

Zur DNA Sequenz eines klinischen Isolaten ähnlichen HCMV (*Humanes  
Cytomegalievirus*) Stammes (*Toledo*):

Veränderte Leserahmen durch zusätzliche Startkodons?

Inaugural-Dissertation  
zur Erlangung des Doktorgrades der  
Mathematisch-Naturwissenschaftlichen Fakultät  
der Universität zu Köln  
vorgelegt von

Holger Brondke  
aus Leverkusen

Erlangen im August 2005

Berichtersteller: Prof. Dr. W. Doerfler  
Prof. Dr. J. C. Brüning

Prüfungsvorsitzender: Prof. Dr. H.W. Klein

Tag der mündlichen Prüfung: 10. November 2005

Christina

Meinen Eltern

---

**Inhaltsverzeichnis**

<b>1</b>	<b>Einleitung</b>	<b>1</b>
1.1	Herpesviridae	1
1.1.1	Genomstruktur und Genomorganisation der Herpesviren	2
1.2	Humane Cytomegalieviren	4
1.2.1	Aufbau des Virus	5
1.2.2	Virale DNA	6
1.2.3	Wachstumszyklus	8
1.2.4	Persistierende Infektion und Latenz	8
1.2.5	Epidemiologie	9
1.2.6	Pathologie	10
1.2.7	Pathogenese	11
1.2.8	Transplantationspatienten	12
1.2.9	HIV	12
1.3	HCMV Stämme	12
1.3.1	Laborstämme	12
1.3.2	Klinische Isolate	13
1.4	Leserahmen	14
1.4.1	ORF (Offene Leserahmen)	14
1.4.2	Translationsinitiation (Kozak-Regeln)	14
1.4.3	Alternative Start Kodons	15
1.5	Zielsetzung der Arbeit	16
<b>2</b>	<b>Material</b>	<b>18</b>
2.1	Arbeitsmaterialien	18
2.2	Allgemeine Chemikalien	18
2.3	Kommerzielle Kits	19
2.4	Enzyme	20
2.5	Desoxyribonukleotide und Nukleinsäuren	20
2.6	Lösungen und Puffer	21
2.7	Bakterien, Zellen und Viren	23
2.8	Software	23
<b>3</b>	<b>Methoden</b>	<b>24</b>
3.1	Arbeiten mit Bakterien	24
3.1.1	Lagerung und Kultivierung von Bakterien	24
3.1.2	Herstellung kompetenter Zellen	24
3.1.3	Transformation kompetenter Zellen	24
3.2	Extraktion und Reinigung von Nukleinsäuren	25
3.2.1	DNA-Schnellpräparation nach Holmes und Quigley (1981)	25
3.2.2	Ethanol-fällung von DNA	25
3.2.3	Phenol/Chloroform Extraktion	25

3.2.4	Aufreinigung von DNA aus Agarosegelen	26
3.2.5	Photometrische Konzentrationsbestimmung von DNA/RNA	26
3.3	Synthetische Desoxyribooligonukleotide	26
3.3.1	Primer	26
3.4	Amplifikation genomischer Fragmente	27
3.4.1	Polymerase Kettenreaktion (PCR) (Saiki et al., 1986)	27
3.5	Restriktionsendonukleolytische Spaltung von DNA und gelelektrophoretische Trennung der Fragmente	27
3.5.1	Spaltung von DNA mit Restriktionsendonukleasen	27
3.5.2	Gelelektrophorese	28
3.5.3	Trennung von DNA Fragmenten in denaturierenden Polyacrylamidgelen	28
3.6	RNA	28
3.6.1	RNA-Isolierung aus HCMV infizierten Zellen	28
3.6.2	Aufreinigung der isolierten DNA	29
3.6.3	RT-PCR	29
3.7	Zellkultur	29
3.7.1	Infektion der HFF-Zellen	29
<b>4</b>	<b>Ergebnisse</b>	<b>30</b>
4.1	Sequenzierung des Genoms von HCMV Toledo	30
4.1.1	Experimenteller Plan	30
4.1.2	Amplifikation und Sequenzierung der DNA-Fragmente von HCMV Toledo	31
4.1.3	Übersicht der PCR-Produkte	31
4.1.3	Sequenzierung der PCR-Produkte	33
4.2	Übersicht der Sequenzen	38
4.3	Analyse der genomischen DNA von HCMV <i>Toledo</i>	39
4.4	Offene Leseraster	47
4.4.1	Alternative Startkodons	48
4.4.2	Untersuchung der DNA Sequenz des HCMV Stammes AD169 auf alternative Startkodons	49
4.5	Evaluierung der alternativen Initiationskodons	54
4.5.1	Promoteranalysen	54
4.5.2	Benutzung der alternativen Startkodons unter Berücksichtigung der Kozak Regeln	54
4.5.3	RT-PCR Kontrolle der alternativen Startkodons in messenger RNA aus HCMV <i>Toledo</i> infizierten menschlichen Vorhautfibroblasten	56
4.6	Analyse der bei <i>Toledo</i> gefundenen ORFs	57
<b>5</b>	<b>Diskussion</b>	<b>62</b>
5.1	Zielsetzung der Arbeit	62
5.2	Mutationsanalysen der DNA von HCMV <i>Toledo</i>	63
5.3	Untersuchungen der ORFs von HCMV <i>Toledo</i>	63

---

5.4	Evaluierung der Benutzung alternativer Startkodons	65
<b>6</b>	<b>Literaturverzeichnis</b>	<b>67</b>
<b>7</b>	<b>Anhang</b>	<b>73</b>
7.1	Kurzzusammenfassung	73
7.2	Abstract	74
7.3	Lebenslauf	76
7.4	Erklärung	77

---

### Abbildungsverzeichnis

Abb. 1: Elektronenmikroskopische Aufnahme eines Herpes simplex Virus	1
Abb. 2: Genomstrukturen der Herpesviren	3
Abb. 3: Schematische Darstellung eines Herpes simplex Virus	5
Abb. 4: Struktur des HCMV Genoms	6
Abb. 5: Schematische Darstellung des Amplifikations- und Sequenzierungsplans	32
Abb. 6: Übersicht der <i>Toledo</i> PCR-Produkte / Erster Teil	34
Abb. 7: Übersicht der <i>Toledo</i> PCR-Produkte / Zweiter Teil	35
Abb. 8: Übersicht problematischer Sequenzen	36
Abb. 9: Contig 1.029	38
Abb. 10: Toledo Sequenzübersicht	39
Abb. 11: Sequenzvergleich zwischen <i>Toledo</i> und <i>AD169</i>	40
Abb. 12: Nukleotiddivergenzen zwischen HCMV <i>AD169</i> und <i>Toledo</i>	41
Abb. 13: Nukleotiddivergenzen zwischen HCMV <i>AD169</i> und <i>Toledo</i>	42
Abb. 14: Nukleotiddivergenzen zwischen HCMV <i>AD169</i> und <i>Toledo</i>	43
Abb. 15: Nukleotiddivergenzen zwischen HCMV <i>AD169</i> und <i>Toledo</i>	44
Abb. 16: Mutationspräferenzen	46
Abb. 17: Mit ATG beginnende ORFs des Fragmentes 17-22 der DNA des HCMV <i>Toledo</i> Stammes	47
Abb. 18: Mit ATG, GTG und CTG beginnende ORFs des Fragmentes 17-22 der DNA des HCMV Stammes <i>Toledo</i>	48
Abb. 19: Die 5' vom Startkodon gelegene Region von UL11	50
Abb. 20: Die 5' vom Startkodon gelegenen Regionen von UL4, UL36 und UL92	

52

### Tabellenverzeichnis

Tabelle 1: Humane Herpesviren	2
Tabelle 2: Eingesetzte Arbeitsmaterialien	18
Tabelle 3: Eingesetzte Chemikalien	19
Tabelle 4: Eingesetzte Kits	19
Tabelle 5: Eingesetzte Enzyme	20
Tabelle 6: Nukleinsäuren	20
Tabelle 7: Lösungen und Puffer	22
Tabelle 8: Bakterien, Zellen und Vieren	23
Tabelle 9: Startkodons im Kozak Kontext	55
Tabelle 10: Evaluierung der alternativen Startkodons	56
Tabelle 11: Vergleich der Aminosäuresequenzen der ORFs der HCMV Stämme <i>AD169</i> und <i>Toledo</i>	58
Tabelle 12: Vergleich der Aminosäuresequenzen der ORFs der HCMV Stämme <i>AD169</i> und	

### Abkürzungsverzeichnis:

Es gelten die SI-Einheiten und die daraus ableitbaren zusammengesetzten Größen.

Abb.	Abbildung
bp	Basenpaare
Et al.	Et alii (und andere)
HCMV	Humanes Cytomegalievirus
h	hour (stunde)
k	Kilo-
kbp	Kilobasenpaare
μ	Mikro-
m	Milli-
min	Minute
mRNA	messenger RNA
ORF	Open reading Frame
PCR	Polymerase Chain Reaction
RNA	Ribonukleinsäure
RT	Raumtemperatur
RT-PCR	Reverse Transcription-PCR
z.B.	zum Beispiel

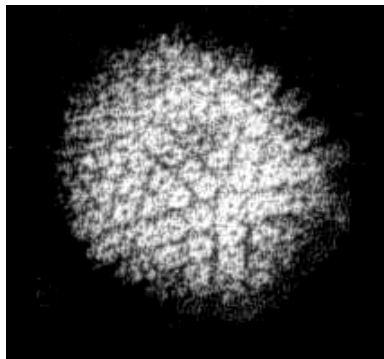


# 1. Einleitung

## 1.1. Herpesviridae

Aus der Familie der Herpesviridae sind heute etwa 100 Mitglieder bekannt, die man aus Säugetieren, Vögeln, Reptilien und Amphibien isoliert hat. Die Charakteristika dieser Virusgattung sind ein lineares, doppelsträngiges Genom, ein ikosaedrisches Capsid von 100 nm Durchmesser, ein das Capsid umgebendes amorphes Material (Tegument), sowie eine Glycoprotein tragende Lipidmembran. Die Größe der Viren reicht von 120-300 nm. Ein charakteristischer Vertreter dieser Gattung, das Herpes simplex Virus, ist in Abbildung 1 in einer elektronenmikroskopischen Aufnahme dargestellt.

Die Familie der Herpesviridae setzt sich aus drei Unterfamilien zusammen, den Alpha-, Beta- und Gammaherpesvirinae. Diese Unterteilung wurde durch die „Herpesvirus Study Group“ vorgenommen und gruppiert die einzelnen Viren entsprechend ihrer biologischen Eigenschaften.



**Abb. 1: Elektronenmikroskopische Aufnahme eines Herpes simplex Virus.** Abbildung entliehen von: <http://www.ncbi.nlm.nih.gov/ICTVdb/Images/Safrica/hsv->

- **Alphaherpesvirinae:** Die Viren dieser Subfamilie besitzen keine ausgeprägte Wirtsstringenz, verfügen über einen kurzen reproduktiven Zyklus und vermehren sich schnell in Kultur. Als Virusreservoir dienen sensorischen Ganglien, in denen das Virus latent persistiert.
- **Betaherpesvirinae:** Diese Viren verfügen über eine geringere Auswahl an Wirtsorganismen als die Alphaherpesvirinae, wachsen langsamer in Kultur und weisen auch einen langsameren Reproduktionszyklus auf. Latent persistierende Viren sind in lymphoretikulären Zellen gefunden worden.

- Gammaherpesvirinae: Viren dieser Unterfamilie besitzen nur wenige Wirte, die zumeist aus der Familie des natürlichen Wirtes stammen. Sie werden in vitro in lymphoblastoiden Zellen vermehrt und persistieren im Wirtsorganismus in lymphoiden Zellen.

Die taxonomische Einordnung der Herpesviren erfolgte nach dem Bekanntwerden von Sequenzinformationen anhand der Genorganisation und des Genarrangements. Da die biologischen Eigenschaften der Viren stark mit ihrer genomischen Struktur korrelieren, hat sich die Zugehörigkeit der Viren zu den Unterfamilien nur marginal verändert.

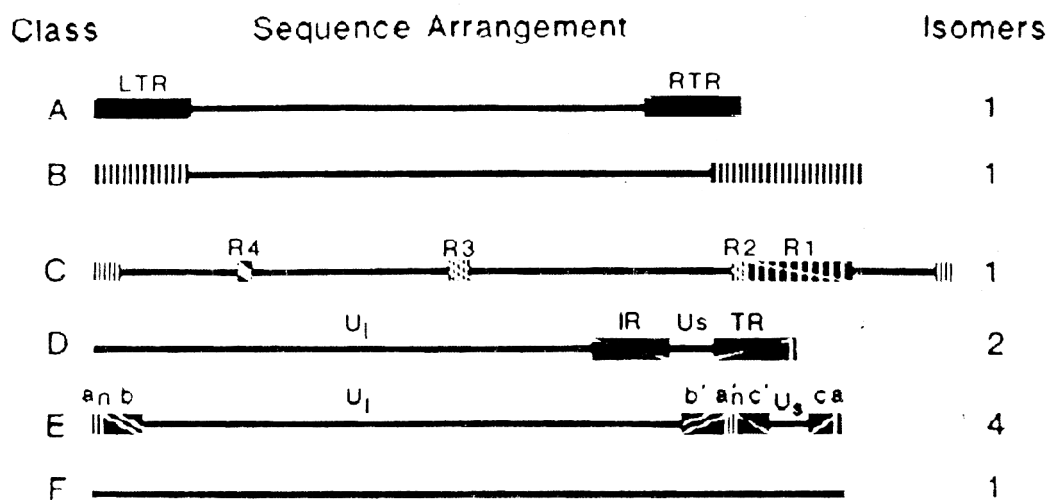
In Tabelle 1 werden die acht Herpesvirengattungen, die bisher aus menschlichen Geweben isoliert werden konnten, ihrer Unterfamilie zugeordnet, unter Angabe der Größe des Genoms, des gebräuchlichen Namens und der hervorgerufenen Erkrankung.

**Tabelle 1: Humane Herpesviren**

Virus	Name	Erkrankung	Größe in bp	Unter- familie
HHV-1	Herpes simplex 1	Oraler Herpes	152.000	a
HHV-2	Herpes simplex 2	Genitaler Herpes	150.000	a
HHV-3	Varizella-Zoster	Windpocken	125.000	a
HHV-4	Epstein-Barr	Mononukleose	172.000	?
HHV-5	Cytomegalievirus	Subklinisch	229.000	β
HHV-6	HHV-6	Roseola infantum	159.000	β
HHV-7	HHV-7	Subklinisch	145.000	β
HHV-8	Karposi's sarcoma associated herpesvirus	Karposisarkom	138.000	?

### 1.1.1. Genomstruktur und Genomorganisation der Herpesviren

Alle Herpesviren besitzen lineare, doppelsträngige DNA-Moleküle, die nach der Freisetzung aus dem Capsid in den Zellkern der infizierten Zelle sofort zirkularisieren. Die viralen Genome unterscheiden sich jedoch beträchtlich in ihrer Größe (125-230 kbp), ihrer Basenkomposition und im Arrangement ihrer Gene.



**Abb. 2: Genomstrukturen der Herpesviren.** Entnommen aus: Fields Virology, Volume 2, 4th Edition

Neben der älteren Unterteilung der Viren aufgrund ihrer biologischen Eigenschaften in drei Unterfamilien, werden die Herpesviren heute entsprechend ihrer Genomstruktur, in 6 Gruppen unterteilt. Eine graphische Darstellung der Genome zeigt Abbildung 2.

- A: Ein ausgedehnter, terminaler Sequenzbereich (LTR) des Virusgenoms wird in gleicher Orientierung am anderen Ende des Genoms (RTR) wiederholt. Ein Vertreter dieser Genomstruktur ist das channel catfish herpesvirus (Abb. 2A).
- B: Eine terminale Sequenz wird in gleicher Orientierung mehrmals an beiden Enden des DNA-Moleküls wiederholt. Ein derart organisiertes Genom liegt beim Herpesvirus saimiri vor (Abb. 2B).
- C: Bei den Epstein-Barr Viren findet man an den Genomenden eine geringere Anzahl an direkten Sequenzwiederholungen, als bei den Typ B Viren. Dafür existieren innerhalb des Genoms Sequenzabschnitte (R1-R4), die in gleicher Orientierung wiederholt werden und die einmaligen Sequenzen, welche nachfolgend nur noch als „unique sequences“ bezeichnet werden, unterteilt. Diese inneren Wiederholungen sind nicht mit den terminalen

---

Wiederholungen verwandt und weisen eine Größe von mindestens 100 bp auf (Abb. 2C).

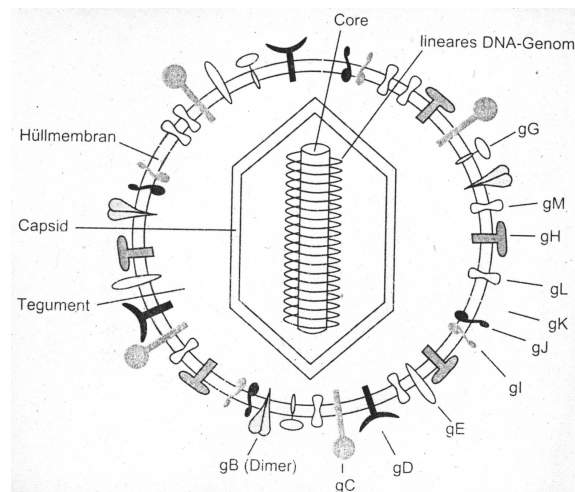
- D: Varizella-Zoster Viren zeichnen sich durch eine indirekte Wiederholung einer terminalen Sequenz innerhalb des Genoms (IR) aus (Abb. 2D).
- E: Das Typ E Genom (Abb. 2E) ist dadurch charakterisiert, daß Sequenzen von beiden Enden des Genoms in einer invertierten Orientierung innerhalb des DNA-Moleküls nebeneinander zu liegen kommen, so daß das Genom in zwei „unique sequences“ ( $U_L$  = unique long;  $U_S$  = unique short) unterteilt wird. Diese beiden unique sequences werden von indirekten Sequenzwiederholungen flankiert, so daß zwei Komponenten entstehen. Diese können relativ zu einander invertieren und somit vier equimolare Genompopulationen bilden. Typische Vertreter für diese Genomstruktur sind die Herpes simplex - und Cytomegalieviren.
- F: Das Genom der Tupaia Herpesviren zeigt keine Sequenzwiederholungen in den terminalen Bereichen oder innerhalb des Genoms.

## 1.2. Humane Cytomegalieviren

Das humane Cytomegalievirus ist ein sehr altes Virus und sehr gut an seinen Wirt angepaßt. Man schätzt die Anzahl der Virusstämme auf über 1000. Auf die primäre Infektion folgt eine persistierende Infektion, die jedoch aufgrund der annähernd symbiotischen Verbindung zwischen Wirt und Virus bei gesunden Menschen keine klinischen Symptome zur Folge hat. Infizierte Zellen zeigen eine Vergrößerung des Zellvolumens und intrazelluläre Einschlüsse.

Dieser histopathologische Befund wurde in Gewebeproben verstorbener Neugeborener gefunden und nach seinem Phänotyp als CID (cytomegalic inclusion disease) bezeichnet. Ähnliche Einschlüsse wie bei CID wurden in Biopsien von Hautläsionen, die durch das Varicella-Zoster Virus hervorgerufen wurden, beobachtet (Tyzzer, 1906). Lipschutz (Lipschutz, 1921) verglich die Phänotypen von CID mit den Beobachtungen von Tyzzer und vermutete ein Virus als Ursache von CID. Gleichsam veränderte Zellen wurden in den Ohrspeicheldrüsen von Säuglingen gefunden, die an einer Vielzahl von Krankheiten verstarben. Aufgrund dieser Entdeckungen wurde dem Virus der

Name „salivary gland virus“ gegeben. Erst später wurde der Name der Viren von Weller (Weller et al., 1960) in Cytomegalievirus umbenannt, um den durch das



**Abb. 3: Schematische Darstellung eines Herpes simplex Virus.** Markiert sind die um das Core gewundene lineare DNA, das Capsid mit Tegument sowie die für Herpes simplex charakteristischen Hüllproteine, die in der Hüllmembran verankert sind. Abbildung entliehen aus: Molekulare Virologie, Susanne Modrow und Dietrich Falke

Virus hervorgerufenen zellulären Veränderungen Ausdruck zu verleihen. Serologische Untersuchungen und Virusisolierungen zeigten, daß durch HCMV (Human Cytomegalovirus) eine häufige, normalerweise subklinisch verlaufende, weltweit auftretende Infektion verursacht wird, die eine ausgeprägte Virulenz in sehr jungen und alten Menschen, in schwächlichen und immunkomprimierten Personen aufweist.

### 1.2.1. Aufbau des Virus

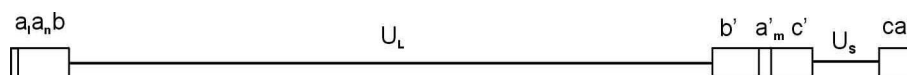
HCMV besitzt ein doppelsträngiges, lineares Genom mit einer Größe von ca. 230 kbp (229354 bei *AD169*). Der Aufbau ist typisch für ein Herpesvirus. Der die DNA enthaltene Kernbereich des Virus liegt in einem ikosaedrischen Capsid von 100 nm Durchmesser, welches in das Tegument eingebettet ist, das wiederum von einer mit viralen Glykoproteinen durchsetzten Lipiddoppelschicht umschlossen ist. Abbildung 3 zeigt eine schematische Darstellung eines Herpes simplex Virus, das starke morphologische Ähnlichkeit mit HCMV besitzt. Das Virus ist nicht sehr stabil und kann durch einen niedrigen pH-Wert, Hitze und Detergenzien inaktiviert werden. Die Halbwertszeit bei 37°C beträgt nur 60 Minuten. Infizierte Zellen setzen drei verschiedene Typen von Viruspartikeln frei (Irmiere und Gibson, 1983). Neben großen Mengen an infektiösen Partikeln werden sogenannte „dense

bodies“ abgegeben, die kein Nukleocapsid enthalten und sich im Zytoplasma infizierter Zellen ansammeln. Zusätzlich werden auch nicht-infektiöse Partikel hergestellt, die ein Capsid ohne DNA Kern enthalten.

### 1.2.2. Virale DNA

Die Genomgröße der Cytomegalieviren variiert zwischen 180 kbp (Pferd) und 240 kbp (Menschen und Meerschweinchen). Das humane Cytomegalievirus ist das einzig bekannte Betaherpesvirus, das ein Klasse E Genom besitzt und somit ein komplexes Arrangement von direkten und indirekten Sequenzwiederholungen aufweist, die zur Ausbildung von 4 Genomisomeren führt. Die Inversion der  $U_L$  und  $U_S$  Bereiche wird hierbei über eine Sequenzwiederholung an den terminalen Enden des Genoms und an der Kreuzung zwischen L und S Bereich gesteuert.

Bei den in Abbildung 4 bezeichneten Regionen „b“ und „c“ handelt es sich um invertierte Wiederholungen, welche die „unique sequences“ Einrahmen. Der Bereich „a“ kennzeichnet eine direkte Sequenzwiederholung an beiden Enden des Genoms und in umgekehrter Orientierung an der L und S Kreuzung. Diese Sequenz beinhalten Spalt- und Verpackungssignale und weist eine Größe von 700-900 bp auf.



**Abb. 4 Struktur des HCMV Genoms**

Das Genom des HCMV-Laborstammes *AD169* wurde 1990 (Chee et al., 1990) komplett sequenziert. 208 offene Leseraster (ORF= Open Reading Frame), die für Proteine mit mehr als 100 Aminosäuren kodieren, wurden errechnet. In 53 dieser ORFs wurden Glykolysierungsstellen entdeckt, und die erzeugten Proteine somit als potentielle Oberflächenantigene charakterisiert. Im Gegensatz zu anderen Herpesviren besitzt das HCMV-Genom ORFs, die eine starke Sequenzhomologie untereinander zeigen und Genfamilien bilden, die vermutlich durch Genduplikationsereignisse entstanden sind. Deletionsmutageneseexperimente haben gezeigt, daß 41 der 208 ORFs bei der in vitro Vermehrung von *AD169* auf humanen Fibroblasten keinerlei Bedeutung haben. 33 ORFs zeigen deutliche Aminosäureähnlichkeiten zu Leserahmen in HSV-1, Varicella-Zoster und Epstein-

---

Barr Virus. Von diesen 33 ORFs scheinen  $\frac{1}{4}$  für DNA Replikations- und Metabolismusproteine zu kodieren und der Rest für Virusreifung und Struktur.

Eine 2002 durchgeführte erneute Analyse des *AD169* Genoms führte zu einer Abänderung der Proteinzugehörigkeiten. Mittels Sequenzvergleichsalgorithmen (MUSCA) und Proteinmustererkennungssoftware (Teiresias Alogorithmus) wurden in den bekannten ORFs 125 Glykolisierungsstellen gefunden. 144 Proteine wurden als potentielle integrale Membranproteine identifiziert, von denen die Hälfte Glykoproteine sind. Des weiteren wurden 49 Signalpeptide und 15 G-Protein-gekoppelte Rezeptor Homologe entdeckt. 9 der vorhergesagten ORFs besaßen Phosphorylierungsstellen und 31 ORFs waren spezifisch für HHV5, welches ein Indiz für das hohe Alter des Virus war (Rigoutsos et al., 2003).

Das Genom zeigt neben den schon erwähnten Sequenzwiederholungen noch weitere, kürzere Wiederholungen und enthält somit eine größere Dichte an Sequenzrepetitionen als jedes andere Herpesvirus. Der G+C Gehalt liegt bei 54-59% und zählt damit in der Gruppe der Herpesviren (G+C = 31-75%) zu den G+C reichen Viren.

Das wiederholte, unverdünnte Passagieren von HCMV in Zellkultur selektiert für mutierte Viren, deren Genom kleiner als das des ursprünglich aus Patienten isolierten Virus ist. Somit haben die Laborstämme *Towne* und *AD169* einige ihrer wirtsselektionierenden Eigenschaften verloren, die in den Genomen von weniger passagierten Stämmen wie dem in dieser Arbeit untersuchten Laborstamm *Toledo*, sowie in klinischen Isolaten noch vorhanden sind. *AD169* weist zum Beispiel eine Deletion von 19 ORFs (Cha et al., 1996) im Vergleich zu klinischen Isolaten auf und der Virusstamm ist nicht mehr in der Lage sich in Endothelzellen zu vermehren. Für *Towne* konnte gezeigt werden, daß von allen 162 einmaligen ORFs nur 45 essentiell für die Vermehrung in Fibroblasten sind und die restlichen 117 bei Einzeldelation keinen nachweisbaren Einfluß auf Infektiösität und Wachstumskinetiken zeigten. Einige der Deletionsmutanten vermehrten sich sogar um die Faktoren 10-500 besser als der Orginalstamm, wobei hier wachstumshemmende Proteine betroffen waren. Folglich kodiert HCMV sowohl unterstützende als auch inhibierende Wachstumsregulatoren, um seine Replikation in Wirtszellen zu optimieren (Dunn et al., 2003). Eine ähnliche Analyse mittels einer Transposon gesteuerten Inaktivierung der ORFs bei *Toledo* ergab

---

das 44 ORFs essentiell für die Vermehrung in Fibroblasten waren, 88 nicht essentiell und 27 wachstumsfördernd (Yu et al., 2003).

Ein Großteil der viralen Gene der Cytomegalieviren enthält keine Spleißsignale. Da auch nur wenige Polyadenylierungssignale vorhanden sind, besitzen ein Großteil der viralen Transkripte eine polycistronische Struktur. Von vielen nicht konservierten Genen konnten jedoch gespleißte mRNAs nachgewiesen werden, die sogar alternativ oder zeitlich reguliert gespleißt werden.

### **1.2.3. Wachstumszyklus**

Im Vergleich zu anderen Herpesviren ist der Replikationszyklus der Cytomegalieviren in Kultur langsam. Er beträgt 48-72 Std., wobei der Wechsel der Expression von frühen zu späten Genen 24-36 Std. p.i. erfolgt. Hierbei können maximale Proteinsynthese, DNA-Replikation und Virusproduktion für mehr als eine Woche aufrecht erhalten werden. Die Replikation der Cytomegalieviren tötet die Wirtszelle nicht direkt. Auch werden im Gegensatz zu anderen Viren die zelleigenen Gene für metabolische Prozesse, Synthese von zellulären Proteinen, DNA- und RNA-Synthese nicht abgeschaltet, sondern stimuliert. Diese Eingriffe in den Metabolismus des Wirtes sind insbesondere deshalb wichtig, da die Synthese viraler DNA von der zellulären DNA-Synthese abhängig ist.

### **1.2.4. Persistierende Infektion und Latenz**

Die Hauptübertragungswege für HCMV beim Menschen sind die Kontamination kleiner infizierter Kinder mit Urin, der bei einer akuten Infektion (viraemische Phase) große Mengen an Virus enthält, sowie der Speichel seropositiver Individuen. Während einer akuten Infektion sind vor allem Epithelzellen, Speicheldrüsen und Nieren Ziele der viralen Replikation. Mittels einer in situ Hybridisierung konnte nachgewiesen werden, daß es sich bei den befallenen Zellen um Epithel- und Endothelzellen sowie um Fibroblasten handelt. Monozyten des peripheren Blutes werden zumeist latent befallen. Hierbei ist der Begriff der Latenz mit Vorsicht zu gebrauchen, da noch nicht ausgeschlossen werden konnte, daß es sich bei der Latenz nicht vielmehr um eine äußerst langsam ablaufende Replikation handelt.



---

Da HCMV die DNA-, RNA- und Proteinsynthese der Wirtszellen stimuliert, ist es naheliegend, daß das Virus in der Literatur mit der Induktion einiger Karzinome und Sarkome (Boldogh et al., 1981) in Verbindung gebracht worden ist. Zu diesen gehören das Prostata Karzinom (Geder et al., 1977), Cervix Karzinom, Adenokarzinom sowie das Kaposi's Sarkom (Giraldo et al, 1980). Neben dieser onkogenen Wirkung besitzen Regulatoren des HCMV die Eigenschaft, die LTR des HI-Virus zu aktivieren, weshalb HCMV die Funktion eines Kofaktors in der Progression von AIDS zugeschrieben wird.

### **1.2.5. Epidemiologie**

Die Übertragung der Humanen Cytomegalieviren kann sowohl direkt als auch indirekt erfolgen. Hierbei enthalten Sputum, Urin, Ausscheidungen aus Cervix und Vagina, Samenflüssigkeit, Muttermilch, Tränen, Kot und Blut das Virus in unterschiedlichen Konzentrationen. Während bei ansonsten gesunden Individuen die akute Infektion nur von kurzer Dauer ist, kann nach fötaler oder konnataler Infektion das Virus über Jahre hinweg in großen Mengen ausgeschieden werden. (Pass, 2001)

Die HCMV Infektion tritt endemisch (Alford und Pass, 1981) und nicht jahreszeitlich gehäuft auf. Amerikanischen Studien zufolge spielen sozioökonomische Faktoren bei der Wahrscheinlichkeit einer vertikalen, intrauterinen oder horizontalen Transmission eine entscheidende Rolle. Die Hauptüberträger bei primären Infektionen von schwangeren Frauen scheinen Säuglinge und Kinder mit subklinischen Infektionen zu sein. In den USA werden 0,2-2,2% aller Neugeborenen im Mutterleib (Stagno et al.,1986) und weitere 8-60% in den ersten 6 Lebensmonaten infiziert (Reynolds et al., 1973). Da bei seropositiven Frauen die Wahrscheinlichkeit einer Reaktivierung der HCMV Infektion im Brustgewebe sehr hoch ist, stellt die Übertragung über die Muttermilch einen wahrscheinlichen Infektionsweg dar. Kinder, für die in Tagesstätten gesorgt wird, weisen einen Durchseuchungsgrad von 80% auf, während gleichaltrige Kinder, die im Elternhaus versorgt werden, nur zu 20% in Kontakt mit dem Virus gekommen sind. In entwickelten Ländern sind 40-80% der Kinder vor dem Erreichen der Pubertät mit dem humanen Cytomegalievirus in Berührung gekommen, wohingegen in anderen Teilen der Welt 90-100% der Bevölkerung in der Kindheit infiziert worden sind.

---

### 1.2.6. Pathologie

Im Wirtsorganismus befällt das Virus, im Gegensatz zur Zellkultur, nur selten Fibroblasten. Bei infizierten Menschen kann man den Krankheitserreger in Monozyten/Makrophagen, glatten Muskelzellen, Stroma- und neuronalen Zellen, Neutrophilen, Hepatozyten, Endothel- und Epithelzellen nachweisen. Diese große Bandbreite an Wirtszellen erschließt sich das Virus über allgemein verbreitete Zielrezeptoren, wie z.B. EGFR ( Epidermaler Wachstums Faktor Rezeptor), an den das Glykoprotein B bindet (Wang et al.,2003). Die infizierten Zellen erscheinen unter dem Lichtmikroskop vergleichsweise groß und zeichnen sich durch wenig Zytoplasma in Verbindung mit großen Zellkernen aus, die zahlreiche Einschlüsse aufweisen.

Die Ausscheidung großer Mengen viraler Partikel über den Urin (Viruria) ist ein häufig angetroffenes Phänomen, das sich über Virusreplikation innerhalb des Urogenitaltraktes und der Nieren erklären läßt.

Außerdem befällt das Virus Gewebe, die endodermalen Ursprungs sind. So wurde bei HIV koinfizierten Personen ein Befall des Gastrointestinaltraktes beobachtet. Am häufigsten ist hierbei der Dickdarm betroffen, in dem man vereinzelt auftretende oberflächliche Geschwüre, aber auch tiefe Geschwüre mit Nekrose des umgebenden Gewebes findet, die sogar zur Perforation des Darms führen können.

Die Lunge, die ebenso wie der Darm endodermalen Ursprungs ist, wird häufig bei Patienten befallen, deren Immunsystem in der Folge von Organtransplantationen unterdrückt wird. Besonders häufig ist dieser Befall bei Knochenmarks-, Nieren- und Herz-Lungentransplantationen zu beobachten, die eine Lungendysfunktion zur Folge haben können.

Infektionen des zentralen Nervensystems treten nur sporadisch, bei AIDS-Patienten und bei im Mutterleib infizierten Föten jedoch häufig auf. Werden neuronale und Gliazellen infiziert, so kommt es zu Kalkablagerungen, die zu einer strukturellen Schädigung des Gehirns führen (Becroft, 1981).

---

### 1.2.7. Pathogenese

HCMV ist ein lytisches Virus, das bei gesunden Personen nur selten klinische Symptome einer Infektion hervorruft. Diese geringe Pathogenität wird mit der langsamen Replikation, dem stringenten Zelltropismus und der stark eingeschränkten interzellulären Ausbreitung erklärt. Der Erreger ist mit 40.000 infiziert geborenen Kindern pro Jahr in den USA, jedoch das am häufigsten übertragene Virus während der Schwangerschaft. Von diesen Kindern zeigen 10-15% neurologische Befunde, womit HCMV die führende infektiöse Ursache für Mißbildungen des zentralen Nervensystems bei Kindern ist (Britt und Alford, 1996).

Eine primäre Infektion der Mutter während der Schwangerschaft führt mit einer 35-50% igen Wahrscheinlichkeit zu einer Übertragung des Erregers auf den Fötus. Ist die Mutter jedoch immunisiert, so verringert sich die Übertragungswahrscheinlichkeit auf 0,2-2%. Aus diesen Daten schließt man, daß die Übertragung von der Mutter auf den Fötus von der Viruslast abhängig ist (Medearis, 1982).

Kleinkinder, die mit einer HCMV-Infektion geboren werden, zeigen in 8-10% der Fälle Symptome der Erkrankung (Demmler, 1991). Ein Großteil der klinischen Befunde ist mild. So weisen viele Kinder eine Vergrößerung von Leber und Milz (Hepatosplenomegalie) oder einen Mangel an Thrombozyten auf. Es kann jedoch auch zu lebensbedrohlichem Organversagen kommen, das zu einer Sterblichkeit von 11-20% führt (Becroft, 1981). 80% der mit einer HCMV-Infektion mit symptomatischem Verlauf geborenen Kinder weisen andauernde neurologische Abnormalitäten auf (Boppana et al., 1992). Zu diesen gehören Schädigungen der Sinnesorgane wie z.B. des Hörvermögens, das in den meisten Fällen betroffen ist und zu einem milden bis ausgeprägten Hörverlust führen kann. Die durch das Virus hervorgerufenen Schädigungen des akustischen Sinnes ist die häufigste Ursache von nicht erblichem Hörverlust bei Kindern. Eine Schädigung des Gehörs tritt auch bei 5-15% der Kinder auf, die ansonsten einen asymptomatischen Verlauf der Infektion aufweisen (Fowler et al., 1992). Das breite Spektrum an Symptomen erklärt sich durch die weit verbreiteten Zielrezeptoren des Virus (siehe 1.2.5. Pathologie).

---

### **1.2.8. Transplantationspatienten**

Infektionen mit dem humanen Cytomegalievirus sind die am häufigsten auftretende Infektion nach Transplantationen. So entwickeln 60-100% der Patienten, die eine Niere erhalten, eine akute HCMV-Infektion.

### **1.2.9. HIV**

HIV infizierte Menschen haben ein erhöhtes Risiko frühzeitig an AIDS zu erkranken, wenn sie mit HCMV koinfiziert sind. HCMV ist eine wichtige opportunistische Infektion bei AIDS Patienten, die, falls sie zuvor schon einmal mit dem Virus in Kontakt gekommen sind, mit einer Wahrscheinlichkeit von 90% eine aktive HCMV Infektion entwickeln werden. Diese Infektion kann sich in 40% der Fälle zu einer augenlicht- oder lebensbedrohenden Krankheit entwickeln. Ein großes Problem bei der Verhinderung des Ausbrechens von AIDS bei HIV-positiven ist der Umstand, daß die HIV Risikogruppen eine nahezu universelle Seropositivität für HCMV aufweisen und sich einer erhöhten Gefahr der Wiederaansteckung mit HCMV ausgesetzt sehen.

Eine aktive Infektion bei an AIDS erkrankten Personen führt zu einem Befall aller Organe mit HCMV. Im besonderen werden die Lunge, das zentrale Nervensystem und der Gastrointestinaltrakt angegriffen. So entwickelt sich bei vielen Patienten eine lebensbedrohende Pneumonitis. Durch das hohe Maß an viraler Replikation entstehen auch Mutanten, die einen anderen Zelltropismus aufweisen als das Wildtypvirus und somit das zentrale Nervensystem angreifen, wo sie zu einer HIV Retinitis führen können (20-25%). Der Befall des Gastrointestinaltraktes kann in 2-13% der Fälle zu einer Ösophagitis, Gastritis oder Enterocolitis führen. Abschließend scheint das Humane Cytomegalievirus auch eine entscheidende Rolle bei der Entwicklung des für AIDS charakteristischen „wasting syndrome“ zu spielen.

## **1.3. HCMV Stämme**

### **1.3.1. Laborstämme**

Der Laborstamm *AD169* wurde ursprünglich aus Lymphgewebe eines infizierten Kindes isoliert. Durch häufiges passagieren in Fibroblastenkulturen wurde versucht das Virus abzuschwächen, um es als Immunisierungsvirus verwenden zu

---

können (Elek und Stern, 1974). *AD169* ist fungiert als klassischer Laborstamme von HCMV, da seine DNA Sequenz bereits seit 1990 zugänglich war (Chee et al, 1990) und es sich sehr gut in Fibroblasten kultivieren ließ. Die ursprüngliche Analyse des Genoms ergab eine Größe von 229354 bp die 208 ORFs enthielten, von denen 14 ORFs innerhalb wiederholter Regionen doppelt vorlagen (siehe: Abbildung 4). In letzter Zeit wurde die original Analyse durch Vergleich mit der DNA Sequenz des Schimpanzen Cytomegalievirus (Davison et al, 2003), einer erneuten Analyse mit leistungsfähigeren Genfindungs-Algorithmen (Murphy et al, 2003a) und einer nochmaligen Sequenzierung von *AD169* mit anschließendem Vergleich der DNA mit Sequenzinformationen anderer Laborstämme und klinischer Isolate (Murphy et al., 2003b) aktualisiert. Der Laborstamm unterscheidet sich jedoch von klinischen Isolaten in Genomstruktur und biologischen Eigenschaften. So fehlen *AD169* 19 ORFs die in klinischen Isolaten noch vorhanden sind (Cha et al., 1996) und das Virus ist nicht mehr in der Lage sich in Endothelzellen zu vermehren (MacCormac und Grundy, 1999).

Ein weiterer Laborstamm zu dem komplette Sequenzinformationen zur Verfügung stehen ist *Towne* (Murphy et al 2003b). Dieser Stamm wurde aus dem Urin eines mit HCMV infizierten Neugeborenen isoliert und auch zur Erzeugung eines Vakzinierungsstammes über viele Generationen in Fibroblasten kultiviert. Auch *Towne* weist in der selben Region wie *AD169* im Vergleich zu der DNA klinischer Isolate eine Deletion auf, die jedoch kleiner ist als in *AD169* (Cha et al., 1996).

### **1.3.2. Klinische Isolate**

Zu den komplett sequenziert klinischen Isolaten von HCMV gehören *Toledo*, *FIX*, *PH* und *TR*. Der in dieser Arbeit verwendete Stamm *Toledo* stammt genau wie *Towne* aus dem Urin eines infizierten Neugeborenen (Quinnan et al., 1984). *FIX* wurde von einer schwangeren Frau mit einer HCMV Primärinfektion (Hahn et al., 2002) isoliert, *PH* stammt von einem Patienten der eine Knochenmarkstransplantation erhalten hatte (Rice et al., 1984) und *TR* von einem AIDS Patienten mit Retinitis (Smith et al., 1998). Vergleiche der DNA Sequenzen untereinander haben ergeben das es zu den ursprünglich vorhergesagten 209 ORFs zusätzlich 29 stark konservierte Bereiche gab, die möglicherweise für zusätzliche Proteine kodieren. Alle diese ORFs sind kürzer als die 300 bp

---

Mindestlänge für einen ORF, die in der ursprünglichen Arbeit von Chee et al. 1990 benutzt worden waren.

## 1.4. Leserahmen

### 1.4.1. ORF (Offene Leserahmen)

Proteinkodierende Bereiche der DNA können mit Hilfe geeigneter Software leicht aus Sequenzdaten ermittelt werden. Hierfür wird standardmäßig als Startkodon ATG verwendet. Der offene Leserahmen sollte eine Mindestlänge von 240 Nukleotiden aufweisen, aus denen ein Protein von 80 Aminosäuren Länge synthetisiert werden kann.

### 1.4.2. Translationsinitiation (Kozak-Regeln)

Die Bindung des Ribosoms an die eukaryotische mRNA erfolgt an der 5' gelegenen 7-Methylguanosiniumkappe der mRNA. Von dort wandert das Ribosom in 3' Richtung bis es auf das erste AUG-Kodon trifft, von dem aus die Translation gestartet wird. Diese als „Scanning Model“ bekannte Theorie trifft für 95% der analysierten mRNAs (Kozak, 1987) zu. Jedoch gibt es auch Abweichungen von dieser Regel. Liegt ein weiteres AUG-Kodon in unmittelbarer Nähe des ersten Kodons so kann über einen Mechanismus, der als „leaky scanning“ bezeichnet wird, dieses zweite Kodon alternativ als Initiationsstart gewählt werden. Dieses Phänomen tritt auf, wenn das erste AUG nicht der optimalen Erkennungskonsensussequenz entspricht. Bei diese Konsensussequenz:

**gccAccAUGG**

sind besonders wichtig das Adenosin in Position -3 und das Guanin in Position +4 relativ zum Translationsstart (Kozak, 1986). Liegt das erste AUG-Kodon der mRNA nicht in diesem Nukleotidzusammenhang vor, so kann die 40S Untereinheiten des Ribosoms dieses Kodon überlesen und das 3' folgende AUG als Translationsstart verwenden. Jedoch funktioniert dieser Mechanismus nur bei nahegelegenen Start-Kodons, da weiter entfernte Kodons scheinbar durch das 80S elongierende Ribosom maskiert werden (Kozak, 1995). Ausnahmen hierzu treten nur auf, wenn kurz nach dem 5' proximalen AUG ein Stop-Kodon folgt

---

(Kozak, 1987). In manchen mRNAs findet man auch RNA-Sekundärstrukturen vor, die als IRES (Internal ribosomal entry sites) bezeichnet werden. An IRES-Strukturen können Ribosomen gezielt binden und somit in großer Entfernung zur mRNA-Kappe und an anderen 5' gelegenen ORFs die Translation beginnen. Diese Regeln zur Wahl des Translationsstarts sind als Kozak-Regeln bekannt.

### 1.4.3. Alternative Start Kodons

Neben dem hauptsächlich genutzten AUG als Start-Kodon der Translation gibt es auch alternative Kodons, die eine Bindung des Ribosoms veranlassen können. 1988 wurde beim Sendai Virus die Benutzung des Start-Kodons ACG beschrieben (Curran und Kolakofsky 1988) und ein Jahr später wurde ein ORF beim Moloney murine leukaemia virus entdeckt, der nicht mit dem postulierten AUG sondern mit einem 5' gelegenen CUG begann (Prats et al., 1989). Die bisher gefundenen alternativen Start-Kodons sind alle einmal Nukleotidabweichungen des Standard AUG Kodons. Es konnte bei Mutageneseversuchen am Dihydrofolat Reduktase ORF gezeigt werden, daß ACG, CUG, GUG, UUG, AUA, AUC und AUU als Translationsstarts akzeptiert wurden. Methionin tauchte in allen Versuchskonstrukten als erste Aminosäure im Protein auf, was auf eine nicht exakte Antikodonerkennung der <sup>Met</sup>tRNA schließen ließ (Peabody, 1987).

Die Bindung der <sup>Met</sup>tRNA an die alternativen Start-Kodons ist sehr schwach, weshalb die Translationseffizienz dieser alternativen Start-Kodons im Vergleich zu AUG gering ist. Die Bedeutung für den Organismus ist jedoch möglicherweise sehr groß, da durch alternative Startkodons Proteinisoformen entstehen, deren zusätzlichen aminoterminalen Domänen potentielle neue Funktionen beinhalten können. Diese Isoformen können zum Beispiel eine spezifische Rolle im Replikationszyklus des Virus übernehmen. Die meisten ORFs mit alternativen Startkodons findet man bei Protoonkogenen, Transkriptionsfaktoren, Kinasen und Wachstumsfaktoren (Touriol et al., 2003). Eine zusätzliche Funktion eines alternativen Startkodons wäre die Regulation des 3' gelegenen AUGs. Durch Initiation am alternativen Startkodon würde die Expression des 3' gelegenen ORFs herabgesetzt. Hierbei müßte das erzeugte Protein auch nicht funktionell sein, und das alternative Startkodon nicht im Leserahmen des 3' gelegenen ORFs liegen (Curran und Kolakofsky 1988).

---

Obwohl die Bindung der <sup>Met</sup>tRNA an nicht AUG Startkodons schwach ist, kann durch ein optimiertes Sequenzumfeld die Initiationsrate erhöht werden. Besonders wichtig ist hierbei das erste Kodon nach dem alternativen Startkodon. Findet man in Position +5 ein Adenosin und +6 ein Uracil, so kann die Expression alternativer Startkodons fast die Expressionsrate normaler AUG Kodons erreichen (Boeck und Kolakofsky 1994).

## 1.5. Zielsetzung der Arbeit

Das Humane Cytomegalievirus ist ein an seinen Wirt gut angepaßtes Virus, das bei gesunden Individuen eine latente oder persistierende Infektion hervorruft, die dem Wirt im allgemeinen nur geringen Schaden zufügt. Aufgrund der enormen Größe des Virusgenomes (~230 kbp) war zu Beginn der Arbeit nur der Laborstamm *AD169* komplett sequenziert worden. Zu den beiden anderen Laborstämmen *Towne* und *Toledo* standen nur begrenzte Sequenzinformationen für die virale DNA zur Verfügung. Es war jedoch bekannt, daß die Laborstämme Veränderungen in ihrem Zelltropismus aufweisen und zahlreiche Mutationen in nicht konservierten Regionen angesammelt haben.

In der vorliegenden Arbeit wurde die virale DNA des klinischen Isolates *Toledo* untersucht. Dieser Stamm sollte repräsentativ für alle klinischen Isolate des Humanen Cytomegalievirus mit der Sequenz des Laborstammes *AD169* verglichen werden. Über diese Vergleiche erhofften wir uns ORFs zu identifizieren, die für die Abwehr des Virus gegen Immunantworten des Wirtes wichtig waren. Aufgrund des mangelnden selektiven Druckes auf *AD169* in Zellkultur wären diese ORFs wahrscheinlich stark mutiert gewesen und hätten somit leicht identifiziert werden können. Die so identifizierten Gene hätten nachfolgend auf ihre Verwertbarkeit als Ziele für Medikamente oder Immunisierungen untersucht werden können.

Mit Hilfe der erarbeiteten Sequenzinformationen wurden Vergleiche zwischen den potentiellen ORFs des Laborstammes *AD169* und *Toledo* angestellt. Hierfür wurden sowohl die DNA-Sequenzen direkt miteinander verglichen als auch die aus den DNA-Sequenzen resultierenden Proteinsequenzen. Von den Arbeitsgruppen von T. Shenk und R. Meyers, von der Stanford sequencing group, mit erheblicher finanzieller Unterstützung des U.S. Department of Energy (DOE),



wurde die komplette DNA-Sequenz von *Toledo*, etwa zur Zeit der Fertigstellung der hier durchgeführten Sequenzierung, veröffentlicht. Daher mußte die Analyse der erhaltenen DNA-Sequenz über das normale Maß an Sequenzvergleich von ORFs erweitert werden. Hierzu sollten für die im Rahmen dieser Dissertation gefundenen ORFs die 5'-nichtkodierenden Bereiche auf alternative Startkodons untersucht werden.

Die experimentelle Arbeit wurde in Zusammenarbeit mit Prof. Dr. Shenk vom „Department of Molecular Biology“ der Universität Princeton, in Princeton, New Jersey, USA durchgeführt. Hierbei wurden virale DNA und Primer aus Princeton bezogen.

## 2. Material

### 2.1. Arbeitsmaterialien

Material	Hersteller
Einwegspritzen und Kanülen	Braun, Melsungen
Filterpipettenspitzen 10-1000µl	SLG, Gauting
Glaswaren	Schott, Mainz
Inkubator für Bakterienkulturen	Memmert
Kimwipes Papiertücher	Kimberly Clark, EEC
Parafilm „M“	American Can Company
PCR Reaktionsgefäße	Perkin Elmer, Norwalk, USA
PCR Cycler PE480, PE9600	Perkin Elmer, Norwalk, USA
pH-Meter	Knick
Photometer	Pharmacia, Uppsala, Schweden
Pipetten	Brang, wertheim
Plastik-Reaktionsgefäße	Eppendorf, Hamburg
Quarzküvetten	Hellma
Reaktionsgefäße, 15ml und 50ml	Greiner Falcon
Sterilfilter, Porengröße 22µm und 45µm	Millipore, Eschborn
Whatman-Filterpapier	Schleicher & Schüll, Dassel
Zentrifugen: Biofuge pico und TC6	Heraeus, Bad Homburg

Tabelle 2: Eingesetzte Arbeitsmaterialien

### 2.2. Allgemeine Chemikalien

Material	Hersteller
Agarose Ultra pure	FMC, Rockland, USA
Ammoniumpersulfat (APS)	Sigma Chemicals, Deisenhofen
Ampicilin (Binotal)	Grünenthal GmbH, Aachen

<b>Material</b>	<b>Hersteller</b>
Bacto Tryptone	Difco, Detroit, USA
Bacto-Yeast-Extract	Difco, Detroit, USA
Bacto Agar	Difco, Detroit, USA
Brom-Phenol-Blau	Sigma Chemicals, Deisenhofen
Dextransulfat	Biomol, Hamburg
DTE (Dithioerythrol)	Sigma Chemicals, Deisenhofen
DTT (Dithiothreitol)	Sigma Chemicals, Deisenhofen
Ethidiumbromid	Calbiochem, Los Angeles, USA
Glyzerin	Merck
H <sub>2</sub> O HPLC-grade	Merck
HEPES	Sigma, Deisenhofen
Milchpulver (Trockenmilch)	Glücksklee, München
Polyacrylamid (PAA-Lösung 38% / 2%)	Carl Roth GmbH, Karlsruhe
Saccharose	Merck, Darmstadt
TEMED	Merck, Darmstadt
Triton X-100	Sigma, Deisenhofen
Xylencyanol	Fluka

**Tabelle 3: Eingesetzte Chemikalien**

### **2.3. Kommerzielle Kits**

Die Kits wurden nach angaben des Hersteller verwendet.

<b>Kit-Bezeichnung</b>	<b>Hersteller</b>
Qiaquick spin Gelextraktions Kit	Qiagen, Hilden
GEM-T PCR Cloning Kit	Promega, Mannheim
RNeasy Midi Kit	Qiagen, Hilden

**Tabelle 4: Eingesetzte Kits**

## 2.4. Enzyme

Alle Enzyme wurden mit den vom dem Vertreiber gelieferten Puffern unter Befolgung der vom Hersteller empfohlenen Reaktionsbedingungen verwendet.

Enzym	Hersteller
<i>Bam</i> HI	Boehringer, Mannheim
<i>Lysozym</i>	Calbiochem, La Jolla, USA
Polynukleotidkinase aus T4 Phagen	Boehringer, Mannheim
<i>Pvu</i> II	Sigma, Deisenhofen
RNase A	Sigma, Deisenhofen
Taq DNA Polymerase	Sigma, Deisenhofen
<i>Xba</i> I	MBI Fermentas

Tabelle 5: Eingesetzte Enzyme

## 2.5. Desoxyribonukletide und Nukleinsäuren

Verdünnungen und Lösungen von Nukleinsäuren wurden mit H<sub>2</sub>O HPLC-grade angesetzt und bei -20°C gelagert.

Nukleinsäuren	Bezugsquelle
Desoxyribonucleotid-5'-triphosphate (dNTPs)	Boehringer, Mannheim
Gene Ruler 100bp ladder	Fermentas MBI
HCMV <i>Toledo</i> genomische DNA	Princeton University, USA
Lambda DNA / <i>Eco</i> 130I / <i>Mlu</i> I, No. 17	Fermentas MBI
Primer	Sigma Genosys
pGEM T Vektor	Promega, Mannheim
pUC Mix, No. 8	Fermentas MBI

Tabelle 6: Nukleinsäuren

## 2.6. Lösungen und Puffer

Die nachfolgend aufgeführten Puffer wurden alle mit deionisiertem Wasser, das über eine Millipore Super-Q Filtrationsanlage aufgereinigt wurde, angesetzt.

Lösungen und Puffer	Zusammensetzung
DNA Ladepuffer für Agarosegele	0,1M EDTA 1% SDS 50% Glycerin 1% Brom-Phenol-Blau 1% Xylencyanol
DnD	1,533g DTT 9ml DMSO 100µl Kaliumacetat (1M)
DYT-Medium	1,6% Bacto-Tryptone 1% Bacto-Yeast-Extract 0,5% NaCl
Ethidiumbromidlösung	0,5mg/ml in 1 X TBE
IPTG-Stammlösung	0,1M IPTG in H <sub>2</sub> O
Ladepuffer für denaturierende PAA Gele	0,3% Brom-Phenol-Blau 0,3% Xylencyanol 10mM EDTA 97,5% deionisiertes Formamid
LB-Agarplatten	1,25% Bacto Agar zum LB-Medium
LysozymbLösung	12,5g/ml Lysozym in 250mM Tris-HCL, pH 7,5
Milchpulver-Mix	½ Vol. 20%iges SDS 5% Milchpulver
Mg <sup>2+</sup> -Mix	1M MgCl <sub>2</sub> 1M MgSO <sub>4</sub>

PAA-Gel (denaturierend)	8% PAA-Lösung 1 X TBE 0,132mg/ml APS 1/1000 Vol. TEMED 8M Harnstoff
RNase A Lösung	5mg/ml RNase A in 10mM Tris HCl, pH 7,5 15 min bei 100°C denaturieren
STET-Lösung	50mM Tris-HCl, pH 8,0 50mM EDTA 0,5% Triton X-100 8% Saccherose
SOB-Medium	2% Bacto-Tryptone 0,5% Bacto-Yeast-Extract 1% 1M NaCl 0,25% 1M KCl
SOC-Medium	SOB –Medium mit 1% 2M Mg <sup>2+</sup> 1% 2M Glukose
TBE	89mM Tris-HCL, pH 7,6 89mM H <sub>3</sub> BO <sub>3</sub> 2mM EDTA
TE	10mM Tris-HCL, pH 7,5 1mM EDTA
TFB	10mM K-MES 45mM MnCl <sub>2</sub> 10mM CaCl <sub>2</sub> 100 mM KCl 3mM Hexamincobalt-Chlorid
X-Gal Lösung	50mg/ml X-Gal in N,N'Dimethylformamid

Tabelle 7: Lösungen und Puffer

## 2.7. Bakterien, Zellen und Viren

Bakterien, Zellen und Viren	Bezugsquelle
HFF-Zellen	Labor Marschall, Erlangen
HCMV <i>Toledo</i>	Labor Marschall, Erlangen
XL1 Blue MRF'	Stratagene, Heidelberg

**Tabelle 8: Bakterien, Zellen und Viren**

## 2.8 Software

Zur Vereinigung der einzelnen Sequenzen zu Contigs, zur Identifizierung von ORFs auf DNA Sequenzen sowie zu Sequenzvergleichen auf Proteinebene wurde das Programm „Vector NTI Suite 7“ vertrieben von Invitrogen verwendet.

Die Untersuchung der DNA auf Promotoren erfolgte über:

NNPP2.2: [http://www.fruitfly.org/seq\\_tools/promoter.html](http://www.fruitfly.org/seq_tools/promoter.html)

und Promoter 2.0 Prediction server: <http://www.cbs.dtu.dk/services/Promoter/>

Der Vergleich der gefundenen ORFs bei Toledo mit der Datenbank erfolgte über blastp von NCBI: <http://www.ncbi.nlm.nih.gov/BLAST/>

---

## Methoden

### 2.8. Arbeiten mit Bakterien

#### 3.1.1 Lagerung und Kultivierung von Bakterien

Frische Bakterienkulturen wurden in 67%igem Glycerol aufgenommen und umgehend bei -80 °C eingefroren. Neue Kulturen von diesen Stocks wurden in 2 ml dYT-Medium ü.N. kultiviert, wobei ein wenig der gefrorenen Substanz mit einer Plastikpipettenspitze abgekratzt und damit das Medium angeimpft wurde.

#### 3.1.2 Herstellung kompetenter Zellen

Von einer antibiotikafreien LB-Agarplatte wurde eine Kolonie des Bakterienstammes XL1Blue MRF' isoliert und in 10ml, auf 37°C vorgewärmtes SOB-Medium gegeben. Die Kultur wurde anschließend in einem Erlenmeyerkolben mit dem 10 fachen Fassungsvermögen der Nährlösung ü.N. kultiviert.

Von dieser Kultur wurden 2ml in 25ml vorgewärmtes SOB-Medium gegeben. Bei einer optischen Dichte von  $OD_{600}=0,7A$  wurde die Kultur in ein gekühltes Falconröhrchen (50ml) überführt und 15 min auf Eis gekühlt. Anschließend wurden die Bakterien bei ein 4 °C für 15 min bei 2500 UpM abzentrifugiert und der Überstand sorgfältig abgenommen. Nach Resuspension des Zellsediments in 1/3 Volumen (8ml) kaltem TFB-Puffers wurden die Bakterien wieder für 15 min auf Eis gelagert. Danach wurden die Zellen erneut abzentrifugiert und diesmal in 1/12,5 Volumen (2ml) kaltem TFB-Puffer resuspendiert, so daß eine finale Konzentration von  $1 \times 10^9$  Zellen pro ml vorlag. Nach Zugabe von 1/125 Volumen DnD-Lösung (140µl) wurden die Zellen für 10 min auf Eis inkubiert. Die vorliegenden kompetenten Zellen sollten nun nach dem unter 3.1.3 angegebenen Protokoll innerhalb der nächsten 20min transformiert werden.

#### 3.1.3 Transformation kompetenter Zellen.

Jeweils 4 µl eines gekühlten Ligationsansatzes wurden in ein Polypropylenröhrchen gegeben und mit 200 µl einer kompetenten Bakteriensuspension für 30 min auf Eis inkubiert. In einem auf 42 °C erhitztem Wasserbad wurden die Bakterien für 90 sek einen Hitzeschock unterzogen und



---

sofort wieder auf Eis abgekühlt. Zu dem Transformationsansatz wurden nun 800 µl vorgewärmtes SOC-Medium gegeben und die Kultur für 1 Std. bei 37 °C unter Schütteln inkubiert. Von diesem Ansatz wurden nun 200 µl auf eine mit Resistenzmarker und Färbemedium versetzte LB-Platte aufgetragen und ü.N. bei 37 °C inkubiert.

## **3.2 Extraktion und Reinigung von Nukleinsäuren**

### **3.2.1 DNA-Schnellpräparation nach Holmes und Quigley (1981)**

1,5 ml einer DNA-ü.N.-Kultur wurde in ein Eppendorfgefäß überführt und für 30 sek bei maximaler Umdrehungszahl in einer Tischzentrifuge sedimentiert. Der Überstand wurde verworfen, das Pellet in 420 µl kalter STET-Lösung resuspendiert und mit 25 µl Lysozymlösung versetzt. Nach kurzem vortexen wurde die Lösung für 90 sek in ein kochendes Wasserbad getaucht und anschließend 10 min bei 13000 UpM abzentrifugiert. Der klare, die DNA enthaltende Überstand wird in ein neues Gefäß gegeben und mit 28 µl 3 M Natriumacetat (pH 5,2) und 400 µl Isopropanol 15 min lang gefällt. Der Überstand wurde wieder verworfen und das DNA Pellet mit 70%igem Ethanol (-20°C) gewaschen. Das Ethanol wurde anschließend verworfen und die DNA in einem Exicator getrocknet und in 30 µl deionisiertem Wasser resuspendiert.

### **3.2.2 Ethanolfällung von DNA**

Das Präzipitieren von DNA aus Lösungen erfolgt durch Zugabe von 1/10 Volumen einer 3M Natriumacetat-Lösung (pH 5,2) und des 3 fachen Vol. an absolutem Ethanol (-20 °C) . Diese Lösung wurde ü.N. bei -20 °C gelagert. Die DNA wurde anschließend für 15 min bei 4 °C und 13000 UpM gefällt und das so erhaltene Pellet danach in 70%igem Ethanol gewaschen. Nach Trocknung der DNA wurde diese entweder in TE oder H<sub>2</sub>O aufgenommen.

### **3.2.3 Phenol/Chloroform Extraktion**

Die Phenol/Chloroform-Extraktion wurde zur Isolierung von Nukleinsäuren aus Lösungen verwendet, die auch Proteine und Lipide enthielten. Die zu reinigende Lösung wurde mit dem gleichen Volumen Phenol versetzt, 5 min durchmischt und anschließend zur besseren Phasentrennung zentrifugiert. Hierbei sammeln sich

---

die Proteine und Lipide in der phenolischen, die DNA in der wäßrigen Phase an. Die phenolische Phase wurde anschließend abgenommen und die DNA-haltige Lösung anschließend zweimal mit dem gleichen Volumen Phenol/Chloroform (1:1) aufgereinigt. Reste von Phenol wurden durch Zugabe von 1 Volumen Chloroform/Isoamylalkohol (24:1) aus der Lösung entfernt. Die Durchmischung der Phasen wurde jeweils 5 min vorgenommen und die Zentrifugation fand bei 3000 UpM (TC6) statt.

### **3.2.4 Aufreinigung von DNA aus Agarosegelen**

Über Agarosegele (siehe 3.5.2) aufgereinigte DNA-Fragmente konnten aus den Gelen ausgeschnitten und anschließend über kommerzielle Säulensysteme aufgereinigt werden. Bei diesem System (Perfectprep Gel Cleanup Kit, Eppendorf) wurde die DNA an eine positiv geladene Membran gebunden auf der sie mehrmals gewaschen und anschließend mit einem geringen Volumen H<sub>2</sub>O oder TE-Puffer eluiert wurde. Diese Säulensysteme wurden entsprechend den Angaben des Herstellers verwendet.

### **3.2.5 Photometrische Konzentrationsbestimmung von DNA/RNA**

Die Bestimmung der DNA/RNA-Konzentration in wäßrigen Lösungen erfolgte durch Detektion der spezifischen Absorption (OD = Optische Dichte) bei Wellenlängen von  $\lambda=260\text{nm}$  (OD<sub>260</sub>) und  $\lambda=280$  (OD<sub>280</sub>) bei DNA und RNA.

Eine OD<sub>260</sub> = 1 entspricht 50µg an doppelsträngiger und 33µg an einzelsträngiger DNA.

Bei RNA gilt: OD<sub>260</sub> = 1 entspricht einer Konzentration von 40µg/ml.

Der Quotient der Messungen OD<sub>260</sub>/ OD<sub>280</sub> gibt die Reinheit der Lösung an. Bei einer unkontaminierten DNA Lösung sollte der Wert zwischen 1,8 und 2,0 liegen.

## **3.3 Synthetische Desoxyribooligonukleotide**

### **3.3.1 Primer**

Die für die PCR Amplifikation und anschließenden Sequenzierungen verwendeten Oligonukleotide (Primer) wurden bei der Firma SIGMA Genosys in gereinigter und entsalzener Form bestellt. Bei der Auswahl der Primer wurde darauf geachtet, daß

---

der G/C Gehalt bei über 50% lag, keine Sekundärstrukturen ausgebildet wurden, am Anfang und Ende des Oligonukleotides möglichst 2 Nukleotide lagen, die drei Wasserstoffbrückenbindungen eingehen konnten. Die Schmelztemperatur sollte über 60°C liegen und die Primer sollten miteinander nicht dimerisieren können. Die gelieferten Primer wurden mit H<sub>2</sub>O (HPLC-grade) auf eine Stammkonzentration von 1µg/µl resuspendiert. Die Konzentration der Arbeitsverdünnungen betrug jeweils 20ng/µl. Lagerung der Primer erfolgte bei -20°C.

Einzelne Primer wurden auf einem denaturierendem PAA-Gel auf ihre Reinheit und mögliche Degradierung überprüft.

## **3.4 Amplifikation genomischer Fragmente**

### **3.4.1 Polymerase-Kettenreaktion (PCR) (Saiki et al., 1986)**

Mit Hilfe der PCR wurden definierte Bereiche des Virusgenoms spezifisch vermehrt und damit der späteren Sequenzierung zugänglich gemacht. Die zu amplifizierende DNA wurde hierbei mit jeweils 70ng eines jeden Primers versetzt und in 1xTaq Reaktionspuffer, 2,5mM MgCl<sub>2</sub>, einem dNTP-Mix (je 10mM jedes Nukleotides), sowie 1 Einheit Taq DNA-Polymerase je 25 µl Ansatz inkubiert. Bei schwierig zu amplifizierenden Bereichen des Virusgenoms wurde dem Ansatz 5% DMSO zugesetzt, um mögliche Sekundärstrukturen der DNA aufzulösen. Wurden unspezifische Produkte, zahlreichen Nebenbanden oder kein PCR-Produkt während der Reaktion gebildet, konnte durch Anpassung der Schmelztemperaturen der Primer oder der MgCl<sub>2</sub> Konzentration, die Reaktion in Hinsicht auf Produktmenge und Spezifität beeinflusst werden.

## **3.5 Restriktionsendonukleolytische Spaltung von DNA und gelelektrophoretische Trennung der Fragmente**

### **3.5.1 Spaltung von DNA mit Restriktionsendonukleasen**

Die DNA wurde jeweils nach dem von Enzymhersteller vorgegebenen Protokoll unter Verwendung der mitgelieferten Puffer gespalten. Es wurden zwischen 1-10 Einheiten Enzym pro µg DNA für jede Reaktion eingesetzt. Die Reaktionszeit

---

betrug mindestens 1 Std. Das Endvolumen entsprach etwa dem 3-5 fachen der eingesetzten DNA-Lösung.

### **3.5.2 Gelelektrophorese**

Die optische Analyse und Aufreinigung von DNA-Fragmenten, die über endonukleolytische Spaltung oder PCR-Amplifikation erzeugt wurden, erfolgte in Agarosegelen. Zur Herstellung der Gele wurde 1 % Agarose in 1 X TBE-Puffer aufgekocht und anschließend in einer Gelkammer aushärten gelassen. Die zu untersuchenden DNA-Fragmente sowie ein kommerzielle Referenz-DNA (DNA-Leiter) wurden in dem Gel bei 100 mA entsprechend ihrer Größe aufgetrennt. Zur Visualisierung der DNA wurde das Gel in eine Ethidiumbromidlösung für 10 Minuten gefärbt und anschließend weitere 10 Minuten in einem Wasserbad entfärbt. Die DNA konnte nun anhand des interkalierenden Farbstoffes Ethidiumbromid unter UV-Licht detektiert werden.

### **3.5.3 Trennung von DNA Fragmenten in denaturierenden Polyacrylamidgelen**

In der vorliegenden Arbeit wurden Polyacrylamidgele (PAA-Gele) zur Überprüfung kurzer, einzelsträngiger DNA Oligomere auf Denaturierung verwendet. Die Gele wurden mit einer frisch angesetzten PAA-Lösung zubereitet, die einen Vernetzungsgrad von 38:2 aufwies und nach der Zugabe des Radikalspenders TEMED, zur Polymerisation zwischen zwei Glasplatten gegossen wurde. Nach der Aushärtung erfolgte ein Vorlauf in einem 1xTBE Laufpuffer für 30 min bei einer Stromstärke von  $I = 5 \text{ mA/cm}$ . Die zu testenden Oligomere wurden dann in Ladepuffer aufgenommen und bei einer Stromstärke von  $I = 15 \text{ mA/cm}$  aufgetrennt. Die Dokumentation erfolgte unter UV-Licht nach vorheriger Färbung der DNA mit Ethidiumbromid.

## **3.6 RNA**

### **3.6.1 RNA-Isolierung aus HCMV infizierten Zellen**

Mit HCMV infizierte HFF-Zellen (human foreskin fibroblasts) wurden nach vier Tagen Inkubationszeit geerntet. Hierzu wurden das Kulturmedium (Dulbecos

---

Medium mit 5% FKS) entfernt, die Zellen zweimal mit PBS-d gewaschen und anschließend mit Trypsin vom Flaschenboden abgelöst. Diese Zellsuspension wurde anschließend zentrifugiert und das Zellsediment zur Auszählung in 1 ml Kulturmedium resuspendiert. Für die RNA-Isolierung wurde das RNeasy Midi Kit (Qiagen) verwendet, das entsprechend Herstellerprotokoll verwendet wurde.

### **3.6.2 Aufreinigung der isolierten RNA**

Aus der mit dem RNeasy Midi Kit (Qiagen) aufgereinigten DNA wurde anschließend über Oligo(dT) Latex Kügelchen die mRNA isoliert. Verwendung fand hierbei das NucleoTrap mRNA Kit von Macherey – Nagel. Das Kit wurde nach den Angaben des Herstellers verwendet.

### **3.6.3 RT-PCR**

Der Nachweis der gesuchten mRNA erfolgte über ein abgewandeltes PCR-Protokoll (siehe 3.4.1). Hierbei wurde die zu amplifizierende RNA mittels einer Reversen Transkriptase in einzelsträngige DNA umgewandelt und anschließend unter normalen PCR Bedingungen amplifiziert.

## **3.7 Zellkultur**

### **3.7.1 Infektion der HFF-Zellen**

Alle nachfolgenden Schritte wurden unter sterilen Bedingungen durchgeführt.

Die zur Infektion verwendeten HFF-Zellen wurden im konfluenten Zustand für die Infektion vorbereitet. Um eine optimale Infektion zu gewährleisten wurde das alte Nährmedium verworfen und durch neues ersetzt. Für die Infektion wurde jedoch nur 1/10 des ursprünglichen Volumens an Nährmedium zugesetzt und mit 10 PFU (Plaque Forming Units) pro Zelle HCMV *Toledo* Virus angereichert. Die Zellen wurden danach für 96h bei 37°C im Inkubator aufbewahrt und danach geerntet.

---

## 3. Ergebnisse

### 3.1. Sequenzierung des Genoms von HCMV Toledo

#### 4.1.1 Experimenteller Plan

Die Sequenzierung hochmolekularer DNA erfolgt normalerweise über ein als „DNA-shotgunning“ bezeichnetes Verfahren. Da aber zu Beginn der Arbeit der Laborstamm *AD169* von HCMV bereits vollständig sequenziert war, entschlossen wir uns statt der „shotgunning“ Methode, eine auf PCR-Fragmenten basierende „chromosomal walking“ Methode zu verwenden. Über überlappende PCR-Fragmente, deren Primer anhand der *AD169* Sequenz erstellt werden konnten, wurde das gesamte Genom von Toledo gezielt in kleinere, sequenzierbare Fragmente unterteilt. Diese direktere Sequenzierung ersparte uns das Anlegen einer viralen DNA-Library in Plasmiden, und deren Anreicherung in Bakterien, wodurch wir hofften potentielle Mutageneseereignisse in unserer DNA während der Amplifikation in Bakterien vermeiden zu können. Zusätzlich erlaubte die PCR gestützte Sequenzierung ein zielgerichteteres Arbeiten. Bestimmte Regionen des Virus konnten amplifiziert und sequenziert werden, ein Verfahren welches bei der „DNA-shotgunning“ Methode nicht möglich gewesen wäre.

Gleichzeitig mit der Vollendung der hier durchgeführten Sequenzierung des Stammes *Toledo* wurden die kompletten Sequenzen von in BAC-Vektoren integrierten Genomen der Laborstämme *AD169* und *Towne*, sowie der klinischen Isolate *Toledo*, *FIX*, *PH* und *TR*, von der Arbeitsgruppe um T. Shenk, mit der Unterstützung des U.S. Department of Energy, veröffentlicht. Daher mußte die Analyse der von mir zunächst sequenzierten Bereiche über das bis dahin angestrebte Vergleichen der DNA und Proteinsequenzen von *Toledo* mit den bereits bekannten ORFs von *AD169* hinausgehen. So wurde die *Toledo* Sequenz nicht nur auf herkömmliche ORFs untersucht, sondern auch auf ORFs, die mit den alternativen Startkodons GUG und CUG initiiert wurden. Zusätzlich wurde die DNA-Sequenz auch noch auf mögliche Promotorbindungsstellen für die alternativen Startkodons untersucht.

---

### 4.1.2 Amplifikation und Sequenzierung der DNA-Fragmente von HCMV Toledo

Die Sequenzierung des HCMV *Toledo* Genoms erfolgte in dieser Arbeit über PCR-Fragmente. Hierzu wurden auf der Basis der bekannten *AD169* Sequenz PCR Primer kreiert, die 2200 bp große Bereiche der viralen DNA amplifizieren konnten. Die Primer wurden dabei so gewählt, daß benachbarte Amplikons jeweils um 250 bp überlappten, sodaß die fertig sequenzierten Fragmente leichter bündig angeordnet werden konnten. Da für die Amplifikation der viralen DNA keine DNA-Polymerase mit 3' Exonukleaseaktivität verwendet wurde, mußten 4 separate PCR Ansätze pro Zielfragment erzeugt werden, um mögliche Fehler der Polymerase zu ausschließen zu können.

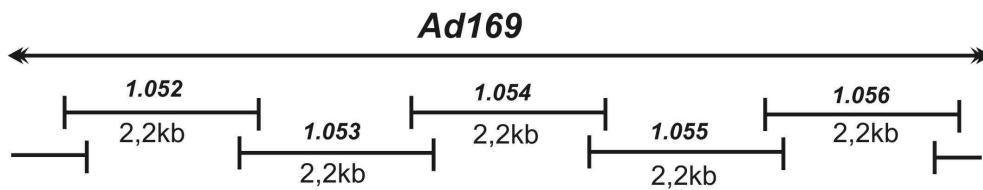
Die durchschnittliche Informationsausbeute eines erfolgreichen Sequenzierungsansatzes bestand in Nukleotidfolgen von ca. 600 bp. Um das gesamte PCR Produkt zu sequenzieren mußten also noch zusätzlich Sequenzierungsprimer erzeugt werden. Diese Primer wurden dann anhand der vorliegenden Sequenzdaten der Toledo DNA so gewählt, daß die neu mit diesen Primern erzeugte Sequenz mit der vorhergehenden um mindestens 50 bp überlappte. Für jeden Strang (+/-) wurden 4 Primer verwendet, wobei jede Sequenz zweimal bestimmt wurde. Somit wurde jedes Nukleotid, unter Verwendung von zwei verschiedenen Primern, 4 mal bestimmt (Abb. 5).

### 4.1.3 Übersicht der PCR-Produkte

Da die zur Erzeugung der PCR-Produkte der *Toledo* DNA verwendeten Primer nur homolog zur *AD169* Sequenz waren, mußte damit gerechnet werden, daß nicht alle PCR-Reaktionen Ergebnisse erbringen konnten. Von den 116 verwendeten Primerpaaren funktionierten nur 59 unter Standard PCR-Bedingungen gut genug, sodaß man die Produkte isolieren konnte (Abb. 6 / Abb. 7, grüne Kästchen). 28 PCR-Produkte konnten erst nach individueller Optimierung der  $MgCl_2$  Konzentration und der Anlagerungstemperaturen der Primer, sowie in einigen Fällen durch Zugabe von DMSO zur Auflösung von DNA-Sekundärstrukturen, gereinigt werden (Abb. 6 / Abb. 7, blaue Kästchen).

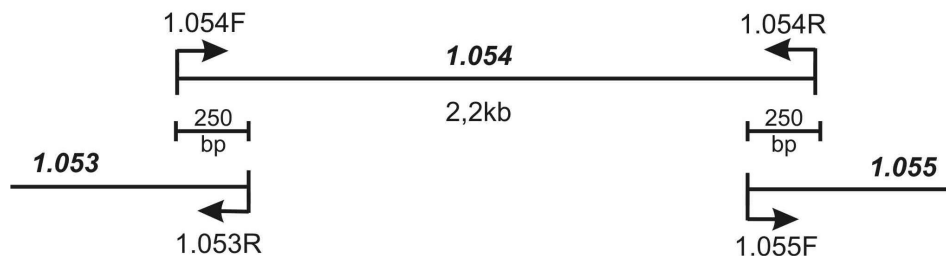
a)

Unterteilung des *Ad169* Genoms in 2,2kb große, überlappende Fragmente



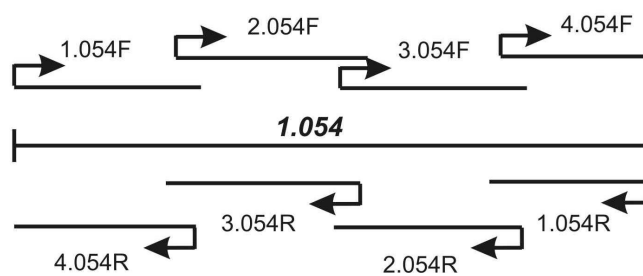
b)

Endsequenzierung eines PCR-Produktes mit den PCR-Primern



c)

Sequenzierung des Fragmentes anhand *Toledo* spezifischer Primer, die von den zuvor ermittelten Sequenzen abgeleitet wurden



**Abb. 5: Schematische Darstellung des Amplifikations- und Sequenzierungsplans.**

a) Exemplarische Bereiche des *AD169* Genoms und die jeweiligen PCR-Produkte basierend auf der DNA des Stammes *Toledo*. b) Das PCR-Produkt 1.054 mit seinen jeweiligen Nachbarn. Die Produkte überlappen jeweils mit ca. 250 bp. Die Pfeile zeigen die jeweiligen PCR-Primer an, die für die initiale Sequenzierung verwendet wurden. c) zeigt das PCR-Fragment 1.054 mit seinen insgesamt 8 Primern. Die Primer 2., 3. und 4. wurden aufgrund der Sequenzdaten der vorherigen Analysen entworfen.

Nomenklatur: die erste Ziffer steht für die jeweilige Primerposition und die Zahl nach dem Punkt gibt das PCR-Fragment ausgehend vom linken Ende des Referenzstammes *AD169* an, z.B. ist 2.054F der Primer der auf der ersten Sequenzierung des Fragmentes 1.054 mit dem „Forward“ Primer 1.054F beruht und ca. 600 Nukleotide in 3' Richtung von dem Ursprungsprimer liegt.

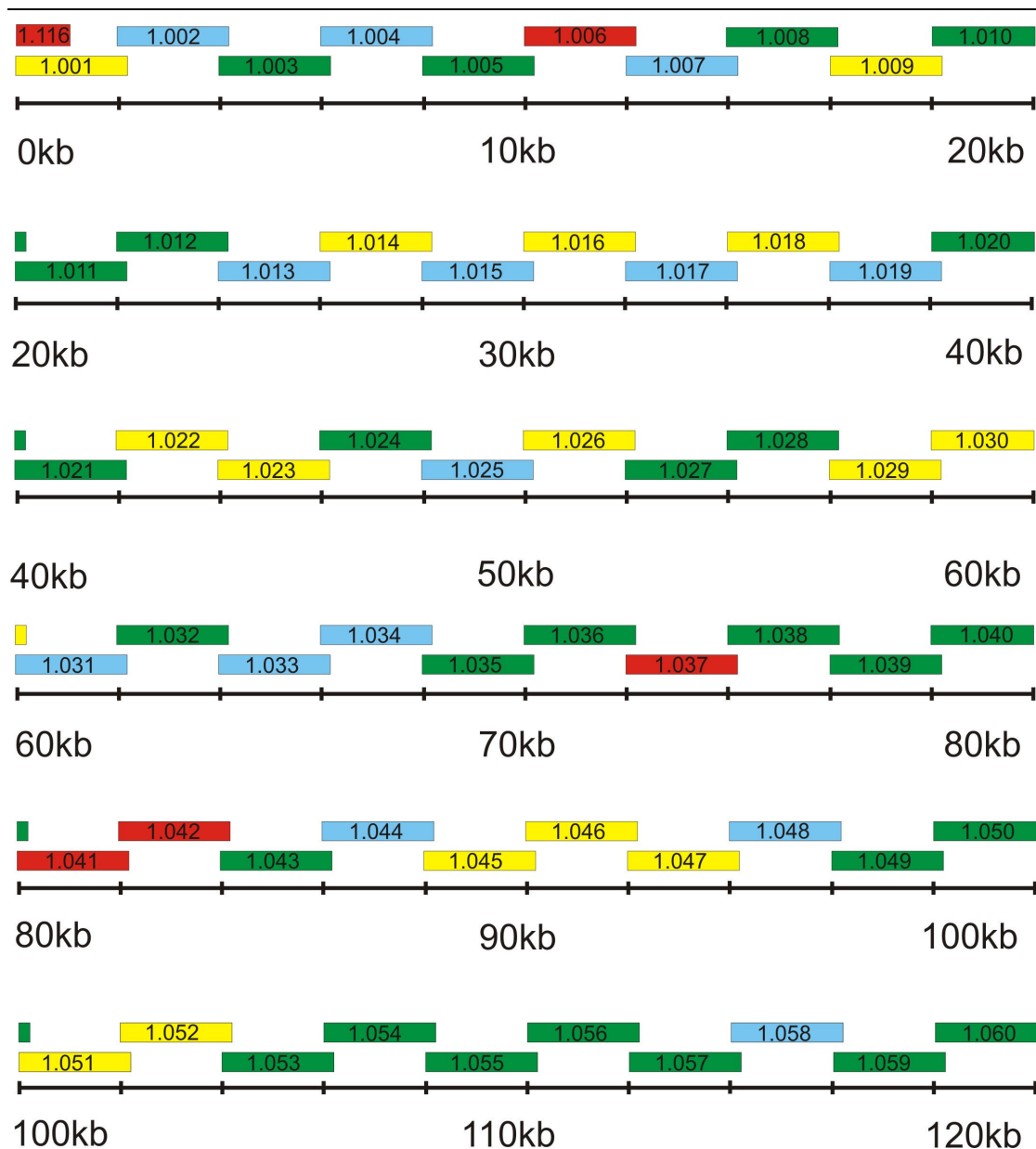


---

Für 19 PCR-Fragmente mußten unter Verwendung bereits sequenzierter Nachbarfragmente neue Primer entworfen werden (Abb. 6 / Abb. 7, gelbe Kästchen), um isolierbare Produkte zu erhalten und 10 Fragmente ließen sich auch mit neuen, *Toledo*-Sequenz homologen, Primern nicht amplifizieren (Abb. 6 / Abb. 7, rote Kästchen). Auch eine Verringerung der Amplifikatgröße auf 1000 bp, oder eine Überbrückung der nicht amplifizierbaren Bereiche mittels einer LA-PCR (Long and accurate – PCR), einer modifizierten PCR-Reaktion, die Amplifikate bis zu einer Größe von 20 kb erzeugen kann, erbrachten keine Resultate. Zum Beispiel konnte das Produkt 1.037 (Abb. 6) nicht amplifiziert werden, obwohl die Sequenzierung der benachbarten PCR-Fragment ergab, das die PCR-Primer homolog zur *Toledo* Sequenz waren. Auch mit zwei alternativen Primerpaaren konnte dieser Bereich nicht amplifiziert werden.

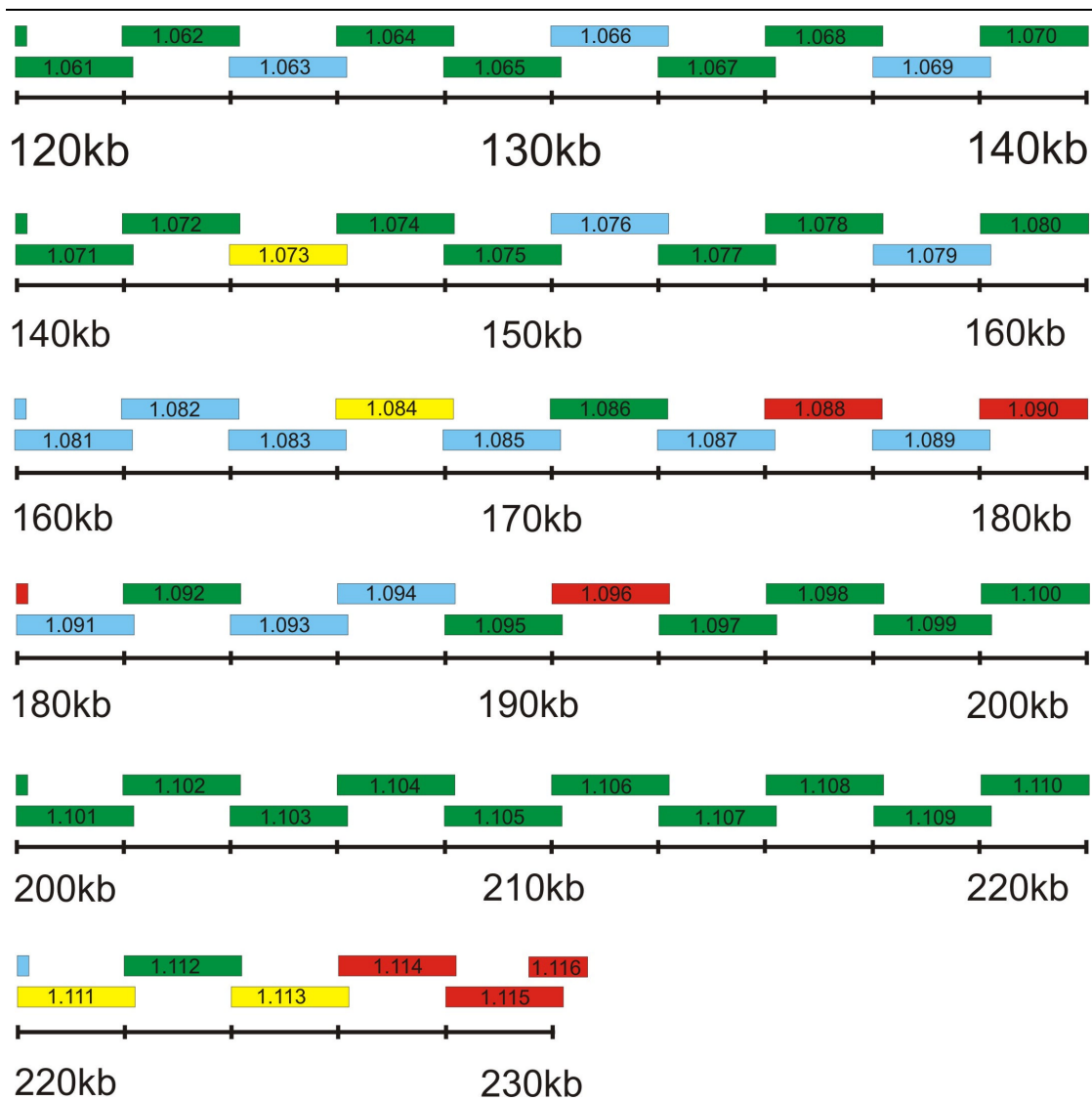
#### 4.1.3 Sequenzierung der PCR-Produkte

Die Sequenzierung der *Toledo* PCR-Fragmente erfolgte sowohl über *AD169*, als auch *Toledo* homologe Primer. Eine gute Sequenzierungsreaktion erbrachte dabei eine Sequenzfolge von 600 – 800 Nukleotiden. Da jedes Nukleotid vier mal validiert werden sollte, umfaßte jedes komplett sequenzierte Fragment mindestens 16 Sequenzierungsansätze. Für das gesamte Genom ergab das 1856 Sequenzreaktionen. Jedoch mußten die Sequenzreaktionen, wie auch schon die PCR-Reaktionen, in etwa der Hälfte der Fälle individuell optimiert werden. Obwohl Standardprotokolle verwendet wurden und die DNA über Agarosegele und kommerzielle Säulensysteme aufgereinigt worden war, traten eine Vielzahl von Problemen auf. So gab es Reaktionen die, obwohl die DNA selektiert worden war, den Eindruck erweckten, als ob die Sequenzierung nicht spezifisch war. Die erste Sequenzierung des Fragmentes 1.038 mit dem ersten Vorwärtsprimer (siehe Abb. 8b) ist dafür beispielhaft. Die Signale der Chromatophoren überlappten in vielen Positionen und eine eindeutige Sequenz war nicht zu ermitteln. Das häufigste Problem war jedoch ein nicht optimales Verhältnis zwischen Primer und Template-DNA. War zu viel Primer während der Reaktion zugegen,



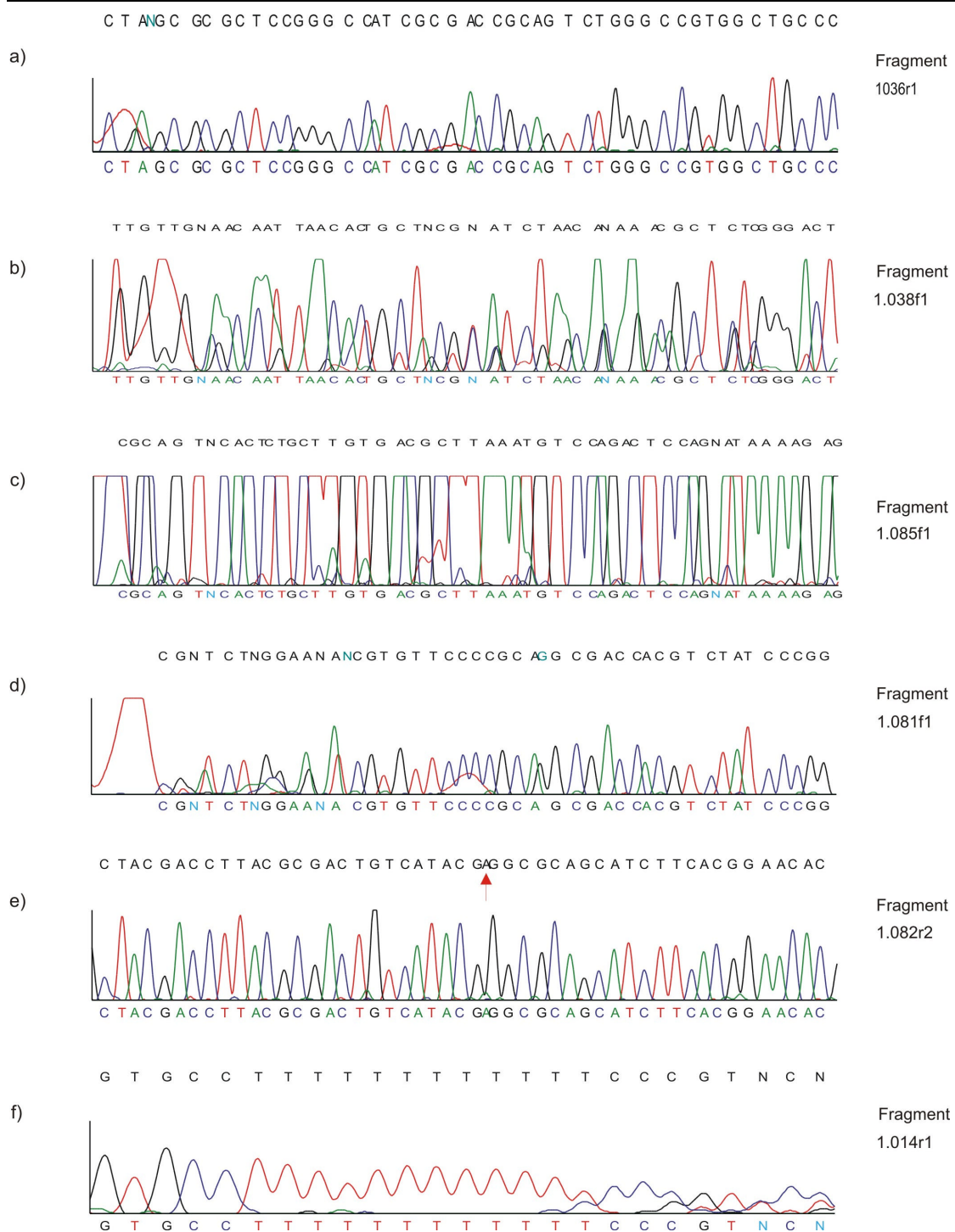
**Abb. 6: Übersicht der Toledo PCR-Produkte / Erster Teil**

so waren die Signale der Nukleotide am Anfang der Sequenz sehr hoch, fielen dann jedoch sehr rasch ab, sodaß nach einer Länge von circa 300 Basen die Signale so stark abgeschwächt erschienen, daß sie nicht mehr gelesen werden konnten (Abb. 8c). Da die sehr starken Signale zu Beginn der Sequenzen eine Kontrolle der computerberechneten Sequenz zudem nicht zuließen, konnten diese kurzen Sequenzen nicht verwertet werden. In diesem Fall mußten unterschiedliche Primerkonzentrationen ausgetestet werden um eine Sequenz von wenigstens 500 Nukleotiden Länge zu erhalten. Der Anfang der Sequenz 1.081f1 zeigt ein weiteres, wenn auch nicht so gravierendes Problem (Abb. 8d). Die ersten ~20 Nukleotide der Sequenz waren fehlerhaft,



**Abb. 7: Übersicht der Toledo PCR-Produkte / Zweiter Teil.** Die Abbildungen 2 und 3 zeigen das *Toledo* Genom mit den Positionen der jeweiligen PCR-Produkte als Blöcke. Grün dargestellt sind PCR-Produkte die unter Standard PCR Bedingungen das gewünschte Produkt spezifisch vermehrt. Die blauen Blöcke zeigen PCR-Produkte, die zwar das gesuchte Fragment amplifizierten, aber entweder zu wenig Produkt lieferten oder Nebenbanden erzeugten, so daß die Reaktion individuell modifiziert werden mußte. Die gelben Blöcke zeigen Fragmente an, bei denen die ursprünglichen *AD169* Primer nicht ausreichende Homologie zur *Toledo* Sequenz hatten. In diesen Fällen mußten Primer neu entworfen werden. Rote Fragmente waren auch nach Erzeugung neuer, *Toledo* Sequenz homologer Primer nicht amplifizierbar.

jedoch war die Sequenz danach absolut perfekt. In den meisten Fällen wurde der betroffene Bereich während der Analyse nicht berücksichtigt, da jedoch die Überlappungsbereiche der einzelnen Sequenzen in diesen Bereichen lagen, konnte es vorkommen, das die benachbarten Fragmente nur über die Sequenzen des Gegenstranges zusammengefügt werden konnten.



**Abb. 8: Übersicht problematischer Sequenzen.** Mit Ausnahme von e) und f) sind bei allen gezeigten Chromatographen die jeweiligen Anfänge der Sequenzen zu sehen. a) stellt eine optimal Sequenz dar. b) ist nicht auswertbar, da hier zwei Sequenzen überlappen. c) ist ein typisches Beispiel von unausgewogenen Primer/Template Verhältnis in der Reaktion, die Chromatophoren werden zu Beginn der Reaktion verbraucht, weshalb die lesbare Länge der Sequenz bei unter 400 bp lag. d) ist eine gute Sequenz, jedoch sind die anfänglichen Basen nicht auswertbar. In e) wurde fälschlicherweise Hintergrundsignal ausgewertet und f) zeigt eine T-Repetition mit nachfolgenden unlesbaren Sequenzbereichen.

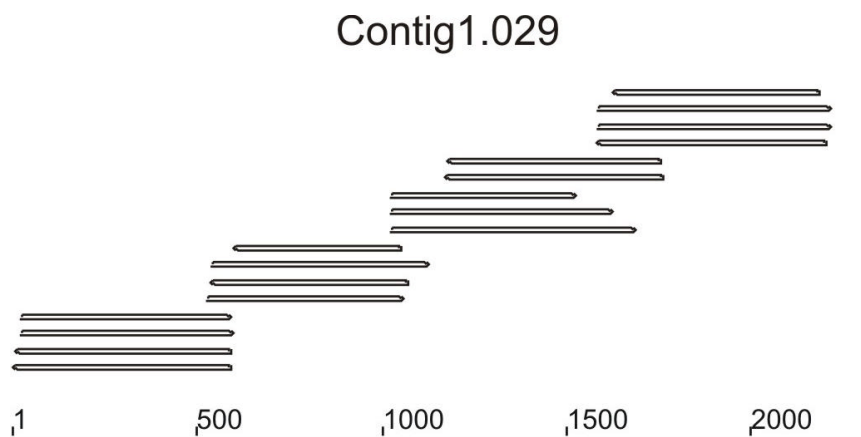
---

Zusätzlich mußten diese Bereiche noch ein weiteres mal sequenziert werden, um eine vierfache Validierung jedes Nukleotides zu gewährleisten. Der in Abb. 8e dargestellte Bereich der Sequenz 1.082r2 zeigt nicht wie bei den vorher behandelten Sequenzen den Anfang der Nukleotidfolge, sondern wurde aus der „Mitte“ der Sequenz entnommen. Obwohl viele Sequenzen sehr gut lesbar waren, deutete die Software „Hintergrundstrahlung“ mitunter als Signal und bettete, wie mit einem roten Pfeil in Abb. 8e dargestellt, eine Nukleotid zusätzlich in die Sequenz ein. Sämtliche Sequenzen mußten deshalb von Hand kontrolliert und mitunter neu interpretiert werden. Hierbei diente einerseits die Intensität des Signals, andererseits der Abstand der Nukleotide voneinander eine Rolle bei der Beurteilung potentieller Software bedingter Sequenzierungsfehler. Ein selteneres Problem bei der Sequenzierung waren längere Folgen eines Nukleotides, wie die in Abb. 8f gezeigte Thymin-Folge in Fragment 1.014r1. Die häufige Wiederholung eines Nukleotides störte die Sequenzierung beträchtlich, da dann die nachfolgenden Nukleotide nicht mehr korrekt erkannt werden konnten. Da die Sequenzierung von beiden Richtungen aus ( +/- Strang) nicht korrekt über diese Stelle hinweglesen konnte, hatte man zwar die Sequenzen zu beiden Seiten der Repetition, aber die genaue Bestimmung der Anzahl der wiederholten Nukleotide war schwierig. Wie Abb. 8f, zeigt ist die Intensität des letzten T-Nukleotides schwächer als die der Vorläufer. An der gleichen Position befindet sich auch noch ein schwaches Signal einer Cytosinbase. Zur genauen Bestimmung solcher Repetitionen wurden diese Bereiche acht mal sequenziert und die Intensitäten der jeweils letzten Basen miteinander verglichen.

Zur Kontrolle der hier vorgestellten Sequenzdaten und der Reproduzierbarkeit der Sequenzierungen, wurden die Konsensussequenzen von 11 DNA-Segmenten von jeweils 2,2 kb Größe, die gleichmäßig über das Genom von HCMV *Toledo* verteilt waren, mit den jeweiligen Einzelsequenzen aus denen die Segmente zusammengesetzt waren, verglichen. Innerhalb dieser 24200 überprüften Nukleotide konnte nur eine einzige Sequenzabweichung zwischen einem Sequenzierungsansatz und der Konsensussequenz festgestellt werden. Hierbei handelte es sich um eine einzelne Basenpaarabweichung. Fälschliche Insertionen oder Deletionen von Nukleotiden, die ORFs zerstören könnten, wurden nicht entdeckt.

## 4.2 Übersicht der Sequenzen

Für die Analyse der Toledo-DNA mußten die einzelnen Sequenzen zu größeren, zusammenhängenden Fragmenten (Contigs) vereinigt werden. Zuerst wurden die Sequenzen der jeweiligen PCR-Produkte zusammengefügt (Abb. 9). Der Überlappungsbereich der einzelnen Sequenzen mußte dabei mindestens 30 bp betragen um eine eindeutige Zuordnung der benachbarten Sequenzen zu ermöglichen. Die so erzeugten Contigs enthielten die

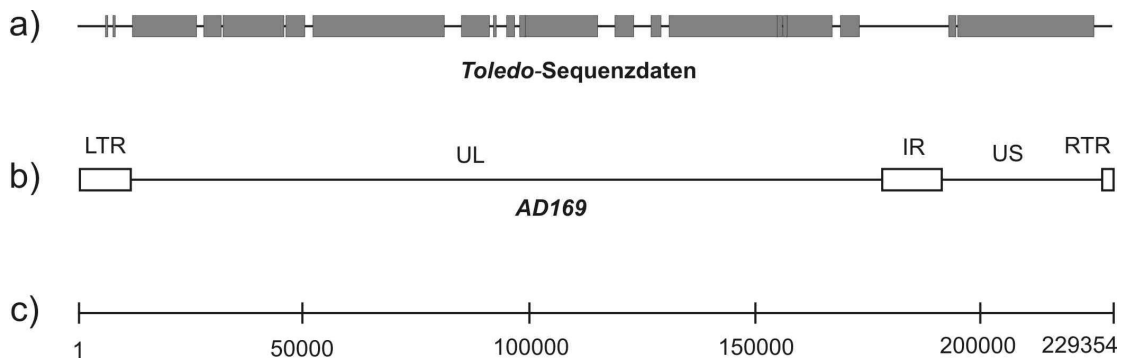


**Abb. 9: Contig 1.029.** Die einzelnen Sequenzen sind als horizontale Blöcke dargestellt. Die Überlappungsbereiche der Sequenzen sind sichtbar. Die Längenvarianz einiger Sequenzen verdeutlicht die Resultate der individuellen Anpassungen der Sequenzierungsreaktionen (dritter Block von links).

komplette Sequenzinformation eines PCR-Fragmentes und konnten wiederum mit den Contigs benachbarter PCR-Produkte verschmolzen werden. Die hieraus resultierenden DNA Abschnitte waren häufig mehrere Kilobasen lang und konnten zur Analyse des Gengehalts und der Genomstruktur des Toledo-Stammes verwendet werden.

Eine Übersicht aller sequenzierter Bereiche von HCMV *Toledo* zeigt Abb. 10. Die DNA Sequenzen des *Toledo* Stammes sind hierbei in Relation zu dem als PCR Matrize dienendem Stamm AD169 gesetzt worden. Die vergleichsweise große Lücke bei der internen Repetition („internal repeat“) wurde nicht sequenziert, da zu Beginn der Arbeit dieser Sequenzbereich bereits veröffentlicht worden war (Cha et. al., 1996). Des weiteren fallen Lücken bei den terminalen Wiederholungen (LTR, RTR) auf. Der RTR konnte nicht sequenziert werden, da die Region über PCR nicht amplifiziert werden konnte (siehe Abb. 7). Für die linke terminale Wiederholung konnten zwar alle PCR-Fragmente erzeugt werden, jedoch traten

bei der Sequenzierung dieser Fragment erhebliche Probleme auf. Somit konnten in dieser Region nur zwei kurze Bereiche sequenziert werden, nämlich die ersten und letzten ~500 Nukleotide des Fragmentes 1.004 und die ersten ~500 Nukleotide von 1.005 (siehe Abb. 10, kurze Blöcke ganz links).



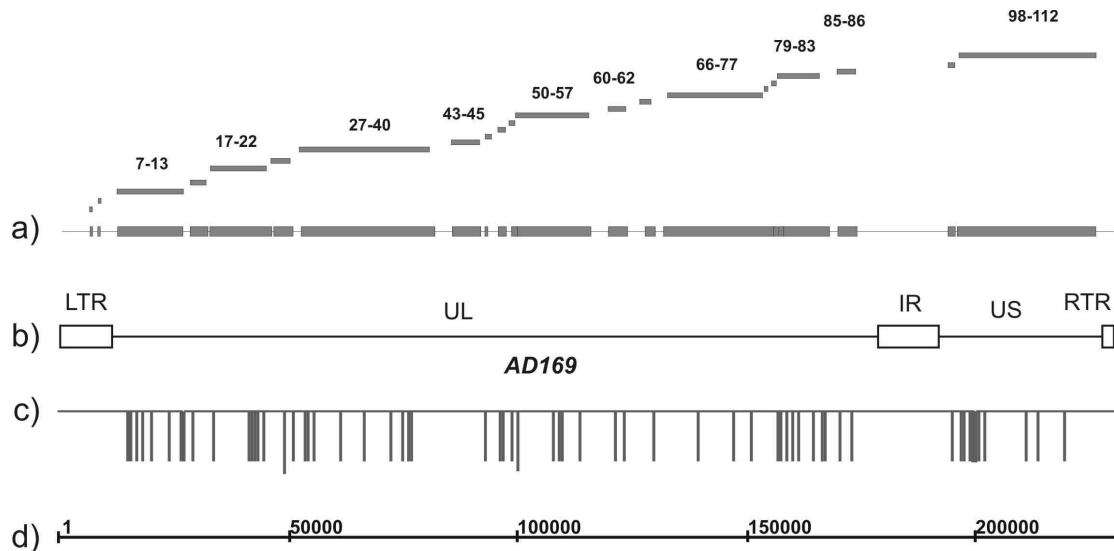
**Abb. 10: Toledo Sequenzübersicht.** Die grauen Blöcke in a) repräsentieren zusammenhängende Nukleotidfolgen von mindestens 500 bp Länge des klinischen Isolates *Toledo* und sind in räumliche Beziehung zur Genomstruktur des Referenzstammes *AD169* gebracht. b) eine graphische Darstellung der Genomstruktur von *AD169*, (unique long (UL), unique short (US), left terminal repeat (LTR), right terminal repeat (RTR) und internal repeat (IR)). c) zeigt eine Größenübersicht des Genoms in Basen.

Die Gesamtsequenz ist in der GenBank unter den folgenden Kennnummern veröffentlicht: AY486470-486480.

### 4.3 Analyse der genomischen DNA von HCMV *Toledo*

Mit Hilfe der Programme „align2sequences“ (blast) und AlignX (Vector NTI Suite7) wurden die zuvor erzeugten Sequenzen mit der veröffentlichten Sequenz des Laborstammes *AD169* (NC\_001347) verglichen. Hierzu wurde zuerst die komplette Sequenz von *AD169* mit allen in dieser Arbeit erzeugten Sequenzen von *Toledo* verglichen (Abb. 11), um eine Übersicht über lokale Häufungen von Sequenzdivergenzen zu bekommen. Hierbei stellte sich heraus, daß das Genom von *Toledo* Regionen enthielt, in denen die Abweichung zum Referenzstamm *AD169* sehr groß waren und Bereiche, die stark konserviert schienen. Auffällig war eine Häufung von Mutationen in den  $U_L$  und  $U_S$  Regionen in relativer Nähe zum Internal Repeat, sowie im  $U_L$  in der Nähe des LTR. Die in der Nähe der Wiederholungen liegenden DNA-Fragment „17-22“, „79-83“ und „98-112“ wiesen im Vergleich mit den restlichen Fragmenten eine deutliche höhere Mutationsdichte auf. Die Fragmente „43-45“ und „66-77“ hingegen waren im Vergleich stärker konserviert als die restlichen Fragmente.

Eine genauere Darstellung der Abweichungen zwischen den beiden Stämmen erhielt man bei Vergleich einzelner Fragmente mit den homologen Regionen von HCMV *AD169* (siehe: Abb. 12 bis Abb. 15). In diesen Abbildungen sind alle Nukleotidabweichungen eingezeichnet und nicht nur wie in Abb. 11 die größeren



**Abb. 11: Sequenzvergleich zwischen *Toledo* und *AD169*.** Die einzelnen Contigs von *Toledo* wurden in Relation zum Referenzstamm *AD169* gesetzt und mit diesem verglichen. a) Die sequenzierten Contigs sind sowohl auf einer horizontalen Geraden, als auch einzeln mit horizontaler und vertikaler Trennung zur besseren Betrachtung und mit Beschriftung der größeren Fragmente dargestellt. Kleinere Contigs wurden aus Platzgründen nicht beschriftet. Nicht benannt sind: 4/1; 4/2; 15-16; 24-25; 46; 48; 49; 64; 78/1; 78/2; 97. b) zeigt die graphische Darstellung der Genomstruktur von *AD169*. Die vertikalen Balken in c) direkt über dem Größenstandard (d)) zeigen die Positionen von Mutationsschwerpunkten an. Sichtbar in dieser Darstellung sind nur Bereiche in den Mutationen gehäuft auftraten, einzelne Mutationen wurden in dieser Darstellung nicht berücksichtigt. Die Länge der vertikalen Balken gibt Auskunft über die relative Anzahl an Mutationen.

Mutationsschwerpunkte, die aus längeren nicht homologen Bereichen sowie größeren Deletionen oder Insertionen bestehen. Ein Vergleich ähnlich großer Fragmente, wie z. B. 7-13 und 17-22 (siehe Abb. 12) zeigte besonders deutlich die unterschiedliche Verteilung von Mutationsereignissen auf dem Genom von *Toledo*. Während 17-22 eine Divergenz von 1.94% (siehe Tabelle 9) aufwies, unterschieden sich *Toledo* und *AD169* im Bereich des Fragmentes 7-13 zu 6,46%. Die Schwerpunkte der Mutationen lagen dabei in den Bereichen, in denen Glycoproteine vorhanden waren. Somit waren auf dem Fragment 7-13 die Glycoproteine UL9 und UL 11, auf 27-40 das Glycoprotein UL45 und das Capsidprotein UL46, auf 50-57 die Glycoproteine UL73 und UL74, auf 79-83 UL116 (unbekannte Funktion), auf 85-86 das Glycoprotein UL120 und der



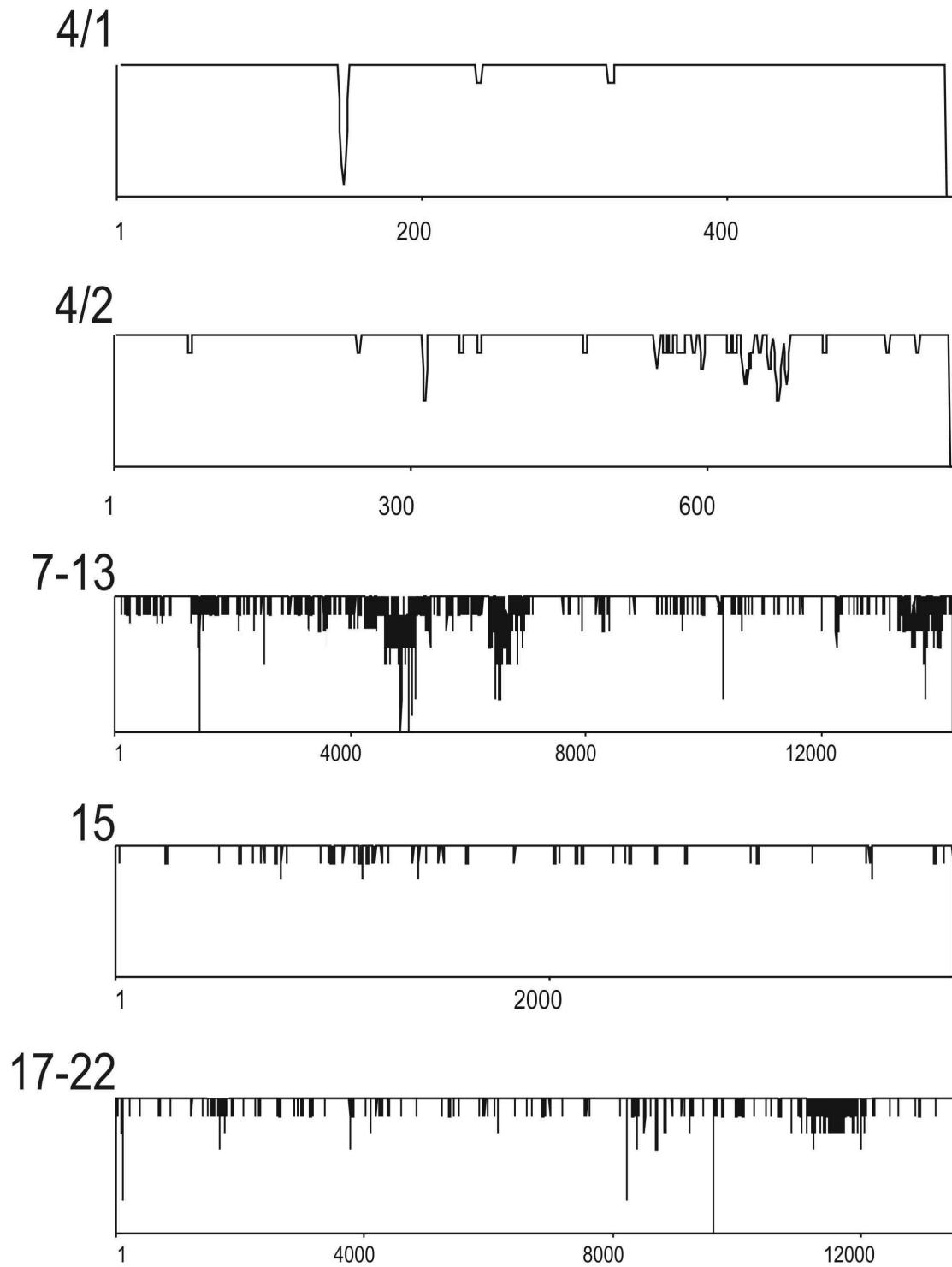
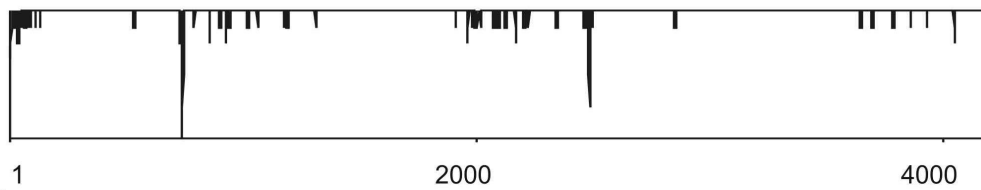
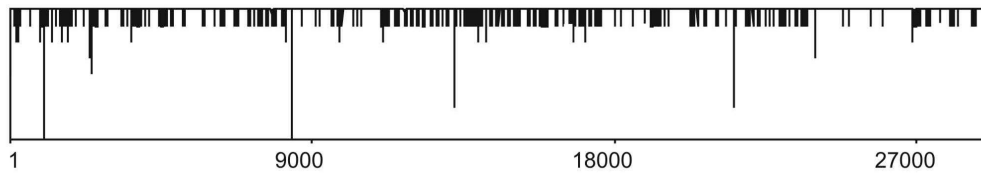


Abb. 12: Nukleotiddivergenzen zwischen HCMV *AD169* und HCMV *Toledo*.

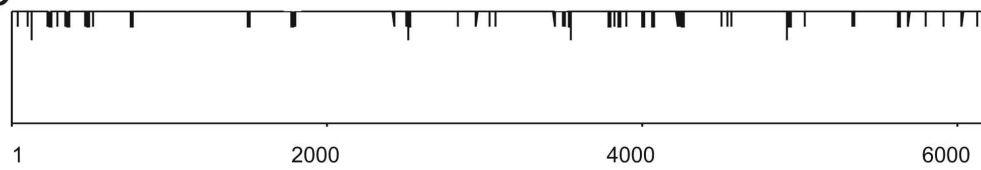
24-25



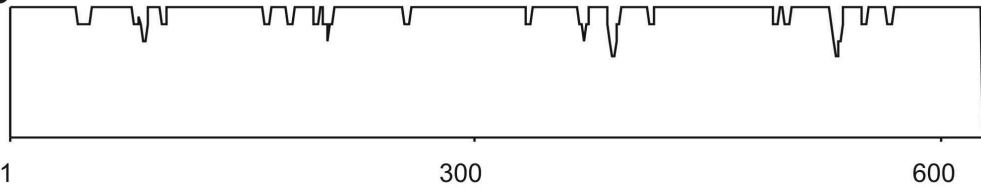
27-40



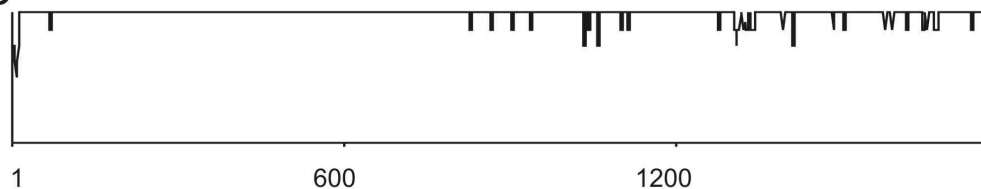
43-45



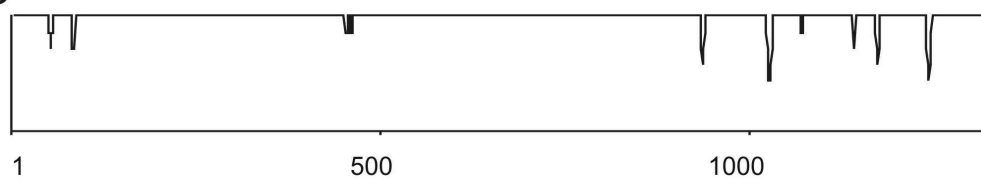
46



48



49



50-57

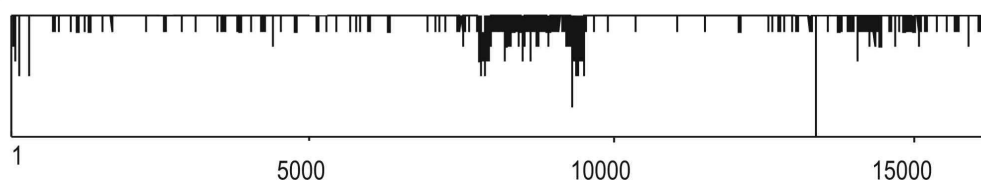
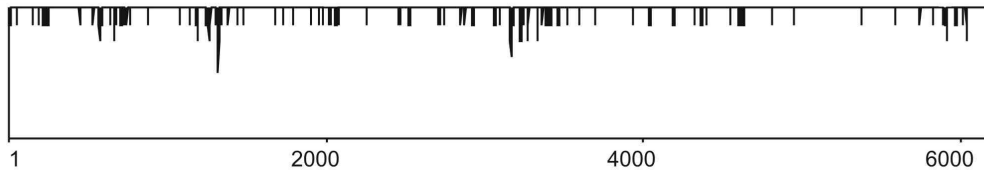


Abb. 13: Nukleotiddivergenzen zwischen HCMV *AD169* und HCMV *Toledo*.

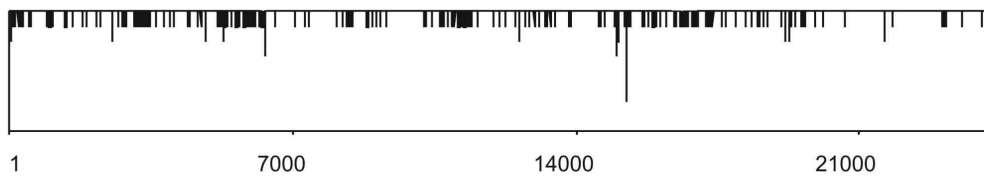
60-62



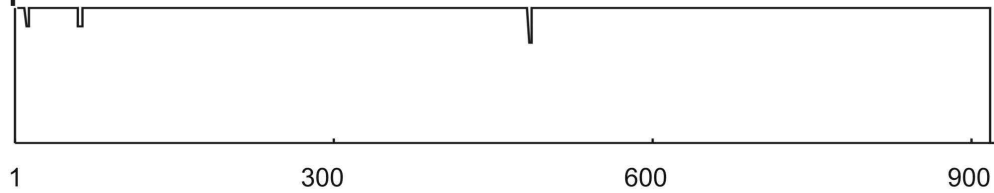
64



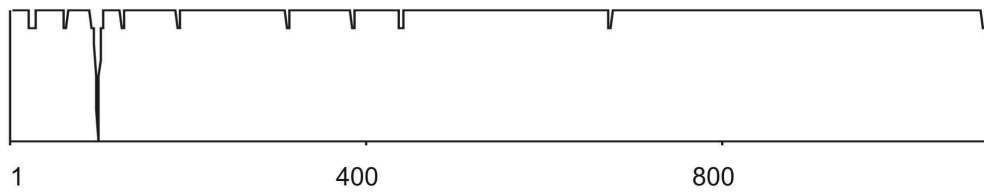
66-77



78/1



78/2



79-83

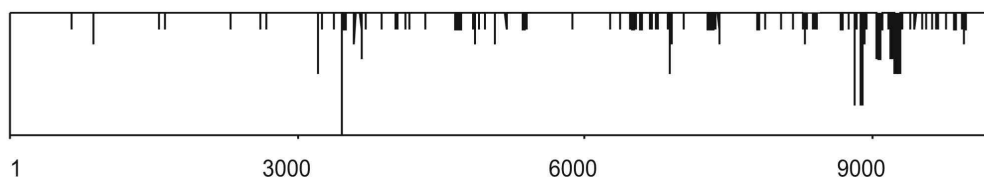
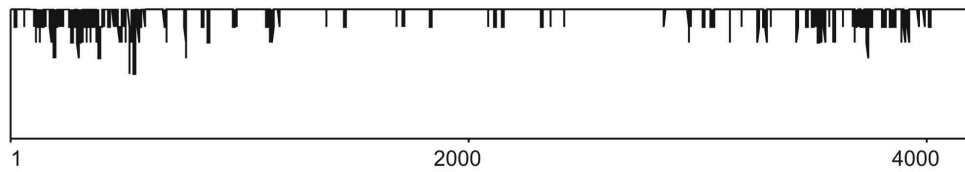
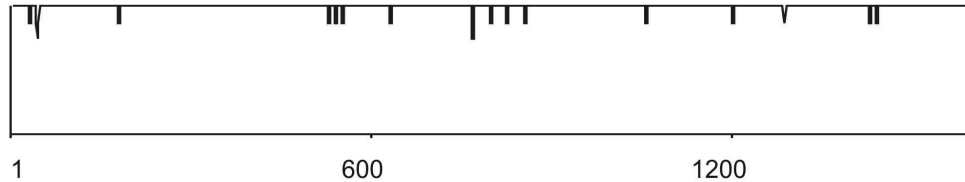


Abb. 14: Nukleotiddivergenzen zwischen HCMV *AD169* und HCMV *Toledo*.

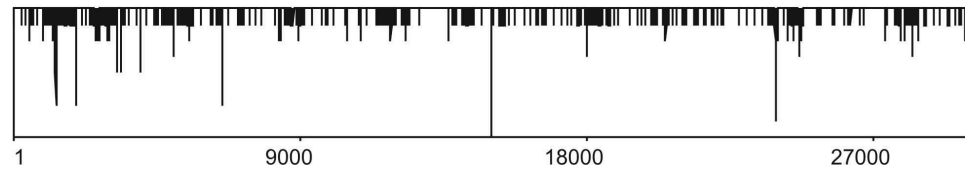
85-86



97



98-112



**Abb. 15: Nukleotiddivergenzen zwischen HCMV *AD169* und HCMV *Toledo*.** Die Abbildungen 8-11 zeigen die einzelnen Sequenzabweichungen der sequenzierten *Toledo*-DNA-Fragmente vom Referenzstamm *AD169*. Die untere Gerade zeigt die Länge der Fragmente in bp, darüber befindet sich eine Kurve, die die jeweiligen relativen Sequenzabweichungen angibt. Die relative Abweichung ergibt sich hierbei aus einem einzelnen Nukleotid und den zwei Nukleotiden 5' und 3'. Aufgrund eines Stauchungseffektes ist die Homologie-Kurve bei größeren Fragmenten nicht mehr maßstabsgerecht und nimmt die Form eines Balkendiagrammes an (siehe 98-112). Das Fragment 78-2 in **Abb. 14** zeigt die ungestauchte Darstellung. Der obere Rand der Kurve entspricht 100% Homologie der Sequenzen.

„immediate early transcriptional regulator“ UL123 und auf 98-112 die Glycoproteine US6 und US8 besonders stark durch Mutationen gekennzeichnet.

Insgesamt ließ sich für die in dieser Arbeit sequenzierten 172.178 Basen des klinischen Isolates *Toledo* eine Abweichung von 2,04% vom Laborstamm *AD169* feststellen. Die bereits erwähnten Mutationsschwerpunkte einzelner Fragmente ließen sich deutlich bei der statistischen Auswertung der Daten belegen (siehe Tabelle 9). Somit wies das Fragment 7-13 mit einer Abweichung von 6,46% eine mehr als dreimal höhere Mutationshäufigkeit auf als der durchschnittliche Mutationswert. Weiter Fragmente mit markanter Mutationshäufigkeit waren 4/2 (4,54%), 46 (4,60%) und 85-86 (4,28%), die immerhin noch das doppelte der durchschnittlichen Abweichung von 2,04% aufwiesen. Zwischen den beiden

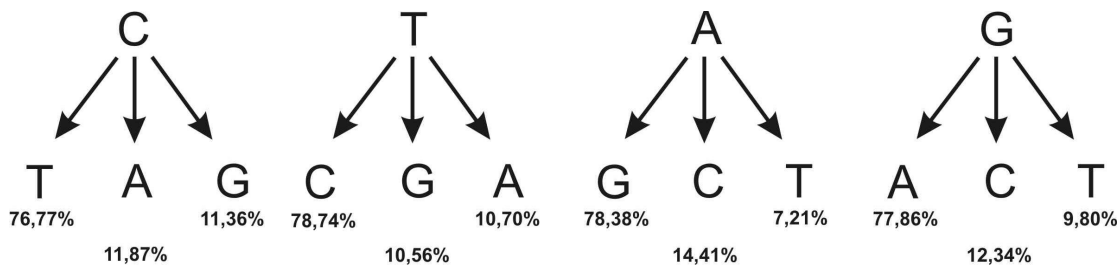
HCMV Stämmen besonders stark konserviert waren hingegen die Fragmente 78/1 (0,32%) und 66-77 (0,88%).

Fragment	C			T			A			G			Del	Ins	Größe (bp)	Unterschied
	T	A	G	C	G	A	G	C	T	A	C	T				
4/1	0	0	0	2	0	0	0	0	0	0	0	0	0	4	526	1,14%
4/2	4	1	0	7	0	0	6	0	1	8	0	0	9	2	837	4,54%
7-13	160	37	26	147	25	37	126	48	29	157	31	17	43	38	14256	6,46%
15	16	2	1	7	4	0	4	0	0	8	1	1	0	3	3875	1,21%
17-22	52	5	4	57	1	2	55	10	1	47	3	6	5	14	13522	1,94%
24-25	14	1	5	10	2	0	11	3	1	8	0	0	1	8	4201	1,52%
27-40	49	11	4	43	5	0	93	10	3	79	6	1	17	20	29168	1,17%
43-45	8	1	0	4	1	0	15	2	0	20	0	0	0	0	6219	0,82%
46	8	0	0	5	2	1	4	2	0	3	1	1	2	0	630	4,60%
48	4	1	1	2	0	10	5	2	0	9	0	1	4	2	1769	2,32%
49	2	0	0	1	0	0	0	0	0	2	0	0	12	7	1327	1,81%
50-57	74	6	16	65	8	12	61	13	14	59	7	23	42	21	16251	2,59%
60-62	10	3	4	26	2	0	15	1	0	23	8	0	5	4	6190	1,63%
64	9	0	0	9	1	1	18	0	0	6	1	0	0	1	2218	2,07%
66-77	50	3	6	60	3	2	40	1	0	34	6	2	6	1	24220	0,88%
78/1	1	0	0	0	0	0	0	0	0	0	0	0	1	1	924	0,32%
78/2	1	3	0	3	0	1	2	0	0	1	0	1	0	4	1101	1,45%
79-83	30	4	1	26	4	2	19	2	2	22	3	4	57	22	10248	1,93%
85-86	34	6	10	26	6	3	29	0	1	37	8	2	18	0	4206	4,28%
97	4	0	3	2	0	0	5	0	0	2	0	0	0	1	1601	1,06%
98-112	82	11	9	81	13	7	107	18	5	95	22	18	7	61	30252	1,77%
Summe	608	94	90	574	77	78	609	112	56	612	97	77	220	208	172178	2,04%

**Tabelle 9: Mutationsübersicht.** In der Tabelle sind für alle Fragmente die jeweiligen Nukleotidveränderung von *Toledo* relative zu *AD169* dargestellt. Hierbei steht in der obersten Zeile das Ursprungsnukleotid von *AD169* und in der darunter liegenden Zeile das bei *Toledo* gefundene abweichende Nukleotid. Zusätzlich sind Deletionen (Del) und Insertionen (Ins), die Größe des Fragmentes und der prozentuale Unterschied angegeben.

Bei der Auszählung der einzelnen Mutationen fiel auf, daß bestimmte Nukleotidaustausche bevorzugt auftraten (siehe Abb. 16). Dabei waren ausgehend von *AD169* 19,71% aller Mutationen Cytosin zu Thymin Veränderungen, 18,61% Austausch von Thymin nach Cytosin, 19,75% Veränderungen von Adenosin nach Guanin und 19,84% Guanin nach Adenosin Veränderungen. Damit machten alleine diese vier Mutationsarten 77,91% aller Nukleotidaustausche aus. Die restlichen Basenveränderungen schwankten zwischen 1,82% (A nach T) und 3,63% (A nach C). Der Vergleich der einzelnen Mutationsgruppen ergab, daß die Nukleotide fast gleichhäufig ausgetauscht worden waren. Das Cytosin bei *AD169* war bei *Toledo* 792 mal, das Thymin 729 mal, das Adenosin 777 mal und das Guanin 786 mal von Basenpaaraustauschen

betroffen. Innerhalb der Gruppe der C-Mutationen machten 76,77% aller Veränderungen Cytosin nach Thymin Mutationen aus. Die restlichen Mutationen zu Adenosin und Guanin traten gleichhäufig mit 11,87% und 11,36% auf. Bei den T-Mutationen zeigte sich ein ähnliches Bild. 78,38% machten Veränderungen



**Abb. 16: Mutationspräferenzen.** Für jedes mutierte Nukleotid in *AD169* ist die prozentuale Häufigkeit angegeben, mit der das abweichende Nukleotid in *Toledo* gefunden wurde.

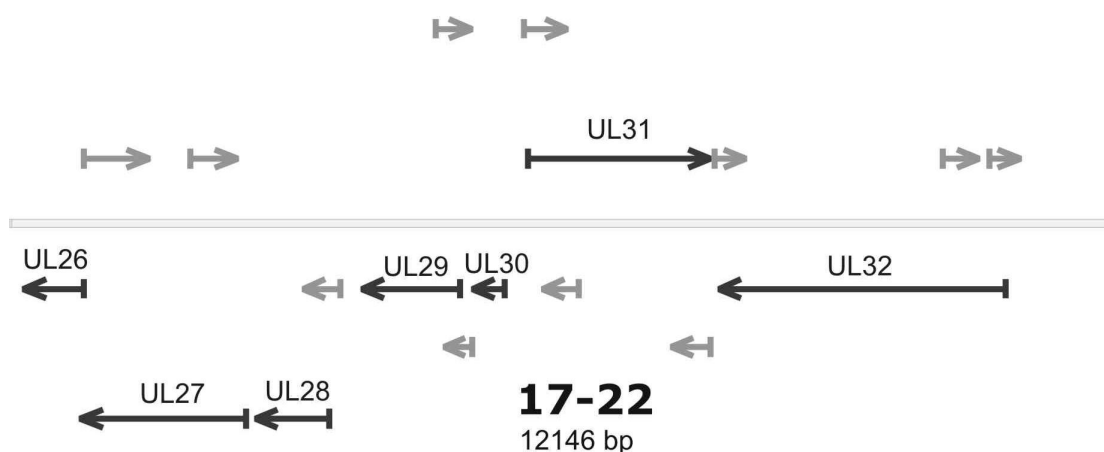
zu Cytosin aus. Austausch zu Guanin und Adenosin traten mit 10,56% und 10,70% seltener auf. Auch die Adenosin und Guanin Mutationen waren ähnlich häufig. Adenosin wurde in 78,38% der Fälle zu Guanin und Guanin in 77,86% der Fälle zu Adenosin verändert. Jedoch wichen diese beiden Mutationsgruppen von dem zuvor entstandenem Bild, daß die weniger häufigen Mutationen mit gleicher Wahrscheinlichkeit auftreten, ab. Die Veränderung Adenosin nach Cytosin trat mit einer Häufigkeit von 14,41% auf wohingegen Adenosin nach Thymin nur 7,21% der Mutationen ausmachte. Bei den Guanin-Mutationen trat die Veränderung zu Cytosin mit 12,34% auf und die Mutation zu Thymin mit 9,80%.

Die Anzahl der Deletionen und Insertionen bei Toledo hielt sich mit 220 Deletionen und 208 Insertionen annähernd die Waage. Die am stärksten betroffenen Fragmente waren 4/2, mit einer Fehlerrate bezüglich der Deletionen und Insertionen von 1,3% und 49 mit einer Fehlerrate von 1,4%. Das am stärksten konservierte Fragment war 43-45 mit keiner Mutation. Prinzipiell ließ sich feststellen, daß Fragmente mit einer hohen Gesamtfehlerrate (siehe Tabelle 9) nicht zwingend eine hohe Deletions- und Insertionsdichte aufwiesen. Das Fragment 4/2 wies zum Beispiel eine Gesamtfehlerrate von 4,54% auf und war mit einem Fehlerwert von 1,3% bei Deletionen und Insertionen das am zweit häufigsten betroffene Fragment. Im Gegensatz dazu wies das am stärksten betroffene Fragment 49 (1,4%) nur eine Gesamtfehlerrate von 1,81% auf, einen

Wert also, der unter der errechneten Abweichung von *AD169* zu *Toledo* von 2,04% (siehe Tabelle 9) lag.

#### 4.4 Offene Leseraster

Die Suche nach offenen Leserastern in den sequenzierten Bereichen der DNA des HCMV *Toledo* Stammes wurde mit Hilfe der Vector NTI Suite 7 Software durchgeführt. Ein Leserahmen wurde dabei entsprechend der Vorgaben der ursprünglichen Sequenzierung und Analyse des HCMV Stammes *AD169* (Chee et. al., 1990) definiert. Folglich mußte ein potentiell Gen mit einem ATG Kodon beginnen und eine Mindestlänge von 300 Nukleotiden aufweisen. Da alle Analysen auf DNA Ebene durchgeführt wurden, wurde Thymin nicht in Uracil abgeändert. Zusätzlich sollten die gefundenen Leserahmen zu nicht mehr als 20% überlappen.



**Abb. 17: Mit ATG beginnende ORFs des Fragmentes 17-22 der DNA des HCMV *Toledo* Stammes.** Graphische Darstellung aller potentieller "open reading frames" mit einer Mindestlänge von 300 Nukleotiden und ATG als Startkodon. Die für die DNA des HCMV Stammes *AD169* beschriebenen ORFs sind schwarz dargestellt und mit den jeweiligen Kennungen versehen. Graue ORFs sind in der Literatur nicht beschrieben, da sie zu mehr als 20% mit den vorherigen ORFs überlappen.

Innerhalb der sequenzierten Bereiche der *Toledo* DNA wurden für die 172.178 Nukleotide 231 potentielle ORFs ermittelt (siehe Abb. 17).

Da in dieser Arbeit aber nicht nur das Standard Startkodon ATG als Translationsinitiator berücksichtigt wurde, sondern auch die alternativen Startkodons CTG und GTG, wurden für den gesamten sequenzierten Bereich 370 potentielle ORFs errechnet. Die Verwendung zusätzlicher Startkodons führte dabei sowohl zur Entstehung neuer ORFs (Vergleich zwischen Abb. 17 und Abb.

18), als auch zu einer 5' Verlängerung bekannter, also mit ATG als Startkodon beschriebener Gene.



**Abb. 18: Mit ATG, GTG und CTG beginnende ORFs des Fragmentes 17-22 der DNA des HCMV Stammes Toledo.** Die für HCMV AD169 beschriebenen, mit ATG anfangenden Gene sind wieder mit der Kennung und in schwarz dargestellt. Die grauen Pfeile sind in der Literatur nicht beschriebene, potentielle ORFs. Der „open reading frame“ UL27 weist eine in dieser Darstellung schwer zu erkennende 5' Extension des ursprünglich mit ATG anfangenden Genes auf. Die Extension beträgt 63 Nukleotide und als Startkodon fungiert CTG.

Alle 370 berechneten ORFs wurden über das Programm „blastp“ (NCBI) mit den genetischen Datenbanken von EMBL und NCBI verglichen. Hierzu wurden die einzelnen ORFs mittels Vector NTI Suite 7 translatiert und dann auf Proteinebene mit den Datenbanken des NCBI und EMBLE verglichen. Auf diese Weise konnten unter den 370 gefundenen ORFs die 127 *AD169* homologen Gene identifiziert werden. Zusätzlich konnten die ORFs, die für *AD169* nicht beschrieben waren auf Ähnlichkeit zu anderen Proteinen untersucht werden. Potentielle „open reading frames“, die kurze homologe Bereiche mit anderen veröffentlichten Genen aufwiesen, wurden als Kandidaten für unbekannte Gene, oder vormals funktionale Gene betrachtet, und auf die Existenz funktionaler Domänen untersucht.

#### 4.4.1 Alternative Startkodons

Die Verwendung der alternativen Startkodons GTG und CTG neben dem Standard Startkodon ATG führte nicht nur zu einer Zunahme potentieller ORFs in der DNA des HCMV Stammes *Toledo*, sondern hatte auch Auswirkungen auf die Länge vorhandener ORFs. Von den insgesamt 121 in der DNA des HCMV Stammes *Toledo* gefundenen ORFs, die auch bei *AD169* beschrieben worden sind (Chee et al., 1990; Davison et al., 2003; Murphy et al., 2003), wiesen 48 ORFs eine Verlängerung beim Vergleich zu den veröffentlichten *AD169* ORFs und 10 eine



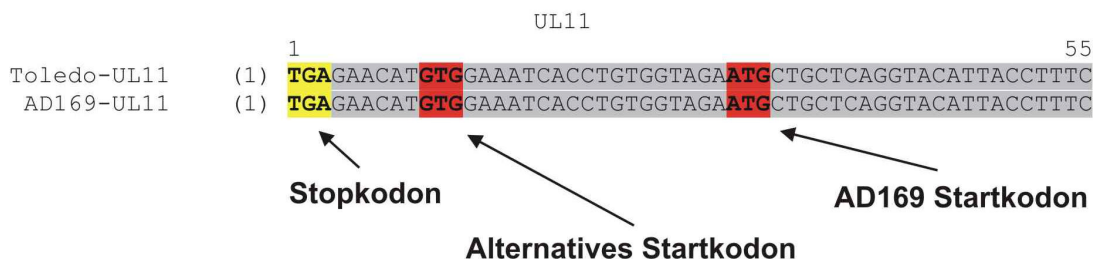
Verkürzung auf (siehe Tabelle 12 und Tabelle 13). Von den 48 verlängerten ORFs (rote gekennzeichnete ORFs in Tabelle 12 und Tabelle 13) begannen 23 mit Valin, 22 mit Leucin und 3 mit Methionin. Die Verlängerungen der ORFs konnten in einigen Fällen sehr ausgeprägt sein. UL92 z.B. (siehe Tabelle 12) besitzt ein Valinstartkodon, daß bei Verwendung eine Verlängerung des Proteins um 120 Aminosäuren bewirken würde. Jedoch finden sich unter den potentiell verlängerten ORFs auch sehr kurze Erweiterungen des Leserahmens. UL121 weist zum Beispiel eine Verlängerung von nur einer Aminosäure (Valin) auf (siehe Tabelle 13). Besonders interessant sind die drei Verlängerungen, die in der DNA des HCMV Stammes *Toledo* eine auf AUG basierende Extension aufweisen. Die drei betroffenen Gene UL14 (+68 AS), UL26 (+34 AS) und UL84 (+39 AS) sind damit, auch wenn man alternative Startkodons außer Acht läßt, mit hoher Wahrscheinlichkeit verlängert, da die Kozak Regeln (Kozak, 1986) besagen, daß das erste AUG Kodon in der Regel zur Translationsinitiation verwendet wird.

Die 10 ORFs, bei denen das erste AUG Kodon in der DNA des HCMV Stammes *Toledo* nicht vorhanden war (blaue ORFs in Tabelle 12 und Tabelle 13), könnten in drei Fällen durch ein Leucin (UL107, US25 und US25A), in drei weiteren Fällen durch Valin (UL108, UL113 und ORF7) und in den restlichen vier Fällen (UL4, UL9, UL15 und UL101) durch ein 3' gelegenes Methionin Kodon initiiert werden. Die meisten der betroffenen ORFs (UL15, UL101, UL107, UL108, UL113, US25) waren dabei in der Gruppe der Gene, die von Chee (1990) für die DNA des HCMV Stammes *AD169* postuliert worden waren, allerdings später in der Neuanalyse des Kodierungspotentiales von *AD169* durch Davison (2003) nicht mehr berücksichtigt wurden.

#### **4.4.2 Untersuchung der DNA Sequenz des HCMV Stammes *AD169* auf alternative Startkodons**

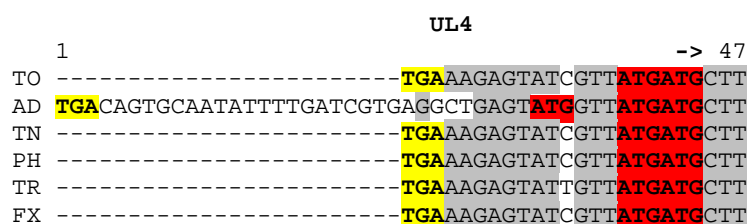
Da die bei der DNA des HCMV Stammes *Toledo* gefundenen alternativen Startkodons bei der DNA des HCMV Stammes *AD169* nicht beschrieben waren, wurden die homologen, bereits veröffentlichten ORFs von HCMV *AD169* auch auf das Vorhandensein alternativer Startkodons untersucht. Um sich ein besseres Bild über den Aufbau der 5' gelegenen Bereiche der AUG Startkodons machen zu können, wurden nicht nur die Gene selbst, sondern auch die 5' des AUG gelegenen Sequenzen verglichen. Der untersuchte

Bereich erstreckte sich dabei vom AUG Kodon bis zum nächsten 5' gelegenen Stopkodon, das im Leseraster des AUG Startkodons war. Das 5' gelegene Stopkodon war dabei nicht zwingend identisch mit dem Stopkodon des 5' gelegenen ORFs (siehe Abb. 19).



**Abb. 19: Die 5' vom Startkodon gelegene Region von UL11.** Vergleich der DNA Sequenz 5' des ATG Startkodons in der DNA der HCMV Stämme *AD169* und *Toledo*. Das ATG Standard Startkodon sowie das alternative Startkodon GTG sind rot hinterlegt. Das relativ zum ATG im Leseraster 5' liegende Stopkodon, das die maximale Ausdehnung des nicht transkribierten Teiles des ORFs darstellt, ist gelb hinterlegt. Da die Sequenzen homolog sind, ist das alternative Startkodon GTG in der DNA des HCMV Stammes *AD169* konserviert.

Von den insgesamt 58 von Extensionen betroffenen ORFs (siehe 3.6.1), die aufgrund alternativer Startkodons oder zerstörter ATG Startkodons in der DNA des HCMV Stammes *Toledo* Veränderungen des Aminoterminus der erzeugten Proteine aufwiesen, verfügten die meisten über eine hohe Homologie zu der DNA des HCMV Stammes *AD169* im Bereich der Regionen 5' des ATG Kodons. Einige wenige ORFs wiesen jedoch wegen Basenpaar Austausch oder Deletionen und Insertionen in diesen Bereichen starke Veränderungen auf. In Abb. 20 sind die 5' des ATG Kodons gelegenen Sequenzen der ORFs UL4, UL 36 und UL92 dargestellt, um die Auswirkungen von Mutationen auf alternative Startkodons zu verdeutlichen.



## UL36

```

1                               55
TO TAGTTCTCCGAAGGCGGACGGAGGAGTCTGTCGACCGCAGCGCTGCTGGAGAAG
AD -----
TN -----
TR -----
PH -----
FX -----

56                               110
TO CGCAGCGTCGGCGAGCGAAGGTAGAGGAGTCCGTCATGACGACCTACGGGACAC
AD -----
TN -----
TR -----
PH -----
FX -----

111                              165
TO GCTGATGGCCTACGGCTGCATCGCCATCCGAGCCGGGGACTTTAACGGTCTCAAC
AD --TGAATGGCCTACGGCTGCATCGCCATCCGAGCCGGGGACTTTAACGGTCTCAAC
TN -----
TR -----
PH --TGAATGGCCTACGGCTGCATCGCCATCCGAGCCGGGGACTTTAACGGTCTCAAC
FX --TGAATGGCCTACGGCTGCATCGCCATCCGAGCCGGGGACTTTAACGGTCTCAAC

166                              220
TO GACTTTCTGGAGCAAGAATGCGGCACCCGGCTGCACGTGGCCTGGCCCGAACGCT
AD GACTTTCTGGAGCAGGAATGCGGCACCCGGCTGCACGTGGCCTGGCCCGAACGCT
TN -----
TR -----
PH GACTTTCTGGAGCAGGAATGCGGCACCCGGCTGCACGTGGCCTGGCCCGAACGCT
FX GACTTTCTGGAGCAGGAATGCGGCACCCGGCTGCACGTGGCCTGGCCCGAACGCT

221                              275
TO GCTTCATCCAGTCCGTTTCGCGCAGCGCCCTGGGGCCCTTCGTGGGCAAGATGGG
AD GCTTCATCCAGTCCGTTTCGCGCAGCGCCCTGGGGCCCTTCGTGGGCAAGATGGG
TN -----
TR -----
PH GCTTCATCCAGTCCGTTTCGCGCAGCGCCCTGGGGCCCTTCGTGGGCAAGATGGG
FX GCTTCATCCAGTCCGTTTCGCGCAGCGCCCTGGGGCCCTTCGTGGGCAAGATGGG

276                              330
TO CACCGTCTGTTTCGCAAGGTAAGCCCCACGTCGTTGAAGACACCTGGAAAGAGGAC
AD CACCGTCTGTTTCGCAAGGTAAGCCCCACGTCGTTGAAGACACCTGGAAAGAGGAC
TN -----TGAAGACACCTGGAAAGAGGAC
TR -----TGAAGACACCTGGAAAGAGGAC
PH CACCGTCTGTTTCGCAAGGTAAGCCCCACGTCGTTGAAGACACCTGGAAAGAGGAC
FX CACCGTCTGTTTCGCAAGGTAAGCCCCACGTCGTTGAAGACACCTGGAAAGAGGAC

331                              385
TO GTTCGCTCGGGCACGTTCTTTCCAGGTGTTTTCAACGTGCGTGGATTTTTTCTA
AD GTTCGCTCGGGCACGTTCTTTCCAGGTGTTTTCAACGTGCGTGGATTTTTT-CTA
TN GTTCGCTCGGGCACGTTCTTTCCAGGTGTTTTCAACGTGCGTGGATTTTTT----
TR GTTCGCTCGGGCACGTTCTTTCCAGGTGTTTTCAACGTGCGTGGATTTTTT----
PH GTTCGCTCGGGCACGTTCTTTCCAGGTGTTTTCAACGTGCGTGGATTTTTT----
FX GTTCGCTCGGGCACGTTCTTTCCAGGTGTTTTCAACGTGCGTGGATTTTTT----

386                              440
TO TTATTTCTCTACCAGGTGCTTACGTCGCTGTCAGGAGTACCTGCACCCCTTCCG
AD TT----CTCTACCAGGTGCTTACGTCGCTGTCAGGAGTACCTGCACCCCTTTGG
TN TTATTTCTCTACCAGGTGCTTACGTCGCTGTCAGGAAATACCTGCACCCCTTTGG
TR TTATTTCTCTACCAGGTGCTTACGTCGCTGTCAGGAAATACCTGCACCCCTTTGG
PH ---C-TCTCTACCAGGTGCTTACGTCGCTGTCAGGAGTACCTGCACCCCTTTGG
FX ---C-TCTCTACCAGGTGCTTACGTCGCTGTCAGGAGTACCTGCACCCCTTTGG

441                               ->468
TO CTTCGTCGAGGGTCCGGGCTTTATGCGC
AD CTTCGTCGAGGGTCCGGGCTTTATGCGC
TN CTTCGTCGAGGGTCCGGGCTTTATGCGC
TR CTTCGTCGAGGGTCCGGGCTTTATGCGC
PH CTTCGTCGAGGGTCCGGGCTTTATGCGC
FX CTTCGTCGAGGGTCCGGGCTTTATGCGC

```

**UL92**

```

1                                                                                               55
TO TGA1AAACCGGAACCAATCCGCCGACACCGGCGCCGCG55GTGACGCGCGCCATAAAA
AD TGA1AAACCGGAACCAATCCGCCGACACCGGCGCCGCG55GTGACGCGCGCCATAAAA
TN TGA1AAACCGGAACCAATCCGCCGACACCGGCGCCGCG55GTGACGCGCGCCATAAAA
PH TGA1AAACCGGAACCAATCCGCCGACACCGGCGCCGCG55GTGACGCGCGCCATAAAA
TR TGA1AAACCGGAACCAATCCGCCGACACCGGCGCCGCG55GTGACGCGCGCCATAAAA
FX TGA1AAACCGGAACCAATCCGCCGACACCGGCGCCGCG55GTGACGCGCGCCATAAAA

56                                                                                               110
TO ACGAAAGTGTCTGTCGTCTGCGACCCGCCACAGCCGCCATGAAC110TCTGTTGCTGGCGG
AD ACGAAAGTGTCTGTCGTCTGCGACCCGCCACAGCCGCCATGAAC110TCTGTTGCTGGCGG
TN ACGAAAGTGTCTGTCGTCTGCGACCCGCCACAGCCGCCATGAAC110TCTGTTGCTGGCGG
PH ACGAAAGTGTCTGTCGTCTGCGACCCGCCACAGCCGCCATGAAC110TCTGTTGCTGACGG
TR ACGAAAGTGTCTGTCGTCTGCGACCCGCCACAGCCGCCATGAAC110TCTGTTGCTGGCGG
FX ACGAAAGTGTCTGTCGTCTGCGACCCGCCACAGCCGCCATGAAC110TCTGTTGCTGGCGG

111                                                                                               165
TO AACTCAACCGACTGGGGGTCTGCGCACGCCACTACGGAGGATGTTTTATC165TTTGT
AD AACTCAACCGACTAGGGGTCTGCGCACGCCACTACGGAGGATGTTTTATC165TTTGT
TN AACTCAACCGACTGGGGGTCTGCGCACGCCACTACGGAGGATGTTTTATC165TTTGT
PH AACTCAACCGACTGGGGGTCTGCGCACGCCACTACGGAGGATGTTTTATC165TTTGT
TR AACTCAACCGACTGGGGGTCTGCGCACGCCACTACGGAGGATGTTTTATC165TTTGT
FX AACTCAACCGACTGGGGGTCTGCGCACGCCACTACGGAGGATGTTTTATC165TTTGT

166                                                                                               220
TO CGACCGCCTCTTTCAACACTTTTCTTCTTTTCCAGGCCGAGGAGT220CAGGCCCG
AD CGACCGCCTCTTTCAACACTTTTCTTCTTTTCCAGGCCGAGGAGT220CAGGCCCG
TN CGACCGCCTCTTTCAACACTTTTCTTCTTTTCCAGGCCGAGGAGT220CAGGCCCG
PH CGACCGCCTCTTTCAACACTTTTCTTCTTTTCCAGGCCGAGGAGT220CAGGCCCG
TR CGACCGCCTCTTTCAACACTTTTCTTCTTTTCCAGGCCGAGGAGT220CAGGCCCG
FX CGACCGCCTCTTTCAACACTTTTCTTCTTTTCCAGGCCGAGGAGT220CAGGCCCG

221                                                                                               275
TO CGCCGCTTGGAAGTGGTCTGCGTCCGTGTTTCGAGCACCTGACGGT275GAGTGCCTCA
AD CGCCGCTTGGAAGTGGTCTGCGTCCGTGTTTCGAGCACCTGACGGT275GAGTGCCTCA
TN CGCCGCTTGGAAGTGGTCTGCGTCCGTGTTTCGAGCACCTGACGGT275GAGTGCCTCA
PH CGCCGCTTGGAAGTGGTCTGCGTCCGTGTTTCGAGCACCTGACGGT275GAGTGCCTCA
TR CGCCGCTTGGAAGTGGTCTGCGTCCGTGTTTCGAGCACCTGACGGT275GAGTGCCTCA
FX CGCCGCTTGGAAGTGGTCTGCGTCCGTGTTTCGAGCACCTGACGGT275GAGTGCCTCA

276                                                                                               330
TO ACGACATCCTGGACGCCTG330CAGTCA330CCCGACGTGAACGTCTGCGGAGACAAGCAA
AD ACGACATCCTGGACGCCTG330CAGTCA330CCCGACGTGAACGTCTGCGGAGACAAGCAA
TN ACGACATCCTGGACGCCTG330CAGCACC330CGACGTGAACGTCTGCGGAGACAAGCAA
PH ACGACATCCTGGACGCCTG330CAGCCACC330CGACGTGAACGTCTGCGGAGACAAGCAA
TR ACGACATCCTGGACGCCTG330CAGCCAT330CCCGACGTGAACGTCTGCGGAGACAAGCAA
FX ACGACATCCTGGACGCCTG330CAGCCACC330CGACGTGAACGTCTGCGGAGACAAGCAA

331                                                                                               385
TO CAC385CTGTCGTCC385CTGCCCTTCTCCTATTCCCTCCGCC385CCCCAAA385ACTGTCAGCGGC
AD CAC385CTGTCGTCC385CTGCCCTTCTCCTGTTCCCTCCGCC385CCCCAAA385ACTGTCAGCGGC
TN CAC385CTGTCGTCC385CTGCCCTTCTCCTGTTCCCTCCGCC385CCCCAAA385ACTGTCAGCGAC
PH CAC385CTGTCGTCC385CTGCCCTTCTCCTGTTCCCTCCGCC385CCCCAAA385ACTGTCAGCGAC
TR CAC385CTGTCGTCC385CTGCCCTTCTCCTGTTCCCTCCGCC385CCCCAAA385ACTGTCAGCGAC
FX CAC385CTGTCGTCC385CTGCCCTTCTCCTGTTCCCTCCGCC385CCCCAAA385ACTGTCAGCGAC

386                                                                                               402
TO GCTCAGACGTCAT402GTGC
AD GCTCAGACGTCAT402GTGC
TN GCTCAGACGTCAT402GTGC
PH GCTCAGACGTCAT402GTGC
TR GCTCAGACGTCAT402GTGC
FX GCTCAGACGTCAT402GTGC

```

**Abb. 20:** Die 5' vom Startkodon gelegenen Regionen von UL4, UL36 und UL92. Die Darstellung zeigt die 5' Regionen der ORFs UL4, UL36 und UL92 in der DNA der HCMV Stämme Toledo (TO), AD169 (AD), Town (TN), PH (PH), TR (TR) und FIX (FX). Die ausgehend vom ATG (->) im Leseraster liegenden Startkodons sind rot hinterlegt. Die im Leseraster liegenden 5' gelegenen Stopkodons sind in der Darstellung gelb hinterlegt. Ein grauer Hintergrund deutet homologe Basen an, wogegen ein weißer Hintergrund eine Mutation hervorhebt.

Der ORF UL4 ist als Beispiel für eine Verkürzung eines für *AD169* veröffentlichten ORFs im Vergleich zur Sequenz der DNA des HCMV Stammes *Toledo* anzusehen. Das bei *AD169* als Startkodon veröffentlichte ATG ist bei allen anderen klinischen Isolaten und Laborstämmen von HCMV mutiert (siehe Abb. 20). Die beiden 3' gelegenen ATGs sind jedoch konserviert. Auch die 5' Region ist bei allen untersuchten Stämmen annähernd identisch, nur *AD169* weist eine Mutation im 5' gelegenen Stopkodon auf, so daß der 5' Bereich leicht verlängert ist. Der ORF UL36 zeigt die Auswirkungen von Deletionen und Insertionen auf die Länge und das Kodierungspotential der 5' des ATG gelegenen DNA Sequenzen. Der Bereich zwischen dem ATG Standard Startkodon und dem ersten CTG Kodon ist mit Ausnahme eines Basenpaaraustausches in der DNA des HCMV Stammes *Toledo* konserviert (Position 438, Abb. 20). In der Umgebung der Nukleotidposition 385 (siehe Abb. 20) häufen sich jedoch Deletionen und Insertionen in den Sequenzen der unterschiedlichen Stämme von HCMV. Aufgrund dieser Mutationen kommt es zu unterschiedlichen Verkürzungen der 5' Region des ORFs. Bei der Nukleotidposition 310 tritt bei *Town* und *TR* ein Stopkodon auf, bei *FIX*, *TR* und *AD169* erst in Position 115. Bis zu dieser Position erscheinen in den DNAs aller untersuchter Stämme, mit Ausnahme von *Toledo*, keine weiteren Startkodons auf. Die längste 5' Region weist mit einer Länge von circa 460 Nukleotiden die DNA des HCMV Stammes *Toledo* auf. In dieser Region liegen zwei CTGs, ein GTG und drei ATG Startkodons. Von diesen Startkodons ist nur das erste CTG bei allen Stämmen konserviert, die restlichen in der DNA des HCMV Stammes *Toledo* gefundenen Startkodons sind aufgrund von Verschiebungen des Leserahmens jedoch zerstört. Die Sequenz des Gens UL 92 hingegen dient als Beispiel für hohe Homologie in der 5' Region zwischen den Sequenzen aller sequenzierter HCMV Stämme. Die Länge dieses Bereiches ist bei allen Stämmen, mit Ausnahme von *AD169*, identisch. In der DNA des HCMV Stammes *AD169* ist aufgrund eines Basenpaaraustausches von Cytosin nach Thymin ein Stopkodon entstanden. Das bei allen anderen Stämmen vorhandene Stopkodon existiert jedoch auch in der DNA des HCMV Stammes *AD169* (siehe Abb. 20). Alle HCMV Stämme verfügen bei diesem ORF über 3 alternative CTG Startkodons. Sämtliche Isolate mit Ausnahme von HCMV *AD169* besitzen zusätzlich noch zwei GTG Startkodons.

---

## 4.5 Evaluierung der alternativen Initiationskodons

Die Evaluierung der alternativen Startkodons erfolgte über drei separate Ansätze. Mittels eines Promoter suchenden Algorithmuses wurde ermittelt, ob die alternativen Startkodons innerhalb der Reichweite der jeweiligen Promotoren waren. Zudem wurden die Startkodons auf Kozak Motive untersucht, um die Wahrscheinlichkeit einer Translationsinitiation an diesen Triplets zu bestimmen. Zuletzt wurde für ausgewählte ORFs mittels RT-PCR überprüft, ob die alternativen Startkodons in der jeweiligen mRNA vorhanden waren.

### 4.5.1 Promoteranalysen

Zur Identifikation potentieller Promotoren wurde mittels der internet-basierenden Software „Promoter 2.0 Prediction“ und „Neural Network Promoter Prediction 2.2“ die in dieser Arbeit sequenzierten HCMV *Toledo* DNA Fragmente auf Promotoren untersucht. Da die beiden Programme eine Vielzahl wahrscheinlicher Promotoren für jedes untersuchte DNA-Fragment vorhersagten, wurden nur die Promotoren berücksichtigt, die von beiden Programmen vorgeschlagen worden waren. Untersucht wurden nur die 48 ORFs für die aufgrund der Benutzung alternativer Initiationskodons eine Verlängerung der Proteinsequenz vorhergesagt worden war (siehe 3.6.1 und rote ORFs in Tabelle 12 und Tabelle 13). Jedem dieser 48 Gene konnte ein Promoter zugeordnet werden, wobei für 35 Gene der Promoter für ATG und CTG,GTG identisch war, und bei 13 ORFs die berechneten Promotoren innerhalb der potentiellen Extensionen der ORFs lagen ( UL27, UL28, UL69, UL73, UL78, ORF8, UL100, UL122, US3, US9, US12, US26 und US31).

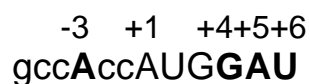
### 4.5.2 Benutzung der alternativen Startkodons unter Berücksichtigung der Kozak Regeln

Da alternative Startkodons, wenn sie im optimalen Sequenzkontext stehen, mit der gleichen Effizienz die Translation initiieren können wie AUG Kodons, wurden die gefundenen, nicht mit AUG anfangenden Extensionen auf ihre Homologie zur Kozak Sequenz untersucht. Extensionen die mit AUG anfangen wurden nicht

ORF	AUG	Ext.	ORF	AUG	Ext.	ORF	AUG	Ext.
UL6	0	-	UL70	0	-	UL117	-	-
UL8	0	0	UL71	-	0	UL121	-	-
UL10	-	-	UL73	0	-	UL122	0	0
UL11	0	0	UL77	0	0	UL123	0	0
UL13	0	0	UL78	-	0	ORF9	0	0
UL14	-	-	ORF6	-	0	ORF10	x	-
UL16	-	0	UL92	-	-	US3	-	0
UL19	x	0	UL94	0	-	US9	-	0
UL27	-	-	UL89	0	-	US12	0	x
UL28	0	0	UL94	x	-	US15	0	0
UL31	-	0	UL96	-	0	US21	-	-
UL36	0	-	UL97	0	0	US23	-	0
UL43	0	0	UL98	0	-	US25	0	-
UL44	0	-	ORF8	0	-	US25A	x	-
UL46	x	-	UL100	0	0	US26	-	0
UL48	-	-	UL103	0	0	US28	0	0
UL49	0	0	UL104	0	-	US30	0	-
UL69	0	-	UL108	x	0	US31	-	0

**Tabelle 10: Startkodons im Kozak Kontext.** Die Tabelle listet das veröffentlichte AUG Startkodon des HCMV Stammes *AD169* auf, sowie die alternativen Startkodons (Ext. = Extensionen). (-) Bedeutet keinerlei Ähnlichkeit zum Kozak Motiv, (0) zeigt eine leichte Ähnlichkeit mit dem Kozak Motiv an und (x) deutet eine starke Homologie zur Kozak Sequenz an.

untersucht, da normalerweise das erste AUG als Translationsstart dient. Die optimale Initiationssequenz lautet (Kozak, 1986; Boeck und Kolakofsky 1994):



Hierbei sind nach Kozak das A in Position -3 und das G in Position +4, nach Boeck und Kolakofsky das A in Position +5 und das U in Position +6 besonders wichtig für eine optimale Initiations-effizienz. Da eines dieser 4 Nukleotide bereits die Initiationswahrscheinlichkeit des betroffenen Startkodons markant erhöhen kann (Boeck und Kolakofsky 1994), wurde die Existenz einer dieser Nukleotide an der benötigten Position als leichte Ähnlichkeit zum optimalen Startkodon bezeichnet. Waren zwei dieser Nukleotide zugegen, so wurde von einer starken Homologie gesprochen. In Tabelle 10 sind die von Extensionen betroffenen ORFs aufgelistet. Für jeden ORF ist angegeben wie ähnlich die Umgebungssequenzen des alternativen und des AUG Startkodons der optimalen Initiationssequenz waren. Die Analyse aller Startkodons ergab, dass keines der untersuchten alternativen und Standard Startkodons der optimalen Initiationssequenz entsprach. Die Startkodons mit der höchsten Ähnlichkeit (UL19, UL108, ORF10,

US12 und US25A) wiesen 2 der insgesamt 4 wichtigen Nukleotide auf. Von diesen fünf Startkodons waren alle mit Ausnahme von US12 AUG Kodons. Von den insgesamt 54 ORFs wiesen die alternativen Startkodons von 25 ORFs keinerlei Ähnlichkeit mit der optimalen Initiationssequenz auf. Jedoch zeigten auch 19 mit AUG beginnenden ORFs keinerlei Ähnlichkeit mit der Kozak Sequenz.

#### 4.5.3 RT-PCR Kontrolle der alternativen Startkodons in messenger RNA aus HCMV *Toledo* infizierten menschlichen Vorhaut-fibroblasten

Ein wichtiges Indiz für die Benutzung alternativer Startkodons konnte uns die von HCMV *Toledo* produzierte mRNA liefern. Die Startkodons konnten nur dann funktional sein, wenn sie auf den mRNAs vorhanden waren. Mittels RT-PCR mit

ORF	Alternatives Startkodon	Extension (AS)	Startkodon auf mRNA
UL14	AUG	68	nein
UL31	CUG	87	nein
UL36	CUG	107	ja
UL89	CUG	115	ja
UL92	GUG	120	nein
UL94	GUG	113	ja
UL96	CUG	47	ja
US15	CUG	123	ja

**Tabelle 11: Evaluierung der alternativen Startkodons.** Acht aufgrund ihrer Größe und des Startkodons ausgesuchte ORFs wurden auf die Existenz des alternativen Startkodons auf der mRNA von HCMV *Toledo* untersucht. Dargestellt sind die jeweiligen Startkodons, die Größe der erzeugten Aminosäureverlängerung (AS) im Protein durch die alternativen Startkodons, sowie die Ergebnisse der RT-PCR (Startkodon auf mRNA).

gereinigter mRNA unter Verwendung spezieller Primer, die die alternativen Startkodons einschlossen, wurde die Existenz der alternativen Startkodons untersucht. Die acht untersuchten mRNAs waren so ausgesucht, das alle alternativen Startkodons vertreten waren, und die erzeugten Verlängerungen durch diese Startkodons eine markante Größe aufwiesen. Auf drei dieser acht mRNAs (UL14, UL31 und UL92) konnten die alternativen Startkodons nicht nachgewiesen werden. Jedoch konnte für die ORFs UL36, UL89, UL94, UL96 und US15 nachgewiesen werden, daß die alternativen Startkodons auf der mRNA vorhanden waren und somit als alternativer Translationsstartpunkte zur Verfügung standen.



#### 4.6 Analyse der bei Toledo gefundenen ORFs

Die über Vector NTI Suite 7 gefundenen und mittels „blastp“ selektierten HCMV *Toledo* ORFs wurden mit dem Programm „AlignX“ (Vector NTI Suite 7 Package) mit den veröffentlichten Sequenzen des HCMV Stammes *AD169* verglichen. Die Gegenüberstellung der ORFs erfolgte über die Aminosäuresequenzen der translatierten Gene der beiden Stämme HCMV *AD169* und *Toledo*, da nur funktionale Basenpaaraustausche berücksichtigt werden sollten. Für die in dieser Arbeit untersuchten DNA Sequenzen des HCMV Stammes *Toledo* waren für den HCMV Stamm *AD169* 127 ORFs vorhergesagt worden (Chee et al., 1990; Davison et al., 2003; Murphy et al., 2003). Von diesen 127 ORFs waren bei HCMV *Toledo* 5 ORFs (UL12, UL109, UL110, UL111 und US4) nicht auffindbar (siehe Tabelle 12 und Tabelle 13, „ORF nicht existent“). Bei diesen 5 ORFs handelte es sich um Gene, die ursprünglich von Chee vorhergesagt worden waren und in der Neubewertung des Kodierungspotentiales des HCMV Stammes *AD169* durch Davison nicht mehr berücksichtigt worden waren. Des weiteren fielen bei dem Vergleich der ORFs der HCMV Stämme *AD169* und *Toledo* 4 Gene auf ( UL9, UL15, UL74 und UL116, siehe Tabelle 12 und Tabelle 13), die eine besonders hohe Anzahl an Aminosäureaustauschen und Deletionen / Insertionen aufwiesen. Die Divergenzen zwischen den Aminosäuresequenzen bei HCMV *AD169* und *Toledo* waren dabei so groß, daß eine Funktionalität der ORFs nicht zwingend gegeben war. Bei UL9 waren zwischen den beiden Stämmen z.B. 10 Aminosäuren von einem Austausch funktional ähnlicher Aminosäuren betroffen und 56 Aminosäuren durch einen Austausch mit einer Aminosäure mit einer anderen Funktionalität. Zusätzlich war ein Bereich von 23 Aminosäuren bei Toledo deletiert, jedoch an anderer Stelle waren Insertionen von insgesamt vier Aminosäuren vorhanden. Der ORF UL15 wies 93 ähnliche und 8 unähnliche Austausche auf wobei das gesamte Protein nur eine Länge von 332 Aminosäuren hatte. Zusätzlich waren 6 Insertionsereignisse vorhanden, über die 74 zusätzliche Aminosäuren in das Protein eingebaut worden wären. Außerdem war das ursprünglich für HCMV *AD169* bestimmte Startkodon bei HCMV *Toledo* nicht vorhanden, so daß der NH<sub>2</sub>-Terminus des Proteins eine Verkürzung um 63 Aminosäuren aufwies. Der dritte stark betroffene ORF (UL116) wies bei einer

ORF <sup>a</sup>	Länge des ORFs in Aminosäuren <i>Toledo/AD169</i> <sup>b</sup>	Austausch ähnlicher Aminosäuren <sup>c</sup>	Austausch nicht ähnlicher Aminosäuren <sup>d</sup>	Deletionen/Insertionen <sup>e</sup>	Vermutete Funktion <sup>f</sup>	Bedeutung des ORF für <i>AD169</i> Wachstum in HFFs <sup>g</sup>
UL4	150/152	4	2	-2 (M)	early glycoprotein	NE
UL5	166/166	2	2	0	RL11 family member	NE
UL6	284/284	11	4	0	RL11 family member	NE
UL7	222/222	16	7	0	RL11 family member	NE
UL8	125/122	19	2	+3 (V)	RL11 family member	NE
UL9	175/228	10	56	-59 (M); -26; +2; +1; +1	RL11 family member	NE
UL10	279/258	5	3	+21 (L)	HHV5gp017	NE
UL11	279/275	57	18	+7 (V); +1; -4	membrane protein	NE
UL12				ORF nicht existent		
UL13	527/473	6	2	+54 (V)	HHV5gp019	NE
UL14	395/327	2	0	+68 (M)	HHV5gp020	NE
UL15A	102/102	2	0	0	HHV5gp021	NE
UL15	332/342	93	8	-63 (M); +1; +3; +4; +8; +4; +53	HHV5gp020	
UL16	250/230	3	1	+20 (L)	immunomodulation	NE
UL17	104/104	2	0	0	HHV5gp023	NE
UL18	368/368	8	2	0	MHC homolog	NE
UL19	107/98	1	0	+9 (V)	HHV5gp025	NE
UL23	332/284	2	2	+48 (V)	tegument protein	NE
UL24	383/358	1	0	+25 (V)	tegument protein	NE
UL26	222/189	1	0	+34 (M)	tegument protein	A
UL27	629/608	8	2	+21 (L)	HHV5gp033	A
UL28	369/341	1	0	+28 (V)	hypothetical protein	A
UL29	360/360	0	0	0	US22 family member	A
UL30	121/121	1	0	0	HHV5gp036	A
UL31	682/595	3	0	+87 (L)	Transcription	NE
UL32	1049/1048	13	8	+2; -1; -1; +1	pp150 tegument protein	E
UL35	641/640	0	1	+1	UL25 family member	NE
UL36	583/476	3	2	+107 (L)	cell death suppressor	NE
UL40	221/221	5	4	0	HHV5gp045	NE
UL41	141/141	1	0	0	unbekannt	
UL41A	141/141	0	0	0	unbekannt	NE
UL42	122/125	4	2	-1; -2	HHV5gp047	NE
UL43	431/423	1	1	+8 (L)	tegument	NE
UL44	455/433	0	0	+22 (V)	DNA replication	E
UL45	906/906	3	1	0	ribonucleosid-diphosphat reductase	NE
UL46	310/290	0	0	+20 (V)	capsid protein	E
UL47	983/983	5	0	0	Tegument DNA release	A
UL48	2347/2241	16	2	+105 (L)	tegument	A
UL48.5	75/75	0	0	0	capsid protein	E
UL49	599/570	3	4	+29 (L)	HHV5gp055	E
UL50	398/397	1	0	+1	egress	E
UL51	157/157	1	0	0	DNA packaging/cleavage	E
UL52	667/668	3	0	-1	DNA packaging/cleavage	E
UL53	376/376	1	0	0	egress	E
UL67	115/113	0	1	+2 (V)	deoxyridine triphosphatase (dUTPase)	
UL69	765/749	1	1	+26 (V); -5	tegument	A
UL70	953/946	1	1	+7 (L)	helicase-primase complex	E
UL71	403/370	1	0	+42 (L)	HHV5gp066 tegument protein	A
UL72	388/388	3	0	0	dUTPase	A
UL73	164/138	20	9	+29 (L); -2; -1	structural glycoprotein gpUL73	E
UL74	464/466	18	50	-1; -2; +1	glycoprotein O	A
UL75	743/743	2	0	0	envelope glycoprot. H	E
UL76	325/325	5	0	0	HHV5gp071	A
UL77	658/642	7	1	+10 (L)	DNA packaging/cleavage	E
UL78	434/431	5	4	+3 (V)	G protein receptor	NE
UL83	561/561	1	1	0	pp65 lower matrix phosphoprotein	NE
UL84	625/587	6	3	+39 (M); -1	DNA replication	E
UL85	306/306	0	0	0	capsid protein	E
UL88	429/429	0	0	0	Tegument	NE
UL89 <sup>f</sup>	759/674	2	1	+115 (L)	conserved herpesvirus spliced gene	E
UL91	121/111	0	1	+10 (V)	HHV5gp085	E
UL92	321/201	0	0	+120 (V)	HHV5gp086	E
UL93	594/594	11	1	0	HHV5gp087	E
UL94	458/345	0	0	+113 (V)	HHV5gp088	A
UL95	531/531	5	1	0	tegument protein	E
UL96	174/127	0	1	+47 (L)	HHV5gp090	A
UL97	722/707	2	1	+15 (V)	phosphorylates ganciclovir	A
UL98	607/584	2	1	+23 (L)	alkaline nuclease	E

**Tabelle 12: Vergleich der Aminosäuresequenzen der ORFs der HCMV Stämme *AD169* und *Toledo*.** Die Tabelle beinhaltet alle für *AD169* veröffentlichten ORFs (<sup>a</sup>), die auf der in dieser Arbeit sequenzierten *Toledo*-DNA gefunden wurden (Chee et al., 1990; Davison et al., 2003; Murphy et al., 2003). Der Vergleich der ORFs von *AD169* und *Toledo* erfolgte auf Proteinebene. Rot markierte ORFs wiesen im Vergleich zu den veröffentlichten ORFs von *AD169* bei *Toledo* eine NH<sub>2</sub>terminale Verlängerung des Proteins auf. Blau markierte ORFs hingegen waren bei *Toledo* am Aminoterminus des Proteins kürzer als bei *AD169*. Die Farbmarkierung hängt dabei von den absoluten Längen der translatierten ORFs (<sup>b</sup>) ab. Die einzelnen Aminosäureaustausche wurden entweder als ähnlich (<sup>c</sup>) (z.B. Lysin (basisch) gegen Arginin (basisch)) oder als unähnlich (<sup>d</sup>) (z.B. Alanin (neutral und hydrophob) gegen Serin (neutral und polar)) klassifiziert. ? Tab. 3

ORF <sup>a</sup>	Länge des ORFs in Aminosäuren Toledo/AD169 <sup>b</sup>	Austausch ähnlicher Aminosäuren <sup>c</sup>	Austausch nicht ähnlicher Aminosäuren <sup>d</sup>	Deletionen/Insertionen <sup>e</sup>	Vermutete Funktion <sup>f</sup>	Bedeutung des ORF für AD169 Wachstum in HFFs <sup>g</sup>
UL99	190/190	3	0	0	myristylated tegument protein	E
UL100	392/372	1	1	+20 (L)	glycoprotein M	E
UL101	102/115	4	1	-13 (M)	unbekannt	
UL102	873/873	0	0	0	component of DNA helicase-primase	E
UL103	255/249	0	0	+6 (V)	tegument protein; HSV-1 UL7 counterpart	A
UL104	697/697	0	1	0	DNA packaging protein	E
UL107	117/150	0	0	-33 (L)	unbekannt	
UL108	112/123	0	0	-11 (V)	unbekannt	
UL109				ORF nicht existent		
UL110				ORF nicht existent		
UL111				ORF nicht existent		
UL111A <sup>h</sup>				nur erstes Exon vorhanden da Contig zu kurz	viral interleukin 10	NE
UL112	268/268	0	0	0	early phosphoprotein p34	A
UL113	484/499	5	1	-15 (V)	major early protein	A
UL112/113 <sup>h</sup>	684/684	4	1	0	HHV5gp100	A
UL114	250/250	1	0	0	uracil-DNA glycosylase	A
UL115	278/278	2	0	0	glycoprotein L precursor	E
UL116	303/314	22	4	-13; -2; -1; +2; +1	HHV5gp103	NE
UL120	197/201	17	10	-4	HHV5gp106	NE
UL121	181/180	6	5	+1 (V)	membrane protein	NE
US1	193/155	0	0	+38 (V)	unbekannt	NE
US3	186/186	1	3	0	immunomodulation	NE
US4				ORF nicht existent		
US5	127/126	2	2	+1	unbekannt	NE
US6	183/183	4	1	0	inhibits TAP-med. pep. translocation	NE
US7	225/225	10	3	0	glycoprotein	NE
US8	227/227	4	1	0	immunomodulation	NE
US9	254/247	3	0	+23 (V); -15	glycoprotein	NE
US10	158/158	4	0	0	glycoprotein	NE
US11	215/215	2	0	0	targets MHC class I molecules...	NE
US12	281/281	2	1	0	multiply hydrophobic protein	NE
US13	261/261	0	0	0	multiply hydrophobic protein	NE
US14	310/310	4	6	0	HVLF4 protein	NE
US15	385/262	1	1	+123 (L)	multiply hydrophobic protein	NE
US16	309/309	1	0	0	multiply hydrophobic protein	NE
US17	293/293	1	3	0	multiply hydrophobic protein	NE
US18	274/274	0	0	0	multiply hydrophobic protein	NE
US19	240/240	2	2	0	multiply hydrophobic protein	NE
US20	346/342	2	0	+4	transmembrane protein HWLF3	NE
US21	243/239	1	0	+4	multiply hydrophobic protein	NE
US22	576/576	0	2	0	tegument protein	NE
US23	597/592	0	0	+6 (L); -1	HHLF7 protein	A
US24	501/501	0	1	0	US22 family member	A
US25	121/179	5	0	-43 (L)	hypothetical protein US25	NE
US25A	107/156	1	0	-26 (L)	hypothetical protein HHRF1	NE
US26	613/603	3	0	+10 (L)	HHV5gp143	A
US27	364/362	16	0	+2	G protein receptor	NE
US28	387/354	2	0	+33 (V)	G protein receptor	NE
US29	462/462	8	0	0	HHV5gp146	NE
US30	363/349	14	3	+14 (V)	HHV5gp147	NE
US31	193/161	0	0	+32 (L)	HHV5gp148	NE
US32	183/183	0	0	0	HHV5gp149	NE

ORF1	83/83	0	0	0	unbekannt	-
ORF2	105/105	2	0	0	unbekannt	-
ORF6	123/123	3	0	0	unbekannt	-
ORF7	185/193	5	1	-9 (V); +1	unbekannt	-
ORF8	165/128	20	4	+37 (V)	unbekannt	-
ORF9	384/305	4	2	+79 (L)	unbekannt	-
ORF10	237/141	0	4	+96 (L)	unbekannt	-

**Tabelle 13: Vergleich der Aminosäuresequenzen der ORFs von AD169 und Toledo.**

Die Spalte Deletionen/Insertionen (<sup>e</sup>) gibt die Anzahl zusammenhängender Deletionen/Insertionen an. Der erste Wert gefolgt von M, V oder L (Aminosäuren der jeweiligen Startkodons) zeigt eine Veränderung des Aminoterminus relativ zur veröffentlichten Proteinsequenz bei AD169 an. Für alle ORFs ist auch die vermutete Funktion angegeben (<sup>f</sup>), wobei die Namen der Proteine nicht ins deutsche übersetzt wurden. Die letzte Spalte (<sup>g</sup>) zeigt die Wichtigkeit des ORFs für die Vermehrung von AD169 in HFF-Zellen an. NE bedeutet NonEssential, E steht für „Essential“ und A für „Groth Augmenting“. Diese Klassifizierung beruht auf einem Mutationsexperiment, bei dem bei AD169 jeder ORF einzeln zerstört und anschließend das Wachstum der Viren gemessen wurde (Yu et al., 2003). Die mit (\*) gekennzeichneten ORFs waren in der Neubewertung des Kodierungspotentials von AD169 (Davison et al., 2003) nicht mehr vorhanden. –Die Tabellen 12 und 13 sind im Anhang ab Seite 79 nochmals vergrößert dargestellt–

Länge von 303 Aminosäuren 22 ähnliche und vier unähnliche Aminosäureaustausche auf, zudem lagen in der Proteinsequenz drei Deletionen und zwei Insertionen vor. Die größte Deletion hatte dabei eine Länge von 13, die restlichen von einer und zwei Aminosäuren. Die Insertionen waren zwei und eine Aminosäure lang. Die ORFs UL9 und UL116 waren dabei für die Vermehrung von HCMV *AD169* in HFF-Zellen (Yu et al., 2003) als nicht wichtig (NE, siehe Tabelle 12 und Tabelle 13) charakterisiert worden. UL15 hingegen war eines der Gene die von Chee ursprünglich postuliert und in der Arbeit von Davison verworfen worden waren. Der vierte, durch starke Mutationen auffallende, ORF (UL 74) ist als einziges Gen dieser Gruppe genauer charakterisiert worden. Die Proteinsequenz von UL74 (Glykoprotein O) weist bei HCMV *Toledo* im Vergleich zu HCMV *AD169* zwei Deletionsstellen mit insgesamt 3 betroffenen Aminosäuren und eine Insertion von einer Aminosäure auf. Zusätzlich sind 18 Aminosäuren durch ähnliche und 50 Aminosäuren durch funktional unähnliche Aminosäuren ersetzt worden. Die hohe Divergenz zwischen den beiden HCMV Stämmen in diesem ORF beruht hierbei auf der Funktion des Proteins als Antigen. Des weiteren unterschied sich der ORF UL74 von den drei anderen in dieser Gruppe besprochenen ORFs (UL9, UL15 und UL116) insofern, als daß die Zerstörung des ORFs zu einer Wachstumsbeschleunigung des Virus in HFF-Zellkultur führte (Yu et al., 2003).

Die restlichen 118 ORFs waren entweder gar nicht oder nur gering von Mutationsereignissen betroffen. Von Deletions- oder Insertionsereignissen waren UL32 (pp150 Tegumentprotein), UL 42 (unbekannte Funktion), UL 50 (Ausschleusungsfunktion), UL52 (DNA Verpackung/Spaltung), UL84 (DNA Replikation), US5 (unbekannte Funktion), US20 (Transmembranprotein HWLF3), US21 (stark hydrophobes Protein), US23 (HHLF7 Protein), US27 (G Protein Rezeptor) und ORF7 (unbekannte Funktion) betroffen. Mit Ausnahme von UL32 und UL42 waren die Deletions- und Insertionsereignisse auf eine Position beschränkt. Für alle 118 ORFs ließ sich des weiteren feststellen, daß hauptsächlich Aminosäureaustausche auftraten, bei denen die Funktionalität der ausgetauschten Aminosäure erhalten wurde.

Innerhalb der Gruppe der 118 ORFs, bei denen die Proteinsequenzen der ORFs von HCMV *Toledo* und HCMV *AD169* sehr ähnlich waren, befanden sich 19 Gene (UL29, UL41A, UL46, UL48.5, UL85, UL88, UL92, UL94, UL102, UL103, UL107, UL108, UL112, US1, US13, US18, US31, US32 und ORF1), deren

Aminosäuresequenzen absolut identisch waren. Diese Gene waren sowohl als essentiell für die Vermehrung des Virus HCMV AD169 in HFF-Zellen, als auch als unwichtig oder Wachstumsbeschleunigend, von Yu beschrieben worden. Besonders interessant in dieser Gruppe ist der ORF UL108, da er in der Neubewertung des Kodierungspotentiales von HCMV *AD169* (Davison et al., 2003) nicht mehr auftaucht, da der ORF nicht mit einem ATG Startkodon anfing. Ein alternatives Valinstartkodon ist jedoch vorhanden, daß als Translationsstartpunkt dienen könnte.

---

## 5 Diskussion

### 5.1 Zielsetzung der Arbeit

Der Laborstamm *AD169* des Humanen Cytomegalievirus wurde in dieser Arbeit als Grundlage für eine Sequenzierung des klinischen Isolates HCMV *Toledo* verwendet. Da HCMV *AD169* zur Gewinnung eines Vakzinestammes häufig in Fibroblasten kultiviert wurde (Elek und Stern, 1974), hat sich das Virus an diese Wirtszellen angepaßt und läßt sich nun, im Gegensatz zu klinischen Isolaten, sehr gut in Fibroblasten kultivieren. Aufgrund dieser Anpassung entwickelte sich HCMV *AD169* zum klassischen HCMV Laborstamm, für den bereits 1990 die komplette DNA Sequenz vorlag (Chee et al., 1990). Über die lange Kultivierung des HCMV Stammes *AD169* unter Laborbedingungen, also ohne größeren selektiven Druck, konnten sich viele Mutationen ausbilden (Cha et al., 1996). Da es nur begrenzte Therapiemöglichkeiten für HCMV gibt, und es nie gelungen ist eine erfolgreiche Vakzinierung mit HCMV durchzuführen, entschlossen wir uns, zusammen mit T. Shenk, Princeton University, Princeton, New Jersey, USA, das klinische Isolat HCMV *Toledo* zu untersuchen. Mittels einer kompletten Sequenzierung des Genoms von HCMV *Toledo* sollten konservierte und variable ORFs des Virus identifiziert werden. Diese ORFs sollten dann in weiterführenden Arbeiten untersucht werden, um sie als Ziele für eine erfolgreiche Immunisierung oder Medikation zu bewerten.

Über eine auf PCR Produkten basierende Sequenzierungsmethode konnten 75% des 230kbp großen Genoms von HCMV *Toledo* entschlüsselt werden. Zu diesem Zeitpunkt veröffentlichte unser Kollaborationspartner T. Shenk, zusammen mit R. Myers von der Stanford sequencing group mit finanzieller Unterstützung des U.S. Department of Energy (DOE), die kompletten DNA Sequenzen der HCMV Stämme *AD169*, *Towne*, *Toledo*, *FIX*, *TR* und *PH*, so daß wir von der Vollendung der Sequenzierung des HCMV Stammes *Toledo* Abstand nehmen mußten. Da sich die Analysen der sequenzierten HCMV Stämme durch die Gruppe um Shenk auf eine Identifizierung konservierter ORFs zwischen den Stämmen beschränkte, wurde der ursprüngliche Arbeitsansatz, der Vergleich der DNA Sequenzen von HCMV *AD169* und *Toledo* zur Identifizierung konservierter und hochgradig variabler ORFS, erweitert. Hierzu wurden die gefundenen ORFs von HCMV *Toledo* auf funktionale Austausche von Aminosäuren untersucht und die Initiation

---

von ORFs durch alternative Startkodons berücksichtigt. Die Verifizierung dieser Startkodons erfolgte über Promotoranalysen, Untersuchungen der alternativen Startkodons auf Kozak Sequenzen und RT-PCR Untersuchungen der mRNA.

## 5.2 Mutationsanalysen der DNA von HCMV *Toledo*

Der Vergleich der sequenzierten DNA Bereiche bei HCMV *AD169* und *Toledo* ergab für die untersuchten 172.178 bp eine Abweichung von insgesamt 2,04%. Die gefundenen Mutationen waren derart, daß das Cytosin bei HCMV *AD169* bei HCMV *Toledo* 792 mal einen Basenpaaraustausch aufwies, das Thymin 729 mal, das Adenosin 777 mal und das Guanin 786 mal. Damit waren alle Basen annähernd gleich häufig von Mutationen betroffen. Des weiteren fiel auf, daß jedes Nukleotid bevorzugt in eine bestimmte Richtung mutiert war. Somit war Cytosin in 76,77% der Fälle durch ein Thymin ersetzt worden, Thymin in 78,74% der Fälle durch Cytosin, Adenosin zu 78,38% durch ein Guanin ersetzt worden, und Guanin mit einer Häufigkeit von 77,86% mit Adenosin ausgewechselt worden. Da die bevorzugten Mutationsrichtungen bei allen Nukleotiden gleich häufig auftraten und die Mutationsrichtungen sich gegenseitig die Waage hielten (C->T, T->C, G->A, A->G), blieb trotz der Differenz von 2,04% zwischen den beiden HCMV Stämmen die Anzahl der Nukleotide innerhalb der Genome von HCMV *Toledo* und *AD169* annähernd konstant. Auch die Anzahl der gefundenen Deletionen und Insertionen in HCMV *Toledo* glich sich mit 220 deletierten zu 208 insertierten Nukleotiden annähernd aus. Ein mögliche Erklärung dieses Phänomens wäre ein für das Virusgenom zu erkennender Selektionsdruck, die Schmelztemperatur und die Länge der DNA konstant zu halten, um eine optimale Virusverpackung und Vermehrung zu gewährleisten.

## 5.3 Untersuchungen der ORFs von HCMV *Toledo*

Die Analyse der auf den sequenzierten DNA Bereichen von HCMV *Toledo* gelegenen ORFs erfolgte auf Proteinebene, da nur funktionale Mutationereignisse berücksichtigt werden sollten. Hierfür wurden alle für HCMV *Toledo* berechneten ORFs mit den entsprechenden ORFs von HCMV *AD169* verglichen. Von den insgesamt 127 ORFs, die für HCMV *AD169* in dem untersuchten DNA Bereich beschrieben waren (Chee et al., 1990; Davison et al., 2003; Murphy et al., 2003), konnten 5 ORFs nicht bei *Toledo* nachgewiesen werden. Diese ORFs, die schon von Davison et al. (2003) als wahrscheinlich nicht

---

kodierend verworfen worden waren, waren aufgrund von Deletions- und Insertionsereignissen innerhalb der Leserahmen nicht mehr als funktionale ORFs zu erkennen. Die restlichen ORFs waren soweit konserviert, daß alle HCMV *Toledo* ORFs, selbst wenn sie von vielen Mutationseignissen betroffen waren, ihren Gegenstücken bei *AD169* zugeordnet werden konnten. Die von Yu et al. vorgenommene Einteilung der ORFs von HCMV *AD169* in drei Gruppen („NonEssential“, „Essential“ und „Growth Augmenting“) bezüglich ihrer Wichtigkeit für die Vermehrung des Virus in HFF-Zellen, konnte dabei nicht in Relation zu den Mutationshäufigkeiten in bestimmten Genen gesetzt werden. Ursprünglich waren wir von der Hypothese ausgegangen, daß Gene, die für die Kontrolle und Abwehr von Immunreaktionen des Wirtes verantwortlich sind, für *AD169* nicht essentiell sind. In diesen Genen hätte sich dann aufgrund fehlenden Selektionsdruckes Mutationen anreichern können, so daß diese Gene leicht hätten identifiziert werden können. Da jedoch auch viele der „Non Essential“ Gene (Yu et al., 2003) konserviert waren, sollte diese Arbeitshypothese nicht weiter verfolgt werden. Obwohl viele dieser als „NonEssential“ charakterisierten Gene Mutationen in den Aminosäuresequenzen aufwiesen, waren diese zumeist derart, daß die Funktionalitäten der ursprünglichen gegenüber den ausgetauschten Aminosäuren identisch waren. Viele der „NonEssential“ Gene waren sogar in ihrer Aminosäuresequenz in den beiden HCMV Stämmen absolut identisch, obwohl sie für die Vermehrung in HFF-Zellen nicht wichtig waren.

Man hätte vermuten können, daß Gene, auf denen kein Selektionsdruck lastet, nach häufiger serieller Passage, wie es bei dem Laborstamm HCMV *AD169* zutrifft, zwangsweise zufällige Mutationen ansammeln würden. Auch die dem Immunsystem exponierten Glykoproteine zeigen kein einheitliches Bild bei der Anhäufung von Mutationen. Die Glykoproteine H, L, M sowie der für ein unbekanntes Glykoproteine kodierende ORF US10 sind stark konserviert. Das „structural glycoprotein gpUL73“ sowie das Glykoprotein O weisen jedoch eine Vielzahl von Mutationen in den Proteinsequenzen von HCMV *AD169* und *Toledo* auf. Eine Erklärung für diese Unterschiede könnte sein, daß die Glykoproteine O und gpUL73 eine Häufung von Mutationen tolerieren können, ohne funktional beeinträchtigt zu werden, oder daß die erhöhte Mutationsrate dieser Gene im veränderten Wirtszelltropismus des Laborstammes HCMV *AD169* begründet liegt, auf den dieses Virus selektioniert worden war. Die konservierten Glykoproteine



---

hingegen erlauben aufgrund ihrer Funktionalität nur einen geringen Spielraum für Mutationen.

#### **5.4 Evaluierung der Benutzung alternativer Startkodons**

Ein weiterer Aspekt, der bei der Analyse des Kodierungspotentiales von HCMV Toledo berücksichtigt wurde, war die Initiation der Translation bekannter HCMV *AD169* ORFs über alternative Startkodons. Über die Benutzung alternativer Startkodons können bekannte Proteine eine neue Funktionalität erhalten, oder die Expression dieser ORFs kann über diese Startkodons reguliert werden (Touriol et al., 2003). Bereits 1989 wurde das erste Virus (Sendai Virus) identifiziert, daß einen ORF besaß, dessen Translation über ein ACG Kodon initiiert wurde (Curran und Kolakofsky, 1988). Ein Jahr später wurde beim Moloney murine leukemia virus ein ORF entdeckt, der nicht über ein vorhandenes AUG Kodon initiiert wurde, sondern über ein 5' gelegenes CUG Kodon (Prats et al., 1989). Die in dieser Arbeit für HCMV untersuchten alternativen Startkodons CUG und GUG konnten bereits in anderen viralen Systemen nachgewiesen werden. Das CUG Kodon wurde z.B. als Translationsstartpunkt für den UL16 ORF des Herpes Simplex Typ 2 Virus diskutiert (Dolan et al., 1998). GUG wiederum wurde als Translationsinitiator beim Pflanzen Virus „Beet Western yellows virus“ (Pfeffer et al., 2002) und im Säugersystem beschrieben (Fuxe et al., 2000). Die c-myc mRNA enthält ein 5' gelegenes CUG Kodon sowie ein 3' gelegenes AUG Kodon, die die Synthese zweier Proteinisoformen steuern (Hann et al., 1994). Ein weiteres Beispiel für die Benutzung alternativer 5' gelegener Startkodons ist das FGF2 Gene. Genprodukte, die über das alternative Startkodon erzeugt wurden, enthalten ein NLS (Nuclear localization signal), daß in den Proteinen fehlt, die über das „normale“ Startkodon initiiert wurden (Bugler et al., 1991). Die Benutzung alternativer AUG Startkodons für HCMV *AD169* wurde bereits nachgewiesen. Der ORF UL26 besitzt ein zusätzliches AUG Startkodon 5' von dem veröffentlichten Startkodon. Dieses AUG besitzt keine Kozak Motiv, jedoch konnten zwei Polypeptide in infizierten HFF Zellen nachgewiesen werden (Stamminger et al., 2002).

Von den insgesamt 217 identifizierten ORFs bei HCMV Toledo konnten für 58 ORFs alternative Startkodons nachgewiesen werden, die bei Benutzung zu einer Verlängerung des NH<sub>2</sub>-terminalen Bereiches des betroffenen Proteins führten. Die

---

gefundenen Verlängerungen konnten dabei bis zu 123 Aminosäuren lang sein (US15). Da die alternativen Startkodons, wenn sie in einem bestimmten Sequenzkontext standen (Kozak, 1987, Boeck und Kolakofsky 1994), mit der annähernd gleichen Effizienz eines ATG Startkodons die Initiation der Translation der mRNA veranlassen konnten, wurden alle identifizierten alternativen Startkodons auf die Existenz eines Kozak Motivs (Kozak, 1987) untersucht. Obwohl keines der alternativen Startkodons eine optimale Kozak Konsensus Sequenz aufwies, zeigten doch viele untersuchte Startkodons die Minimalanforderung für eine potentielle Initiation. Für jedes gefundene alternative Startkodon wurde auch die Position der jeweiligen Promotoren untersucht. Hierbei stellte sich heraus, daß bei 72% der verlängerten ORFs die Promotoren 5' des ATG- als auch des alternativen Startkodons lagen. Des weiteren wurde untersucht, ob die alternativen Startkodons auf der mRNA der entsprechenden ORFs vorhanden waren. Mittels RT-PCR wurde bei acht ORFs die Existenz der alternativen Startkodons auf den mRNAs in von HCMV *Toledo* infizierten HFF-Zellen untersucht. Die analysierten ORFs hatten dabei NH<sub>2</sub>-terminale Verlängerungen von 47-123 Aminosäuren. Für fünf der acht ORFs konnte auch auf mRNA Ebene die Existenz der alternativen Startkodons nachgewiesen werden. Der entscheidende Versuch, ein Nachweis der Extensionen in den durch HCMV erzeugten Proteinen selbst, konnte zum Ende der Arbeit noch nicht erbracht werden. Obwohl die durchgeführten Untersuchungen nicht ausreichen um definitiv die Benutzung der alternativen Startkodons zu beweisen, liefern sie doch starke Indizien dafür, das eine Verwendung der alternativen Startkodons gegeben ist. Zur Zeit arbeite ich weiter am experimentellen Nachweis der verlängerten ORFs auch auf Proteinebene.

---

## 6 Literaturverzeichnis

**Alford C.A., Pass R.F., 1981:**

Epidemiology of chronic congenital and perinatal infections of man.

in: Clin Perinatol. 8(3):397-414

**Becroft D.M.O., 1981:**

Prenatal cytomegalovirus infection: Epidemiology, pathology, and pathogenesis.

in: Rosenberg H.S., Bernstein J., eds. Perspective in Pediatric Pathology. New York: Masson Press: 203-241

**Boeck R., Kolakofsky D., 1994:**

Positions +5 and +6 can be major determinants of the efficiency of non-AUG initiation codons for protein synthesis.

in: EMBO J. 1;13(15):3608-17

**Boldogh I., Beth E., Huang E.S., Kyalwazi S.K., Giraldo G., 1981:**

Kaposi's sarcoma. IV. Detection of CMV DNA, CMV RNA and CMNA in tumor biopsies.

in: Int J Cancer. 15;28(4):469-74

**Boppana S.B., Pass R.F., Britt W.J., Stagno S., Alford C.A., 1992:**

Symptomatic congenital cytomegalovirus infection: neonatal morbidity and mortality.

in: Pediatr Infect Dis J. 11(2):93-9

**Britt W.J., Alford C.A., 1996:**

Cytomegalovirus

in: Fields B.N., Knipe D.M., Howley P.M., eds. Fields Virology, Third Edition. Philadelphia: Lippincott-Raven Publishers: 2494-2501

**Bugler, B., Amalric, F., Prats H., 1991:**

Alternative initiation of translation determines cytoplasmic or nuclear localization of basic fibroblast growth factor.

in: *Mol. Cell. Biol.* 11, 573-577.

**Cha T.A., Tom E., Kemble G.W., Duke G.M., Mocarski E.S., Spaete R.R., 1996:**

Human cytomegalovirus clinical isolates carry at least 19 genes not found in laboratory strains.

in: J Virol. 70(1):78-83

---

**Chee M.S., Bankier A.T., Beck S., Bohni R., Brown C.M., Cerny R., Horsnell T., Hutchison C.A. 3rd, Kouzarides T., Martignetti J.A., et al., 1990:**

Analysis of the protein-coding content of the sequence of human cytomegalovirus strain *AD169*.

in: *Curr Top Microbiol Immunol.* 154:125-69

**Curran j., Kolakofsky D., 1988:**

Ribosomal initiation from an ACG codon in the Sendai virus P/C mRNA.

in: *EMBO J.* 7(1):245-51

**Davison A.J., Dolan A., Akter P., Addison C., Dargan D.J., Alcendor D.J., McGeoch D.J., Hayward G.S., 2003:**

The human cytomegalovirus genome revisited: comparison with the chimpanzee cytomegalovirus genome.

in: *J Gen Virol.* 84(Pt 1):17-28

**Demmler G.J., 1991:**

Infectious Diseases Society of America and Centers for Disease Control.

Summary of a workshop on surveillance for congenital cytomegalovirus disease.

in: *Rev Infect Dis.* 13(2):315-29

**Dolan, A., Jamieson, F.E., Cunningham, C., Barnett, B.C., McGeoch, D.J., 1998:**

The genome sequence of herpes simplex virus type 2.

in: *J. Virol.* 72, 2010-2021.

**Dunn W., Chou C., Li H., Hai R., Patterson D., Stolc V., Zhu H., Liu F., 2003:**

Functional profiling of a human cytomegalovirus genome.

in: *Proc Natl Acad Sci U S A.* 25;100(24):14223-8

**Elek S.D., Stern H., 1974:**

Development of a vaccine against mental retardation caused by cytomegalovirus infection in utero.

in: *Lancet.* 1(7845):1-5

**Fowler K.B., Stagno S., Pass R.F., Britt W.J., Boll T.J., Alford C.A., 1992:**

The outcome of congenital cytomegalovirus infection in relation to maternal antibody status.

in: *N Engl J Med.* 326(10):663-7

**Fuxe, J., Raschperger, E., Pettersson, R.F., 2000:**

Translation of p15.5INK4B, an N-terminally extended and fully active form of p15INK4B, is initiated from an upstream GUG codon.

in: *Oncogene* 19, 1724-1728.

**Geder L., Sanford E.J., Rohner T.J., Rapp F., 1977:**

Cytomegalovirus and cancer of the prostate: in vitro transformation of human cells.

in: *Cancer Treat Rep.* 61(2):139-46

**Giraldo G., Beth E., Huang E.S., 1980:**

Kaposi's sarcoma and its relationship to cytomegalovirus (CMNV). III. CMV DNA and CMV early antigens in Kaposi's sarcoma.

in: *Int J Cancer.* 15;26(1):23-9

**Hahn G., Khan H., Baldanti F., Koszinowski U.H., Revello M.G., Gerna G., 2002:**

The human cytomegalovirus ribonucleotide reductase homolog UL45 is dispensable for growth in endothelial cells, as determined by a BAC-cloned clinical isolate of human cytomegalovirus with preserved wild-type characteristics.

in: *J Virol.* 76(18):9551-5

**Hann, S.R., Dixit, M., Sears, R.C., Sealy L., 1994:**

The alternatively initiated c-Myc proteins differentially regulate transcription through a noncanonical DNA-binding site.

in: *Genes Dev.* 8, 2441-2452.

**Irmiere A., Gibson W., 1983:**

Isolation and characterization of a noninfectious virion-like particle released from cells infected with human strains of cytomegalovirus.

in: *Virology.* 15;130(1):118-33

**Jarvis M.A., Nelson J.A., 2002:**

Human cytomegalovirus persistence and latency in endothelial cells and macrophages.

in: *Curr Opin Microbiol.* 5(4):403-7

**Kozak M., 1986:**

Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes;

in: *Cell* 31;44(2):283-92.

**Kozak M., 1987:**

An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs;  
in: *Nucleic Acids Res.* 26;15(20):8125-48.

**Kozak M.;1995:**

Adherence to the first-AUG rule when a second AUG codon follows closely upon  
the first;  
in: *Proc Natl Acad Sci U S A.* 28;92(7):2662-6

**MacCormac L.P., Grundy J.E., 1999:**

Two clinical isolates and the Toledo strain of cytomegalovirus contain endothelial  
cell tropic variants that are not present in the AD169, Towne, or Davis strains.  
in: *J Med Virol.* 57(3):298-307

**Medearis D.N. Jr., 1982:**

Cmv immunity: imperfect but protective.  
in: *N Engl J Med.* 306(16):985-6

**Murphy E., Rigoutsos I., Shibuya T., Shenk T.E., 2003:**

Reevaluation of human cytomegalovirus coding potential.  
in: *Proc Natl Acad Sci U S A.* 100(23):13585-90

**Murphy E., Yu D., Grimwood J., Schmutz J., Dickson M., Jarvis M.A., Hahn  
G., Nelson J.A., Myers R.M., Shenk T.E., 2003:**

Coding potential of laboratory and clinical strains of human cytomegalovirus.  
*Proc Natl Acad Sci U S A.* 100(25):14976-81

**Peabody D.S., 1987:**

Translation initiation at an ACG triplet in mammalian cells.  
in: *J Biol Chem.* 25;262(24):11847-51

**Pfeffer, S., Dunoyer, P., Heim, F., Richards, K.E., Jonard, G., Ziegler-Graff, V.,  
2002:**

P0 of Beet Western yellows virus is a suppressor of posttranscriptional gene  
silencing.  
in: *J. Virol.* 76, 6815–6824.

**Prats A.C., De Billy G., Wang P., Darlix J.L., 1989:**

CUG initiation codon used for the synthesis of a cell surface antigen coded by the murine leukemia virus.

in: J Mol Biol. 20;205(2):363-72

**Quinnan G.V. Jr, Delery M., Rook A.H., Frederick W.R., Epstein J.S., Manischewitz J.F., Jackson L., Ramsey K.M., Mittal K., Plotkin S.A., et al., 1984:**

Comparative virulence and immunogenicity of the Towne strain and a nonattenuated strain of cytomegalovirus.

in: Ann Intern Med. 101(4):478-83

**Reynolds D.W., Stagno S., Hosty T.S., Tiller M., Alford C.A. Jr., 1973:**

Maternal cytomegalovirus excretion and perinatal infection.

in: N Engl J Med. 289(1):1-5

**Rice G.P., Schrier R.D., Oldstone M.B., 1984:**

Cytomegalovirus infects human lymphocytes and monocytes: virus expression is restricted to immediate-early gene products.

in: Proc Natl Acad Sci U S A. 81(19):6134-8

**Rigoutsos I., Novotny J., Huynh T., Chin-Bow S.T., Parida L., Platt D., Coleman D., Shenk T.:**

In silico pattern-based analysis of the human cytomegalovirus genome.

in: J Virol. 77(7):4326-44

**Smith I.L., Taskintuna I., Rahhal F.M., Powell H.C., Ai E., Mueller A.J., Spector S.A., Freeman W.R., 1998:**

Clinical failure of CMV retinitis with intravitreal cidofovir is associated with antiviral resistance.

in: Arch Ophthalmol. 116(2):178-85

**Stagno S., Pass R.F., Dworsky M.E., Alford C.A., 1983:**

Congenital and perinatal cytomegalovirus infections.

in: Semin Perinatol. 7(1):31-42

**Stamminger, T., Gstaiger, M., Weinzierl, K., Lorz, K., Winkler, M., Schaffner, W. 2002:**

Open reading frame UL26 of human cytomegalovirus encodes a novel tegument protein that contains a strong transcriptional activation domain.

in: *J. Virol.* 76, 4836-4847.

**Touriol C., Bornes S., Bonnal S., Audigier S., Prats H., Prats A.C., Vagner S., 2003:**

Generation of protein isoform diversity by alternative initiation of translation at non-AUG codons.

in: *Biol Cell.* 95(3-4):169-78

**Wang X., Huong S.M., Chiu M.L., Raab-Traub N., Huang E.S., 2003:**

Epidermal growth factor receptor is a cellular receptor for human cytomegalovirus.

in: *Nature.* 24;424(6947):456-61

**Weller T.H., Hanshaw J.B., Scott D.E., 1960:**

Serologic differentiation of viruses responsible for cytomegalic inclusion disease.

in: *Virology.* 12:130-2

**Yu D., Silva M.C., Shenk T., 2003:**

Functional map of human cytomegalovirus AD169 defined by global mutational analysis.

in: *Proc Natl Acad Sci U S A.* 100(21):12396-401



---

## 7 Anhang

### 7.1 Kurzzusammenfassung

Die DNA Sequenzen klinischer Isolate des Humanen Cytomegalievirus (HCMV) und des klassischen Laborstammes *AD169* können stark voneinander abweichen. Da es für HCMV nur begrenzte Therapiemöglichkeiten gibt und eine Immunisierung bis heute nicht möglich war, sollten über Sequenzvergleiche zwischen HCMV *AD169* und HCMV *Toledo* Gene identifiziert werden, die Ansatzpunkte für eine Therapie der Virusinfektion bieten. In der vorliegenden Arbeit haben wir einen Großteil der DNA Sequenz des klinischen HCMV Isolates *Toledo* bestimmt, die Aminosäuresequenzen von 127 Offenen Leserahmen (ORFs) ermittelt und mit den Sequenzen des Laborstammes HCMV *AD169* verglichen. Ausgehend von aufgereinigter, aus Virionen isolierter DNA des HCMV Stammes *Toledo*, wurde die DNA Sequenz des klinischen Isolates ähnlichen Stammes über eine „genome walking“ Methode erschlossen. Für die Sequenzierung wurden *AD169* Sequenz homologe Primer verwendet. Ein Vergleich der DNA Sequenzen von HCMV *AD169* und HCMV *Toledo* zeigte Basenpaaraustausche, sowie Deletionen und Insertionen. Mehrere ORFs wiesen NH<sub>2</sub>-terminale Extensionen der kodierenden Regionen auf, die bis zu 123 Aminosäuren lang sein konnten. Einige dieser Extensionen waren hochgradig homolog zwischen den beiden untersuchten HCMV Stämmen. Sechs dieser Extensionen begannen mit AUG, 24 mit CUG und 21 mit GUG, die alle als Translationsstartpunkte in Pro- und Eukaryoten bereits nachgewiesen worden sind. Einige der gefundenen verlängerten Sequenzen könnten veränderte Funktionalitäten in den betroffenen Proteinen bewirken und waren ebenfalls nachweisbar in drei anderen klinischen Isolaten und dem Laborstamm *Towne*. Es konnte gezeigt werden, daß einige der alternativen Startkodons in einem Sequenzumfeld lagen, das ihre Benutzung begünstigt, da das postulierte Kozak-Motiv vorkam. Zudem konnte nachgewiesen werden, daß einige der alternativen Startkodons auf den mRNAs vorhanden waren, die in von HCMV infizierten HFF Zellen produziert wurden. Der Beweis der Proteinsynthese über die alternativen Startkodons konnte bis jetzt nicht definitiv erbracht werden, da durchgeführte Expressionsstudien aufgrund technischer Schwierigkeiten bisher noch keine eindeutigen Daten erbracht haben.

---

## 7.2 Abstract

The nucleotide sequences of the DNAs from clinical isolates of human cytomegalovirus (HCMV) and from the standard laboratory strain *AD169* can differ considerably. Since in case of an infection with HCMV only limited medical countermeasures are available, we were interested in identifying potential immunomodulating genes in the genome of HCMV which could be targets for further therapeutic studies. We have determined the nucleotide sequences, derived the amino acid sequences of 127 open reading frames (ORFs) in the low-passage *Toledo* strain of HCMV and compared them to the *AD169* sequences. Purified virion-isolated DNA from the *Toledo* strain has been used in a genomic walking approach with *AD169* DNA sequence-derived PCR primers. A comparison of the DNA of the HCMV strains *AD169* and *Toledo* revealed multiple nucleotide mismatches as well as deletions and insertions. There are several potential amino-terminal extensions of coding regions which range in lengths from a few to 123 amino acids. Some of these extended amino acid sequences are very similar, and can occur in either HCMV strain relative to the other. Six extensions start with an AUG, 24 with a CUG, and 21 with a GUG which can serve as translational start codons in pro- and eukaryotes. Some of the extended sequences might bestow altered, biologically functional properties upon the affected proteins, and they are conserved in three clinical isolates and the *Towne* strain. In support of these notions the analysis of the DNA surrounding the alternative start codons showed that most of these codons displayed a weak homology to the Kozak consensus sequence. These weak homologies could promote the use of the alternative start codons to a certain degree. In addition I could show that some of the postulated alternative start codons were present in mRNAs, which were produced in and

---

isolated from HCMV *Toledo* infected HFF cells. Due to technical difficulties, searches for extended proteins have not yielded conclusive results.

### 7.3 Lebenslauf:

Name: Brondke

Vorname: Lars Holger

Geburtsdatum: 27.12.1972

Geburtsort: Leverkusen

Staatsbürgerschaft: deutsch

Ausbildungsweg:

Grundschule: 01.08.1977 – 06.06.1983; Städtische Gemeinschaftsgrundschule  
Witzhelden

Gymnasium: 13.08.1983 – 11.06.1993; Städtisches Gymnasium Leichlingen

Abschluß: Abitur

Studium: 01.10.1993;

Immatrikulation an der Universität zu Köln

an der biologischen Fakultät der Universität zu Köln,  
Fachrichtung Genetik.

Hauptfächer: Genetik, Entwicklungsbiologie, Biochemie.

Diplomarbeit: im Institut für Genetik der Universität zu Köln,  
Abteilung med. Genetik und Virologie, Prof. Dr. W. Doerfler

Thema: Untersuchungen an der DNA des humanen  
Cytomegalievirus

Abschluß: Diplombiologe

Promotion: Juni 2000 – August 2005:

Promotion als wissenschaftlicher Mitarbeiter am Institut für  
Genetik der Universität zu Köln, Abteilung med. Genetik und  
Virologie, Prof. Dr. W. Doerfler

Wissenschaftliche Projekte:

Sequenzierung und Analyse der DNA eines HCMV Stammes

Expressionsuntersuchungen an viral transformierten  
Hamsterzellen

Erlangen, den 30.08.2005

---

## 7.4 Erklärung

Ich versichere, daß ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit - einschließlich Tabellen, Karten und Abbildungen -, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; daß diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; daß sie - abgesehen von unten angegebenen Teilpublikationen - noch nicht veröffentlicht worden ist sowie, daß ich eine solche Veröffentlichung vor Abschluß des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen dieser Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Dr. W. Doerfler betreut worden.

Teilpublikationen:

Palka-Santini M, Schwarz-Herzke B, Hosel M, Renz D, Auerochs S, Brondke H,

Doerfler W. (2003):

The gastrointestinal tract as the portal of entry for foreign macromolecules: fate of DNA and proteins.

in: Mol Genet Genomics. 270(3):201-15

Holger Brondke, Birgit Schmitz, Thomas Shenk, and Walter Doerfler.(2005):

Differences in the Nucleotide Sequences between the *AD169* and *Toledo* Strains of Human Cytomegalovirus. *submitted*

Ich versichere, daß ich alle Angaben wahrheitsgemäß nach bestem Wissen und Gewissen gemacht habe und verpflichte mich, jedmögliche, die obrigen Angaben betreffenden Veränderungen, dem Dekanat unverzüglich mitzuteilen.

Erlangen, den 30.08.2005

Danksagung:

An dieser Stelle möchte ich mich bei meinem Doktorvater, Herrn Prof. Dr. Walter Doerfler, für die Bereitstellung eines Arbeitsplatzes in seinem Labor sowie für die Vergabe der Doktorarbeit bedanken. Für die angenehme Arbeitsatmosphäre und die hilfreichen Ratschläge und Diskussionen bin ich ihm sehr dankbar.

Auch bei Herrn Prof. Dr. Fleckenstein möchte ich mich für die Bereitstellung eines Arbeitsplatzes in Erlangen bedanken. Der Aufenthalt am Institut für Klinische und Molekulare Virologie war für meine Doktorarbeit von großem Vorteil, da ich hier mit einer Vielzahl von Kollegen mit ähnlichen Arbeitsgebieten diskutieren konnte.

Birgit Schmidt möchte ich für die Hilfen beim Sequenzieren danken.

Mein weiterer Dank gilt meinen Arbeitskollegen in Köln und Erlangen. Besonders möchte ich mich bei Andreas Dorn, Norbert Hochstein, Sabrina Auerochs, Dennis Webb und Oliver Neven bedanken, die so manche Inkubationspause schneller vergehen ließen, als mir manchmal lieb war.

Für Hilfe bei organisatorischen und studientechnischen Fragen möchte ich mich bei Susanne Scheffler und Petra Böhm bedanken. Auch die eine oder andere gepflegte Skatrunde wird nicht in Vergessenheit geraten.

Den Spülfrauen aus Köln und Erlangen für ihren unermüdlichen Einsatz. Ohne ihre gewissenhafte Arbeit wäre kein Experimentieren möglich gewesen.

Natürlich möchte ich auch meiner Familie danken, die mich während meines Studiums nicht nur finanziell unterstützt hat und Frau Christina Deiniger für ihre Unterstützung während des Verfassens dieser Arbeit.

ORF	Länge in Aminosäuren Toledo/AD169	Austausch funktional ähnlicher Aminosäuren	Austausch funktional unähnlicher Aminosäuren	Deletionen/Insertionen	Vermutete Funktion	Funktion des ORFs von AD169 für den Wachstum in HFFs
UL4	150/152	4	2	-2 (M)	early glycoprotein	NE
UL5	166/166	2	2	0	RL11 family member	NE
UL6	284/284	11	4	0	RL11 family member	NE
UL7	222/222	16	7	0	RL11 family member	NE
UL8	125/122	19	2	+3 (V)	RL11 family member	NE
UL9	175/228	10	56	-59 (M); -26; +2; +1; +1	RL11 family member	NE
UL10	279/258	5	3	+21 (L)	HHV5gp017	NE
UL11	279/275	57	18	+7 (V); +1; -4	membrane protein	NE
UL12				no ORF present		
UL13	527/473	6	2	+54 (V)	HHV5gp019	NE
UL14	395/327	2	0	+68 (M)	HHV5gp020	NE
UL15A	102/102	2	0	0	HHV5gp021	NE
UL15	332/342	93	8	-63 (M); +1; +3; +4; +8; +4; +53	HHV5gp020	NE
UL16	250/230	3	1	+20 (L)	Immunomodulation	NE
UL17	104/104	2	0	0	HHV5gp023	NE
UL18	368/368	8	2	0	MHC homolog	NE
UL19	107/98	1	0	+9 (V)	HHV5gp025	NE
UL26	222/189	1	0	+34 (M)	tegument protein	A
UL27	629/608	8	2	+21 (L)	HHV5gp033	A
UL28	369/341	1	0	+28 (V)	hypothetical protein	A
UL29	360/360	0	0	0	US22 family member	A
UL30	121/121	1	0	0	HHV5gp036	A
UL31	682/595	3	0	+87 (L)	Transcription	NE
UL32	1049/1048	13	8	+2; -1; -1; +1	pp150 tegument protein	E
UL35	641/640	0	1	+1	UL25 family member	NE
UL36	583/476	3	2	+107 (L)	cell death suppressor	NE
UL40	221/221	5	4	0	HHV5gp045	NE
UL41	141/141	1	0	0	unnamed	NE
UL41A	141/141	0	0	0	unnamed	NE
UL42	122/125	4	2	-1; -2	HHV5gp047	NE
UL43	431/423	1	1	+8 (L)	Tegument	NE
UL44	455/433	0	0	+22 (V)	DNA replication	E
UL45	906/906	3	1	0	ribonucleosid-diphosphat reductase	NE
UL46	310/290	0	0	+20 (V)	capsid protein	E
UL47	983/983	5	0	0	Tegument DNA release	A
UL48	2347/2241	16	2	+105 (L)	Tegument	A
UL48.5	75/75	0	0	0	capsid protein	E
UL49	599/570	3	4	+29 (L)	HHV5gp055	E

ORF	Länge in Aminosäuren Toledo/AD169	Austausch funktional ähnlicher Aminosäuren	Austausch funktional unähnlicher Aminosäuren	Deletionen/Insertionen	Vermutete Funktion	Funktion des ORFs von AD169 für den Wachstum in HFFs
UL50	398/397	1	0	+1	Egress	E
UL51	157/157	1	0	0	DNA packaging/cleavage	E
UL52	667/668	3	0	-1	DNA packaging/cleavage	E
UL53	376/376	1	0	0	Egress	E
UL69	765/749	1	1	+26 (V); -5	Tegument	A
UL70	953/946	1	1	+7 (L)	Helicase-primase complex	E
UL71	403/370	1	0	+42 (L)	HHV5gp066 tegument protein	A
UL72	388/388	3	0	0	dUTPase	A
UL73	164/138	20	9	+29 (L); -2; -1	structural glycoprotein gpUL73	E
UL74	464/466	18	50	-1; -2; +1	glycoprotein O	A
UL75	743/743	2	0	0	envelope glycoprot. H	E
UL76	325/325	5	0	0	HHV5gp071	A
UL77	658/642	7	1	+10 (L)	DNA packaging/cleavage	E
UL78	434/431	5	4	+3 (V)	G protein receptor	NE
UL84	625/587	6	3	+39 (M); -1	DNA replication	E
UL85	306/306	0	0	0	capsid protein	E
UL88	429/429	0	0	0	Tegument	NE
UL89 <sup>#</sup>	759/674	2	1	+115 (L)	conserved herpesvirus spliced gene	E
UL91	121/111	0	1	+10 (V)	HHV5gp085	E
UL92	321/201	0	0	+120 (V)	HHV5gp086	E
UL93	594/594	11	1	0	HHV5gp087	E
UL94	458/345	0	0	+113 (V)	HHV5gp088	A
UL95	531/531	5	1	0	tegument protein	E
UL96	174/127	0	1	+47 (L)	HHV5gp090	A
UL97	722/707	2	1	+15 (V)	phosphorylates ganciclovir	A
UL98	607/584	2	1	+23 (L)	Alkaline nuclease	E
UL99	190/190	3	0	0	myristylated tegument protein	E
UL100	392/372	1	1	+20 (L)	glycoprotein M	E
UL101 <sup>*</sup>	102/115	4	1	-13 (M)		
UL102	873/873	0	0	0	component of DNA helicase-primase	E
UL103	255/249	0	0	+6 (V)	tegument protein; HSV-1 UL7 counterpart	A
UL104	697/697	0	1	0	DNA packaging protein	E



ORF	Länge in Aminosäuren Toledo/AD169	Austausch funktional ähnlicher Aminosäuren	Austausch funktional unähnlicher Aminosäuren	Deletionen/Insertionen	Vermutete Funktion	Funktion des ORFs von AD169 für den Wachstum in HFFs
UL108*	112/123	0	0	-11 (V)		
UL109*				no ORF present		
UL110*				no ORF present		
UL111*				no ORF present		
UL111A#			only first exon detected		viral interleukin 10	NE
UL112*	268/268	0	0	0	early phosphoprotein p34	A
UL113*	484/499	5	1	-15 (V)	Major early protein	A
UL112/113#	684/684	4	1	0	HHV5gp100	A
UL114	250/250	1	0	0	uracil-DNA glycosylase	A
UL115	278/278	2	0	0	glycoprotein L precursor	E
UL116	303/314	22	4	-13; -2; -1; +2; +1	HHV5gp103	NE
UL120	197/201	17	10	-4	HHV5gp106	NE
UL121	181/180	6	5	+1 (V)	membrane protein	NE
US3	186/186	1	3	0	Immunomodulation	NE
US4*				no ORF present		
US5*	127/126	2	2	+1		NE
US6	183/183	4	1	0	inhibits TAP-med. pep. Translocation	NE
US7	225/225	10	3	0	glycoprotein	NE
US8	227/227	4	1	0	Immunomodulation	NE
US9	254/247	3	0	+23 (V); -15	glycoprotein	NE
US10	158/158	4	0	0	glycoprotein	NE
US11	215/215	2	0	0	targets MHC class I molecules...	NE
US12	281/281	2	1	0	multiply hydrophobic protein	NE
US13	261/261	0	0	0	multiply hydrophobic protein	NE
US14	310/310	4	6	0	HVLF4 protein	NE
US15	385/262	1	1	+123 (L)	multiply hydrophobic protein	NE
US16	309/309	1	0	0	multiply hydrophobic protein	NE
US17	293/293	1	3	0	multiply hydrophobic protein	NE
US18	274/274	0	0	0	multiply hydrophobic protein	NE
US19	240/240	2	2	0	multiply hydrophobic protein	NE
US20	346/342	2	0	+4	Transmembrane protein HWLF3	NE

ORF	Länge in Aminosäuren Toledo/AD169	Austausch funktional ähnlicher Aminosäuren	Austausch funktional unähnlicher Aminosäuren	Deletionen/Insertionen	Vermutete Funktion	Funktion des ORFs von AD169 für den Wachstum in HFFs
US21	243/239	1	0	+4	multiply hydrophobic protein	NE
US22	576/576	0	2	0	tegument protein	NE
US23	597/592	0	0	+6 (L); -1	HHLF7 protein	A
US24	501/501	0	1	0	US22 family member	A
US25*	121/179	5	0	-43 (L)	hypothetical protein US25	NE
US25A	107/156	1	0	-26 (L)	hypothetical protein HHRF1	NE
US26	613/603	3	0	+10 (L)	HHV5gp143	A
US27	364/362	16	0	+2	G protein receptor	NE
US28	387/354	2	0	+33 (V)	G protein receptor	NE
US29	462/462	8	0	0	HHV5gp146	NE
US30	363/349	14	3	+14 (V)	HHV5gp147	NE
US31	193/161	0	0	+32 (L)	HHV5gp148	NE
US32	183/183	0	0	0	HHV5gp149	NE
ORF1	83/83	0	0	0	-	-
ORF2	105/105	2	0	0	-	-
ORF6	123/123	3	0	0	-	-
ORF7	185/193	5	1	-9 (V); +1	-	-
ORF8	165/128	20	4	+37 (V)	-	-
ORF9	384/305	4	2	+79 (L)	-	-
ORF10	237/141	0	4	+96 (L)	-	-

