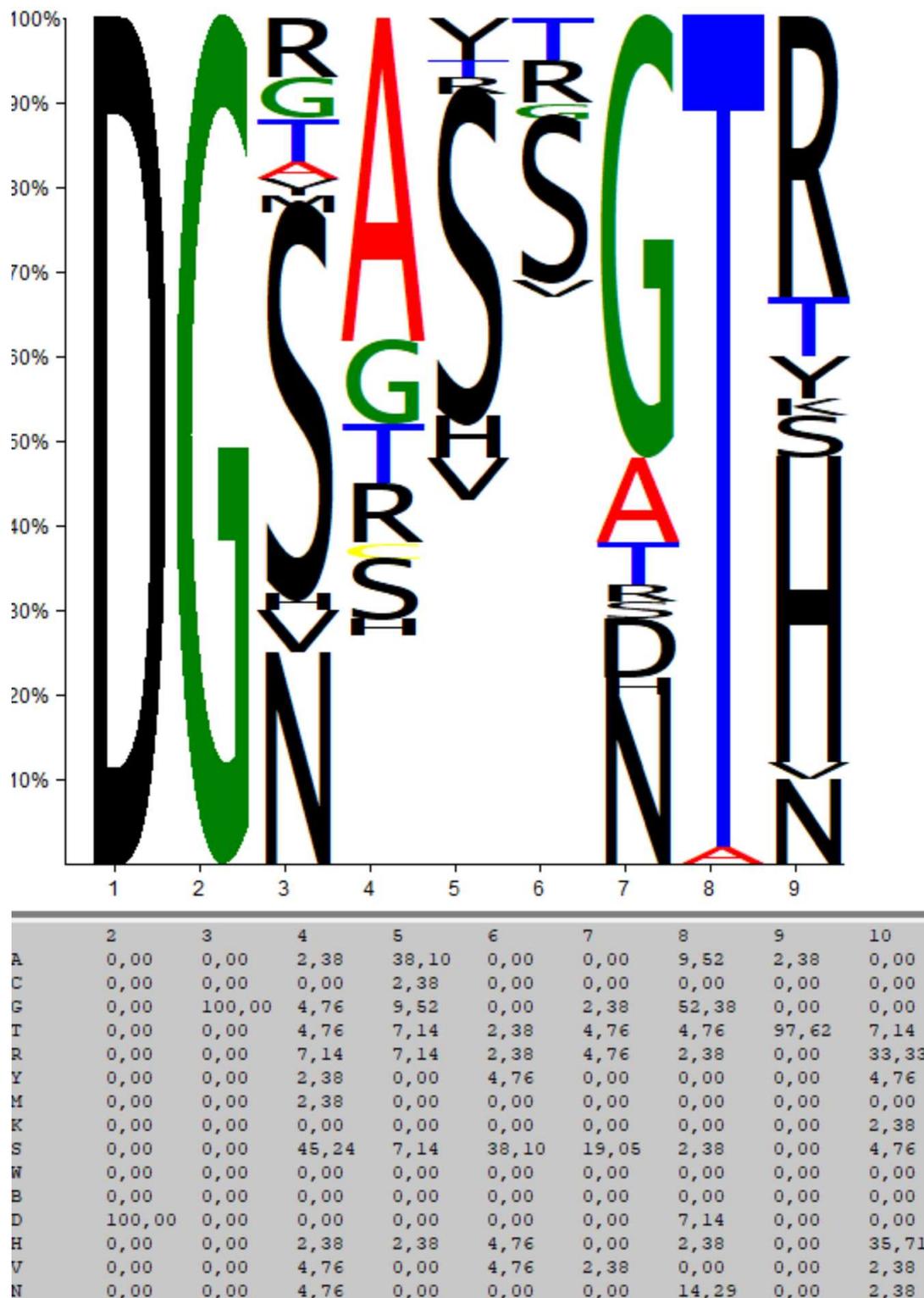


Sequenzlogo und Häufigkeiten der inserierten Aminosäuren in Prozent (alle Proben)

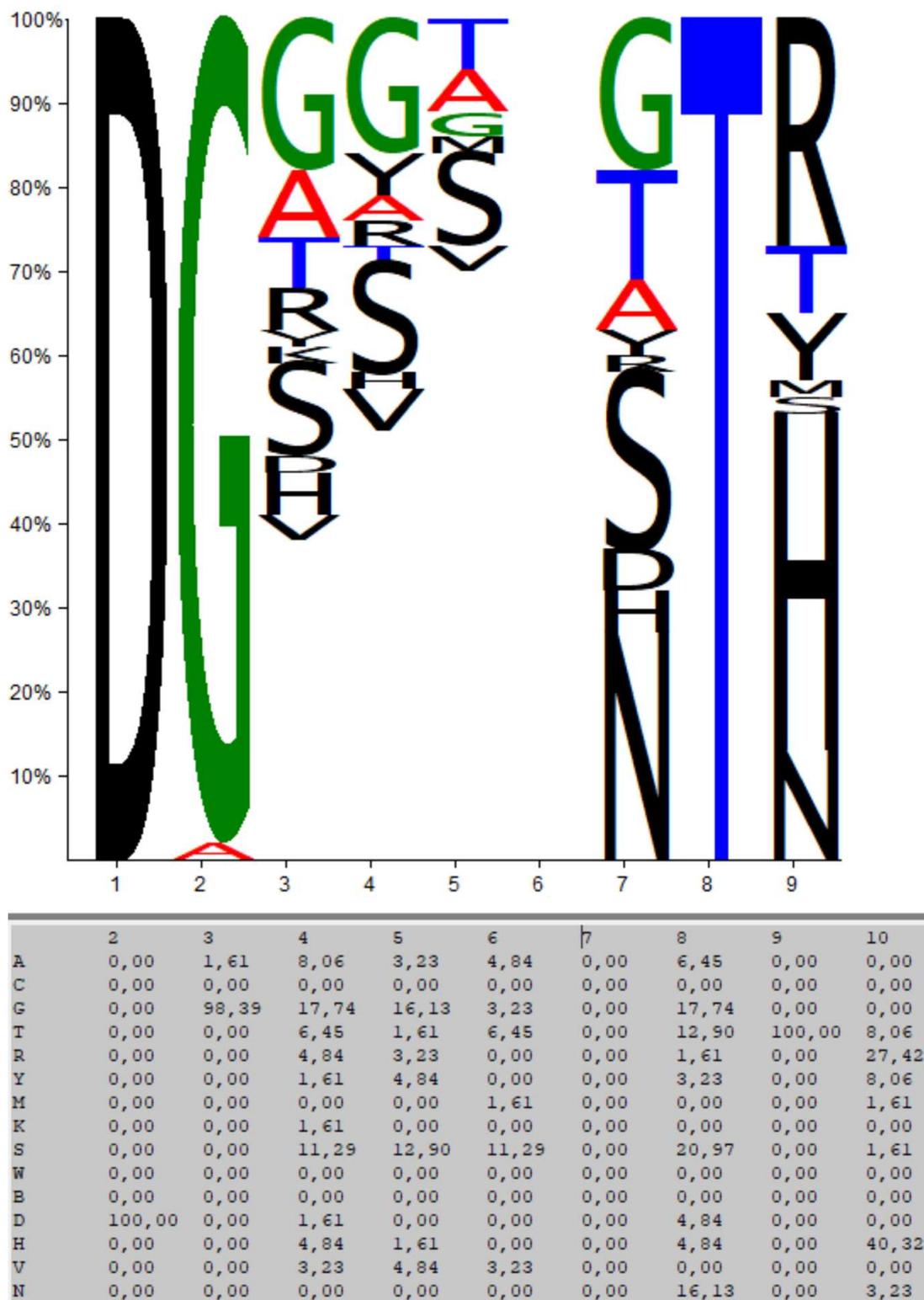


Sequenzlogo und Häufigkeiten der inserierten Aminosäuren in Prozent  
(Proben mit Insertion)

Sequenzlogo und Häufigkeiten der inserierten Aminosäuren in Prozent (*Clade A*)



Sequenzlogo und Häufigkeiten der inserierten Aminosäuren in Prozent (*Clade B*)



Sequenzlogo und Häufigkeiten der inserierten Aminosäuren in Prozent (*Clade C*)

# Danksagung

Der experimentelle Teil der vorliegenden Arbeit wurde im Zeitraum April 2018 bis Oktober/November 2018 am Institut für Virologie der Heinrich-Heine-Universität Düsseldorf durchgeführt.

Herzlich danken möchte ich an dieser Stelle:

Herrn **Professor Dr. Jörg Timm**, für die Möglichkeit, in seiner Arbeitsgruppe zu promovieren, das Festhalten an der Arbeit auch trotz anfänglicher Rückschläge des Projekts sowie für die Unterstützung während meiner Zeit im Labor und der Korrekturphase meiner Dissertation.

Herrn **Professor Dr. Hans Bock**, für die Übernahme der Zweitbetreuung und -begutachtung.

Herrn **Dr. Andreas Walker**, ohne den die Fertigstellung dieser Arbeit nicht möglich gewesen wäre. Insbesondere gilt ihm mein großer Dank für die herzliche Betreuung meiner Arbeit, die zahlreichen Überlegungen sowie Ratschläge und vor allem für den Optimismus und die motivierenden Worte in Anbetracht negativer PCRs und degradierter Proben.

Frau **Dr. Tatjana Schwarz**, die die undankbare Herausforderung angenommen hat, mich mit in die Programme Geneious und SeqFeatR einzuarbeiten, und bei jeglichen Fragen jederzeit mit einem offenen Ohr und vielen hilfreichen Tipps zur Stelle war.

Den weiteren Mitgliedern der damaligen „AG Timm“, **Wiebke, Christopher, Ramona, Ralf, Tina, Christine, Eugen, Alexandra, Albert** und **Anja** für die herzliche Aufnahme in die Arbeitsgruppe, das angenehme Arbeitsklima im Labor und Doktoranden-Büro, die Unterstützung bei Fragen und die schönen gemeinsamen Erlebnisse bei Sommer- oder Weihnachtsfesten sowie dem AG-Ausflug.

Herrn **Johannes Ptok** für die Durchführung der PCA und vor allem deren ausführliche Erklärung, auch noch Jahre nach der ursprünglichen Durchführung, sowie Frau **Anna-Theresa Christ** für die Durchführung der Viruslast-Bestimmung.

Frau **Christiane Cramer** für die freundliche Hilfe bei der Navigation jeglicher bürokratischer Hürden am Anfang der Zeit im Labor.

Meinen Eltern, **Dr. Bernhard Kortenbruck** und **Dr. Dorothee Ahrens-Kortenbruck**, sowie meinen Geschwistern **Hannah, Noah** und **Esther**, ohne deren Unterstützung mein bisheriger Lebensweg nicht möglich und vor allem nicht so erfüllt gewesen wäre, die stets an mich geglaubt und mich bei dem einen oder anderen Kaffee im Garten aufs Neue ermutigt haben. Ich kann euch nicht genug danken!

Meiner Freundin, **Franziska**, die mich in den letzten Jahren ebenfalls unermüdlich begleitet, unterstützt und bestärkt hat. Vielen Dank dir für den Ausgleich, aber auch den Ansporn, den du immer jeweils zur richtigen Zeit zu vermitteln wusstest.

Meinen Kommilitonen und Freunden aus der Zeit in Düsseldorf, insbesondere **Robin, Aglaia, Hellen, Elisa, Isa, Anica, Chris** und **Dominik** für die schöne Studienzeit, die vielen unvergesslichen Erlebnisse sowie eure fortbestehende Freundschaft und motivierenden Worte, vor allem in der Fertigstellungsphase dieser Arbeit.