

Uniklinik Ulm
Institut für Humangenetik
Prof. Dr. med. Walther Vogel

VARIABILITÄTEN IM PROMOTORBEREICH DES GROWTH HORMONE 1-GENS BEI PRIMATEN

Dissertation
zur Erlangung des Doktorgrades der Medizin
der Medizinischen Fakultät der Universität Ulm

von Dorothea Kumpf
Geboren in Backnang

2008

Amtierender Dekan: Prof. Dr. Thomas Wirth
1. Berichterstatter: PD Dr. Kehrer-Sawatzki
2. Berichterstatter: Prof. Dr. Wellinghausen
Tag der Promotion: 18.11.2010

Meinen Eltern

Inhaltsverzeichnis

Inhaltsverzeichnis.....	I
Abkürzungen.....	II
1 Einleitung.....	1
1.1 Evolution des Somatotropins.....	1
1.2 <i>GH1</i> -Gen des Menschen.....	4
1.3 Aufgabenstellung.....	6
2 Material und Methoden.....	8
2.1 Material.....	8
2.2 Methoden.....	13
3 Ergebnisse.....	20
3.1 Promotorregion des <i>cGH1</i> -Gens des Schimpansen.....	21
3.2 Promotorregion des <i>mGH1</i> -Gens des Makaken.....	26
3.3 Vergleich der <i>GH1</i> -Promotorbereiche von Mensch, Schimpanse und Makake.....	28
3.4 Organisation der <i>GH</i> -Genfamilien von Mensch, Schimpanse und Makake.....	29
4 Diskussion.....	37
4.1 Diversität im Promotorbereich des <i>GH1</i> -Gens bei Schimpanse und Makake:.....	37
4.2 Evolution der <i>GH</i> -Gencluster von Mensch und Schimpanse.....	41
5 Zusammenfassung.....	50
6 Literaturverzeichnis.....	52
Anhang.....	56

Abkürzungen

A	Adenin
bp	Basenpaare
c-	chimpanzee, Schimpansen-
C	Cytosin
cen	centromerwärts
Chi	Cross-over hotspot instigator
Chr.	Chromosom
CSH	Chorionic Somatomammotropin Hormone
CSHL1	Chorionic Somatomammotropin Hormone-Like 1
dATP	Desoxyadenosintriphosphat
dCTP	Desoxycytidintriphosphat
dGTP	Desoxyguanosintriphosphat
DNA	desoxyribonucleic acid, Desoxyribonukleinsäure
dNTPs	Desoxyribonukleosidtriphosphate
dTTP	Desoxythymidintriphosphat
EDTA	Ethylendiamintetraessigsäure
ET	Elongationszeit
EtOH	Ethanol
f	weiblich
G	Guanin
GCG	Genetics Computer Group
GH	Growth Hormone, Somatotropin
GHDTA	Growth Hormone-gene derived transcriptional activator
GTP	Guanosintriphosphat
h-	human, menschliches-
H ₂ O	Wasser
HCl	Salzsäure
HGNC	HUGO Gene Nomenclature Committee
HPLC	High pressure liquid chromatography
HSA	<i>Homo sapiens</i> , Mensch
kb	Kilobase
kbp	Kilobasenpaare
KCl	Kaliumchlorid
m-	macaque, Makaken-
m	männlich
M	Mol
MgCl ₂	Magnesiumchlorid
MMU	<i>Macaca mulatta</i> , Rhesusaffe, Makake

NaAc	Natrium-Acetat
NCBI	National Center for Biotechnology Information
OD	optische Dichte
P	Puffer
PCR	polymerase chain reaction, Polymerase-Ketten-Reaktion
PL	Plazentares Laktogen
<i>PTR</i>	<i>an troglodytes</i> , (Gemeiner) Schimpanse
RNA	ribonucleid acid, Ribonukleinsäure
rpm	runds per minute, Umdrehungen pro Minute
SNP	Single Nucleotide Polymorphism
STH	somatotropes Hormon
T	Thymin
TAE	TRIS-EDTA-Acetat
Taq	<i>Thermus aquaticus</i>
TBE	TRIS-Borsäure-EDTA
TE	TRIS-EDTA
tel	telomerwärts
Temp.	Temperatur
TRIS	Trishydroxymethylaminomethan
U	Unit
UCSC	University of California Santa Cruz
UV	Ultraviolett
V	Volt
λ	Wellenlänge

1 Einleitung

Das Somatotropin, oder auch „Growth Hormone“ (GH) genannt, ist ein Wachstumshormon, das im Hypophysenvorderlappen produziert wird. Es gehört zu einer Proteinfamilie zusammen mit Prolaktin, plazentaren Laktogenen und weiteren in Hypophyse und Plazenta synthetisierten Hormonen (Wallis 1994). Die meisten Wirbeltiere bilden Somatotropin, das in der Regel aus ca. 190 Aminosäuren besteht (George et al. 1981, Goffin et al. 1996) und eine wichtige Rolle in der Entwicklung spielt. Seine Funktionen sind unter anderem die Stimulation von Wachstum und Regulation des Metabolismus von Muskel-, Knochen- und Knorpelzellen. Plazentare Laktogene (PL; synonym dazu verwendet wird CSH für Chorionic Somatomammotropin Hormone) finden sich in der Plazenta einiger Primaten und haben vor allem die Aufgabe, den maternalen Stoffwechsel an die Anforderungen des Feten anzupassen (Golos et al. 1993, Li et al. 2005).

1.1 Evolution des Somatotropins

Die Veränderungen des Somatotropins über die Jahrtausende und die speziellen Entwicklungen in unterschiedlichen Tierordnungen und -arten sind schon vielfach untersucht worden (Chen et al. 1989, Wallis 1981, 1994, 1996, 2008, Adkins et al. 2001, Revol de Mendoza et al. 2004). Durch den Vergleich der Aminosäuresequenzen des hypophysären Wachstumshormons zahlreicher Säugetiere war es möglich, eine Art „Ur-GH“ des gemeinsamen Vorfahren der Säugetiere zu rekonstruieren. Die Hormone, die die heutigen Säuger produzieren, stammen alle von dieser ancestralen Version ab (Wallis 1994). Die Kenntnis dieses Ur-Wachstumshormons erlaubt die genauere Analyse der molekularen Entwicklung des Hormons.

Die Weiterentwicklung des Somatotropins seit diesem Ur-Wachstumshormon der Säugetiervorfahren verlief im Allgemeinen sehr langsam. Im Durchschnitt veränderte sich eine Aminosäure in fünf Milliarden Jahren. So bilden Schwein und Hund beispielsweise heute immer noch das ursprüngliche Peptidhormon. (Wallis

1994, Wallis 2008). Das von Mensch und Rhesusaffe (= Makake, ein Vertreter der Altweltaffen) gebildete Hormon hingegen weicht mit 62 bzw. 64 ausgetauschten Aminosäuren ganz besonders ab (Wallis 1994, 1996). Denn es gab mindestens zwei Etappen, in denen die Aminosäuresubstitutionen mit 25 bis 50-facher Geschwindigkeit abliefen: Eine in der Entwicklung der Paarhufer; die andere bei den Primaten, nach Abzweigung der Tarsier, die zwischen Halbaffen und Affen anzusiedeln sind (*Abbildung 1*). Noch vor der Divergenz der Neuweltaffen von Altwelt- und Menschenaffen war diese ereignisreiche Phase abgeschlossen, und die „alte“ Geschwindigkeit wurde wieder eingenommen (Liu et al. 2001, Wallis 1996, Wallis et al. 2001). So lässt sich erklären, warum das hypophysäre GH von Halbaffen und Tarsiern noch dem von Nicht-Primaten ähnelt, während das Somatotropin der Neuweltaffen dem von Rhesusaffe, Schimpanse und Mensch gleicht (Wallis et al. 2001, Li et al. 2005).

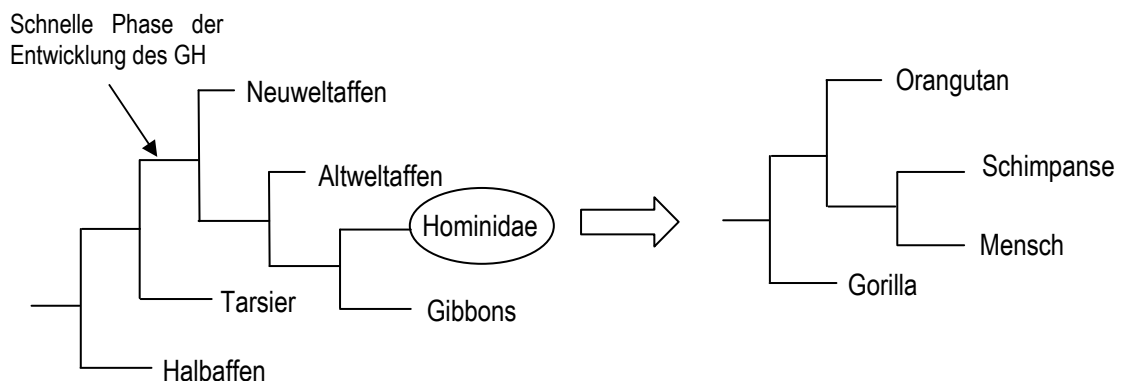


Abbildung 1: zwei phylogenetische Bäume; links: Evolution der Primaten, mit einem Pfeil markiert die schnelle Phase der Entwicklung des Growth Hormone (GH) nach dem Abzweigen der Tarsier; rechts: detaillierterer Stammbaum der heutigen Menschenaffen (Hominidae)

Unter den Nicht-Primaten verfügen fast alle Säuger nur über ein einziges GH(-ähnliches)-Protein, außerdem über Prolaktin und Prolaktin-ähnliche plazentare Hormone, nicht aber über plazentare Laktogene (Wallis 1992, Soares 2004). Bei den Primaten ist es unterschiedlich – während Halbaffen (Prosimier) wie Bushbaby und Pluimplori ebenfalls nur ein hypophysäres *GH*-Gen besitzen (Adkins et al. 2001, Wallis et al. 2001), haben die Affen und Menschenaffen durch Duplikationen einen komplexeren Genlocus entwickelt. Hier liegt das Gen für das Somatotropin zusammen mit weiteren *GH*-Duplikaten vor: vier beim Menschen

und fünf bei Schimpanse und Rhesusaffe (Chen et al. 1989, Wallis et al. 2001; Wallis u. Wallis 2002, Revol de Mendoza et al. 2004, Gonzalez Alvarez et al. 2006). Bei Kapuzineraffen, die zu den Neuweltaffen gehören, besteht die *GH*-Genfamilie sogar aus über 40 Mitgliedern, sowohl funktionellen Genen, als auch Pseudogenen (Wallis u. Wallis 2006).

Die Duplikationen des ursprünglichen *GH*-Gens, aus welchen die weiteren Mitglieder der *GH*-Familie hervorgingen, geschahen nach der schnellen Entwicklungsphase des *GH*-Gens, doch ereigneten sie sich bei den Neuweltaffen unabhängig von denen der Altwelt- und Menschenaffen (Wallis et al. 2001, Li et al. 2005, Gonzalez Alvarez et al. 2006).

Die Entstehung von Genfamilien ist ein häufig beobachteter Vorgang während der Genom-Evolution von Säugern. Die Gründe für die Erweiterung des *GH*-Genlocus durch Duplikationen sind noch etwas umstritten, es ist jedoch anzunehmen, dass Duplikate bereits vorhandener Gene neue Funktionen übernehmen können, was von adaptivem Vorteil für den jeweiligen Träger sein kann (Wallis et al. 2001). Man geht bei der Duplikation von Genen von einem „birth-and-death“-Prozess aus. Demnach können manche Gen-Duplikationen, als funktionelle Gene behalten werden, während andere zu Pseudogenen degeneriert werden, oder aber durch nachträgliche Deletionen wieder aus dem Genom verschwinden (Ye et al. 2005).

Bei Mensch, Schimpanse und Makake sind jeweils zwei *GH*-Gene bekannt. Der Mensch besitzt außerdem drei *CSH*-Gene, eines davon ein Pseudogen. Im Gegensatz dazu haben Rhesusaffe und Schimpanse vier funktionelle *CSH*-Varianten. Nicht nur die *GH*-Gene entstammen wohl einer phylogenetischen Linie, sondern auch die *CSH*-Gene der drei Spezies scheinen einen gemeinsamen Vorgänger zu haben (Chen et al. 1989, Golos et al. 1993, Wallis 1996, Revol de Mendoza et al. 2004, Gonzalez Alvarez et al. 2006).

1.2 GH1-Gen des Menschen

1.2.1 Menschlicher GH-Genlocus

Das hypophysäre GH des Menschen, auch bekannt als Somatotropin, somatotropes Hormon (STH) oder pituitary growth hormone, wird offiziell als hGH1 (für englisch: human Growth Hormone 1) abgekürzt. Es besteht aus 191 Aminosäuren und wird in den α -Zellen des Hypophysenvorderlappens gebildet. Der GH-Genlocus befindet sich auf dem langen Arm von Chromosom 17q22 bis 24 und umfasst insgesamt fünf strukturell verwandte Gene, aus jeweils fünf Exons und vier Introns, die alle mit der selben Orientierung auf dem DNA-Strang liegen. Am 5'-Ende liegt das *hGH1*, dann folgen der Reihe nach *hCSHL1*, *hCSH2*, *hGH2* und *hCSH1* (George et al. 1981, Harper et al. 1982, Chen et al. 1989).

Trotz der sehr hohen Homologie (über 90%) der Nukleotidsequenzen der Gene untereinander werden sie sehr gewebsspezifisch exprimiert (siehe *Abbildung 2*). Bis auf das hypophysäre hGH1-Hormon werden alle Peptide ausschließlich in der Plazenta produziert (Nachtigal et al. 1993).

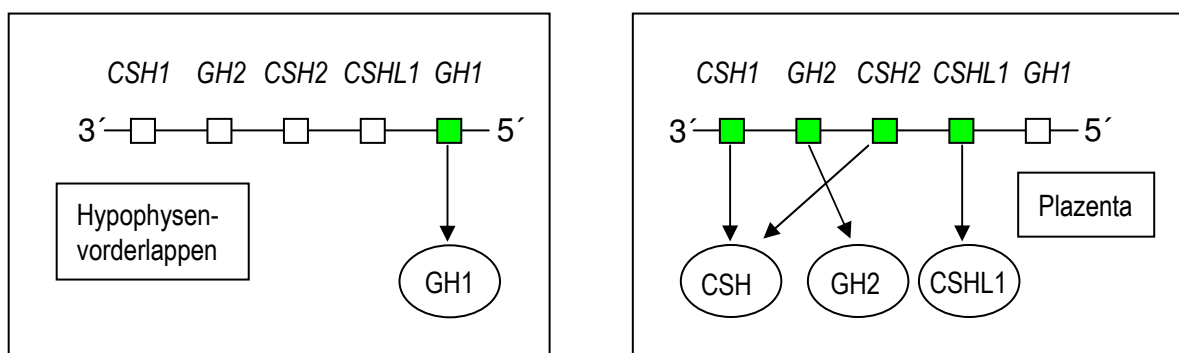


Abbildung 2: Schematische Darstellung des menschlichen GH-Genlocus und der Expression der fünf Gene. Anordnung der Gene (als Rechtecke angedeutet) auf dem Minusstrang von 3' nach 5', grün markiert jeweils exprimierte Gene im Hypophysenvorderlappen (links) beziehungsweise in der Plazenta (rechts). Genbezeichnungen sind im Gegensatz zu den Proteinen kursiv gedruckt. Die Pfeile deuten auf das jeweilige Produkt hin. CSH(L) = Choric Somatotropin Hormone (-like), GH = Growth Hormone (Somatotropin).

Das *hGH2*-Gen ist eine Variante des *hGH1*, die im Synzytiotrophoblasten exprimiert wird und für ein mütterliches Protein steht, das während der Schwangerschaft als GH im Serum überwiegt. Außerdem liegen hier das *hCSH1*-Gen und das *hCSH2*-Gen, die beide für dasselbe Hormon humanes plazentares Laktogen (hPL) oder auch hCSH (human Chorionic Somatomammotropin

Hormon) kodieren und ein Pseudogen *hCSHL1* (L steht für like). Vor allem in der zweiten Hälfte der Schwangerschaft werden *hCSH1* und *hCSH2* exprimiert und erfüllen ihre Funktion als Haupt-Plazenta-Laktogene des Menschen. *HCSHL1* scheint ein Pseudogen zu sein, obwohl es zahlreich in der Plazenta transkribiert wird (Wallis 1996, Giordano et al. 1997).

Bei der näheren Analyse des *hGH1* fiel auf, dass die oben genannten schnellen Veränderungen während der Evolution des *GH* von (höheren) Primaten spezifisch in der kodierenden Region stattgefunden haben. Sowohl das Signalpeptid als auch die 5'-untranslatierte Region oder die Introns waren davon nicht betroffen (Wallis et al. 2001, Wallis 2008).

1.2.2 *GH1*-Promotorbereich des Menschen

Trotzdem zeigt der Promotorbereich des menschlichen *hGH1* sehr interessante Eigenschaften. Er enthält mit 16 Einzelbasen-Polymorphismen (SNPs = single nucleotide polymorphisms) innerhalb einer Region von 535 Basenpaaren etwa 100-mal so viele Polymorphismen wie andere nicht-kodierende Stellen im menschlichen Genom (Giordano et al. 1997).

SNPs sind Abweichungen der DNA an einzelnen Nukleotidpositionen, die innerhalb derselben Spezies gefunden werden, zum Beispiel im Vergleich zwischen einem väterlichem und dem entsprechenden mütterlichen Chromosom eines Individuums. Die Häufigkeit jedes der sogenannten Allele, also der möglichen Base an der betreffenden Stelle, muss über 1% in der Bevölkerung betragen, damit die Variante als Polymorphismus eingestuft wird.

Die vielen Polymorphismen, die vor allem im proximalen Promotorbereich des *hGH1* lokalisiert sind, beeinflussen vermutlich die Stärke der Genexpression (Horan et al. 2003). Fast alle SNPs im Promotorbereich finden sich an Stellen, an denen sich das *hGH1* von einem oder mehreren der verwandten Genen unterscheidet, manchmal trifft das sogar auf ganze Cluster von Nukleotidsequenzen zu (Giordano et al. 1997, Horan et al. 2003). Man glaubt deshalb, dass die Polymorphismen durch Basenübertragung von den *CSH*-Genen auf das *GH1*-Gen entstanden sind. Solch eine „Übertragung von Basen“ erfolgt

während der Rekombination und wird als Genkonversion bezeichnet. Genkonversion ist zwischen sehr ähnlichen Sequenzen möglich, und in der Tat sind die Mitglieder der *GH/CSH*-Familie sehr homolog zueinander (s.o.). Genkonversion kann einerseits die Diversität innerhalb des Genoms erhöhen, indem sich beispielsweise mutierte Basen weiter ausbreiten, andererseits kann ein älteres Allel nach und nach auch ganz verloren gehen.

Die Theorie der hohen Rekombinationsaktivität im Promotorbereich des *hGH1*-Gens wird außerdem durch das Vorhandensein einer Chi-ähnlichen Sequenz untermauert (Chi = cross-over hotspot instigator). Chi- und Chi-ähnliche Sequenzen wurden bei Bakterien und Hefen ausgiebig untersucht und als bevorzugte Stellen für Rekombinationsereignisse, dem reziproken Austausch von Erbmaterial während der Meiose erkannt (Giordano et al. 1997).

Neben der hohen Diversität ist eine weitere Besonderheit des proximalen Promotorbereichs des *hGH1*-Gens, dass er nicht nur die Funktion als Promotor erfüllt, sondern gleichzeitig die kodierende Region eines anderen Gens, *GH-gene derived transcriptional activator (GHDTA)*, unbekannter Funktion enthält (Labarriere et al. 1995).

1.3 Aufgabenstellung

Die Diversität im proximalen Promotorbereich des *GH1*-Gens des Menschen wirft viele Fragen auf. So bleibt zu klären, wie spezifisch die Diversität ist, also ob sie nur im *GH1*-Promotor des Menschen beobachtet wird, oder ob sie auch bei nicht humanen Primaten entdeckt werden kann.

Meine Aufgabe war es deshalb, den *GH1*-Promotorbereich von Rhesusaffen (*Macaca mulatta*), als Vertreter der Altweltaffen, und von Schimpansen (*Pan troglodytes*), den evolutionär gesehen nächsten Verwandten des Menschen, zu sequenzieren und nach SNPs zu forschen. Die Promotorbereiche sollten sowohl innerhalb einer Spezies auf Diversität hin untersucht, als auch mit der menschlichen Sequenz verglichen werden.

Eine zweite Aufgabe meiner Dissertation war es dann, die Organisation des *GH*-Genlocus von Mensch, Schimpanse und Makake zu vergleichen. Auf den ersten

Blick schienen die Genloci der drei Spezies recht ähnlich, doch die in öffentlichen Datenbanken zur Verfügung stehenden Versionen der Primatengenome, wie auch die Publikationen zur genomischen Organisation des *GH*-Genlocus bei Schimpansen sind lückenhaft. So ist der Aufbau des *GH*-Locus beim Schimpansen noch unbekannt, seine Rekonstruktion war Aufgabe meiner Arbeit, die dazu verhelfen sollte, die Entstehung des *GH*-Genclusters bei den drei Spezies Mensch, Schimpanse und Makake näher zu beleuchten.

Zusammengefasst waren im Rahmen meiner Dissertation also folgende zwei Punkte zu bearbeiten:

1. Wie sehen die *GH1*-Promotorbereiche von Schimpanse und Makake im Vergleich zu dem des Menschen aus? Finden sich auch Polymorphismen?
2. Wie sind die *GH*-Genloci der drei Primaten organisiert?

2 Material und Methoden

2.1 Material

2.1.1 Chemikalien

Agarose	Roth, Karlsruhe
Betain	Fluka, Steinheim
Borsäure	AppliChem, Darmstadt
Bromphenolblau	Merck, Darmstadt
Desoxyribonukleosidtriphosphate (dNTPs)	MBI, St. Leon-Rot
Dimethylformamid	Roth, Karlsruhe
Ethanol (EtOH)	Merck, Darmstadt
Ethidiumbromid	Roth, Karlsruhe
Ethylendiamintetraessigsäure (EDTA)	Serva, Heidelberg
HIDI-Puffer	Applied Bioscience
HPLC-H ₂ O	Applied Bioscience
Kaliumchlorid (KCl)	Merck, Darmstadt
Magnesiumchlorid-Hexahydrat	Merck, Darmstadt
Natriumacetat (NaAc)	Merck, Darmstadt
Salzsäure (1M)	J.T. Baker, Deventer (Niederlande)/ Merck, Darmstadt
Trishydroxymethylaminomethan (TRIS)	USB, Cleveland
Xylencyanol	Sigma-Aldrich, Steinheim

2.1.2 Lösungen und Puffer

ABI-Gebrauchspuffer	2,5 ml 10x ABI-Puffer mit EDTA (Applied Biosystems) ad 25,0 ml HPLC-H ₂ O
dNTP-Stammlösung für PCR	1,25 mM dATP

	1,25 mM dCTP
	1,25 mM dGTP
	1,25 mM dTTP
dNTP-Stammlösung für Expand Long	20,0 mM
Template PRC	
Ethidiumbromid-Stammlösung	10,0 mg/ml Ethidiumbromid
Ladepuffer	50% Glycerin
	0,25% Bromphenolblau
	0,25% Xylencyanol
	10,0 mM EDTA, pH 8
1 kb-Leiter	100,0 µl TE
	100,0 µl H ₂ O bidest
	10,0 µl 1 kb-Ladder
	25,0 µl Ladepuffer
PCR-Puffer 15 (10x)	500,0 mM KCl
	100,0 mM TRIS/HCl, pH 8,3
	15,0 mM MgCl ₂
PCR-Puffer 20 (10x)	500,0 mM KCl
	100,0 mM TRIS/HCl, pH 8,3
	20,0 mM MgCl ₂
Primer-Stammlösung	100,0 µM
Primer-Gebrauchslösung (Hin- und Rückprimer)	5,0 µM
Primer-Gebrauchslösung für Expand Long Template PCR	5,0 µM
Primer-Gebrauchslösung für ABI	10,0 µM
PCR-Enhancer	6,76 g Betain
	500,0 µl gesättigte TRIS/HCl
	ad 30,0 ml H ₂ O bidest, filtrieren
TAE (50x)	242,0 g TRIS
	57,1 ml Eisessig
	100,0 ml 0,5 M EDTA pH 8
	ad 1,0 l H ₂ O bidest

TBE (5x)	270,0 g TRIS/HCl 137,5 g Borsäure 18,6 g EDTA ad 5,0 l H ₂ O bidest, pH 8,3
TE-Puffer(10/1)	10,0 ml 1,0 M TRIS/HCl, pH 8,3 2,0 ml 0,5 M EDTA ad 1,0 l H ₂ O bidest., pH 8,3

2.1.3 Standards und Enzyme

Taq-Polymerase	Amersham Bioscience, Freiburg
1 kb Plus DNA Leiter	Invitrogen

2.1.4 Reagentiensysteme, Kits

ABI PRISM® BigDye™ Terminator Cycle Sequencing Ready Reaction Kit	Applied Biosystems, Foster City USA
Expand Long Template PCR System	Roche Diagnostics
GFX™ PCR DNA and Gel Band Purification Kit	Amersham Pharmacia Biotech

2.1.5 Laborgeräte, Hilfsmittel

ABI PRISM® 3100 Genetic Analyzer	Applied Biosystems
Electrophorese Power PAC 300	Bio Rad, München
Gel-Dokumentationssystem CS 1	Cybertech, Berlin
GeneAmp PCR System 9600	Perkin&Elmer
Gene Quant, RNA/DNA Calculator	Pharmacia, Cambridge UK
Mastercycler gradient	Eppendorf
Zentrifugen	Eppendorf, Heraeus

2.1.6 Herkunft der untersuchten DNA-Proben

Zur Verfügung standen die DNA-Proben von 25 Rhesusaffen (*Macaca mulatta*) und 16 Schimpansen (*Pan troglodytes*).

Die DNA-Proben der 16 unverwandten Schimpansen (Liste in *Tabelle 1*) lagen im Labor der Arbeitsgruppe bereits vor, sie wurden dankenswerterweise zur Verfügung gestellt von Prof. Dr. Werner Schempp, Abteilung Humangenetik der Universität Freiburg.

Tabelle 1: Zusammenfassung der 16 DNA-Proben von Schimpansen; In der ersten Spalte stehen die kodierten Bezeichnungen der verwendeten Schimpansen-DNA-Proben, in der zweiten das Geschlecht (m – männlich, f – weiblich). Falls bekannt wurde noch der Name des Individuums notiert.

Schimppanse	Geschlecht	Name (falls bekannt)
PTR 8	f	Chita
PTR 46	f	Mixi
PTR 86	m	
PTR 88	f	
PTR 89	f	
PTR 90	f	
PTR 91	m	
PTR 92	m	
PTR 94	f	
PTR 96	f	Charlotte
PTR 133	?	Katasche
PTR 519	f	Babsy
PTR 582	f	Epona
PTR 694	f	Josi
PTR 828	f	Exota
PTR 835	f	Susi

In *Tabelle 2* sind die Kodierungen der 25 Makaken-Individuen aufgelistet, deren DNA mir vorlag. Bis auf die Halbgeschwister 09G und 10G (gleicher Vater i47) stammen alle analysierten Makaken von verschiedene Eltern. Auch diese DNA-Proben wurden von Prof. Dr. Werner Schempp freundlicherweise zur Verfügung gestellt.

Tabelle 2: Zusammenfassung der 25 DNA-Proben von Makaken; In der ersten Spalte stehen die kodierten Bezeichnungen der verwendeten Makaken-DNA-Proben. In den folgenden drei Spalten stehen Geschlecht (m – männlich, f – weiblich) und die Kodierungen der Eltern.

Makake	Geschlecht	Mutter	Vater
00A	m	971	H86
00B	f	E65	H87
00G	m	57A	E94
01F	m	I70	K32
02D	m	D98	717
04A	f	O56	742
04F	m	G43	E94

Makake	Geschlecht	Mutter	Vater
05C	m	678	717
05D	f	L97	E05
05E	m	931	D71
06A	f	N16	i85
06D	f	L37	O34
06E	f	V58	L01
06G	m	G43	650
07A	f	J89	F32
07C	f	O35	L34
07F	f	N13	G37
08A	m	C12	E54
08F	f	J28	O81
08G	f	56A	B09
09G	f	Z11	i47
10G	f	X86	i47
11H	f	K80	O23
12D	m	G11	L35
13C	f	O56	J26

2.1.7 Elektronische Datenbanken und Software:

Ensembl-Datenbank

<http://www.ensembl.org>

Datenbank des NCBI (National Center for Biotechnology Information)

<http://www.ncbi.nlm.nih.gov>

Datenbank der UCSC (University of California Santa Cruz)

<http://www.genome.ucsc.edu>

Alle drei Datenbanken nutzte ich zum inter- und intraspeziellen Vergleich der untersuchten Sequenzen. Außerdem konnte ich schon vorhandene Sequenzen von Interesse anhand deren Zugangscodes herunterladen.

ClustalW: Das ClustalW Sequenzvergleich-Programm erlaubt den Vergleich mehrerer Sequenzen untereinander durch Erstellen eines sogenannten Multisequenz-Alignments. Außerdem errechnet das Programm phylogenetische Bäume der zu analysierenden Sequenzen und stellt sie graphisch dar.

Genetics Computer Group: Die FASTA-Software des Programmpaketes GCG 10.2 ermöglicht Vergleichsanalysen von Nukleotidsequenzen und im Falle von

sequenzierten PCR-Produkten die korrekte Fusion von Hin- und Rücksequenz. (Genetics Computer Group, University of Wisconsin, Madison, USA)

RepeatMasker: Der RepeatMasker ist ein Programm, das Nukleotidsequenzen auf das Vorkommen von repetitiven Sequenzen überprüft.

2.2 Methoden

2.2.1 Photometrische DNA Konzentrationsbestimmung

Mit Hilfe des Pharmacia Genequant Photometer wurden die DNA-Konzentrationen nach dem Lambert Beer'schen Gesetz bei einer Wellenlänge $\lambda=260$ nm bestimmt. Die Extinktion bei dieser Wellenlänge wird durch RNA, DNA und Oligonukleotide verursacht, während Proteine für die Abschwächung bei $\lambda=280$ nm verantwortlich sind. So kann über Berechnung des Quotienten $OD_{260} \text{ nm}/OD_{280} \text{ nm}$ auf die Reinheit der Probe geschlossen werden. Im Idealfall liegt der Wert zwischen 1,8 und 2,0. Werte unter 1,6 geben Hinweis auf eine Kontamination mit Proteinen, Ergebnisse über 2,0 auf einzelsträngige DNA

2.2.2 Polymerase-Ketten-Reaktion (PCR)

Für die Entwicklung dieser Methode zur in-vitro-Amplifikation spezifischer DNA-Sequenzen wurde Karl Mullis 1993 mit dem Nobelpreis für Chemie ausgezeichnet. Charakteristisch sind der zyklische Ablauf dreier Abschnitte: Denaturierung, Annealing und Elongation.

Während der Denaturierung bei 94 °C werden die DNA-Stränge der Templates zu Einzelsträngen getrennt. Nur so können im nachfolgenden Annealingschritt die Primer spezifisch binden. Bei dieser Stufe herrschen ca. 55 °C, was je nach G/C-Gehalt des Primerpaars variiert werden sollte, da die Spezifität der PCR davon stark beeinflusst wird. Die Primer werden so gewählt, dass sie auf Plus- und Minusstrang liegend, jeweils mit ihrem 3'-Ende in Richtung Zielsequenz zeigen. Generell funktioniert die PCR leichter, wenn der zu amplifizierende Abschnitt kurz ist (unter 800 bp). Die Elongation findet beim Temperaturoptimum der hitzestabilen Taq-Polymerase, 72 °C statt. Gewonnen wird diese DNA-Polymerase aus

einem in heißen Quellen vorkommendem Bakterium namens *Thermus aquaticus*. Ihre Aufgabe ist es, an die Primer anzuknüpfen und den restlichen Einzelstrang des Templates wieder zum Doppelstrang zu vervollständigen. Pro Zyklus verdoppelt sich also die Template DNA. Die Dauer der Elongationszeit steigt mit der Produktgröße und muss somit ebenfalls dem Einzelfall angepasst werden.

Beispiel für ein Standard-PCR-Programm:

1. Initiale Denaturierung: 5 min bei 94 °C
2. 35 s Denaturierung bei 93 °C
3. 1 min Annealing bei 55 °C
4. 1 min Elongation bei 72 °C

34 Zyklen von Schritt 2 – 4

Beendigung der PCR mit einer längeren Elongationsphase, z.B. 5 min

Mein gewählter Standardansatz betrug insgesamt 25 µl und setzte sich aus den in *Tabelle 3* aufgeführten, in dieser Reihenfolge pipettierten Reagentien zusammen:

Tabelle 3: Standardansatz für PCR (dNTPs = Desoxyribonucleosidtriphosphate, P = Puffer, M = Mol, U = Unit)

Reagenz	Pipettiervolumen pro Reaktion	Konzentration der Stammlösung
PCR-Enhancer	15,0 µl	5,0 M
dNTPs	4,0 µl	1,25 mM
P15 bzw. P 20-Puffer	2,5 µl	10-fach
Primermix	2,5 µl	5,0 µM
Taq-Polymerase	0,15 µl	5 U/µl
Template	x µg	100 –200 ng

Eine spezielle Form der PCR, das Expand Long Template PRC System, erlaubt es auch viel größere DNA-Fragmente (bis zu 27 kbp) zu vervielfältigen. Bei dieser PCR wird die thermostabile Taq-Polymerase durch eine weitere Polymerase ergänzt, die eine sogenannte „proofreading“-Funktion vermittelt. Im Rahmen meines Versuches benutzte ich das Kitsystem der Firma Roche.

Bei der Expand Long Template PCR werden folgende Phasen durchlaufen:

1. 2 Min. Denaturierung bei 94 °C
2. 15 s Denaturierung bei 93 °C
3. 1 min Annealing (Temperaturwahl ist abhängig vom Primerpaar)

4. Elongation bei 68 °C, die Dauer ist abhängig von der Größe des zu amplifizierenden DNA-Abschnittes, bei meinen Produkten mit Längen unter 3 kb betrug sie 2 min

34 Zyklen von Schritt 2 bis 4

zum Schluss 5-minütige finale Elongationsphase bei 68 °C

Die Bezeichnung der im Rahmen meiner Arbeit benützten Primer, sowie die Angaben ihrer Sequenzen, Produktgrößen und Annealing-Temperaturen werden in *Tabelle 4*, *Tabelle 5* und *Tabelle 6* zusammengefasst. Dort finden sich auch die jeweils verwendeten Puffer und Elongationszeiten. Je nach Größe des erwarteten PCR-Produkts kam entweder die Standard- oder die Expand Long Template PCR zum Einsatz.

Tabelle 4: Primer zur Amplifizierung des *GH1*-Promotorbereichs der Makaken. Spalten von links nach rechts: Name des Primers, Nukleotidabfolge in 5'-3'-Richtung, Lage auf Chromosom (Chr.) 16 des Makaken (UCSC, MMUL 1.0, Februar 2006), Produktgröße in Basenpaaren (bp), verwendete Temperatur, Puffer (P) und Elongationszeit (ET)

Primer	Sequenz in Richtung 5' - 3'	Lage auf Chr. 16	Produktgröße	Temperatur	Puffer	ET
makgh5	GAAAGGGGAGAGCAAAGTGG	59475130-59475149	730 bp	58°C	P15	1 min
makgh6	CTACAGGGTGCCACCTCTCC	59474422-59474441	730 bp	58°C	P15	1 min

Tabelle 5: Primer für Sequenzierung der Lücke auf Chromosom 17 im Schimpansengenom der Gendatenbanken, eines der placentaren Laktogene liegt vermutlich dort. Spalten von links nach rechts: Name des Primers, Nukleotidabfolge in 5'-3'-Richtung, Lage auf Chromosom 17 (Chr.) des Schimpansen (UCSC, CHIMP 2.1, Mai 2007), Produktgröße in Basenpaaren (bp), verwendete Temperatur, Puffer (P) und Elongationszeit (ET)

Primer	Sequenz in Richtung 5' - 3'	Lage auf Chr 17	Produktgröße	Temperatur	Puffer	ET
dk1	GGGGGCTCATGACACATACA	63187186-63187205	1900 bp	54 °C	P2, P3	2 min
dk2	AGCACCACACTACCTGTTCTG	63188865-63188885	1900 bp	54 °C	P2, P3	2 min
dk3	ATCGCTGTCCGAGGTCATAC	63186953-63186972	2150 bp	54 °C	P1	2 min
dk4	ATGAGGCGGGAGAATCACTT	63189082-63189101	2150 bp	54 °C	P1	2 min
dk5	ACGGTCGTCACTCCATCATC	63187062-63187081	1850 bp (mit dk2)	52 °C, 54 °C, 56 °C, 58 °C	P1, P2	2 min
dk7	ATCCTGAAGGAACGGTCAGT	63187035-63187054	1900 bp (mit dk2)	52 °C, 54 °C, 56 °C, 58 °C	P1, P2	2 min

Tabelle 6: Primer zur Amplifizierung des *GH1*-Promotorbereichs des Schimpansen; Spalten von links nach rechts: Name des Primers, Nukleotidabfolge in 5'-3'-Richtung, Lage auf Chromosom (Chr.) 17 des Schimpansen (UCSC, CHIMP 2.1, Mai 2007), Produktgröße in Basenpaaren (bp), verwendete Temperatur, Puffer (P) und Elongationszeit (ET)

Primer	Sequenz in Richtung 5' - 3'	Lage auf Chr. 17	Produktgröße	Temperatur	Puffer	ET
ptrgh1f	GGTTTGTGTCTCTGCTGCAA	63200862-63200881	790 bp	54-58°C	P20	1 min
ptrgh1r	TGTGCCCAAAGGGATTTAG	63200090-63200109	790 bp	54-58°C	P20	1 min

2.2.3 Agarosegelelektrophorese

Bei neutralem pH-Wert ist DNA negativ geladen. Diesen Umstand macht man sich zur Auftrennung, Identifizierung und Quantifizierung von DNA-Fragmenten zu Nutze. Wenn in der Gelmatrix ein elektrisches Feld angelegt wird, wandert DNA mit zu ihrer Größe umgekehrt proportionaler Geschwindigkeit in Richtung der positiv geladenen Kathode. Kleinere DNA-Fragmente legen in derselben Zeit also weitere Strecken wie größere DNA-Fragmente zurück. Durch den Vergleich der zurückgelegten Entfernung der DNA mit genormten Größenmarkern, die gleichzeitig aufgetragen wurden, lässt sich die Größe des analysierten DNA-Abschnitts abschätzen.

Die Wanderung der DNA-Moleküle hängt außerdem von der angelegten Stromstärke, den Pufferbedingungen und der Agarose-Konzentration ab. Die Elektrophorese ist für Fragmente mit einer Größe von 0,5 bis 12 kb geeignet und wird in 1x TBE-Puffer bei einer Feldstärke von 10 V/cm durchgeführt. Für kleinere Fragmente (100 bp bis 2 kb) benutzt man am besten 1-2%ige Agarosegele während sich größere (1 bis 15 kb) besser in 0,8%iger Agarose auftrennen lassen.

Für meine Analysen habe ich standardmäßig 1,2 %ige Agarosegele hergestellt. Dazu wird Agarose in 1x TBE-Puffer eingewogen und anschließend aufgeköcht. Nach dem Abkühlen auf ca. 60 °C wird die Lösung mit 1 µg/ml Ethidiumbromid versetzt und in horizontale Schlitten gegossen. Durch Kämme, die in die noch flüssige Lösung eingesetzt werden, entstehen beim Erkalten Auftragsaschen im Gel. Das Ethidiumbromid interkaliert mit der DNA-Doppelhelix und fluoresziert bei Bestrahlung mit UV-Licht (254 nm), so dass die aufgetrennten DNA-Fragmente schließlich wieder identifiziert und analysiert werden können.

Um den Erfolg einer PCR zu überprüfen werden die amplifizierte DNA-Proben mit Ladepuffer von 1/10 des Probenvolumens versetzt und anschließend in die Auftragsaschen eingebracht. Im Ladepuffer ist Glycerin enthalten, so dass die Proben in die Auftragsaschen einsinken. Zudem markieren die Farbstoffe Bromphenolblau und Xylencyanol die Lauffront und ermöglichen die sofortige visuelle Erfolgskontrolle beim Befüllen der Taschen. Für die spätere Auswertung wird zwischendurch auch immer wieder ein Größenmarker auf das Gel pipettiert.

Eine Sonderform der Agarosegelelektrophorese ist die Anwendung präparativer Gele zur Isolierung von DNA-Fragmenten, beispielsweise wenn eine PCR mehrere Produkte ergibt. In diesem Fall wurde standardmäßig eine 1 %ige Gelmatrix gegossen, doch anstatt des oben erwähnten TBE-Puffers muss 1xTAE-Puffer verwendet werden, da dieser nicht mit üblichen Aufreinigungsmethoden interferiert und die DNA nicht angreift. Nach Aufbringen der Proben und Wanderung im elektrischen Feld kann aus den aufgetrennten Fragmenten die Bande des gewünschten Fragments ausgeschnitten werden. Die Identifizierung der richtigen Bande erfolgt auch hierbei unter UV-Licht durch Abgleich mit dem Größenmarker.

2.2.4 Aufreinigung von PCR-Produkten

Um die Produkte einer PCR sequenzieren zu können, müssen diese erst von Oligonukleotiden (Primer), Proteinen, Ethidiumbromid, Agarose etc. gereinigt werden. Bei mehreren Produkten sollten diese vorher mittels eines präparativen Gels aufgetrennt und eventuelle unbrauchbare Nebenprodukte verworfen werden.

Zur Aufreinigung wand ich das GFXTM PCR DNA and Gel Band Purification Kit der Firma Amersham Pharmacia Biotech entsprechend dem Begleitprotokoll an.

2.2.5 Sequenzanalyse

Die Didesoxy- oder Kettenabbruchmethode funktioniert im Prinzip wie eine PCR, beruht aber zusätzlich auf dem zufälligen Einbau von fluoreszenzmarkierten Didesoxynukleotiden neben den normalen Desoxynukleotiden. Die markierten Didesoxynukleotide führen durch die fehlende 3'OH-Gruppe beim Einbau zum

Kettenabbruch. So entsteht nicht immer das gleiche PCR-Produkt, sondern verschieden lange Fragmente, jeweils mit einem Didesoxynukleotid am 3' Ende.

Durch das Sortieren der Fragmente nach der Länge im ABI PRISM® 3100 Genetic Analyzer kann die komplette DNA-Sequenz rekonstruiert werden: Jede der vier Basen, die zum Abbruch der Kette führte, ist mit einem anderen Fluoreszenzmarker gekennzeichnet. Für die Sequenzreaktion wurde der BigDye™ Terminator Cycle Sequencing Ready Reaction Kit gemäß Gebrauchsanweisung angewendet.

Der Standardansatz setzt sich wie folgt zusammen (*Tabelle 7*):

Tabelle 7: Standardansatz für die Sequenzreaktion (HPLC: High pressure liquid chromatography, M = Mol)

Reagenz	Volumen/Konzentration
Premix	1,0 µl
5x Puffer	1,0 µl
Primer (10 µM)	0,5 µl
Template-DNA	1,5 - 2,5 µl
HPLC-H ₂ O	ad 5,0 µl

Beim Einsetzen der Template DNA kann man die Menge zwischen 1,5 und 2,5 µl variieren, um so ein zu starkes oder schwaches Beladen der Reaktion zu vermeiden. Das Endvolumen ist aber immer auf 5,0 µl einzustellen.

Das Standard Cycle Sequencing Programm wurde auf einer Perking Elmer durchgeführt und sieht wie folgt aus:

1. 1 min Denaturierung bei 96 °C
2. 10 s Denaturierung bei 96 °C
3. 5 s Annealing bei 55 °C
4. 4 min Extension bei 60 °C;

25 Zyklen von Schritt 2 bis 4

(anschließend bei 4 °C ≤ 24h lagerbar)

Bevor die Proben im ABI PRISM® 3100 Genetic Analyzer analysiert werden können, müssen sie durch Fällung aufgereinigt werden. Das Produkt wird mittels einer Ethanol-fällung aufgereinigt, indem 90 µl HPLC-H₂O (HPLC = High pressure liquid chromatography) zu dem Extensionsprodukt gegeben werden und beides in ein Eppendorfgefäß überführt wird. Anschließend werden 250 µl EtOH und 10 µl

NaAc zugesetzt und die Reaktion durch Vortexen gut vermischt. Nach 15 Minuten bei Raumtemperatur und dem folgenden Zentrifugationsschritt (13000 rpm, 15 min) kann der Überstand mit einer Pipette entfernt werden. Dann wird 200 µl 70% EtOH zugesetzt und erneut (13000 rpm, 15 min) zentrifugiert. Abschließend wird der Überstand entfernt, danach muss die Reaktion im Eppendorfgefäß trocknen. Die Proben können jetzt entweder bei -20°C gelagert werden oder in 20 µl Formamid gelöst und bei 37°C für 10 Minuten inkubiert werden.

Um schließlich die Sequenzierung zu ermöglichen muss die im Moment doppelsträngig vorliegende DNA in zwei Einzelstränge getrennt werden, dies geschieht bei einem Denaturierungsschritt (95°C , 3 min). Danach werden die Analysate sofort auf Eis gestellt, um diesen Zustand bis zum Einsetzen der Proben in den ABI PRISM® aufrechtzuerhalten.

2.2.6 Statistik

In dieser Arbeit wurden keine statistischen Verfahren angewendet, da dies im Rahmen der Fragestellung nicht sinnvoll erschien. Wegen der kleinen Fallzahlen sind keine objektiven Aussagen über die Polymorphismen der Makaken- bzw. Schimpansenpopulation zu erwarten. Vielmehr soll durch die Untersuchungen die Möglichkeit von Variabilitäten im Promotorbereich des *GH1*-Gens von Makake und Schimpanse geprüft werden und durch Interspeziesvergleich der Gene evtl. Rückschlüsse auf die phylogenetische Verwandtschaft mit dem *GH1*-Gen des Menschen gezogen werden. Falls bei einer Spezies Variationen gefunden werden, können Studien mit einer größeren Zahl an untersuchten Individuen Aufschluss über deren Häufigkeit innerhalb der betreffenden Primatenpopulation geben und gegebenenfalls in dieser Arbeit noch nicht entdeckte polymorphe Stellen aufdecken.

3 Ergebnisse

Die proximale Promotorregion des menschlichen *GH1*-Gens erweist sich als überdurchschnittlich polymorph (Giordano et al. 1997). Meine Aufgabe bestand darin, die dazu orthologe Region von Schimpanse und Makake via PCR zu amplifizieren und dann zu sequenzieren, um sie mit der des menschlichen *GH1*-Promotors zu vergleichen. Dazu standen mir die DNA-Proben von 24 nicht verwandten Makaken der Art *Macaca mulatta* (Rhesusaffe) und von 16 Schimpansen (*Pan troglodytes*) zur Verfügung.

Beim Schimpansen wird das hypophysäre *GH*-Gen, das dem *GH1*-Gen des Menschen entspricht, manchmal als *GHN* (N = normal) bezeichnet, im Gegensatz zur *GH*-Variante, *GHV*, der synonymen Bezeichnung für das *GH2*-Gen (Revol de Mendoza et al 2004). Doch gibt es in diesem Fall keine geregelte Bezeichnung des HGNC (HUGO gene nomenclature committee) wie beim menschlichen Somatotropin. Letzteres wird offiziell als *hGH1* entsprechend seinem englischen Namen „human Growth Hormone 1“ abgekürzt. Um die Orthologie zu den entsprechenden menschlichen Genen zu unterstreichen und die Nomenklatur möglichst einfach zu halten, werde ich im Folgenden *cGH1* und *cGH2* für die entsprechenden Gene des Schimpansen verwenden (c = chimpanzee) und *mGH1* und *mGH2* für die des Makaken (m = macaque). Analog dazu werde ich auch die anderen Genbezeichnungen jeweils mit den Buchstaben h, c oder m kennzeichnen.

Bei der Auswahl der Primer für die PCRs orientierte ich mich an den Angaben der Publikation von Horan et al. (2003). In dieser Publikation sind Positionsangaben in Bezug auf den Startpunkt der Transkription des *hGH1*-Gens enthalten. Hierauf beziehen sich auch die Angaben zu den polymorphen Stellen in der Literatur (Giordano et al. 1997, Horan et al. 2003). Diese Positionsangaben habe ich berücksichtigt, um später einen effizienten Vergleich der Promotorregionen der verschiedenen Spezies zu ermöglichen. Es ist üblich, den Transkriptionsstart eines Gens als +1 zu bezeichnen, davor kommt Position -1, in Ableserichtung des Gens wird aufwärts nummeriert. Der PCR-Primer ghf1 aus Horan et al. (2003) liegt demnach an Position -615 bis -599 des menschlichen *GH1*, die bisher

dokumentierten 16 Polymorphismen beim Menschen finden sich zwischen den Positionen -476 und +59 (Wagner 1997, Horan et al. 2003).

3.1 Promotorregion des *cGH1*-Gens des Schimpansen

Der Genom-Browser der UCSC (University of Santa Cruz California) stellt die Referenzsequenzen zahlreicher Genome zur Verfügung. Die Genomsequenz des gemeinen Schimpansen (*Pan troglodytes*) liegt zurzeit als Version CHIMP 2.1 (May 2007) vor. Mit BLAT wird ein Verfahren bezeichnet, das zwei ähnliche Sequenzen nach Übereinstimmungen absucht und miteinander vergleicht. Der UCSC-Browser zeigt daraufhin an, mit welcher Übereinstimmung (in Prozent) und mit welcher Häufigkeit die eingespeiste Sequenz im Genom der ausgewählten Spezies zu finden ist. So kann man beispielsweise die Sequenz eines PCR-Produkts einspeisen und herausfinden, ob der gewünschte Teil des entsprechenden Genoms amplifiziert wurde, beziehungsweise wo die vorliegende Sequenz im Genom lokalisiert ist.

Zwischen dem Genom von Mensch und Schimpanse besteht im Durchschnitt eine sehr hohe Sequenzhomologie von fast 99% (Chimpanzee Sequencing and Analysis Consortium 2005). Deshalb hoffte ich, mit Hilfe der schon vorliegenden menschlichen Sequenz des polymorphen *mGH1*-Promotors die dazu orthologe Region im Schimpansengenom via „BLAT“ im UCSC bestimmen zu können.

Ich benutzte für meine Analyse das Stück des menschlichen *GH1*-Promotors von -615 bis +59 [entspricht Position 4549 bis 5222 der Sequenz mit der Zugangsnummer: *J03071* (gesamter menschlicher *GH*-Locus)] Die Sequenz „*J03071*“ ist aus dem Jahr 1989 und ein Abgleich mit der aktuellen Version des menschlichen Genoms (NCBI 36.1, März 2006) ergibt eine fehlende Base. An der betreffenden Stelle (-33) gibt es einen Polymorphismus im menschlichen Genom hier kann ein Guanin liegen, die Base kann aber auch fehlen (siehe Horan et al. 2003). Die Base wird eingefügt um das Zählraster nicht zu verschieben.

Die BLAT-Analyse auf dem UCSC-Server mit der aktualisierten Version der menschlichen Promotorsequenz (Position –615 bis +59) als Sucheingabe ergibt als bestes Ergebnis eine Region im Schimpansengenom mit sehr großer Ähnlichkeit (*Abbildung 3*). Die Nukleotidabfolgen der beiden Sequenzen stimmen zu 97,1 % überein.

Eingabe: Promotor des hGH1-Gens, 674 Basen ab Position –615						
GGGAGCCCCa	GCAAtGC	TCA	GGGAAAGGGG	AGAGCAAAGT	GTGGGGTTGG	50
TTCTCTCTAG	TGGTCAGTGT	TGGAACTGCA	TCCAGCTGAC	TCAGGCTGAC		100
CCAGGAGTCc	TCAGCAGAAG	TGGAATTCAG	GACTGAATCg	TGCTCACAAC		150
CCCCACAATC	TATTGGCTGT	GCTTGGCCCC	TTTTCCCAAC	ACACACATTC		200
TGTCTGGTGG	GTGGAGGttA	AACATGCGGG	GAGGAGGAAA	GGgATAGGAT		250
AGAGAaTGGG	ATGTGGTCGG	TAgGGGGTCT	CAAGGACTGG	CTATCCTGAC		300
ATCCTTctCC	GCGTtCAGGT	TGGCCACCAT	GGCCTGCgGC	CAGAGGGCAC		350
CCACgTGaCC	CTTAAAGAGA	GGACAAGTTG	GGTGGTATCT	CTGGCTGACA		400
CTCTGTGCAC	AACCCTCACA	ACaCTGGTGA	CGGTGGGAAG	GGAAAGATGA		450
CAAGCCAGGG	GGCATGATCC	CAGCATGTGT	GGGAGGAGCT	TCTAAATTAT		500
CCATTAGCAC	AAGCCCCTCA	GTGGCCCCAT	GCATAAATGT	ACACAGAAAC		550
AGGTGGGGGC	AACAGTGgGA	GAGAAGGGGC	CAGGGTATAA	AAAGGGCCCCA		600
CAAGAGACCA	GCTCAAGGAT	CCCAAGgCCC	AACTCCCCgA	ACCACTCAGG		650
GTCTGTGGA	CAGCTCACCT	AGCg				
Ausgabe: Schimpanse, Chromosom 17						
ctgctcccgg	gatttctggt	tctt	ggtttg	tgtctctgct	gcaa	gtcaaa 63200856
ggagctgggg	caataccttg	agtctgggt	cttcgtcctc	agggacctgg		63200806
GGGAGCCCCc	GCAAcGC	TCA	GGGAAAGGGG	AGAGCAAAGT	GTGGGGTTGG	63200756
TTCTCTCTAG	TGGTCAGTGT	TGGAACTGCA	TCCAGCTGAC	TCAGGCTGAC		63200706
CCAGGAGTCg	TCAGCAGAAG	TGGAATTCAG	GACTGAATCc	TGCTCACAAC		63200656
CCCCACAATC	TATTGGCTGT	GCTTGGCCCC	TTTTCCCAAC	ACACACATTC		63200606
TGTCTGGTGG	GTGGAGGggA	AACATGCGGG	GAGGAGGAAA	GGaATAGGAT		63200556
AGAGAgTGGG	ATGTGGTCGG	TAGGGGTCTC	AAGGACTGGC	TATCCTGACA		63200506
TCCTTcCCG	CGTgCAGGTT	GGCCACCATG	GCCTGCtGCC	AGAGGGCACC		63200456
CACcTgccc	TTAAAGAGAG	GACAAGTTGG	GTGGTATCTC	TGGCTGACAC		63200406
TCTGTGCACA	ACCCTCACAA	CgCTGGTGAC	GGTGGGAAGG	GAAAGATGAC		63200356
AAGCCAGGGG	GCATGATCCC	AGCATGTGTG	GGAGGAGCTT	CTAAATTATC		63200306
CATTAGCACA	AGCCCGTCAG	TGGCCCCATG	CATAAATGTA	CACAGAAAACA		63200256
GGTGGGGGCA	ACAGTgaGAG	AGAAGGGGCC	AGGGTATAAA	AAGGGCCAC		63200206
AAGAGACCAG	CTCAAGGATC	CCAAGaCCCA	ACTCCCCaAA	CCACTCAGGG		63200156
TCCTGTGGAC	AGCTCACCTA	Gctgcaatgg	ctccaggtaa	gcaccctaa		63200106
aatccctttg	ggcaca	atgt	gtcctgaggg	gagagggcggc	accctgtaga	63200056
tgggacgggg	gcactaacc	tcag				

Abbildung 3: oben: Eingabesequenz in den UCSC-Genom Browser: Promotorregion des menschlichen Growth Hormone 1 (*hGH1*) von Position –615 bis +59; unten: bestes Suchergebnis im Schimpansengenom (CHIMP 2.1, May 2007) auf Chromosom 17, Minusstrang. Seitlich ist die Position der jeweils äußersten Base auf dem DNA-Strang angegeben. Angabe der Basen: A(denosin), C(ytosin), G(uanin), T(hymin), dunkelblaue Großbuchstaben bedeuten exakte Übereinstimmung mit der Suchsequenz, hellblaue Buchstaben heben bei lückenhaften Übereinstimmung das erste und letzte passende Nukleotid um die Lücken hervor, schwarz und kleingedruckt erscheinen abweichende Basen und flankierende Sequenzen. Lage folgender PCR-Primer bunt markiert: **gelb**: ghf1 aus Horan et al. (2003), **grün**: ptrgh1f, **rot**: ptrgh1r.

Dieses Ergebnis mit einer Lage im Schimpansengenom auf Chromosom 17 von Position 63.200,805 bis 63.200,134 lässt sich durch die Lokalisierung des gesamten *cGH1*-Gens wie folgt verifizieren: Die Nukleotidsequenz des *cGH1* des Schimpansen ist in der Genomdatenbank des NCBI unter der Zugangsnummer *AF374232* erhältlich. Über eine BLAT-Suche im UCSC-Genom Browser lässt sich das Gen eindeutig im Schimpansengenom in einer angrenzenden Region lokalisieren (Position 63.198,648 bis 63.200,503).

Meine Aufgabe war es, die *cGH1*-Promotorsequenz des Schimpansen über PCR zu amplifizieren, um dann Polymorphismen zu identifizieren. Dazu habe ich die spezifischen Primer *ptrgh1f* und *ptrgh1r* ausgewählt, die diese Region flankieren (Lage der Primer siehe *Abbildung 3*).

In *Abbildung 3* ist die Position des *ghf1* Primers eingezeichnet, die als Orientierung dient. Im Folgenden werde ich die Positionsangaben des *hGH1*-Promotors auch für die des Schimpansen übernehmen. Einerseits weil der Transkriptionsstart des *cGH1* unbekannt ist, andererseits weil ich später Base für Base menschliche und Schimpansen-Sequenz aneinanderlegen und vergleichen möchte. So werde ich immer den menschlichen Promotor als Referenzsequenz verwenden, um Basenangaben zu machen, auch wenn bei Schimpanse oder Makake ab und zu die Anzahl der Basen abweichen sollte.

Mir standen DNA-Proben von 16 Schimpansen zur Verfügung, mit denen ich die PCR mit den Primern *ptrgh1f* und *ptrgh1r* durchgeführt habe. Die PCR-Produkte wurden anschließend sequenziert.

Die Betrachtung der einzelnen Ergebnisse der Sequenzierungen übermittelt graphisch folgende Informationen über Polymorphismen: Ist ein Doppelpeak zu sehen, befinden sich an entsprechender Stelle eines PCR-Produkts zwei verschiedene Basen, was bedeutet, dass zwei verschiedene Allele vorhanden sind. Von väterlicher und mütterlicher Seite wurden in diesem Falle also unterschiedliche Allele weitervererbt. An dieser Stelle des Genoms wird Heterozygotie beobachtet.

Sechs der 16 Schimpansen-DNAs (Bezeichnungen der DNA-Proben: PTR 46, 86, 89, 96, 694 und 828) zeigen keine Polymorphismen in ihrem *cGH1*-

Promotorbereich. Die restlichen zehn enthalten zwischen einer (PTR 133) und fünf (PTR 94) polymorphen Stellen. Insgesamt ergaben sich elf mögliche Lokalisationen für die Polymorphismen, wie in *Tabelle 8* näher beschrieben wird.

Tabelle 8: Beschreibung der elf entdeckten Polymorphismen für die Promotorregion des *cGH1* beim Schimpansen; Die Basenangaben [A(denin), C(ytosin), G(uanin), T(hymin)] gelten für den Minusstrang, die Positionen beziehen sich auf das Chromosom 17 des aktuellen Schimpansengenoms CHIMP2.1, May 2007.

Position der polymorphen Stellen auf Chromosom 17	Mögliche Allele	Heterozygotie beobachtet bei den Schimpansen Individuen mit der Kodierung
63.200,168	A/G	8, 88, 90, 91, 92, 94, 133, 519, 582, 835
63.200,241	T/C	8, 88, 90, 91, 92, 94, 519, 582, 835
63.200,355	A/C	94
63.200,447	G/C	835
63.200,449	A/C	8
63.200,505	T/G	94, 582
63.200,522	A/G	133, 835
63.200,535	T/C	91, 92
63.200,666	C/G	8, 88, 90, 91, 92, 519, 582
63.200,732	C/A	94
63.200,734	G/A	94

Da unter den fünf Polymorphismen von PTR 94 gleich drei sind, die nur bei diesem Schimpansen auftraten, habe ich mehrere unabhängige PCRs durchgeführt, ebenso bei PTR 8 und PTR 835, die ebenfalls beide einen „eigenes“ Allel zeigen. Diese Wiederholungen haben die Ergebnisse und die Angaben in *Tabelle 8* bestätigt.

Im nächsten Schritt meiner Analysen habe ich die Sequenzen der Schimpansen untereinander und mit der Sequenz des *hGH1*-Promotors verglichen. Hierfür verwendete ich das Sequenzvergleichsprogramm ClustalW. Mit diesem Programm ist es möglich, Sequenzen exakt passend übereinander zu setzen und zu vergleichen (Originaldatei siehe *Anhang 1*).

Nur einer der Polymorphismen der Schimpansen befindet sich an einer Stelle, an der der Mensch ebenfalls polymorph ist (Position -476 des Menschen, gelb markiert in *Tabelle 9*). An einer Stelle mit zwei möglichen Basen für den Menschen findet sich beim Schimpansen eine dritte Base (Position -50) (siehe *Tabelle 9*).

Es wird natürlich kein Anspruch auf Vollständigkeit der SNPs erhoben, da die Anzahl der untersuchten Individuen mit 16 Stück recht gering war. Es konnte jedoch gezeigt werden, dass die zur Verfügung Schimpansen in der Tat Polymorphismen im *cGH1*-Gen Promotor haben.

3.2 Promotorregion des *mGH1*-Gens des Makaken

Das Makakengenom und das menschliche Genom stimmen auf Nukleotidebene im Durchschnitt zu etwa 93% überein (Rhesus Macaque Genome Sequencing and Analysis Consortium 2007). Das sequenzierte Genom des Makaken heißt MMUL und ist derzeit in der Version 1.0 vom Februar 2006 auf dem UCSC-Server verfügbar.

Um die Promotorregion des *mGH1* in der publizierten Sequenz des Makaken-Genoms zu identifizieren, habe ich BLAT-Analysen mit der menschlichen homologen Sequenz, dem Promotorstück des *hGH1*-Gens, von Position -615 bis 59 (Zugangsnummer: *J03071*, Position 4549 bis 5222, [Version aktualisiert durch Einfügen einer Base, siehe oben]), unternommen.

Das *mGH1*-Gen aus der NCBI-Datenbank mit der Zugangsnummer DQ002799 (*mGH1* von 376 bis 1843) lässt sich eindeutig auf Chromosom 16 von 59.473,027 bis 59.474,494 lokalisieren. Gleich im Anschluss daran liegt – von 59.474,499 bis 59.475,156 – das beste Ergebnis des BLAT via UCSC mit dem menschlichen *GH1*-Promotor (*Abbildung 4*). Die Sequenz des Makaken *GH1*-Promotors, dessen Echtheit ich auf diese Weise verifizieren konnte, weist eine Übereinstimmung von 92,2% mit der entsprechenden Sequenz des Menschen auf.

Für die Amplifizierung des *GH1*-Promotors der Makaken wählte ich die Primer makgh5 und makgh6 aus, die diese Region flankieren (Lage siehe *Abbildung 4*).

Eingabe: Promotor des hGH1-Gens, 674 Basen ab Position -615							
gggagccccca	gcaat	CTCA	GGGAAAGGGG	AGAGCAAAGT	GtGGgGTTGG	50	
TTCtCTCTAG	TGGTCAGTGT	TgGaACTGCA	TCCAGCTGac	TCAGGctgac		100	
ccAGGAGTcc	tcagcag	AAG	TGGAATTcAG	GACTGAATCG	TgCtCAcAAC	150	
CCCCACAATC	TATTGgCTGT	GCTTGGCCCC	TTTTccCAac	ACACACATTC		200	
TgTCTGgTGG	GTGGAGGttA	AACATGcGGG	GAGGAGGAAA	GGGATAGGAT		250	
AGAGaTGGG	ATGtGGTCCG	TAgGGGGTCT	CAAGGACTGG	CTATCCTGAC		300	
ATCCTTctCC	GCGTTCAGGT	TGGCCACCAT	GGCctTGCgGC	CAGAGGgCAC		350	
CCACgTgacc	CTTAAaGAGA	GGACAAGTTG	GGTGGTATCT	CTGGctGACA		400	
CTCTGTGCAC	AaCCCTCACA	AcaCTGGTGA	CGGTGgGAAG	GGAAAGATGA		450	
CAAGCCAGGG	GGCATGATCC	CAGCATGTGT	GGGAGGAGCT	TCTaAATTAT		500	
CCATTAGCac	AaGccCGTCA	GTgCCCCAT	GCATAAATGT	ACAcAGAAAC		550	
AGTGGGGgc	aacagtggGA	GAgAAGGGCC	CAGGGTATAA	AAAGGGCCCA		600	
CAAGAGACCA	GCTCaAGGAT	cCCAAGGCC	AACTCCCCGA	ACCACTCAGG		650	
GTCCTGTGGA	CAGctCACCT	AGCg					
Ausgabe: Makake, Chromosom 16							
tctgctgcag	gtccaaggag	ctggggcaat	accctgagtc	tgggttcttt		59475207	
gtccccagag	acctggggga	gcgcccccca	ccccacccc	atcccccaag		59475157	
GCTCAGG	AAA	AGGGAGAGG	AAAGTGG	GGA	GTTGGTTCcC	TCTAGTGGTC	59475107
AGTGTtaGcA	CTGCATCCAG	CTGcCTCAGG	Cgggcccagg	AGGAGTcaac		59475057	
aaAAGTGGAA	TTCAGGACTG	AATCGTtCcC	AgAACCCCA	CAATCTATTG		59475007	
cCTGTGCTTG	GCCCtTTTC	aCagCACACA	CATTCTcTCT	GcTGGGTGGA		59474957	
GGggAAACAT	GtGGGGAGGA	GGAAAGGGAT	AGGATAGAGA	gTGGGATGgG		59474907	
GTCGGTAgGG	GTCTCAAGGA	CTGGCTATCC	TGACATCCTT	CcCCGCGTTC		59474857	
AGGTTGGCCA	CCATGGCgTG	CtGCCAGAGG	aCACCCACcT	GctCCTTAAg		59474807	
GAGAGGACAA	GTTGGGTGGT	ATCTCTGGC	GACACTCTGT	GCACAgCCCT		59474757	
CACAAtgCTG	GTGACGGTga	GAAGGAAAG	ATGACAAGCC	AGGGGGCATG		59474707	
ATCCCAGCAT	GTGTGGGAGG	AGCTTCTcAA	TTATCCATTA	GCACAgGCac		59474657	
GTCAGTaGCC	CCATGCATAA	ATGTACAgag	AAACAGGTGG	Ggtcacgcag		59474607	
cgaGAGAtAA	GGGGCCAGG	TATAAAAAGG	GCCCACAAGA	GACCAGCTC		59474557	
AGGATtCCAA	GGCCCAACTC	CCCAACCAC	TCAGGGTCTCT	GTGGACAGC		59474507	
CACCTAGctg	caatggctgc	aggtaagtgc	ccctaaaatc	cctttccgca		59474457	
tgatgtgtcc	cgagg	ggaga	ggtggcacc	tgtag	atggg	acggcagcac	59474407
taaccctc							

Abbildung 4: oben: Eingabesequenz in den UCSC-Genom Browser: Promotorregion des menschlichen Growth Hormone 1 (*hGH1*) von Position -615 bis +59; unten: bestes Suchergebnis im Makakengenom (MMUL 1.0, Februar 2006) auf Chromosom 16, Minusstrang. Seitlich ist die Position der jeweils äußersten Base auf dem DNA-Strang angegeben. Angabe der Basen: A(denosin), C(ytosin), G(uanin) und T(hymin), dunkelblaue Großbuchstaben bedeuten exakte Übereinstimmung mit der Suchsequenz, hellblaue Buchstaben heben bei lückenhaften Übereinstimmung das erste und letzte passende Nukleotid um die Lücken hervor, schwarz und klein erscheinen abweichende Basen und flankierende Sequenz. Lage folgender PCR-Primer bunt markiert: **gelb**: ghf1 aus Horan et al. (2003), **grün**: makgh5, **rot**: makgh6.

Ich machte PCRs mit den 24 Makaken DNA-Proben und sequenzierte die Produkte. Bei den Sequenzanalysen konnte ich keine Polymorphismen erkennen. Auch beim Vergleich der einzelnen Sequenzen untereinander zeigen sich keine Unterschiede zwischen den Tieren. Die amplifizierten Bereiche sind identisch (siehe *Anhang 2*).

3.3 Vergleich der *GH1*-Promotorbereiche von Mensch, Schimpanse und Makake

Ich stellte die sequenzierten Bereiche der Makaken und der Schimpansen dem menschlichen *GH1*-Promotorbereich gegenüber, um die drei Spezies zu vergleichen (siehe *Tabelle 10*). Auch wenn es ein paar numerische Abweichungen der Nukleotidsequenzen von Rhesusaffe und Mensch gibt orientieren sich die Positionsangaben in jedem Fall an der menschlichen Sequenz (siehe auch *Anhang 2*).

Tabelle 10: Polymorphe Stellen (Polym.) der *GH1*-Promotorregion von Schimpanse (*PTR*) und Mensch (*HSA*) mit Positionen; Angabe der möglichen Allele: A(denin), C(ytosin), G(uanin), T(hymin), dazu Basen des jeweils anderen und der Makaken (*MMU*) im Vergleich. * Positionsangaben im Bezug auf Position +1 = Transkriptionsstart des *hGH1*-Gens. Daten für den Menschen aus Horan et al. (2003).

Position*	Polym. HSA			-476	-364		-339			-308	-301	-278			-168		-75	-57		-31	-6	-1	3	16		25	59
	Polym. PTR	-544	-542	-476		-345		-331	-314				-258	-256	-164				-50						24		
Base(n)	HSA	G	A	G/A	G/?	T	G/-	G	T	G/T	G/T	T/G	A	C	T/C	A	A/G	T/G	T	G/-	A/G	A/T	G/C	A/G	G	A/C	T/G
	PTR	G/A	A/C	G/C	G	T/C	-	G/A	T/G	C	G	T	A/C	C/G	T	A/C	A	G	T/C	G	A	A	G	A	G/A	A	T
	MMU	G	A	G	G	T	-	G	T	C	T	T	C	C	T	A	A	T	C	G	C	C	G	A	G	A	T

An 19 der 26 polymorphen Stellen verhält es sich folgendermaßen: Dasselbe Allel kommt bei allen drei Spezies vor, ein weiteres kann bei Schimpanse oder Mensch auftreten (im Falle von Position -476 kann bei Schimpanse und Mensch auch ein anderes Allel vorliegen, das zusätzlich mögliche Allel ist bei den beiden Spezies jedoch verschieden). Einen weiteren Sonderfall stellt Position -339 dar. Hier kann der Mensch ein Guanin haben, es ist aber auch möglich keine Base an dieser Stelle zu tragen. Letzteres findet sich sowohl beim Makaken als auch beim Schimpansen.

Die anderen sieben beschriebenen polymorphen Stellen weisen kein einheitliches Muster auf. In vier Fällen entsprechen die beiden an der polymorphen Stelle möglichen Basen bei Schimpanse oder Mensch den Allelen der jeweils anderen beiden Spezies. Beispielsweise an Position -301 kann der Mensch ein Guanin oder ein Thymin tragen, im Schimpansengenom findet sich an analoger Stelle ein Guanin, beim Makaken ein Thymin. An zwei für den Menschen polymorphen Stellen (Positionen -1 und -6) entspricht eines der beiden möglichen Allele der

Variante im Schimpansengenom, während das zweite mögliche Allel nicht mit dem des Makaken übereinstimmt. An Position -308 haben Schimpanse und Makake die selbe Base (Cytosin), der Mensch kann zwei andere (Guanin oder Thymin) tragen.

Nun ergibt sich folgende Frage: Warum gibt es beim Makaken keine Polymorphismen wie bei den zwei anderen Spezies, Schimpanse und Mensch? Wenn als Ursache der Entstehung verschiedener Allele Genkonversion über verwandte Gene vermutet wird, könnte die Erklärung dafür in der Organisation des *GH*-Genclusters und seiner Evolution gefunden werden. Um nähere Aussagen darüber machen zu können, war es deshalb meine zweite Aufgabe, die Lage der *GH*-Gene und deren Familie zu rekonstruieren.

3.4 Organisation der *GH*-Genfamilien von Mensch, Schimpanse und Makake

Die Organisation des menschlichen *GH*-Genlocus ist bekannt (Chen et al. 1989), ebenso sieht die Situation beim Makaken aus (Gonzalez Alvarez et al. 2006). Beim Schimpansen ist erst die Anzahl der Gene des *GH*-Locus bekannt: Es handelt sich um sechs Stück, zwei *GH*- und vier *CSH*-Gene, also ebenso viele wie der Makake, ein *CSH* mehr als der Mensch (*Tabelle 11*).

Tabelle 11: Übersicht über die Anzahl der Mitglieder des *GH*-Genlocus bei Mensch, Makake und Schimpanse und die jeweils verwendeten Bezeichnungen. GH = Growth Hormone, CSH(L) = Corionic Somatomammotropin Hormone (-Like), PL = Plazentares Laktogen, h = human, m = macaque, c = chimpanzee; * Pseudogen;

Spezies	Bezeichnungen der Mitglieder der <i>GH</i> -Genfamilie:	
	<i>GH</i> -Gene	Plazentare Laktogene
Mensch	<i>hGH1, hGH2</i>	<i>hCSH1, hCSH2, hCSHL1*</i>
Makake	<i>mGH1, mGH2</i>	<i>mCSH1, mCSH2, mCSH3, mCSH4</i>
Schimpanse	<i>cGH1, cGH2</i>	<i>cPLA, cPLB, cPLC, cPLD</i>

Die Sequenzen der einzelnen Gene (*cGH1, cGH2, cPLA, cPLB, cPLC* und *cPLD*, alle funktionell) liegen zwar vor (Revol de Mendoza et al. 2004), eine klare Aussage zur Anordnung in der Literatur fehlt bis jetzt aber noch. Meine Aufgabe war es nun, anhand der in Datenbanken verfügbaren Sequenzen des Schimpansengenoms die Anordnung der *GH*- und *PL*-Gene festzustellen. Diese Analyse wurde dadurch erschwert, dass sich in der entsprechenden Region im aktuellen Schimpansengenom (CHIMP 2.1, May 2007) mehrere Lücken befinden,

so dass nur die Sequenzen von *cGHN*, *cGHV* und drei der vier *PL*-Gene darin enthalten sind.

Meine Bestimmung der Lokalisation der *cPL*-Gene beruht hauptsächlich auf dem Vergleich der vollständigen Gensequenzen [aus Revol de Mendoza et al. (2004)] und dem Schimpansengenom der Datenbanken der UCSC, Ensembl und NCBI. Verwundert hat die Tatsache, dass *cPLD* mit fast 100% Übereinstimmung, bis auf einen wegen einer Lücke fehlenden Teil, in den aktuellen Gendatenbanken zu finden ist, die anderen *PL*-Gene jedoch viel schlechter zuzuordnen sind.

Nur zwei der sechs Gene, *cGH1* und *cPLD*, ließen sich leicht lokalisieren, während die anderen vier Mitglieder der Genfamilie schlechter zuzuordnen war, bzw. gar nicht. Das lag zum einen an einigen Lücken in der Schimpansensequenz der Genomdatenbanken, zum anderen an der nicht völligen Übereinstimmung der Gensequenzen, die aus der Publikation von Revol de Mendoza et al. (2004) vorliegen, mit der Gensequenz des Schimpansengenoms. So liegt *cGH2* in einer Sequenzlücke und allem Anschein nach ebenso *cPLA*. Sowohl die Sequenz von *cPLA* (Zugangscode AY146625) wie auch von *cPLB* (Zugangscode AY146626) zu 96% mit einer im Genom enthaltenen Sequenz überein, aber eine manuelle Überprüfung ergab eine höhere Ähnlichkeit mit *cPLB*. Das *cPLC*-Gen ließ sich ebenfalls nur mit 96% Identität platzieren, doch war es in diesem Fall das beste Ergebnis für diese Stelle.

Zusammen mit den bekannten Orthologiebeziehungen von *hGH1* und *cGHN*, *hGH2* und *cGHV* und *hCSHL1* und *cPLB* und dem Verdacht, auch *cPLD* und *cPLC* könnten zu *hCSH2* respektive *hCSH1* ortholog sein (Revol de Mendoza et al. 2004) ergab sich ein erster Überblick (siehe auch *Abbildung 6*).

Das Gen, dass in der aktuellen Version des Schimpansengenoms (UCSC, CHIMP 2.1., May 2007) nicht vorhanden ist, müsste demnach *cPLA* sein. Es liegt vermutlich im Bereich einer Sequenz-Lücke Chromosom 17, Basenposition 63,187.292 – 63,188.749), denn an deren Rand ließ sich ein Teil des *cPLA*-Gens mit Hilfe der BLAT-Analyse lokalisieren.

3.4.1 Sequenzierung der Stelle des Schimpansen Chromosom 17, an der das *cPLA*-Gen liegen könnte

Um zu bestätigen, dass das *cPLA*-Gen in der noch unbekanntenen Region inmitten des *GH*-Genclusters (im Moment mit ca. 1400 bp angegeben) lokalisiert ist, habe ich versucht diesen Bereich zu amplifizieren und darauf folgend zu sequenzieren.

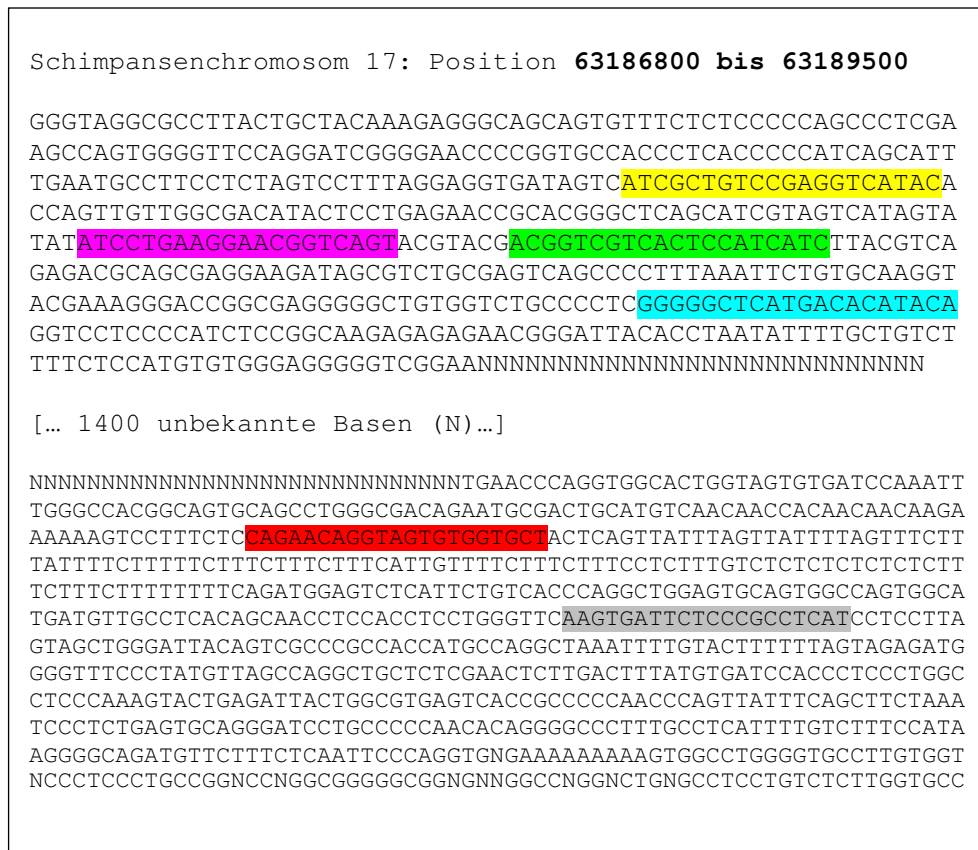


Abbildung 5: Darstellung der Lage der Primer die benutzt wurden, um die noch unsequenzierte Region auf Chromosom 17 des Schimpansen bei 63,187.292 – 63,188.749 (Ensembl, CHIMP 2.1, May 2007) durch PCR zu amplifizieren. Darstellung der Basen: A(denin), C(ytosin), G(uanin) und T(hymin), unbekannte Basen werden im Genome-Browser mit „N“ angezeigt. Bezeichnung der Primer siehe Farbmakierungen (Basenfolge angegeben in 3'-5'-Richtung):

DK1: GGGGGCTCATGACACATACA
DK2: AGCACCACACTACCTGTTCTG
DK3: ATCGCTGTCCGAGGTCATACA
DK4: ATGAGGCGGGAGAATCACTT
DK5: ACGGTCGTCACCTCCATCATC
DK7: ATCCTGAAGGAACGGTCAGT

Da es sich unter Umständen um einen mehrere Kilobasen großen Bereich handelt, den es zu amplifizieren gilt, führte ich eine Expand Long Template PCR durch. Die Primer hierfür sind in den Randbereichen der „Lücke“ lokalisiert, also in den schon bekannten Nukleotidsequenzen in den Datenbanken. Um passende

Primer designen zu können, analysierte ich zuerst besagte Randregionen mit dem RepeatMasker Programm, das repetitive Sequenzen erkennen kann. Diese Stellen eignen sich nicht, um dort PCR-Primer zu positionieren und müssen hierfür umgangen werden. Außerdem müssen die Primer so weit entfernt von der unbekannt Region binden, dass die zu amplifizierenden PCR-Produkte später durch Sequenzierung und Analyse eben dieser Randbereiche als echt-positiv identifiziert werden können.

Um diese Sequenzlücke, in welcher das *cPLA*-Gen lokalisiert zu sein scheint, durch PCR zu amplifizieren, habe ich verschiedene Primerkombinationen verwendet, deren Position in *Abbildung 5* wiedergegeben ist. Insgesamt testete ich sechs verschiedene Primer und benutzte sie auch in verschiedenen Kombinationen von Hin- und Rückprimer. Leider konnte kein verwertbares Ergebnis erzielt werden. Beim Überprüfen der sequenzierten Abschnitte ergaben sich schon in der Gelchromatographie Banden in falschen Größen, teilweise auch mehrere Banden. Manchmal stellte es sich heraus, dass die Amplifizierung gar nicht funktioniert hat. Die genaueren Daten dazu finden sich in *Tabelle 12*.

Tabelle 12: Auflistung der unternommenen Versuche, um die ca. 1400 bp-Lücke in der aktuellen Version des Schimpansengenoms CHIMP 2.1 (May 2007) auf Chromosom 17 (Basenposition 63,187.292 – 63,188.749) zu sequenzieren: Benutzte Primer in der ersten Spalte, dazu Temperatur und verwendete Puffer (P). Das gewünschte, wie auch das tatsächliche Produkt (oder mehrere) sind in Basenpaaren (bp) angegeben, wie sie im Bild nach der Auftrennung in der Agarosegelelektrophorese dargestellt wurden.

Primerpaar	Bedingungen (Temperatur, Puffer)	Erwartetes Produkt (Länge in bp)	Tatsächliche(s) Produkt(e) (Länge in bp)
dk1/dk2	54 °C, P2/P3	1900 bp	1100 bp
dk3/dk2	54°C, P2	2000 bp	Hauptbande: 2500 bp
dk3/dk4	54°C, P1	2150 bp	600 bp, 800 bp, 1400 bp
dk5/dk2	52°C/54°C/56°C, P2	1850 bp	Leiter von 100 bis 2500 bp
	56°C/58°C, P1		Kein Produkt
dk7/dk2	52°C/54°C/56°C, P2	1900 bp	Leiter von 100 bis 2000 bp
	58°C, P2		Hauptbande: 1400bp
	56°C/58°C, P1		Kein Produkt

Nach der Sequenzierung, der teilweise die Isolierung einzelner Banden durch präparative Gele vorausging, konnte ich die Produkte durch Vergleiche mit den Gendatenbanken meist auf anderen Chromosomen lokalisieren, andere waren nicht sequenzierbar. Die Ursache für das Fehlschlagen sämtlicher Versuche, die gesuchte Sequenz zu amplifizieren und analysieren, könnte in einer noch nicht

ganz korrekten Darstellung der Basenabfolge der Regionen vor und nach der Lücke in der aktuellen Sequenz des Schimpansengenoms (CHIMP 2.1, May 2007) liegen. Mangels exakter Informationen war es dann nicht möglich, wirklich passende Primer zu finden.

3.4.2 Vergleich des GH-Genlocus von Mensch, Schimpanse und Makake

Vergleich der GH-Gencluster bei den drei verschiedenen Spezies (*Abbildung 6*) ergibt sehr große Ähnlichkeiten zwischen dem menschlichen und dem der Schimpansen.

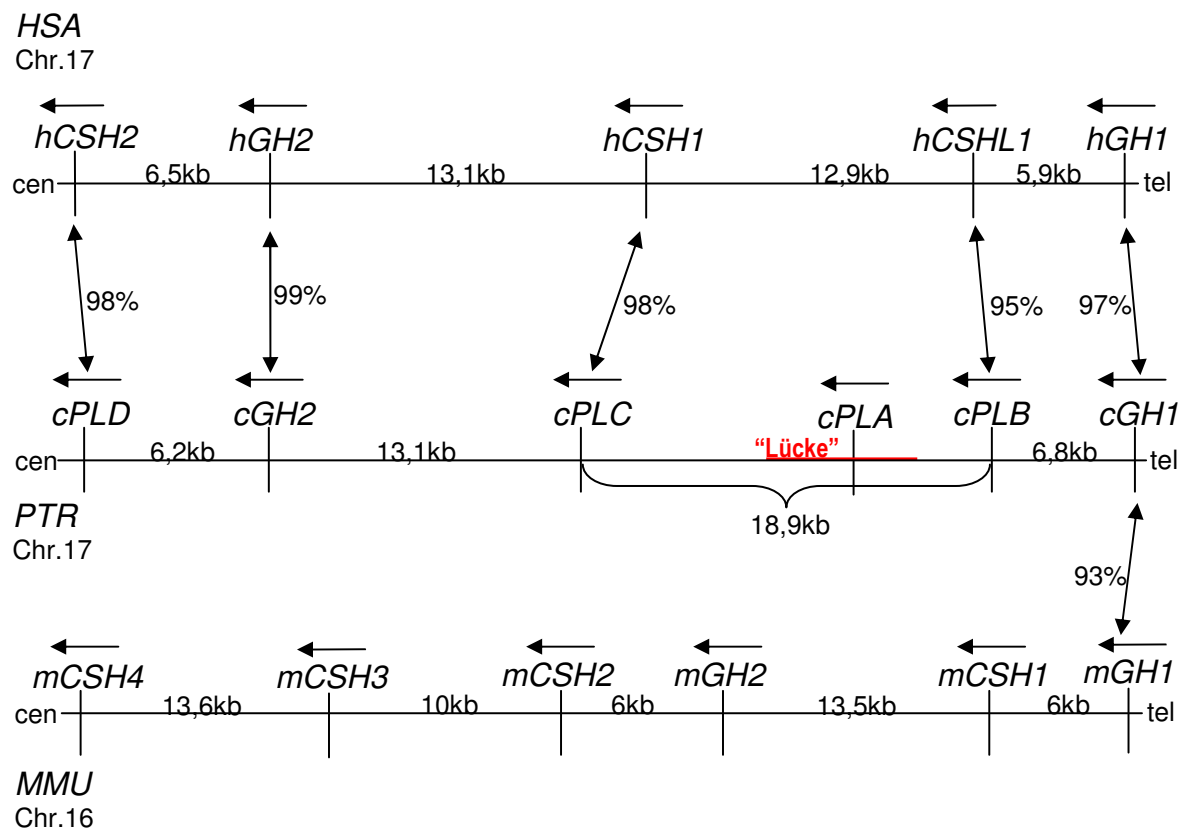


Abbildung 6: Vergleich zwischen den GH/CSH-Genloci von Mensch (HSA), Schimpanse (PTR) und Makake (MMU), auf Chromosom (Chr.) 17 bzw. 16 lokalisiert; Die Orientierung der Gene wird jeweils durch kleine dicke Pfeile angezeigt, die intergenischen Regionen sind in Kilobasen (kb) angegeben. Die langen Pfeile deuten auf die laut Revol de Mendoza et al. (2004) orthologen Gene, die Prozentzahlen geben Übereinstimmungen der Nukleotidsequenzen der Gene an (Daten aus Tabelle 14). Ein roter Balken markiert die ca. 1400 bp umfassende „Lücke“ der Genomdatenbanken, in der vermutlich das *cPLA* lokalisiert ist. Die Daten für Mensch und Schimpanse stammen aus Recherchen im jeweils aktuellem Genbild (NCBI 36.1, CHIMP 2.1), für Makaken wurden sie aus Gonzalez Alvarez et al. (2006) übernommen. GH = Growth Hormon, CSH(L) = Corionic Somatomammotropin Hormone (-Like), PL = Plazentares Laktogen, h = human, m = macaque, c = chimpanzee; cen = Zentromer, tel = Telomer.

Bei der evolutionären Entwicklung gab es allem Anschein nach mehr gemeinsame Schritte als bisher angenommen. Nicht nur drei, sondern fünf Gene scheinen bei einem gemeinsamen Vorfahren von Mensch und Schimpanse schon existiert zu haben. Zu unwahrscheinlich ist eine getrennte Entwicklung, die zu so ähnlichen Resultaten geführt haben könnte, zumal sogar die intergenischen Regionen fast gleich groß sind. Der Makake hat zwar die gleiche Anzahl an Genen wie der Schimpanse, doch sehen die Gencluster bis auf das Richtung Telomer zeigende Ende recht unterschiedlich aufgebaut aus.

Mehr Informationen über die Homologiebeziehungen der *GH*- und *CSH*-Gene der drei Spezies Mensch, Schimpanse und Makake versprach ein Vergleich der Nukleotidsequenzen. Um die Identität der Gene in Prozent zu ermitteln, habe ich die Sequenzen unter zu Hilfenahme des Programms ClustalW miteinander verglichen.

Um mit diesem Programm sinnvolle Ergebnisse zu erzielen, sollten die benutzten Nukleotidsequenzen möglichst gleich lang sein. Doch die verfügbaren Gensequenzen der *GH*- und *CSH*-Gene in den Datenbanken variieren teilweise erheblich in der Länge. Deshalb benutzte ich das primäre Transkript des menschlichen *hGH1*-Gen (GenBank Zugangsnummer: J03071) als Suchsequenz und führte eine Analyse mittels UCSC-Blat durch. Das Blat-Programm bestimmt möglichst ähnliche Sequenzen im ausgewählten Genom und gleicht sie mit der Suchsequenz, also dem *hGH1*, ab. So bekam ich als Ergebnis die paraloge Gene im menschlichen Genom, bzw. ortho- und homologe Gene im Schimpansen- und Makakengenom – soweit sie in den aktuellen Datenbanken vorhanden sind – in passender Länge angezeigt. Die Sequenzen wurden manuell überprüft und den ersten und letzten beiden Basentriplets der *hGH1*-Sequenz angepasst. Letztere sind mit nur wenigen Abweichungen in alle Genomen gut lokalisierbar.

Wie schon erwähnt ist die verfügbare Schimpansensequenz lückenhaft, das *cPLA*-Gen steht hier nicht zur Verfügung, das *cPLD*-Gen nur teilweise. Ich benutze die Gene *cPLA* und *cPLD* aus der Publikation von Revol de Mendoza et al. (2004) (Zugangsnummern AY146625 und AY146628). Auch im Makakengenom gibt es noch Lücken im GH-Locus, deshalb taucht das *mGH2*-Gen gar nicht in der Makakensequenz auf. Ich verwendete deshalb das *mGH2* aus der NCBI-

Datenbank mit dem Zugangscode DQ002799. Das *mCSH2* wird mit Hilfe der Sequenz aus DQ002802 ergänzt. Eine Auflistung der letztendlich verwendeten Daten findet sich in *Tabelle 13*.

Tabelle 13: Zusammenfassung der verwendeten Daten für den Vergleich aller Mitglieder der *GH/CSH*-Genfamilie von Mensch, Schimpanse und Makake. Die Basenpositionen beziehen sich auf die aktuellen Daten der UCSC-Genomdatenbank: NCBI 36.1 (März 2006), CHIMP 2.1 (May 2007), MMUL 1.0 (Februar 2006), das entsprechende Chromosom (Chr.) ist davor angegeben, Länge der Sequenzen in Basenpaaren (bp), falls verwendet wurde der betreffende Zugangscode einer Sequenz für die Genom-Datenbanken angegeben. GH = Growth Hormone, CSH(L) = Chorionic Somatomammotropin Hormone (-Like), PL = Plazentares Laktogen, h = human, m = macaque, c = chimpanzee; das einzige Pseudogen (das menschliche *CSHL1*) ist mit einem Stern gekennzeichnet.

Spezies	Gen/ Pseudogen*	Basenposition im jeweils aktuellen Genbild des UCSC-Genom Browsers	Länge der Sequenz	Zugangscode aus Genom- Datenbanken
Mensch	<i>hGH1</i>	Chr. 17: 59348295 – 59349930	1636 bp	J03071
	<i>hGH2</i>	Chr. 17: 59311307 – 5932955	1649 bp	
	<i>hCSH1</i>	Chr. 17: 59326010 – 59327665	1656 bp	
	<i>hCSH2</i>	Chr. 17: 59303109 – 59304768	1660 bp	
	<i>hCSHL1*</i>	Chr. 17: 59340699 – 59342350	1652 bp	
Schimpanse	<i>cGH1</i>	Chr. 17: 63198556 - 63200191	1636 bp	
	<i>cGH2</i>	Chr. 17: 63154295 - 63155946	1652 bp	
	<i>cPLA</i>	Nicht lokalisierbar wegen Sequenzlücke	1583 bp	AY146625
	<i>cPLB</i>	Chr. 17: 63189850 - 63191508	1659 bp	
	<i>cPLC</i>	Chr. 17: 63169034 - 63170692	1654 bp	
	<i>cPLD</i>	Chr. 17: 63146713 - 63147900	1542 bp	AY146628
Makake	<i>mGH1</i>	Chr. 16: 59472922 - 59474556	1635 bp	
	<i>mGH2</i>	Nicht lokalisierbar wegen Sequenzlücke	1500 bp	DQ002799
	<i>mCSH1</i>	Chr. 16: 59466069 - 59467740	1656 bp	
	<i>mCSH2</i>	Chr. 16: 59453611 - 59437238 (Lücke)	1646 bp	DQ002802
	<i>mCSH3</i>	Chr. 16: 59424188 – 59422532	1644 bp	
	<i>mCSH4</i>	Chr. 16: 59392972 – 59394631	1660 bp	

Wie man *Tabelle 13* entnehmen kann war für das *mGH2*-Gen eine Sequenz von nur 1500 bp Länge verfügbar. Auch *cPLA* und *cPLD* sind um einiges kürzer als die restlichen Gen-Sequenzen. Darum wurde die Analyse im ClustalW für die Identitäten, die in *Tabelle 14* dargestellt sind, für diese drei Gene jeweils nochmals mit verkürzten Sequenzen wiederholt. Dazu wurden jeweils alle anderen Sequenzen an die kürzeren Stücke angepasst.

Die paralogenen Gene des GH-Genclusters des Schimpansen zeigen im Intraspeziesvergleich mit zwischen 91 und 98% Sequenzidentität, ebenso hohe Homologien, wie der Mensch. Der Vergleich der 6 Makakengene untereinander führt zu etwas niedrigeren Resultaten: Alle Gene gleichen *mGH1* zu 90% (*mGH2*, *mCSH1*, *mCSH3*, *mCSH4*) oder zu 91% (*mCSH2*). Eine größere

Übereinstimmung zwischen *mGH2* und *mGH1* wäre zu erwarten gewesen. Am meisten Homologie weisen *mCSH1* und *mCSH2* mit 94% auf.

Tabelle 14: Resultate des Vergleichs der verschiedenen Mitglieder der *GH/CSH*-Familie von Mensch, Schimpanse und Makake, jeweils untereinander und mit denen der anderen zwei Spezies. Dazu wurde das Programm ClustalW benutzt, die Zahlenangaben sind in Prozent und beziehen sich auf übereinstimmende Nukleotide. Grün markiert sind die Ergebnisse (vermutlich) orthologer Gene (nach Revol de Mendoza et al. 2004).

GH = Growth Hormon, PL = Placentares Laktogen, CSH(L) = Corionic Somatomotropin Hormone (-Like).

Spezies	Mensch					Schimpanse						Makake					
	Gen	GH1	GH2	CSH1	CSH2	CSHL1	GH1	GH2	PLA	PLB	PLC	PLD	GH1	GH2	CSH1	CSH2	CSH3
Mensch	GH1	-															
	GH2	94	-														
	CSH1	93	92	-													
	CSH2	93	92	98	-												
	CSHL1	91	91	95	95	-											
Schimpanse	GH1	97	94	93	93	91	-										
	GH2	94	99	91	91	90	94	-									
	PLA	92	93	96	96	94	92	93	-								
	PLB	91	91	96	96	95	91	90	95	-							
	PLC	92	91	98	98	96	92	91	96	97	-						
	PLD	92	91	97	98	95	92	90	96	96	98	-					
Makake	GH1	93	92	91	91	89	93	91	91	91	90	90	-				
	GH2	89	89	88	88	87	89	89	88	87	88	87	90	-			
	CSH1	90	88	91	91	90	89	88	89	91	89	89	90	87	-		
	CSH2	90	89	90	90	89	90	88	89	90	89	89	91	87	94	-	
	CSH3	90	89	90	90	89	90	88	89	89	88	90	90	87	89	88	-
	CSH4	91	90	91	91	90	91	89	90	91	89	90	90	88	91	91	91

Ein Vergleich der Mitglieder der *GH*-Gencluster von Schimpanse und Mensch mit den Genen des Makaken ergibt die höchsten Ähnlichkeit all dieser Gene jeweils mit dem *mGH1*. Das *mGH2* des Makaken zeigt nur jeweils 89% Übereinstimmung zum *GH2* der anderen beiden Primaten.

Die *GH*- und *CSH*-Gene von Mensch und Schimpanse zeigen sehr hohe Übereinstimmungen der Nukleotidsequenzen ihrer jeweils orthologen Gene: 97% für *cGH1* und *hGH1*, 99% für *cGH2* und *hGH2*. Für die (vermuteten) orthologen Paare unter den *CSH*-Genen ergeben sich 95% Identität für *cPLB* und *hCSHL* und jeweils 98% für *cPLC* und *hCSH1* bzw. *cPLD* und *hCSH2*. Die Ergebnisse orthologer Paare erscheinen grün markiert in *Tabelle 14*.

4 Diskussion

4.1 Diversität im Promotorbereich des *GH1*-Gens bei Schimpanse und Makake:

Die proximale Promotorregion des menschlichen *GH1*-Gens weist eine hohe Polymorphismus-Rate auf, die die Expression des *GH1*-Gens beeinflusst und möglicherweise auch zur genetischen Disposition der Körpergröße beiträgt (Horan et al. 2003). In der europäischen Bevölkerung manifestiert sich diese Diversität in mindestens 40 verschiedenen *GH1*-Promotor Haplotypen, wobei man annimmt, dass diese Vielfalt größten Teils durch Genkonversion entstanden ist (Horan et al. 2003).

Ich untersuchte die Promotorregionen des *GH1*-Gens von Makaken und Schimpansen auf Polymorphismen, um zu prüfen, ob auch diese Spezies erhöhte Variabilität aufweisen. Dazu machte ich PCRs mit den DNA-Proben von 24 Makaken und 16 Schimpansen und sequenzierte die Produkte. In der Auswertung zeigten sich elf Polymorphismen bei den Schimpansen, interessanterweise nur einer davon an einer ebenfalls polymorphen Stelle beim Menschen (siehe *Tabelle 9*). Aufgrund der kleinen Versuchszahl erhebe ich keinen Anspruch auf Vollständigkeit der entdeckten Polymorphismen, auch bleibt zu prüfen, ob es sich bei den Variationen, die nur bei einzelnen Individuen aufgetreten sind wirklich um Polymorphismen handelt oder um spontane Mutationen. Dazu müsste eine größere Anzahl an Schimpansen untersucht werden.

Bei den Makaken hingegen fand ich in der entsprechenden Region des *GH1*-Gens keine SNPs, die *GH1*-Promotoren aller Makaken waren in ihrer Sequenz identisch. Das überraschte mich sehr, denn die *GH1*-Promotoren von Makake und Mensch sind sich mit 92% Identität recht ähnlich. [Die durchschnittliche Übereinstimmung von menschlichem und Makakengenom beträgt etwa 93% (Rhesus Macaque Genom Sequencing and Analysis Consortium 2007).] Auch an der Orthologie-Beziehung zwischen den *GH1*-Genen von Schimpanse, Makake und Mensch besteht kein Zweifel. Weshalb aber zeigen Schimpanse und Mensch eine sehr hohe Polymorphismusrate im *GH1*-Promotor, der Makake aber nicht?

Wenn als Ursache der Entstehung der Polymorphismen der *GH1*-Promotoren Genkonversion von den verwandten Genen angenommen wird, wie das beim Menschen der Fall ist (Giordano et al. 1997, Horan et al. 2003), muss das *GH*-Gencluster als Ganzes betrachtet werden um Rückschlüsse ziehen zu können.

Eine weitere Aufgabe meiner Dissertation war es deshalb, die in der Literatur publizierten Ergebnisse zur Organisation des *GH*-Genlocus bei Mensch, Schimpanse und Makake aufzuarbeiten und die Ergebnisse, die in diesen Studien veröffentlicht wurden, mit den mittlerweile sequenzierten Genomen von Makake und Schimpanse zu vergleichen.

Detailles über die Organisation der *GH*-Genloci von Mensch und Makake lagen schon vor (Chen et al. 1989, Golos et al. 1993, Gonzalez Alvarez et al. 2006). Der entsprechende Genlocus des Schimpansen musste erst rekonstruiert werden. Schwierigkeiten bereiteten mir die Sequenzlücken in eben dieser Region des Chromosom 17 des Schimpansen und auch die Differenzen zwischen den publizierten *PL*-Genen (Revol de Mendoza et al. 2004) und der Sequenz des Schimpansengenoms in den Datenbanken der UCSC und Ensembl. Letztendlich ließen sich *cPLC* und *cPLB* nur mit 96% Übereinstimmung zuordnen, während ich *cPLA* in einer Sequenzlücke vermute (UCSC: CHIMP 2.1, May 2007, Chromosom 17, Basenposition 63,187.292 – 63,188.749; siehe unten).

Eine zu weite Entfernung der duplizierten Gene oder die entgegengesetzte Orientierung könnten die Möglichkeit zur Genkonversion zunichte machen (Chen et al. 2007). Diese beiden Punkte können jedoch in diesem Fall nicht als Erklärung dienen, da sich herausstellte, dass die zum jeweiligen *GH*-Gencluster zugehörigen Gene sowohl bei Mensch und Schimpanse, als auch beim Makaken ausnahmslos in dieselbe Richtung abzulesen sind und die Entfernungen zwischen den Mitgliedern der Gencluster recht ähnlich sind (siehe auch *Abbildung 6*). Eine unterschiedliche Entwicklung beim Makaken im Vergleich zu Schimpanse und Mensch ist hieraus nicht abzuleiten.

Auch eine zu rasche Divergenz der Mitglieder eines durch Duplikationen entstandenen Genclusters könnte Genkonversionsereignisse verhindern. Denn Genkonversion scheint nur möglich, wenn die beiden beteiligten Sequenzen mindestens 92%, besser 95% Sequenzhomologie aufweisen (Chen et al. 2007).

Dies trifft innerhalb der *GH*-Genfamilien von Mensch und auch Schimpanse zu, wie sich im Intraspeziesvergleich zeigte (siehe Tabelle 14).

Zur Divergenz der sechs *GH/CSH*-Gene des Makaken ist folgendes zu sagen: Vor allem *mGH2* und die *CSH*-Gene des Makaken haben sich in der Tat, verglichen mit den menschlichen *GH*- bzw. *CSH*-Genen, schnell verändert. Die größte Weiterentwicklung, im Sinne von Nukleotidsubstitutionen, weist *mGH2* auf (Gonzales Alvarez et al. 2006). Darauf lassen auch meine Ergebnisse des Sequenzvergleichs der verschiedenen Mitglieder der *GH/CSH*-Familie von Schimpanse, Mensch und Makake, jeweils untereinander und mit denen der anderen zwei Spezies schließen (siehe Tabelle 14). *mGH2* ähnelt demnach mit 90% Übereinstimmung noch am meisten dem *mGH1*-Gen, mit jeweils 89% stimmt es mit den Sequenzen jeweils beider *GH*-Gene, sowohl von Mensch als auch Schimpanse, überein. Weder *mGH2* noch eines der *mCSH*-Gene ähnelt *mGH1* also so weit, dass die Möglichkeit zur Genkonversion bestünde.

Doch die Publikation von Gonzalez Alvarez et al. (2006) zeigt, dass sich beim Makaken sehr wohl Genkonversion ereignet hat: im zweiten Exon der Gene *mCSH2* und *mCSH4*, im ersten Intron zwischen *mGH1* und *mCSH1* und im dritten Exon von *mGH1* und *mGH2*.

Eine Möglichkeit wäre, dass auch beim Makaken, zu einem früheren Zeitpunkt nach den Duplikationsereignissen, Genkonversion stattgefunden hat, die zur Diversität der verschiedenen Mitglieder der *GH*-Genfamilie geführt hat. Doch im Laufe der Zeit könnten sich einzelne Allele durchgesetzt haben, was in kleineren Populationen durch „Inzucht“ begünstigt wird. Letztlich blieb nur noch einen Genotyp übrig, wie es bei den *mGH1*-Promotoren der Makaken der Fall ist. Im Nachfolgenden müssen sich die Mitglieder der *GH*-Genfamilie der Makaken sehr schnell entwickelt haben, was zu einer Akkumulation an Sequenzunterschieden führte, die heute in Form einer geringen Ähnlichkeit der *GH*-Gene innerhalb der Makaken sichtbar wird. Durch die geringere Homologie der *GH*-Gene zueinander, war dann kein weiteres Austauschen von Sequenzen durch Genkonversion mehr möglich (außer zwischen *mCSH1* und *mCSH2*, deren Sequenzen sich heute noch zu 94% gleichen).

Was bedeutet es überhaupt, wenn Polymorphismen in einen Promotorbereich eingeführt werden? Vielleicht ist es im Falle des *GH1*-Promotorbereichs „normal“, keine Polymorphismen zu haben, bis jetzt wurden jedenfalls nichts dergleichen bei weiteren Säugetieren beobachtet (Krawczak et al. 1999). Dann wären Mensch und Schimpanse die Ausnahme.

Das Variieren von Nukleotiden in einem Promotor könnte einerseits funktionelle Unterschiede bewirken, oder aber die entsprechenden Stellen des Promotors sind für die Funktion des Gens nicht wichtig und waren so frei vom Selektionsdruck für kleine Änderungen (Krawczak et al. 1999).

Große funktionelle Unterschiede zwischen dem *GH1* beim Menschen im Gegensatz zu anderen Säugetieren sind nicht bekannt, noch weniger Informationen gibt es über den Schimpansen. So haben doch die *GH1*-Gene aller Spezies gemeinsam, dass sie in der Hypophyse exprimiert werden, während evtl. vorhandene weitere *GH*-Gene und auch die eng verwandten *CSH*-Gene immer in der Plazenta vorkommen (Nachtigal et al. 1993, Wallis 1994, 1996). Eine Besonderheit des menschlichen *hGH1*-Promotorbereichs ist freilich, dass er von der kodierenden Region eines anderen Gens, *GHDTA* (growth hormone-derived transcriptional activator), überlappt wird. Dessen Rolle ist nicht bekannt, doch es könnte als alternativer Transkriptionsfaktor dienen (Labarriere et al. 1995). Einige der Polymorphismen des Menschen ließen sich also eventuell durch Selektionsdruck als Folge der zwei überlappende Gene erklären (Krawczak et al. 1999).

Dann bliebe noch folgende Erklärung: Polymorphismen in Promotorbereichen können die Expression eines Gens beeinflussen. Dieses Phänomen hat Horan et al. (2003) für den menschlichen *GH1*-Promotorbereich untersucht und er kam zu dem Schluss, dass SNPs in einem Promotor „die Genexpression in einer hoch komplexen und interaktiven Weise beeinflussen“. Die Kombination verschiedener Polymorphismen in einem Haplotyp wirken dabei auf nicht-additive Weise zusammen (Horan et al. 2003). So könnte Promotorvariabilität zu Expressionsvariabilität führen. Wie meine Ergebnisse zeigen gibt es diese Promotorvariabilität interessanterweise auch beim Schimpansen, so dass man annehmen muss, dass diese Mechanismen auch bei nicht-humanen Primaten aktiv sind.

4.2 Evolution der *GH*-Gencluster von Mensch und Schimpanse

Die meisten Säugetiere verfügen nur über ein hypophysäres Wachstumshormon, bei Affen und Menschenaffen allerdings hat sich das *GH*-Gen zu einem ganzen Gencluster paraloger Gene weiterentwickelt. So entstanden bei höheren Primaten weitere GHs und die plazentaren Laktogene (PL), oder auch Chorionic Somatomammotropin Hormones (CSH) genannt, die alle in der Plazenta synthetisiert werden. Noch nicht vollständig geklärt ist die Frage, welche Duplikations-Schritte Altwelt- und Menschenaffen bei der Entstehung der *GH*- und *CSH*-Gene gemeinsam haben, sich also bei einem gemeinsamen Vorgänger ereigneten und welche Duplikations- bzw. Deletionsereignisse speziesspezifisch stattfanden. Stellvertretend für die Altweltaffen war der Makake (*Macaca mulatta*) Gegenstand meiner Untersuchung, als Menschenaffe diente der (Gemeine) Schimpanse (*Pan troglodytes*), der als nächster Verwandter des Menschen zählt und dessen Genom auf Nukleotidebene in der Tat durchschnittlich nur etwas mehr als 1% Abweichungen zum menschlichen Genom aufweist (Chimpanzee Sequencing and Analysis Consortium 2005).

Um eine Aussage über die Beziehung der einzelnen Gene eines Genclusters zueinander machen zu können ist nicht nur die Anordnung wichtig. Ebenso notwendig ist der Überblick über die Sequenzhomologien. Dafür verglich ich alle Gene der *GH/CSH*-Familien von Mensch, Schimpanse und Makake jeweils sowohl innerhalb der Spezies, als auch zwischen den drei Spezies (*Tabelle 14*).

Die paralogen Mitglieder der *GH/CSH*-Genfamilie des Menschen sind sich alle sehr ähnlich, sie weisen eine Sequenzhomologie von zwischen 91 und 98% auf. Ähnlich verhält es sich beim Schimpansen, dessen Gene sich zu 90 bis 98% gleichen. Bei den Makaken fällt eine geringere Übereinstimmung der paralogen Gene auf (87 bis 94%).

Nach Barsh et al. (1983) sind *hGH1* und *hGH2*, die 94% Identität aufweisen, durch die Duplikation eines gemeinsamen Vorläufers entstanden. Davor war schon eine andere Kopie jenes Ur-*GH*-Gens entstanden, die den Vorläufer der *CSH*-Gene lieferte. Die beiden Vorläufergene, Ur-*GH*- und Ur-*CSH*-Gen waren schon beim gemeinsamen Vorgänger von Neu- und Altweltaffen vorhanden, doch die Duplikation die zum *GH2* führte, hat sich später in der gemeinsamen

Entwicklungslinie von Schimpanse und Mensch ereignet (Revol de Mendoza et al. 2004, Gonzalez Alvarez et al. 2006). Das würde heißen, *hGH1*, *cGH1* und *mGH1* wie auch die Nachkommen des *CSH*-Vorläufers sind orthologe Gene. Ebenso wären *hGH2* und *cGH2* ortholog, während das *GH2* des Makaken in einem speziesspezifischen Ereignis entstand.

Die Sequenzvergleiche scheinen diese These zu unterstützen, denn die als orthologe Paare bezeichneten Gene sind sich sehr ähnlich: 97% Übereinstimmung zwischen *cGH1* und *hGH1*, 99% zwischen *cGH2* und *hGH2*. Die orthologen Paare unter den *CSH*-Genen zeigen 95% identische Nukleotide bei *cPLB* und *hCSHL1* und jeweils 98% bei *cPLC* und *hCSH1* bzw. *cPLD* und *hCSH2*. Das *GH1* des Makaken weist zwar nur 93% Identität jeweils mit den orthologen Genen von Schimpanse und Mensch auf, doch ist dies das Ergebnis mit der größten Übereinstimmung zwischen einem Gen des Makaken und einem Gen einer der anderen beiden Spezies.

Das *mGH2* ist mit 98% Sequenzidentität sowohl zum menschlichen und zum Schimpansen-*GH1*, wie auch zu beiden *GH2*-Genen nicht klar zuzuordnen. Auch zwischen den *CSH*-Genen des Makaken und denen der anderen beiden Spezies lässt sich keine enge Homologiebeziehung feststellen. Das wiederum wirft folgende Frage auf: Welches Makakengen stammt vom *CSH*-Vorgänger ab? Der Vorläufer der *CSH*-Linie soll schließlich noch vor der Trennung der Linie des Makaken von Schimpanse und Mensch aus einer Duplikation des ursprünglichen *GH*-Gens entstanden sein, darin stimmen die meisten Autoren überein (Golos et al. 1993, Wallis 1996, Revol de Mendoza et al. 2004, Gonzalez Alvarez et al. 2006).

Die Abkömmlinge des Ur-*CSH* sind vermutlich das heutige *hCSHL1* des Menschen, das *cPLB* des Schimpansen und damit bei Gegenüberstellung der drei Genloci am ehesten das *mCSH1* des Makaken. Dieser Zusammenhang des *mCSH1*-Gens mit *hCSHL1* und *cPLB* spiegelt sich mit 91% Identität jeweils zwischen dem Makaken-Gen und einem der beiden anderen Genen nur sehr bedingt in den Ergebnissen der Sequenzvergleiche wieder.

An dieser Stelle sei jedoch erwähnt, dass die vier *CSH*-Gene des Makaken und vor allem das *mGH2* eine sehr rasche Weiterentwicklung durchgemacht haben

(Ye et al. 2005, Gonzales Alvarez et al. 2006). Daher rührt auch die oben genannte auffallend geringe Übereinstimmung der Gene im Interspeziesvergleich. Das könnte die Ergebnisse dieser Analyse in dem Sinne verfälschen, dass dann Orthologiebeziehungen nicht mehr so ohne weiteres offensichtlich wären und auf Grund der niedrigen Homologie fälschlicherweise nicht erkannt würden.

Um der Frage nach der Entwicklung des *GH*-Genclusters des Menschen weiter nachgehen zu können, wollte ich also den Makaken wegen der etwas schwierigen Situation zurückstellen und mich auf die anderen beiden untersuchten Spezies konzentrieren. Wie oben erwähnt teilten Mensch und Schimpanse zumindest zu Beginn der Erweiterung des Genlocus mehrere Duplikationsschritte.

Auf der Basis der neuen Erkenntnisse über die Organisation der *GH*-Gencluster von Schimpansen und Mensch die ich den aktuellen Genomdatenbanken entnehmen konnte, war es mir möglich, eine Theorie über die Entstehung des *GH*-Genclusters aufzustellen. Diese Theorie beruht auf dem Prinzip der maximalen Parsimonie, also unter der Annahme möglichst weniger Ereignisse im Zuge der Evolution.

Meine erste Arbeitshypothese ist in *Abbildung 7* veranschaulicht und beinhaltet folgende Evolutionsschritte:

Das erste ancestrale Ur-*GH*-Gen wird dupliziert, wird selbst der *GH*-Vorläufer „*vGH*“, während das andere Gen zum Vorgänger der *CSH*-Gene „*vCSH*“ wird. Beide Vorläufergene werden zusammen im Cluster dupliziert, die „Originale“ werden sich zu *hCSHL1* (heute ein Pseudogen) und *hGH1* entwickeln, die neuen Gene werden zu *hCSH2* und *hGH2*. Abschließend kommt es zur Duplikation vom zukünftigen *hCSH2*, aus der noch das *hCSH1* entsteht, was die starke Ähnlichkeit von *hCSH1* und *hCSH2* erklären würde und zusätzlich die Möglichkeit zur Genkonversion zwischen den beiden Genen. Diese kann nur zwischen sehr ähnlichen Genen stattfinden, z.B. Gen-Duplikationen, die sich noch nicht stark auseinander entwickelt haben.

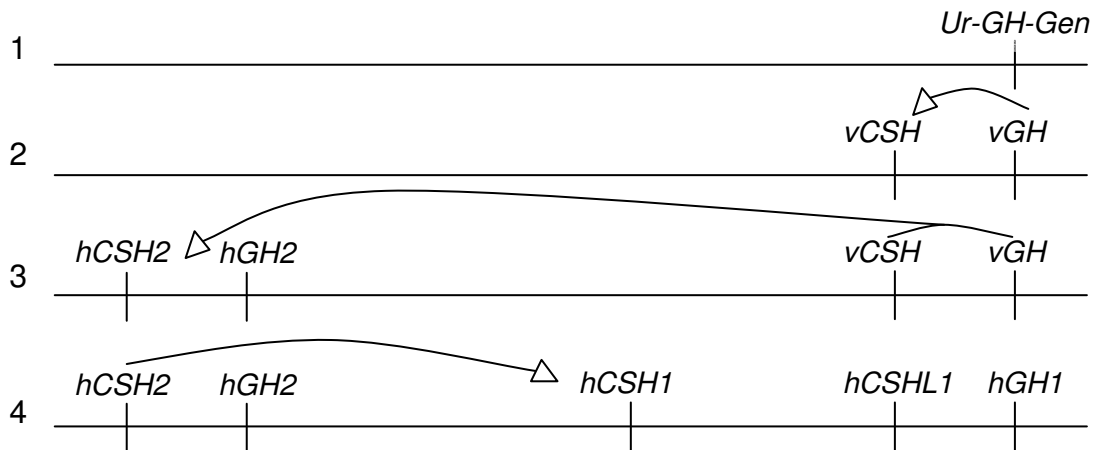


Abbildung 7: Erste Theorie zur Entstehung des GH-Genclusters des Menschen bestehend aus 4 Schritten:

1. Zuerst besteht nur ein ancestrales *Ur-GH-Gen*. Dieses wird dupliziert,
2. so entstehen „*vGH*“, ein Vorläufer für die *GH*-Gene und „*vCSH*“, für die *CSH*-Familie.
3. Duplikation von einem Cluster, der sowohl *vCSH* als auch *vGH* enthält. Die „Originale“ werden sich schließlich zu *hCSHL1* und *hGH1* entwickeln, die neuen Gene zu *hCSH2* und *hGH2*.
4. Die letzte Duplikation ist die von *hCSH2*, dadurch entsteht das heutige *hCSH1*-Gen.

GH = Growth Hormone, CSH(L) = Chorionic Somatomammotropin Hormone (-Like), v = Vorläufer. Die Pfeile deuten den Ursprung der Duplikation bzw. deren Richtung an.

Freilich muss man bedenken, dass auch die Entstehung des *GH*-Genclusters beim Schimpanse mit diesem Modell in *Abbildung 7* vereinbar sein muss. Man geht nämlich davon aus, dass sich die Trennung der Linien von Mensch und Schimpanse erst nach der Duplikation die zum *GH2*-Gen führte, ereignet hat (Revol de Mendoza et al. 2004). Das bedeutet aber nach dem Modell in *Abbildung 7*, dass nicht nur ein *GH2* entstanden ist, sondern zusätzlich auch das orthologe Gen zum menschlichen *hCSH2*, das *cPLD* des Schimpansen.

Ein Sequenz-Vergleichen zeigt für *cPLD* und *hCSH2* bzw. *cPLC* und *hCSH1* 98% Identität. Doch auch *hCSH1* und *hCSH2* zeigen 98% Sequenzidentität und bei phylogenetischen Analysen werden sie oft als engste Verwandte (Schwesterngruppe) dargestellt (Revol de Mendoza et al. 2004, *Abbildung 3*, Li et al. 2005, *Abbildung 1*).

Nun ist die Frage, ob der Schimpanse die Duplikation, die beim Menschen zu *hCSH1* geführt hat noch geteilt hat und deswegen sind die folglich orthologen Gene zwischen beiden Spezies sehr ähnlich, während Genkonversion für immer noch weitere intraspeziespezifische Angleichung aber auch zu leichten Differenzen zur jeweils anderen Art geführt hat. Es wäre aber auch möglich, dass Mensch und Schimpanse, wie von Revol de Mendoza et al. (2004) vorgeschlagen

diese Verdopplung jeweils eigenständig gemacht haben. Demzufolge wären die paralogen Gene nur deshalb im Vergleich der Spezies so ähnlich, weil sie ja alle von einem Gen (dem gemeinsamen Vorgänger des heutigen *hCSH2*- bzw. *cPLD*-Gens) abstammten und dann jeweils innerhalb jeder Spezies durch Genkonversion immer wieder an die ursprüngliche Form angeglichen wurden.

Da die Gene auf ihren jeweiligen Chromosomen bis auf das zusätzliche sechste Gen bei Schimpansen total gleich angeordnet sind und sogar die intergenischen Regionen, soweit beurteilbar, fast gleich lang sind (siehe auch *Abbildung 6*), ist die erste Möglichkeit wahrscheinlicher. Diese beinhaltet, dass *hCSH1* ortholog ist zu *cPLC* und *hCSH2* das orthologe Gen des *cPLB* ist.

Nach dem Prinzip der maximalen Parsimonie sind immer noch zwei verschiedene Wege denkbar, die zur unterschiedlichen Anzahl der Mitglieder im Gencluster bei Schimpanse und Mensch geführt haben: Eine weitere, linienspezifische Duplikation des Schimpansen oder eine Deletion im menschlichen *GH/CSH*-Locus. Doch muss noch bedacht werden, dass das *cPLA* vermutlich ganz früh entstanden ist, eventuell aus dem ancestralen *CSH*, jedenfalls noch vor den anderen drei *CSH*-Genen (Revol de Mendoza et al. 2004, Ye et al. 2005).

Die wahrscheinlichste aller Möglichkeiten zur Evolution des *GH*-Genclusters beim Schimpansen, die dem Parsimonie-Prinzip folgt, ist in *Abbildung 8* wiedergegeben. Dieses Modell habe ich auf der Basis der Homologie-Beziehungen der einzelnen Gene zueinander und der Homologien zu den orthologen Genen des Menschen entwickelt. Das Modell beinhaltet, dass die erste Duplikation von einem ancestralen *GH*-Gen zum Vorläufer der *CSH*-Familie führt, der sich beim Schimpansen letztendlich zum *cPLB* entwickeln wird. Dann wird zweimal das jetzt bestehende Gencluster dupliziert. Die ersten beiden entstehenden Gene werden zu *cPLA* und *cPLC*, die anderen beiden zu *cGHV* und *cPLD* (*Abbildung 8*).

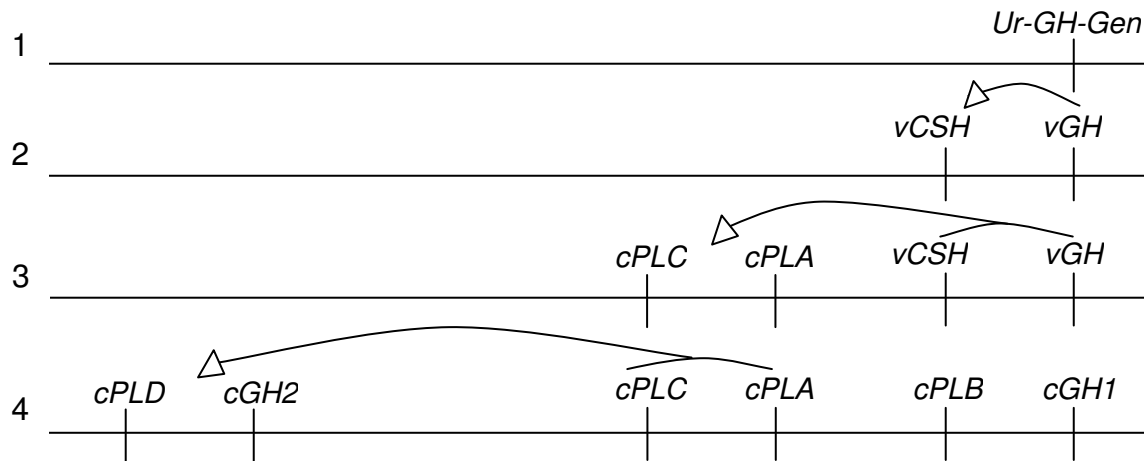


Abbildung 8: Theorie zur Entstehung des *GH*-Genclusters beim Schimpanse in vier Schritten:

1. Zuerst besteht nur ein ancestrales *Ur-GH-Gen*. Dieses wird dupliziert,
2. so entstehen „*vGH*“, ein Vorläufer für die *GH*-Gene und „*vCSH*“, für die *CSH*-Familie.
3. Die erste Duplikation der zwei Vorläufergene im Cluster führt zu *cPLA* und *cPLC*.
4. Die beiden Gene werden noch ein weiteres Mal zusammen dupliziert, der Genlocus besteht jetzt aus sechs Genen: *cPLD*, *cGH2*, *cPLC*, *cPLA*, *cPLB* und *cGH1*.

GH = Growth Hormone, *CSH* = Chorionic Somatomammotropin Hormone, *PL* = Plazentares Laktogen, *v* = Vorläufer, *c* = chimpanzee; (*CSH* und *PL* sind Synonyme). Die Pfeile deuten den Ursprung der Duplikation bzw. deren Richtung an.

Für den Menschen würde dies bedeuten, dass das dem *cPLA* entsprechende Gen deletiert wurde.

Wenn man sich die Ergebnisse der Intraspezies-Vergleiche der fünf *GH*- und *CSH*-Gene des Menschen anschaut (*Tabelle 14*) unterstützt meiner Meinung nach die „nur“ 94%ige Übereinstimmung zwischen *hGH1* und *hGH2* die in *Abbildung 8* veranschaulichte Theorie.

Die meisten bis jetzt publizierten phylogenetischen Bäume, die schon die vier *CSH*-Gene des Schimpanse berücksichtigen zeigen keinen besonders engen Zusammenhang zwischen *cPLA* und den beiden *GHs* des Schimpanse. Sie siedeln das *cPLA* eher bei den *PL*-Genen an (Revol de Mendoza et al. 2004, Ye et al. 2005). Doch haben sich die Gene ja alle in unterschiedlichen Raten weiterentwickelt, verschiedenste Genkonversions-Ereignisse sind bekannt, die das Erstellen von phylogenetischen Bäumen schwierig machen. Genkonversion hat auch zwischen *cGH1* und *cPLA* stattgefunden, dabei war das *GH*-Gen der „Spender“, das *CSH*-Gen der „Empfänger“ (Revol de Mendoza et al. 2004). Dies ist wiederum ein Indiz dafür, dass sie in einem sehr engen Zusammenhang stehen und vielleicht früher noch ähnlicher waren, denn Genkonversion findet gewöhnlich zwischen Sequenzen statt, die über 95% Identität aufweisen (Chen et al. 2007).

Heute jedenfalls stimmen die beiden Nukleotidsequenzen nur noch zu 92% überein.

Ich ließ im Programm Clustalw einen phylogenetischen Baum (Cladogram) berechnen, der die Verwandtschaftsbeziehungen der Schimpansengene verdeutlichen soll. Im Gegensatz zu oben genannten phylogenetischen Bäumen in anderen Publikationen, verwendete ich aber nicht das *cPLC* aus Revol de Mendoza et al. (2004), sondern die korrespondierende Stelle aus dem Genom. (Gründe waren die Unvollständigkeit der aus der genannten Publikation zur Verfügung stehenden Sequenz, die außerdem mit ca. 4% doch recht stark von der Schimpansengenom-Sequenz abweicht.)

Das Ergebnis (*Abbildung 9*) untermauert meine Theorie von der Evolution des *GH*-Locus der Schimpansen, ist aber, wegen der ungeklärten Datenlage im Bezug auf das *cPLA* und das *cPLB*-Gen unter Vorbehalt zu betrachten.

Cladogram

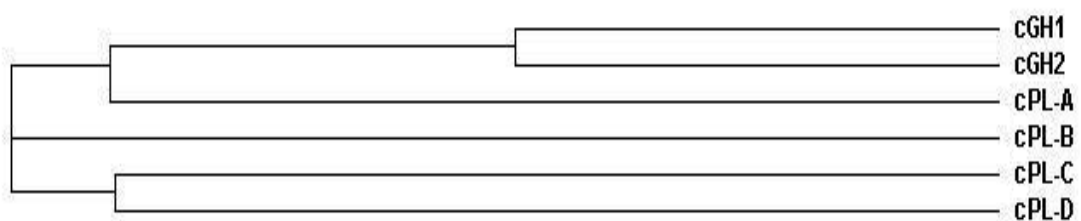


Abbildung 9: phylogenetischer Baum mit den Mitgliedern der *GH*-Familie des Schimpansen, erstellt im Programm Clustalw. GH = Growth Hormone, PL = Plazentares Laktogen, c = chimpanzee. Verwendete Daten: AF374232 (*cGH1*), AF 374233 (*cGH2*), AY146625 (*cPLA*), AY146626 (*cPLB*), AY146628 (*cPLD*) und *cPLC* aus dem Schimpansengenom des UCSC, Chromosom 17, 63.169,085 bis 63.171,007

Meine Ergebnisse siedeln das *cPLA* deutlich auf der Seite der beiden *GH*-Gene an. Das spricht für eine nähere Verwandtschaft des *cPLA*-Gens mit *cGH1* und *cGH2* als mit den anderen *PL*-Genen.

Hilfreich wären zur weiteren Klärung der Entwicklungsgeschichte der *GH/CSH*-Gencluster bei Mensch und Schimpanse die vollständige Sequenzierung der betreffenden Region beim Schimpansen, um überhaupt die Lage und die bisher publizierte Sequenz des *cPLA*-Gens zu überprüfen. Das *cPLA*-Gen liegt sehr wahrscheinlich in einer noch nicht sequenzierten bzw. veröffentlichten Region im

GH-Gencluster des Schimpansen. Es ist mir im Rahmen meiner Doktorarbeit nicht gelungen, diese noch unbekannt Sequenz durch PCR zu amplifizieren und anschließend zu sequenzieren. Der Grund für das Fehlschlagen aller Versuche könnte darin liegen, dass die Randsequenz der momentan bestehenden Sequenzlücke im Schimpansengenom sämtlicher Genomdatenbanken noch Fehler enthält. Somit ist es auf der Basis der aktuell verfügbaren Daten nicht möglich passende PCR-Primer zu designen, um die Lücke zu sequenzieren.

Das Erstellen der Sequenz, die die ancestrale *CSH*-Version hatte und aus der die einzelnen *CSH*-Gene hervorgegangen sind, wäre sicherlich informativ. So könnte gesichert werden, ob das *cPLA* aus einer Duplikation des Ur-*GH*-Gens oder doch aus der *CSH*-Linie entstanden ist und in welchem Zeitrahmen dies geschehen sein könnte.

Beim Betrachten der Theorie für die Entstehung des *GH*-Genlocus bei Mensch und Schimpanse fällt auf, dass sie vielleicht auch im Bezug auf die noch nicht geklärte Entwicklung beim Makaken weiterhelfen kann: Das Ur-*GH*-Gen und das daraus durch Duplikation entstandene ancestrale *CSH*-Gen sollen ja auch schon beim Makaken existiert haben. Das sind also vermutlich die heutigen Gene *mGH1* und *mCSH1*. Wenn man sich die vorgeschlagene zweite Duplikation meiner Theorie, die mit beiden Vorläufergenen als Cluster geschieht, vor Augen hält, so scheint es nicht abwegig, dass der Makake doch eine weitere Duplikation mit den anderen beiden Spezies geteilt hat. [Dies bedeutet jedoch nicht, dass das *mGH2* zu den *GH2*-Genen von Mensch und Schimpanse ortholog ist, wie oft in älteren Publikationen als scheinbar zwingende Folge angenommen wird (Golos et al. 1993, Krawczak et al. 1999, Mendoza et al. 2004), sondern vielmehr zum *cPLA*.] Für ein weiteres gemeinsames Duplikationsereignis von Mensch, Schimpanse und Makake spräche jedenfalls die kleine, jeweils ca. sechs Kilobasen umfassende intergenische Region jeweils zwischen den Genen, die sich aus den zwei ancestralen Genen entwickelt haben, also zwischen *hGH1* und *hCSH1*, bzw. *cGH1* und *cPLB*, bzw. *mGH1* und *mCSH1*, wie auch zwischen den möglicherweise beim Makaken entstandenen Duplikaten *mGH2* und *mCSH2*. Außerdem findet sich die höchstmögliche Übereinstimmung aller *GH*- und *CSH*-Gene des Makaken zwischen *mCSH1* und *mCSH2*. Eine ähnlich hohe

Sequenzhomologie mit dem *mGH1* lässt sich für das *mGH2* nicht beobachten, doch hat sich letzteres wie schon erwähnt am schnellsten von allen Genen des Clusters weiterentwickelt, so dass ein solcher Zusammenhang nun nicht mehr so leicht erkennbar sein dürfte.

Ein vollständiges Modell für die Entwicklung der *GH*-Genfamilie beim Makaken fehlt noch. Vielleicht kann an dieser Stelle eine Analyse der *GH*-Gencluster anderer Altweltaffen mehr Klarheit bringen. Weitere Informationen, die helfen könnten meine Theorie zu bestätigen, erhoffe ich mir außerdem von der vollständigen Sequenzierung der bisher unbekannt Region im *GH*-Gencluster der Schimpansen. Folgende Fragen bleiben offen: Wie waren die speziesspezifischen Duplikationsschritte des Makaken? Warum haben sich die *GH*- und *CSH*-Gene des Makaken im Vergleich mit den Genen der anderen beiden Spezies Schimpanse und Mensch so stark auseinander entwickelt?

5 Zusammenfassung

In der menschlichen Hypophyse wird das Wachstumshormon Somatotropin (hGH1 = human Growth Hormone 1) gebildet. Der proximale Promotorbereich des *hGH1*-Gens des Menschen erweist sich mit 16 SNPs (single nucleotide polymorphisms) auf 535 bp (Basenpaare) als überdurchschnittlich polymorph. Meine Aufgabe war es herauszufinden, wie sich die dazu orthologe Region von Schimpanse und Makake verhält. Zur Verfügung standen die DNA-Proben von 24 Makaken und von 16 Schimpansen. Via PCR amplifizierte ich die jeweiligen *GH1*-Promotorbereiche (*mGH1*-Promotor, m = macaque, bzw. *cGH1*-Promotor, c = chimpanzee), sequenzierte die Produkte und konnte sie so mit den für den Menschen bekannten Daten vergleichen. Bei den Makaken fanden sich überhaupt keine SNPs in der besagten Region, während ich bei den Schimpansen elf polymorphe Stellen innerhalb einer Spanne von 568 bp des *cGH1*-Promotors entdeckte.

Vom Menschen ist bekannt, dass die Polymorphismen im proximalen Promotorbereich des *GH1*-Gens auf komplexe Weise die Genexpression beeinflussen. Die Frage nach einer ähnlichen Funktion der polymorphen Stellen im *GH1*-Promotor des Schimpansen ist noch nicht geklärt. Als Erklärung für die SNPs der *GH1*-Promotorbereiche von Mensch und Schimpanse gelten Genkonversionsereignisse zwischen den nah verwandten *CSH*- (Chorionic Somatomammotropin Hormone) und *GH*-Genen. Die Möglichkeit, Genmaterial auszutauschen, hängt unter anderem von einer hohen Homologie (normalerweise mindestens 95%) von Spender- und Empfängersequenz ab. Dies ist im Falle der Mitglieder der *GH/CSH*-Familie von Mensch und Schimpanse gegeben.

Auch während der Entwicklung der *GH*- und *CSH*-Gene des Makaken fand einst Genkonversion statt, doch nach einer schnellen Weiterentwicklung der einzelnen Gene war Genkonversion wegen der geringeren Sequenzhomologie nicht mehr möglich. Einzelne Allele haben sich dann wohl im Laufe der Zeit durchgesetzt, so dass für den *mGH1*-Promotor heute nur noch einen Genotyp existiert.

Höhere Primaten (Affen und Menschenaffen) haben gemeinsam, dass sie im Laufe der Evolution aus einem ancestralen *GH*-Gen einen ganzen Gencluster

verwandter Gene entwickelt haben. Mich interessierte dabei vor allem, wie die Entstehung des menschlichen *GH*-Genlocus vonstatten ging und in wieweit sie gemeinsam mit Schimpanse und Makake durchlaufen wurde. Die Organisation des *GH*-Genlocus des Schimpansen war nicht bekannt, so stand zuerst dessen Rekonstruktion an. Wegen lückenhafter Informationen im betreffenden Bereich des Schimpansengenoms der Genom-Datenbanken war dies unter Vorbehalt möglich. Demnach sehe ich die Anordnung der Mitglieder des *GH*-Genclusters von 5' zu 3' wie folgt: *cGH1*, *cPLB*, *cPLA*, *cPLC*, *cGH2*, *cPLD* (c = chimpanzee-, PL = Plazentares Laktogen, synonym zu CSH).

Das *cPLA*-Gen liegt meiner Meinung nach in einer noch nicht sequenzierten Region, die in den aktuellen Genom-Datenbanken mit ca. 1400 bp Größe angegeben ist. Alle meine Versuche diesen Teil des Schimpansenchromosoms 17 via PCR zu amplifizieren schlugen fehl.

Nach meiner Theorie über die Entwicklung des *GH*-Genlocus von Mensch und Schimpanse, haben die beiden Spezies alle Duplikationsschritte geteilt. Die Gegenüberstellung der Genloci und die durchgeführten Sequenzvergleiche der *GH/CSH*-Gene bestätigen die bisher fragliche Orthologiebeziehung zwischen *hCSHL1* und *cPLB* und zwischen *hCSH2* und *cPLD*. Ich komme zu dem Schluss, dass außerdem *hCSH1* und *cPLC* ortholog sind. Das *cPLA*-Gen entstand durch eine Duplikation des Ur-*GH*-Gens. Das orthologe Gen des Menschen wurde nachträglich deletiert. Sobald das Schimpansengenom in der Region des *GH*-Genclusters vollständig sequenziert ist, wäre es interessant, die Sequenzen und die Lage der einzelnen *CSH*-Gene nochmals zu verifizieren, um meine Theorie von der Entstehung der Duplikationen bei Mensch und Schimpanse zu überprüfen.

Ich halte es außerdem für möglich, dass der Makake, die ersten zwei Duplikationsschritte mit den anderen beiden Spezies geteilt hat. Ein vollständiges Modell für die Entwicklung des *GH*-Genlocus des Makaken fehlt noch, hier sind noch einige Fragen offen: Wie waren die speziesspezifischen Duplikationsschritte des Makaken? Warum haben sich die Mitglieder der *GH*-Genfamilie des Makaken im Vergleich mit den anderen beiden Spezies so stark auseinander entwickelt? Vielleicht könnte an dieser Stelle die Analyse von *GH*-Genloci weiterer Altweltaffen weiterhelfen.

6 Literaturverzeichnis

Adkins RM, Nekrutenko A, Li WH.

Bushbaby growth hormone is much more similar to nonprimate growth hormones than to rhesus monkey and human growth hormones. Mol Biol Evol. 18: 55-60 (2001)

Barsh GS, Seeburg PH, Gelinas RE.

The human growth hormone gene family: structure and evolution of the chromosomal locus. Nucleic Acids Res. 11: 3939-3958 (1983)

Chen EY, Liao YC, Smith DH, Barrera-Saldana HA, Gelinas RE, Seeburg PH.

The human growth hormone locus: nucleotide sequence, biology, and evolution. Genomics. 4: 479-497 (1989)

Chen JM, Cooper DN, Chuzhanova N, Férec C, Patrinos GP.

Gene conversion: mechanisms, evolution and human disease. Nat Rev Genet. 8: 762-775 (2007)

Chimpanzee Sequencing and Analysis Consortium

Initial sequence of the chimpanzee genome and comparison with the human genome. Nature. 437: 69-87 (2005)

George DL, Phillips JA 3rd, Francke U, Seeburg PH.

The genes for growth hormone and chorionic somatomammotropin are on the long arm of human chromosome 17 in region q21 to qter. Hum Genet. 57: 138-141 (1981)

Giordano M, Marchetti C, Chiorboli E, Bona G, Momigliano Richiardi P.

Evidence for gene conversion in the generation of extensive polymorphism in the promoter of the growth hormone gene. Hum Genet. 100: 249-255 (1997)

Goffin V, Shiverick KT, Kelly PA, Martial JA.

Sequence-function relationships within the expanding family of prolactin, growth hormone, placental lactogen, and related proteins in mammals. Endocr Rev. 17: 385-410 (1996)

Golos TG, Durning M, Fisher JM, Fowler PD.

Cloning of four growth hormone/chorionic somatomammotropin-related complementary deoxyribonucleic acids differentially expressed during pregnancy in the rhesus monkey placenta. Endocrinology. 133: 1744-1752 (1993)

- Gonzalez Alvarez R, Revol de Mendoza A, Esquivel Escobedo D, Corrales Felix G, Rodriguez Sanchez I, Gonzalez V, Davila G, Cao Q, de Jong P, Fu YX, Barrera Saldana HA.
Growth hormone locus expands and diverges after the separation of New and Old World Monkeys. Gene. 380: 38-45 (2006)
- Harper ME, Barrera-Saldaña HA, Saunders GF.
Chromosomal localization of the human placental lactogen-growth hormone gene cluster to 17q22-24. Am J Hum Genet. 34: 227-234 (1982)
- Horan M, Millar DS, Hedderich J, Lewis G, Newsday V, Mo N, Fryklund L, Procter AM, Krawczak M, Cooper DN.
Human growth hormone 1 (GH1) gene expression: complex haplotype-dependent influence of polymorphic variation in the proximal promoter and locus control region. Hum Mutat. 21: 408-423 (2003)
- Krawczak M, Chuzhanova NA, Cooper DN.
Evolution of the proximal promoter region of the mammalian growth hormone gene. Gene. 237: 143-151 (1999)
- Labarriere N, Selvais PL, Lemaigre FP, Michel A, Maiter DM, Rousseau GG.
A novel transcriptional activator originating from an upstream promoter in human growth hormone gene. J Biol Chem. 270: 19205-19208 (1995)
- Li Y, Ye C, Shi P, Zou XJ, Xiao R, Gong YY, Zhang YP.
Independent origin of the growth hormone gene family in New World monkeys and Old World monkeys/hominoids. J Mol Endocrinol. 35: 399-409 (2005)
- Liu JC, Makova KD, Adkins RM, Gibson S, Li WH.
Episodic evolution of growth hormone in primates and emergence of the species specificity of human growth hormone receptor. Mol Biol Evol. 18: 945-953 (2001)
- Nachtigal MW, Nickel BE, Cattini PA.
Pituitary-specific repression of placental members of the human growth hormone gene family. A possible mechanism for locus regulation. J Biol Chem. 268: 8473-8479 (1993)
- Revol De Mendoza A, Esquivel Escobedo D, Martínez Dávila I, Saldaña H.
Expansion and divergence of the GH locus between spider monkey and chimpanzee. Gene. 336: 185-193 (2004)

- Rhesus Macaque Genome Sequencing and Analysis Consortium
Evolutionary and biomedical insights from the rhesus macaque genome.
Science. 316: 222-234 (2007)
- Soares MJ.
The prolactin and growth hormone families: pregnancy-specific hormones/cytokines at the maternal-fetal interface. Reprod Biol Endocrinol. 5: 2-51 (2004)
- Wallis M.
The Molecular Evolution of Pituitary Growth Hormone Prolactin and Placental Lactogen: A Protein Family Showing Variable Rates of Evolution.
J Mol Evol 17: 10-18 (1981)
- Wallis M.
Variable evolutionary rates in the molecular evolution of mammalian growth hormones. J Mol Evol. 38: 619-627 (1994)
- Wallis M.
The molecular evolution of vertebrate growth hormones: a pattern of near-stasis interrupted by sustained bursts of rapid change. J Mol Evol. 43: 93-100 (1996)
- Wallis M.
Function switching as a basis for bursts of rapid change during the evolution of pituitary growth hormone. J Mol Evol. 44: 348-350 (1997)
- Wallis OC, Zhang YP, Wallis M.
Molecular evolution of GH in primates: characterisation of the GH genes from slow loris and marmoset defines an episode of rapid evolutionary change. J Mol Endocrinol. 26: 249-258 (2001)
- Wallis OC, Wallis M.
*Characterisation of the GH gene cluster in a new-world monkey, the marmoset (*Callithrix jacchus*).* J Mol Endocrinol. 29: 89-97 (2002)
- Wallis OC, Wallis M.
*Evolution of growth hormone in primates: the GH gene clusters of the New World monkeys marmoset (*Callithrix jacchus*) and white-fronted capuchin (*Cebus albifrons*).* J Mol Evol. 63: 591-601 (2006)
- Wallis M.
*Mammalian genome projects reveal new growth hormone (GH) sequences. Characterization of the GH-encoding genes of armadillo (*Dasypus**

novemcinctus), hedgehog (*Erinaceus europaeus*), bat (*Myotis lucifugus*), hyrax (*Procavia capensis*), shrew (*Sorex araneus*), ground squirrel (*Spermophilus tridecemlineatus*), elephant (*Loxodonta africana*), cat (*Felis catus*) and opossum (*Monodelphis domestica*). *Gen Comp Endocrinol.* 155: 271-279 (2008)

Ye C, Li Y, Shi P, Zhang YP.

Molecular evolution of growth hormone gene family in old world monkeys and hominoids. *Gene.* 350: 183-192 (2005)

Anhang

- A1: Vergleich aller *GH1*-Promotorregionen der untersuchten Schimpansen mit der orthologen Region des Menschen
- A2: Vergleich aller *GH1*-Promotorregionen der untersuchten Makaken mit der orthologen Region beim Menschen

A1:

Vergleich aller *GH1*-Promotorregionen der untersuchten Schimpansen (Angabe der Kodierungen) mit der orthologen Region des Menschen mit Hilfe des Programms ClustalW. Basenangaben: A(denin), C(ytidin), G(uanin), T(hymin). **Rot** markiert sind alle polymorphen Stellen des Schimpansen (als N angezeigt), **grün** die des Menschen. Positionsangaben beziehen sich auf den Transkriptionsstart des menschlichen *GH1* bei +1. In der jeweils untersten Zeile bedeuten die Sternchen Übereinstimmung aller Sequenzen an dieser Stelle, Lücken Abweichungen einer oder mehrerer Sequenzen. Verwendete Sequenz des Menschen: Genom Browser der UCSC (University of Santa Cruz California) Chromosom 17 (Minusstrang), Position 59.350,595 bis 59.349,846 (Version NCBI 36.1, März 2006).

```
CLUSTAL W (1.83) multiple sequence alignment
ptr8      -----TTCGTCCTCAGGGACCTGGGGGAGCCCC
ptr46     -----CCTTGAGTCTGGGTTCTTCGTCCTCAGGGACCTGGGGGAGCCCC
ptr86     -----TCTTCGTCCTCAGGGACCTGGGGGAGCCCC
ptr88     -----TTCGTCCTCAGGGACCTGGGGGAGCCCC
ptr89     -----TGGGTTCTTCGTCCTCAGGGACCTGGGGGAGCCCC
ptr90     -----TTCGTCCTCAGGGACCTGGGGGAGCCCC
ptr91     -----GGAGCCCC
prt92     -----TTCGTCCTCAGGGACCTGGGGGAGCCCC
ptr94     -----TGGGTTCTTCGTCCTCAGGGACCTGGGGGAGCCCC
ptr96     -----CTGGGGCAATCCTTGAGTCTGGGTTCTTCGTCCTCAGGGACCTGGGGGAGCCCC
ptr133    -----CTGGGGCAATCCTTGAGTCTGGGTTCTTCGTCCTCAGGGACCTGGGGGAGCCCC
ptr519    -----CCTTGAGTCTGGGTTCTTCGTCCTCAGGGACCTGGGGGAGCCCC
ptr582    -----CCTTGAGTCTGGGTTCTTCGTCCTCAGGGACCTGGGGGAGCCCC
ptr694    -----CCTTGAGTCTGGGTTCTTCGTCCTCAGGGACCTGGGGGAGCCCC
ptr828    -----GGGGCATCCTTGAGTCTGGGTTCTTCGTCCTCAGGGACCTGGGGGAGCCCC
ptr835    -----CCTTGAGTCTGGGTTCTTCGTCCTCAGGGACCTGGGGGAGCCCC
Mensch   GGAGCTGGGGCAATACCTTGAGTCTGGGTTCTTCGTCCTCAGGGACCTGGGGGAGCCCCA
                                                *****
```

```
ptr8      GCAACGCTCAGGGAAAGGGGAGAGCAAAGTGTGGGGTTGGTTCTCTCTAGTGGTCAGTGT
ptr46     GCAACGCTCAGGGAAAGGGGAGAGCAAAGTGTGGGGTTGGTTCTCTCTAGTGGTCAGTGT
ptr86     GCAACGCTCAGGGAAAGGGGAGAGCAAAGTGTGGGGTTGGTTCTCTCTAGTGGTCAGTGT
ptr88     GCAACGCTCAGGGAAAGGGGAGAGCAAAGTGTGGGGTTGGTTCTCTCTAGTGGTCAGTGT
ptr89     GCAACGCTCAGGGAAAGGGGAGAGCAAAGTGTGGGGTTGGTTCTCTCTAGTGGTCAGTGT
ptr90     GCAACGCTCAGGGAAAGGGGAGAGCAAAGTGTGGGGTTGGTTCTCTCTAGTGGTCAGTGT
ptr91     GCAACGCTCAGGGAAAGGGGAGAGCAAAGTGTGGGGTTGGTTCTCTCTAGTGGTCAGTGT
prt92     GCAACGCTCAGGGAAAGGGGAGAGCAAAGTGTGGGGTTGGTTCTCTCTAGTGGTCAGTGT
ptr94     GCAACGCTCAGGGAAAGGGGAGAGCAAAGTGTGGGGTTGGTTCTCTCTAGTGGTCAGTGT
ptr96     GCAACGCTCAGGGAAAGGGGAGAGCAAAGTGTGGGGTTGGTTCTCTCTAGTGGTCAGTGT
ptr133    GCAACGCTCAGGGAAAGGGGAGAGCAAAGTGTGGGGTTGGTTCTCTCTAGTGGTCAGTGT
ptr519    GCAACGCTCAGGGAAAGGGGAGAGCAAAGTGTGGGGTTGGTTCTCTCTAGTGGTCAGTGT
ptr582    GCAACGCTCAGGGAAAGGGGAGAGCAAAGTGTGGGGTTGGTTCTCTCTAGTGGTCAGTGT
ptr694    GCAACGCTCAGGGAAAGGGGAGAGCAAAGTGTGGGGTTGGTTCTCTCTAGTGGTCAGTGT
ptr828    GCAACGCTCAGGGAAAGGGGAGAGCAAAGTGTGGGGTTGGTTCTCTCTAGTGGTCAGTGT
ptr835    GCAACGCTCAGGGAAAGGGGAGAGCAAAGTGTGGGGTTGGTTCTCTCTAGTGGTCAGTGT
Mensch   GCAATGCTCAGGGAAAGGGGAGAGCAAAGTGTGGGGTTGGTTCTCTCTAGTGGTCAGTGT
****  *****
```

ptr8 TGGAACTGCATCCAGCTGACTCAGGCTGACCCAGGAGTCGTCAGCAGAAGTGGAAATTCAG
ptr46 TGGAACTGCATCCAGCTGACTCAGGCTGACCCAGGAGTCGTCAGCAGAAGTGGAAATTCAG
ptr86 TGGAACTGCATCCAGCTGACTCAGGCTGACCCAGGAGTCGTCAGCAGAAGTGGAAATTCAG
ptr88 TGGAACTGCATCCAGCTGACTCAGGCTGACCCAGGAGTCGTCAGCAGAAGTGGAAATTCAG
ptr89 TGGAACTGCATCCAGCTGACTCAGGCTGACCCAGGAGTCGTCAGCAGAAGTGGAAATTCAG
ptr90 TGGAACTGCATCCAGCTGACTCAGGCTGACCCAGGAGTCGTCAGCAGAAGTGGAAATTCAG
ptr91 TGGAACTGCATCCAGCTGACTCAGGCTGACCCAGGAGTCGTCAGCAGAAGTGGAAATTCAG
prt92 TGGAACTGCATCCAGCTGACTCAGGCTGACCCAGGAGTCGTCAGCAGAAGTGGAAATTCAG
ptr94 TGGAACTGCATCCAGCTGACTCAGGCTGACCCAGGAGTCGTCAGCAGAAGTGGAAATTCAG
ptr96 TGGAACTGCATCCAGCTGACTCAGGCTGACCCAGGAGTCGTCAGCAGAAGTGGAAATTCAG
ptr133 TGGAACTGCATCCAGCTGACTCAGGCTGACCCAGGAGTCGTCAGCAGAAGTGGAAATTCAG
ptr519 TGGAACTGCATCCAGCTGACTCAGGCTGACCCAGGAGTCGTCAGCAGAAGTGGAAATTCAG
ptr582 TGGAACTGCATCCAGCTGACTCAGGCTGACCCAGGAGTCGTCAGCAGAAGTGGAAATTCAG
ptr694 TGGAACTGCATCCAGCTGACTCAGGCTGACCCAGGAGTCGTCAGCAGAAGTGGAAATTCAG
ptr828 TGGAACTGCATCCAGCTGACTCAGGCTGACCCAGGAGTCGTCAGCAGAAGTGGAAATTCAG
ptr835 TGGAACTGCATCCAGCTGACTCAGGCTGACCCAGGAGTCGTCAGCAGAAGTGGAAATTCAG
Mensch TGGAACTGCATCCAGCTGACTCAGGCTGACCCAGGAGTCCTCAGCAGAAGTGGAAATTCAG
* * *****

-544/-542

ptr8 GACTGAATCTGCTCACAACCCCCACAATCTATTGGCTGTGCTTGGCCCCTTTTCCCAAC
ptr46 GACTGAATCTGCTCACAACCCCCACAATCTATTGGCTGTGCTTGGCCCCTTTTCCCAAC
ptr86 GACTGAATCTGCTCACAACCCCCACAATCTATTGGCTGTGCTTGGCCCCTTTTCCCAAC
ptr88 GACTGAATCTGCTCACAACCCCCACAATCTATTGGCTGTGCTTGGCCCCTTTTCCCAAC
ptr89 GACTGAATCTGCTCACAACCCCCACAATCTATTGGCTGTGCTTGGCCCCTTTTCCCAAC
ptr90 GACTGAATCTGCTCACAACCCCCACAATCTATTGGCTGTGCTTGGCCCCTTTTCCCAAC
ptr91 GACTGAATCTGCTCACAACCCCCACAATCTATTGGCTGTGCTTGGCCCCTTTTCCCAAC
prt92 GACTGAATCTGCTCACAACCCCCACAATCTATTGGCTGTGCTTGGCCCCTTTTCCCAAC
ptr94 GACTGAATCTGCTCACAACCCCCACAATCTATTGGCTGTGCTTGGCCCCTTTTCCCAAC
ptr96 GACTGAATCTGCTCACAACCCCCACAATCTATTGGCTGTGCTTGGCCCCTTTTCCCAAC
ptr133 GACTGAATCTGCTCACAACCCCCACAATCTATTGGCTGTGCTTGGCCCCTTTTCCCAAC
ptr519 GACTGAATCTGCTCACAACCCCCACAATCTATTGGCTGTGCTTGGCCCCTTTTCCCAAC
ptr582 GACTGAATCTGCTCACAACCCCCACAATCTATTGGCTGTGCTTGGCCCCTTTTCCCAAC
ptr694 GACTGAATCTGCTCACAACCCCCACAATCTATTGGCTGTGCTTGGCCCCTTTTCCCAAC
ptr828 GACTGAATCTGCTCACAACCCCCACAATCTATTGGCTGTGCTTGGCCCCTTTTCCCAAC
ptr835 GACTGAATCTGCTCACAACCCCCACAATCTATTGGCTGTGCTTGGCCCCTTTTCCCAAC
Mensch GACTGAATCTGCTCACAACCCCCACAATCTATTGGCTGTGCTTGGCCCCTTTTCCCAAC

-476

ptr8 ACACACATTCTGTCTGGTGGGTGGAGGGGAAACATGCGGGGAGGAGGAAAGGAATAGGAT
ptr46 ACACACATTCTGTCTGGTGGGTGGAGGGGAAACATGCGGGGAGGAGGAAAGGAATAGGAT
ptr86 ACACACATTCTGTCTGGTGGGTGGAGGGGAAACATGCGGGGAGGAGGAAAGGAATAGGAT
ptr88 ACACACATTCTGTCTGGTGGGTGGAGGGGAAACATGCGGGGAGGAGGAAAGGAATAGGAT
ptr89 ACACACATTCTGTCTGGTGGGTGGAGGGGAAACATGCGGGGAGGAGGAAAGGAATAGGAT
ptr90 ACACACATTCTGTCTGGTGGGTGGAGGGGAAACATGCGGGGAGGAGGAAAGGAATAGGAT
ptr91 ACACACATTCTGTCTGGTGGGTGGAGGGGAAACATGCGGGGAGGAGGAAAGGAATAGGAT
prt92 ACACACATTCTGTCTGGTGGGTGGAGGGGAAACATGCGGGGAGGAGGAAAGGAATAGGAT
ptr94 ACACACATTCTGTCTGGTGGGTGGAGGGGAAACATGCGGGGAGGAGGAAAGGAATAGGAT
ptr96 ACACACATTCTGTCTGGTGGGTGGAGGGGAAACATGCGGGGAGGAGGAAAGGAATAGGAT
ptr133 ACACACATTCTGTCTGGTGGGTGGAGGGGAAACATGCGGGGAGGAGGAAAGGAATAGGAT
ptr519 ACACACATTCTGTCTGGTGGGTGGAGGGGAAACATGCGGGGAGGAGGAAAGGAATAGGAT
ptr582 ACACACATTCTGTCTGGTGGGTGGAGGGGAAACATGCGGGGAGGAGGAAAGGAATAGGAT
ptr694 ACACACATTCTGTCTGGTGGGTGGAGGGGAAACATGCGGGGAGGAGGAAAGGAATAGGAT
ptr828 ACACACATTCTGTCTGGTGGGTGGAGGGGAAACATGCGGGGAGGAGGAAAGGAATAGGAT
ptr835 ACACACATTCTGTCTGGTGGGTGGAGGGGAAACATGCGGGGAGGAGGAAAGGAATAGGAT
Mensch ACACACATTCTGTCTGGTGGGTGGAGGTTAAACATGCGGGGAGGAGGAAAGGGATAGGAT

ptr8 AGAGAGTGGGATGTGGTCCGTAGGGG-TCTCAAGGACTGGCTATCCTGACATCCTTCCCC
ptr46 AGAGAGTGGGATGTGGTCCGTAGGGG-TCTCAAGGACTGGCTATCCTGACATCCTTCCCC
ptr86 AGAGAGTGGGATGTGGTCCGTAGGGG-TCTCAAGGACTGGCTATCCTGACATCCTTCCCC
ptr88 AGAGAGTGGGATGTGGTCCGTAGGGG-TCTCAAGGACTGGCTATCCTGACATCCTTCCCC
ptr89 AGAGAGTGGGATGTGGTCCGTAGGGG-TCTCAAGGACTGGCTATCCTGACATCCTTCCCC
ptr90 AGAGAGTGGGATGTGGTCCGTAGGGG-TCTCAAGGACTGGCTATCCTGACATCCTTCCCC
ptr91 AGAGAGTGGGATGTGGTCCGTAGGGG-TCTCAAGGACTGGCTATCCTGACATCCTTCCCC

prt92 AGAGAGTGGGATGTGGTCGGGAGGGG-TCTCAAGGACTGGCTATCCTGACATCCTTCCCC
ptr94 AGAGAGTGGGATGTGGTCGGTAGGGG-TCTCAAGGACTGGCTATCCTGACACCTTCCCC
ptr96 AGAGAGTGGGATGTGGTCGGCAGGGG-TCTCAAGGACTGGCTATCCTGACATCCTTCCCC
ptr133 AGAGAGTGGGATGTGGTCGGTAGGGG-TCTCAAGACTGGCTATCCTGACATCCTTCCCC
ptr519 AGAGAGTGGGATGTGGTCGGTAGGGG-TCTCAAGGACTGGCTATCCTGACATCCTTCCCC
ptr582 AGAGAGTGGGATGTGGTCGGTAGGGG-TCTCAAGGACTGGCTATCCTGACACCTTCCCC
ptr694 AGAGAGTGGGATGTGGTCGGTAGGGG-TCTCAAGAACTGGCTATCCTGACATCCTTCCCC
ptr828 AGAGAGTGGGATGTGGTCGGTAGGGG-TCTCAAGGACTGGCTATCCTGACATCCTTCCCC
ptr835 AGAGAGTGGGATGTGGTCGGTAGGGG-TCTCAAGACTGGCTATCCTGACATCCTTCCCC
Mensch AGAATGGGATGTGGTCGGTAGGGGTCTCAAGGACTGGCTATCCTGACATCCTTCCCC

-364 -345 -339 -331 -314 -308

ptr8 GCGTGCAGGTTGGCCACCATGGCCTGCTGCCAGAGGGCACCACCTGCCCTTAAAGAGA
ptr46 GCGTGCAGGTTGGCCACCATGGCCTGCTGCCAGAGGGCACCACCTGCCCTTAAAGAGA
ptr86 GCGTGCAGGTTGGCCACCATGGCCTGCTGCCAGAGGGCACCACCTGCCCTTAAAGAGA
ptr88 GCGTGCAGGTTGGCCACCATGGCCTGCTGCCAGAGGGCACCACCTGCCCTTAAAGAGA
ptr89 GCGTGCAGGTTGGCCACCATGGCCTGCTGCCAGAGGGCACCACCTGCCCTTAAAGAGA
ptr90 GCGTGCAGGTTGGCCACCATGGCCTGCTGCCAGAGGGCACCACCTGCCCTTAAAGAGA
ptr91 GCGTGCAGGTTGGCCACCATGGCCTGCTGCCAGAGGGCACCACCTGCCCTTAAAGAGA
prt92 GCGTGCAGGTTGGCCACCATGGCCTGCTGCCAGAGGGCACCACCTGCCCTTAAAGAGA
ptr94 GCGTGCAGGTTGGCCACCATGGCCTGCTGCCAGAGGGCACCACCTGCCCTTAAAGAGA
ptr96 GCGTGCAGGTTGGCCACCATGGCCTGCTGCCAGAGGGCACCACCTGCCCTTAAAGAGA
ptr133 GCGTGCAGGTTGGCCACCATGGCCTGCTGCCAGAGGGCACCACCTGCCCTTAAAGAGA
ptr519 GCGTGCAGGTTGGCCACCATGGCCTGCTGCCAGAGGGCACCACCTGCCCTTAAAGAGA
ptr582 GCGTGCAGGTTGGCCACCATGGCCTGCTGCCAGAGGGCACCACCTGCCCTTAAAGAGA
ptr694 GCGTGCAGGTTGGCCACCATGGCCTGCTGCCAGAGGGCACCACCTGCCCTTAAAGAGA
ptr828 GCGTGCAGGTTGGCCACCATGGCCTGCTGCCAGAGGGCACCACCTGCCCTTAAAGAGA
ptr835 GCGTGCAGGTTGGCCACCATGGCCTGCTGCCAGAGGGCACCACCTGCCCTTAAAGAGA
Mensch GCGTGCAGGTTGGCCACCATGGCCTGCTGCCAGAGGGCACCACCTGCCCTTAAAGAGA

-301 -278 -258/-256

ptr8 GGACAAGTTGGGTGGTATCTCTGGCTGACACTCTGTGCACAACCCTCACAACGCTGGTGA
ptr46 GGACAAGTTGGGTGGTATCTCTGGCTGACACTCTGTGCACAACCCTCACAACGCTGGTGA
ptr86 GGACAAGTTGGGTGGTATCTCTGGCTGACACTCTGTGCACAACCCTCACAACGCTGGTGA
ptr88 GGACAAGTTGGGTGGTATCTCTGGCTGACACTCTGTGCACAACCCTCACAACGCTGGTGA
ptr89 GGACAAGTTGGGTGGTATCTCTGGCTGACACTCTGTGCACAACCCTCACAACGCTGGTGA
ptr90 GGACAAGTTGGGTGGTATCTCTGGCTGACACTCTGTGCACAACCCTCACAACGCTGGTGA
ptr91 GGACAAGTTGGGTGGTATCTCTGGCTGACACTCTGTGCACAACCCTCACAACGCTGGTGA
prt92 GGACAAGTTGGGTGGTATCTCTGGCTGACACTCTGTGCACAACCCTCACAACGCTGGTGA
ptr94 GGACAAGTTGGGTGGTATCTCTGGCTGACACTCTGTGCACAACCCTCACAACGCTGGTGA
ptr96 GGACAAGTTGGGTGGTATCTCTGGCTGACACTCTGTGCACAACCCTCACAACGCTGGTGA
ptr133 GGACAAGTTGGGTGGTATCTCTGGCTGACACTCTGTGCACAACCCTCACAACGCTGGTGA
ptr519 GGACAAGTTGGGTGGTATCTCTGGCTGACACTCTGTGCACAACCCTCACAACGCTGGTGA
ptr582 GGACAAGTTGGGTGGTATCTCTGGCTGACACTCTGTGCACAACCCTCACAACGCTGGTGA
ptr694 GGACAAGTTGGGTGGTATCTCTGGCTGACACTCTGTGCACAACCCTCACAACGCTGGTGA
ptr828 GGACAAGTTGGGTGGTATCTCTGGCTGACACTCTGTGCACAACCCTCACAACGCTGGTGA
ptr835 GGACAAGTTGGGTGGTATCTCTGGCTGACACTCTGTGCACAACCCTCACAACGCTGGTGA
Mensch GGACAAGTTGGGTGGTATCTCTGGCTGACACTCTGTGCACAACCCTCACAACGCTGGTGA

ptr8 CGGTGGGAAGGGAAAGATGACAAGCCAGGGGGCATGATCCCAGCATGTGTGGGAGGAGCT
ptr46 CGGTGGGAAGGGAAAGATGACAAGCCAGGGGGCATGATCCCAGCATGTGTGGGAGGAGCT
ptr86 CGGTGGGAAGGGAAAGATGACAAGCCAGGGGGCATGATCCCAGCATGTGTGGGAGGAGCT
ptr88 CGGTGGGAAGGGAAAGATGACAAGCCAGGGGGCATGATCCCAGCATGTGTGGGAGGAGCT
ptr89 CGGTGGGAAGGGAAAGATGACAAGCCAGGGGGCATGATCCCAGCATGTGTGGGAGGAGCT
ptr90 CGGTGGGAAGGGAAAGATGACAAGCCAGGGGGCATGATCCCAGCATGTGTGGGAGGAGCT
ptr91 CGGTGGGAAGGGAAAGATGACAAGCCAGGGGGCATGATCCCAGCATGTGTGGGAGGAGCT
prt92 CGGTGGGAAGGGAAAGATGACAAGCCAGGGGGCATGATCCCAGCATGTGTGGGAGGAGCT
ptr94 CGGTGGGAAGGGAAAGATGACAAGCCAGGGGGCATGATCCCAGCATGTGTGGGAGGAGCT
ptr96 CGGTGGGAAGGGAAAGATGACAAGCCAGGGGGCATGATCCCAGCATGTGTGGGAGGAGCT
ptr133 CGGTGGGAAGGGAAAGATGACAAGCCAGGGGGCATGATCCCAGCATGTGTGGGAGGAGCT
ptr519 CGGTGGGAAGGGAAAGATGACAAGCCAGGGGGCATGATCCCAGCATGTGTGGGAGGAGCT
ptr582 CGGTGGGAAGGGAAAGATGACAAGCCAGGGGGCATGATCCCAGCATGTGTGGGAGGAGCT
ptr694 CGGTGGGAAGGGAAAGATGACAAGCCAGGGGGCATGATCCCAGCATGTGTGGGAGGAGCT

ptr828 CGGTGGGAAGGGAAAGATGACAAGCCAGGGGGCATGATCCCAGCATGTGTGGGAGGAGCT
 ptr835 CGGTGGGAAGGGAAAGATGACAAGCCAGGGGGCATGATCCCAGCATGTGTGGGAGGAGCT
 Mensch CGGTGGGAAGGGAAAGA GACAAGCCAGGGGGCATGATCCCAGCATGTGTGGGAGGAGCT

 -168 -164

ptr8 TCTAAATTATCCATTAGCACAAAGCCCCTCAGTGGCCCCATGCATAAAATGTACACAGAAAC
 ptr46 TCTAAATTATCCATTAGCACAAAGCCCCTCAGTGGCCCCATGCATAAAATGTACACAGAAAC
 ptr86 TCTAAATTATCCATTAGCACAAAGCCCCTCAGTGGCCCCATGCATAAAATGTACACAGAAAC
 ptr88 TCTAAATTATCCATTAGCACAAAGCCCCTCAGTGGCCCCATGCATAAAATGTACACAGAAAC
 ptr89 TCTAAATTATCCATTAGCACAAAGCCCCTCAGTGGCCCCATGCATAAAATGTACACAGAAAC
 ptr90 TCTAAATTATCCATTAGCACAAAGCCCCTCAGTGGCCCCATGCATAAAATGTACACAGAAAC
 ptr91 TCTAAATTATCCATTAGCACAAAGCCCCTCAGTGGCCCCATGCATAAAATGTACACAGAAAC
 prt92 TCTAAATTATCCATTAGCACAAAGCCCCTCAGTGGCCCCATGCATAAAATGTACACAGAAAC
 ptr94 TCTAAATTATCCATTAGCACAAAGCCCCTCAGTGGCCCCATGCATAAAATGTACACAGAAAC
 ptr96 TCTAAATTATCCATTAGCACAAAGCCCCTCAGTGGCCCCATGCATAAAATGTACACAGAAAC
 ptr133 TCTAAATTATCCATTAGCACAAAGCCCCTCAGTGGCCCCATGCATAAAATGTACACAGAAAC
 ptr519 TCTAAATTATCCATTAGCACAAAGCCCCTCAGTGGCCCCATGCATAAAATGTACACAGAAAC
 ptr582 TCTAAATTATCCATTAGCACAAAGCCCCTCAGTGGCCCCATGCATAAAATGTACACAGAAAC
 ptr694 TCTAAATTATCCATTAGCACAAAGCCCCTCAGTGGCCCCATGCATAAAATGTACACAGAAAC
 ptr828 TCTAAATTATCCATTAGCACAAAGCCCCTCAGTGGCCCCATGCATAAAATGTACACAGAAAC
 ptr835 TCTAAATTATCCATTAGCACAAAGCCCCTCAGTGGCCCCATGCATAAAATGTACACAGAAAC
 Mensch TCTAAATTATCCATTAGCACAAAGCCCCTCAGTGGCCCCATGCATAAAATGTACACAGAAAC

 -75

ptr8 AGGTGGGGGCAACAG GAGAGAGAAGGGGCCAGGGTATAAAAAGGGCCCACAAGAGACCA
 ptr46 AGGTGGGGGCAACAGT GAGAGAGAAGGGGCCAGGGTATAAAAAGGGCCCACAAGAGACCA
 ptr86 AGGTGGGGGCAACAGCGAGAGAGAAGGGGCCAGGGTATAAAAAGGGCCCACAAGAGACCA
 ptr88 AGGTGGGGGCAACAG GAGAGAGAAGGGGCCAGGGTATAAAAAGGGCCCACAAGAGACCA
 ptr89 AGGTGGGGGCAACAGCGAGAGAGAAGGGGCCAGGGTATAAAAAGGGCCCACAAGAGACCA
 ptr90 AGGTGGGGGCAACAG GAGAGAGAAGGGGCCAGGGTATAAAAAGGGCCCACAAGAGACCA
 ptr91 AGGTGGGGGCAACAG GAGAGAGAAGGGGCCAGGGTATAAAAAGGGCCCACAAGAGACCA
 prt92 AGGTGGGGGCAACAG GAGAGAGAAGGGGCCAGGGTATAAAAAGGGCCCACAAGAGACCA
 ptr94 AGGTGGGGGCAACAG GAGAGAGAAGGGGCCAGGGTATAAAAAGGGCCCACAAGAGACCA
 ptr96 AGGTGGGGGCAACAGCGAGAGAGAAGGGGCCAGGGTATAAAAAGGGCCCACAAGAGACCA
 ptr133 AGGTGGGGGCAACAGT GAGAGAGAAGGGGCCAGGGTATAAAAAGGGCCCACAAGAGACCA
 ptr519 AGGTGGGGGCAACAG GAGAGAGAAGGGGCCAGGGTATAAAAAGGGCCCACAAGAGACCA
 ptr582 AGGTGGGGGCAACAG GAGAGAGAAGGGGCCAGGGTATAAAAAGGGCCCACAAGAGACCA
 ptr694 AGGTGGGGGCAACAGT GAGAGAGAAGGGGCCAGGGTATAAAAAGGGCCCACAAGAGACCA
 ptr828 AGGTGGGGGCAACAGT GAGAGAGAAGGGGCCAGGGTATAAAAAGGGCCCACAAGAGACCA
 ptr835 AGGTGGGGGCAACAGT GAGAGAGAAGGGGCCAGGGTATAAAAAGGGCCCACAAGAGACCA
 Mensch AGGTGGGG CAACAGTGGGAGAGAAGGGGCCAGG TATAAAAAGGGCCCACAAGAGACCA

 -57 -50 -31 -6

ptr8 GCTCAAGGATCCCAAGACCCAACCTCCCA NAACCACTCAGGGTCTGTGGA-----
 ptr46 GCTCAAGGATCCCAAGACCCAACCTCCCA NAACCACTCAGGGTCTGTGGACAGCTCACCT
 ptr86 GCTCAAGGATCCCAAGACCCAACCTCCCA NAACCACTCAGGGTCTGTGGACAGCTCACCT
 ptr88 GCTCAAGGATCCCAAGACCCAACCTCCCA NAACCACTCAGGGTCTGTG-----
 ptr89 GCTCAAGGATCCCAAGACCCAACCTCCCA NAACCACTCAGGGTCTGTG-----
 ptr90 GCTCAAGGATCCCAAGACCCAACCTCCCA NAACCACTCAGGGTCTGTGG-----
 ptr91 GCTCAAGGATCCCAAGACCCAACCTCCCA NAACCACTCAGGGTCTGTGGAC-----
 prt92 GCTCAAGGATCCCAAGACCCAACCTCCCA NAACCACTCAGGGTCTGTGGACA-----
 ptr94 GCTCAAGGATCCCAAGACCCAACCTCCCA NAACCACTCAGGGTCTGTGGACAG-----
 ptr96 GCTCAAGGATCCCAAGACCCAACCTCCCA NAACCACTCAGGGTCTGTGGACAGCTCACCT
 ptr133 GCTCAAGGATCCCAAGACCCAACCTCCCA NAACCACTCAGGGTCTGTGGACAGCTCAC--
 ptr519 GCTCAAGGATCCCAAGACCCAACCTCCCA NAACCACTCAGGGTCTGTGGACAGCTCA---
 ptr582 GCTCAAGGATCCCAAGACCCAACCTCCCA NAACCACTCAGGGTCTGTGGACAGCTCAC--
 ptr694 GCTCAAGGATCCCAAGACCCAACCTCCCA NAACCACTCAGGGTCTGTGGACAGCTCACCT
 ptr828 GCTCAAGGATCCCAAGACCCAACCTCCCA NAACCACTCAGGGTCTGTGGACAGCTCACCT
 ptr835 GCTCAAGGATCCCAAGACCCAACCTCCCA NAACCACTCAGGGTCTGTGGACAGCTCACCT
 Mensch GCTCAAGGATCCCAAGGCCC ACTCCCG NAACCACTCAGGGTCTGTGGACAGCTCACCT

 -1 +3 +16 24/25

```

ptr8      -----
ptr46     AGCTGCAATG-----
ptr86     -----
ptr88     -----
ptr89     -----
ptr90     -----
ptr91     -----
prt92     -----
ptr94     -----
ptr96     -----
ptr133    -----
ptr519    -----
ptr582    -----
ptr694    A-----
ptr828    AGCTGCA-----
ptr835    AGCTGCAA-----
Mensch    AGC[G]CAATGGCTACAGGTAAGCGCCCTA
          +59

```

A2:

Vergleich aller *GH1*-Promotorregionen der untersuchten Makaken (Angabe der Kodierungen) mit der orthologen Region des Menschen mit Hilfe des Programms ClustalW. Basenangaben: A(denin), C(ytidin), G(uanin), T(hymin). Grün markiert alle polymorphen Stellen des Menschen. Positionsangaben beziehen sich auf den Transkriptionsstart des menschlichen *GH1* bei +1. mit der orthologen Region beim Menschen mit Hilfe des Programms ClustalW. Verwendete Sequenz des Menschen: Genom Browser der UCSC (University of Santa Cruz California), Chromosom 17 (Minusstrang), Basen 59.350,595 bis 59.349,846 (Version NCBI 36.1, März 2006).

```

CLUSTAL W (1.83) multiple sequence alignment
00A      -----TGCATCCAGCTGCCTCAGGCGGGCCCAGGAGGAGTCAACAAAAGTGGGAATTCAG
00B      -----CCTCAGGCGGGCCCAGGAGGAGTCAACAAAAGTGGGAATTCAG
00G      -----TCCAGCTGCCTCAGGCGGGCCCAGGAGGAGTCAACAAAAGTGGGAATTCAG
01F      -----CCAGCTGCCTCAGGCGGGCCCAGGAGGAGTCAACAAAAGTGGGAATTCAG
02D      -----CATCCAGCTGCCTCAGGCGGGCCCAGGAGGAGTCAACAAAAGTGGGAATTCAG
04A      TAGCACTGCATCCAGCTGCCTCAGGCGGGCCCAGGAGGAGTCAACAAAAGTGGGAATTCAG
04F      -----CCAGCTGCCTCAGGCGGGCCCAGGAGGAGTCAACAAAAGTGGGAATTCAG
05C      -----GCATCCAGCTGCCTCAGGCGGGCCCAGGAGGAGTCAACAAAAGTGGGAATTCAG
05D      -----CCAGCTGCCTCAGGCGGGCCCAGGAGGAGTCAACAAAAGTGGGAATTCAG
05E      -----TCCAGCTGCCTCAGGCGGGCCCAGGAGGAGTCAACAAAAGTGGGAATTCAG
06A      -----TGCATCCAGCTGCCTCAGGCGGGCCCAGGAGGAGTCAACAAAAGTGGGAATTCAG
06D      -----TGCATCCAGCTGCCTCAGGCGGGCCCAGGAGGAGTCAACAAAAGTGGGAATTCAG
06E      -----TCCAGCTGCCTCAGGCGGGCCCAGGAGGAGTCAACAAAAGTGGGAATTCAG
06G      -----CTGCATCCAGCTGCCTCAGGCGGGCCCAGGAGGAGTCAACAAAAGTGGGAATTCAG
07A      -----TGCATCCAGCTGCCTCAGGCGGGCCCAGGAGGAGTCAACAAAAGTGGGAATTCAG
07C      -----CAGCTGCCTCAGGCGGGCCCAGGAGGAGTCAACAAAAGTGGGAATTCAG
07F      -----CCAGCTGCCTCAGGCGGGCCCAGGAGGAGTCAACAAAAGTGGGAATTCAG
08A      -----TCCAGCTGCCTCAGGCGGGCCCAGGAGGAGTCAACAAAAGTGGGAATTCAG
08F      -----AGCTGCCTCAGGCGGGCCCAGGAGGAGTCAACAAAAGTGGGAATTCAG
08G      -----CAGCTGCCTCAGGCGGGCCCAGGAGGAGTCAACAAAAGTGGGAATTCAG
09G      -----CTGCCTCAGGCGGGCCCAGGAGGAGTCAACAAAAGTGGGAATTCAG
10G      -----AGCTGCCTCAGGCGGGCCCAGGAGGAGTCAACAAAAGTGGGAATTCAG
11H      -----TGCATCCAGCTGCCTCAGGCGGGCCCAGGAGGAGTCAACAAAAGTGGGAATTCAG
12D      -----AGCTGCCTCAGGCGGGCCCAGGAGGAGTCAACAAAAGTGGGAATTCAG
13C      ----ACTGCATCCAGCTGCCTCAGGCGGGCCCAGGAGGAGTCAACAAAAGTGGGAATTCAG
HSA      TGGAACTGCATCCAGCTGACTCAGGCTGACCCAGGAGTCTCAGCAGAAGTGGGAATTCAG
          ***** * ***** ** ** *****

```

00A GACTGAATCGTTCCCAGAACCCCCACAATCTATTGCCTGTGCTTGGCCCCTTTTTCACAGC
00B GACTGAATCGTTCCCAGAACCCCCACAATCTATTGCCTGTGCTTGGCCCCTTTTTCACAGC
00G GACTGAATCGTTCCCAGAACCCCCACAATCTATTGCCTGTGCTTGGCCCCTTTTTCACAGC
01F GACTGAATCGTTCCCAGAACCCCCACAATCTATTGCCTGTGCTTGGCCCCTTTTTCACAGC
02D GACTGAATCGTTCCCAGAACCCCCACAATCTATTGCCTGTGCTTGGCCCCTTTTTCACAGC
04A GACTGAATCGTTCCCAGAACCCCCACAATCTATTGCCTGTGCTTGGCCCCTTTTTCACAGC
04F GACTGAATCGTTCCCAGAACCCCCACAATCTATTGCCTGTGCTTGGCCCCTTTTTCACAGC
05C GACTGAATCGTTCCCAGAACCCCCACAATCTATTGCCTGTGCTTGGCCCCTTTTTCACAGC
05D GACTGAATCGTTCCCAGAACCCCCACAATCTATTGCCTGTGCTTGGCCCCTTTTTCACAGC
05E GACTGAATCGTTCCCAGAACCCCCACAATCTATTGCCTGTGCTTGGCCCCTTTTTCACAGC
06A GACTGAATCGTTCCCAGAACCCCCACAATCTATTGCCTGTGCTTGGCCCCTTTTTCACAGC
06D GACTGAATCGTTCCCAGAACCCCCACAATCTATTGCCTGTGCTTGGCCCCTTTTTCACAGC
06E GACTGAATCGTTCCCAGAACCCCCACAATCTATTGCCTGTGCTTGGCCCCTTTTTCACAGC
06G GACTGAATCGTTCCCAGAACCCCCACAATCTATTGCCTGTGCTTGGCCCCTTTTTCACAGC
07A GACTGAATCGTTCCCAGAACCCCCACAATCTATTGCCTGTGCTTGGCCCCTTTTTCACAGC
07C GACTGAATCGTTCCCAGAACCCCCACAATCTATTGCCTGTGCTTGGCCCCTTTTTCACAGC
07F GACTGAATCGTTCCCAGAACCCCCACAATCTATTGCCTGTGCTTGGCCCCTTTTTCACAGC
08A GACTGAATCGTTCCCAGAACCCCCACAATCTATTGCCTGTGCTTGGCCCCTTTTTCACAGC
08F GACTGAATCGTTCCCAGAACCCCCACAATCTATTGCCTGTGCTTGGCCCCTTTTTCACAGC
08G GACTGAATCGTTCCCAGAACCCCCACAATCTATTGCCTGTGCTTGGCCCCTTTTTCACAGC
09G GACTGAATCGTTCCCAGAACCCCCACAATCTATTGCCTGTGCTTGGCCCCTTTTTCACAGC
10G GACTGAATCGTTCCCAGAACCCCCACAATCTATTGCCTGTGCTTGGCCCCTTTTTCACAGC
11H GACTGAATCGTTCCCAGAACCCCCACAATCTATTGCCTGTGCTTGGCCCCTTTTTCACAGC
12D GACTGAATCGTTCCCAGAACCCCCACAATCTATTGCCTGTGCTTGGCCCCTTTTTCACAGC
13C GACTGAATCGTTCCCAGAACCCCCACAATCTATTGCCTGTGCTTGGCCCCTTTTTCACAGC
HSA GACTGAATCGTTGCTCACAACCCCCACAATCTATTGGCTGTGCTTGGCCCCTTTTCCCAAC
***** * ** ***** ***** ***** ***** ***** ***** ***** *****

-476

00A ACACACATTCTCTCTGCTGGGTGGAGGGGAAACATGTGGGGAGGAGGAAAGGGATAGGAT
00B ACACACATTCTCTCTGCTGGGTGGAGGGGAAACATGTGGGGAGGAGGAAAGGGATAGGAT
00G ACACACATTCTCTCTGCTGGGTGGAGGGGAAACATGTGGGGAGGAGGAAAGGGATAGGAT
01F ACACACATTCTCTCTGCTGGGTGGAGGGGAAACATGTGGGGAGGAGGAAAGGGATAGGAT
02D ACACACATTCTCTCTGCTGGGTGGAGGGGAAACATGTGGGGAGGAGGAAAGGGATAGGAT
04A ACACACATTCTCTCTGCTGGGTGGAGGGGAAACATGTGGGGAGGAGGAAAGGGATAGGAT
04F ACACACATTCTCTCTGCTGGGTGGAGGGGAAACATGTGGGGAGGAGGAAAGGGATAGGAT
05C ACACACATTCTCTCTGCTGGGTGGAGGGGAAACATGTGGGGAGGAGGAAAGGGATAGGAT
05D ACACACATTCTCTCTGCTGGGTGGAGGGGAAACATGTGGGGAGGAGGAAAGGGATAGGAT
05E ACACACATTCTCTCTGCTGGGTGGAGGGGAAACATGTGGGGAGGAGGAAAGGGATAGGAT
06A ACACACATTCTCTCTGCTGGGTGGAGGGGAAACATGTGGGGAGGAGGAAAGGGATAGGAT
06D ACACACATTCTCTCTGCTGGGTGGAGGGGAAACATGTGGGGAGGAGGAAAGGGATAGGAT
06E ACACACATTCTCTCTGCTGGGTGGAGGGGAAACATGTGGGGAGGAGGAAAGGGATAGGAT
06G ACACACATTCTCTCTGCTGGGTGGAGGGGAAACATGTGGGGAGGAGGAAAGGGATAGGAT
07A ACACACATTCTCTCTGCTGGGTGGAGGGGAAACATGTGGGGAGGAGGAAAGGGATAGGAT
07C ACACACATTCTCTCTGCTGGGTGGAGGGGAAACATGTGGGGAGGAGGAAAGGGATAGGAT
07F ACACACATTCTCTCTGCTGGGTGGAGGGGAAACATGTGGGGAGGAGGAAAGGGATAGGAT
08A ACACACATTCTCTCTGCTGGGTGGAGGGGAAACATGTGGGGAGGAGGAAAGGGATAGGAT
08F ACACACATTCTCTCTGCTGGGTGGAGGGGAAACATGTGGGGAGGAGGAAAGGGATAGGAT
08G ACACACATTCTCTCTGCTGGGTGGAGGGGAAACATGTGGGGAGGAGGAAAGGGATAGGAT
09G ACACACATTCTCTCTGCTGGGTGGAGGGGAAACATGTGGGGAGGAGGAAAGGGATAGGAT
10G ACACACATTCTCTCTGCTGGGTGGAGGGGAAACATGTGGGGAGGAGGAAAGGGATAGGAT
11H ACACACATTCTCTCTGCTGGGTGGAGGGGAAACATGTGGGGAGGAGGAAAGGGATAGGAT
12D ACACACATTCTCTCTGCTGGGTGGAGGGGAAACATGTGGGGAGGAGGAAAGGGATAGGAT
13C ACACACATTCTCTCTGCTGGGTGGAGGGGAAACATGTGGGGAGGAGGAAAGGGATAGGAT
HSA ACACACATTCTGTCTGGTGGGTGGAGGTTAAACATGCGGGGAGGAGGAAAGGGATAGGAT
***** ***** ***** ***** ***** ***** ***** ***** ***** *****

00A AGAGAGTGGGATGGGGTTCGGTAGGGG-TCTCAAGGACTGGCTATCCTGACATCCTTCCCC
00B AGAGAGTGGGATGGGGTTCGGTAGGGG-TCTCAAGGACTGGCTATCCTGACATCCTTCCCC

00G AGAGAGTGGGATGGGGTTCGGTAGGGG-TCTCAAGGACTGGCTATCCTGACATCCTTCCCC
01F AGAGAGTGGGATGGGGTTCGGTAGGGG-TCTCAAGGACTGGCTATCCTGACATCCTTCCCC
02D AGAGAGTGGGATGGGGTTCGGTAGGGG-TCTCAAGGACTGGCTATCCTGACATCCTTCCCC
04A AGAGAGTGGGATGGGGTTCGGTAGGGG-TCTCAAGGACTGGCTATCCTGACATCCTTCCCC
04F AGAGAGTGGGATGGGGTTCGGTAGGGG-TCTCAAGGACTGGCTATCCTGACATCCTTCCCC
05C AGAGAGTGGGATGGGGTTCGGTAGGGG-TCTCAAGGACTGGCTATCCTGACATCCTTCCCC
05D AGAGAGTGGGATGGGGTTCGGTAGGGG-TCTCAAGGACTGGCTATCCTGACATCCTTCCCC
05E AGAGAGTGGGATGGGGTTCGGTAGGGG-TCTCAAGGACTGGCTATCCTGACATCCTTCCCC
06A AGAGAGTGGGATGGGGTTCGGTAGGGG-TCTCAAGGACTGGCTATCCTGACATCCTTCCCC
06D AGAGAGTGGGATGGGGTTCGGTAGGGG-TCTCAAGGACTGGCTATCCTGACATCCTTCCCC
06E AGAGAGTGGGATGGGGTTCGGTAGGGG-TCTCAAGGACTGGCTATCCTGACATCCTTCCCC
06G AGAGAGTGGGATGGGGTTCGGTAGGGG-TCTCAAGGACTGGCTATCCTGACATCCTTCCCC
07A AGAGAGTGGGATGGGGTTCGGTAGGGG-TCTCAAGGACTGGCTATCCTGACATCCTTCCCC
07C AGAGAGTGGGATGGGGTTCGGTAGGGG-TCTCAAGGACTGGCTATCCTGACATCCTTCCCC
07F AGAGAGTGGGATGGGGTTCGGTAGGGG-TCTCAAGGACTGGCTATCCTGACATCCTTCCCC
08A AGAGAGTGGGATGGGGTTCGGTAGGGG-TCTCAAGGACTGGCTATCCTGACATCCTTCCCC
08F AGAGAGTGGGATGGGGTTCGGTAGGGG-TCTCAAGGACTGGCTATCCTGACATCCTTCCCC
08G AGAGAGTGGGATGGGGTTCGGTAGGGG-TCTCAAGGACTGGCTATCCTGACATCCTTCCCC
09G AGAGAGTGGGATGGGGTTCGGTAGGGG-TCTCAAGGACTGGCTATCCTGACATCCTTCCCC
10G AGAGAGTGGGATGGGGTTCGGTAGGGG-TCTCAAGGACTGGCTATCCTGACATCCTTCCCC
11H AGAGAGTGGGATGGGGTTCGGTAGGGG-TCTCAAGGACTGGCTATCCTGACATCCTTCCCC
12D AGAGAGTGGGATGGGGTTCGGTAGGGG-TCTCAAGGACTGGCTATCCTGACATCCTTCCCC
13C AGAGAGTGGGATGGGGTTCGGTAGGGG-TCTCAAGGACTGGCTATCCTGACATCCTTCCCC
HSA AGAAGTGGGATGTGGTTCGGTAGGGG-TCTCAAGGACTGGCTATCCTGACATCCTTCCC

-364 -339 -308

00A GCGTTCAGGTTGGCCACCATGGCGTGTGCCAGAGGACACCCACCTGCTCCTTAAGGAGA
00B GCGTTCAGGTTGGCCACCATGGCGTGTGCCAGAGGACACCCACCTGCTCCTTAAGGAGA
00G GCGTTCAGGTTGGCCACCATGGCGTGTGCCAGAGGACACCCACCTGCTCCTTAAGGAGA
01F GCGTTCAGGTTGGCCACCATGGCGTGTGCCAGAGGACACCCACCTGCTCCTTAAGGAGA
02D GCGTTCAGGTTGGCCACCATGGCGTGTGCCAGAGGACACCCACCTGCTCCTTAAGGAGA
04A GCGTTCAGGTTGGCCACCATGGCGTGTGCCAGAGGACACCCACCTGCTCCTTAAGGAGA
04F GCGTTCAGGTTGGCCACCATGGCGTGTGCCAGAGGACACCCACCTGCTCCTTAAGGAGA
05C GCGTTCAGGTTGGCCACCATGGCGTGTGCCAGAGGACACCCACCTGCTCCTTAAGGAGA
05D GCGTTCAGGTTGGCCACCATGGCGTGTGCCAGAGGACACCCACCTGCTCCTTAAGGAGA
05E GCGTTCAGGTTGGCCACCATGGCGTGTGCCAGAGGACACCCACCTGCTCCTTAAGGAGA
06A GCGTTCAGGTTGGCCACCATGGCGTGTGCCAGAGGACACCCACCTGCTCCTTAAGGAGA
06D GCGTTCAGGTTGGCCACCATGGCGTGTGCCAGAGGACACCCACCTGCTCCTTAAGGAGA
06E GCGTTCAGGTTGGCCACCATGGCGTGTGCCAGAGGACACCCACCTGCTCCTTAAGGAGA
06G GCGTTCAGGTTGGCCACCATGGCGTGTGCCAGAGGACACCCACCTGCTCCTTAAGGAGA
07A GCGTTCAGGTTGGCCACCATGGCGTGTGCCAGAGGACACCCACCTGCTCCTTAAGGAGA
07C GCGTTCAGGTTGGCCACCATGGCGTGTGCCAGAGGACACCCACCTGCTCCTTAAGGAGA
07F GCGTTCAGGTTGGCCACCATGGCGTGTGCCAGAGGACACCCACCTGCTCCTTAAGGAGA
08A GCGTTCAGGTTGGCCACCATGGCGTGTGCCAGAGGACACCCACCTGCTCCTTAAGGAGA
08F GCGTTCAGGTTGGCCACCATGGCGTGTGCCAGAGGACACCCACCTGCTCCTTAAGGAGA
08G GCGTTCAGGTTGGCCACCATGGCGTGTGCCAGAGGACACCCACCTGCTCCTTAAGGAGA
09G GCGTTCAGGTTGGCCACCATGGCGTGTGCCAGAGGACACCCACCTGCTCCTTAAGGAGA
10G GCGTTCAGGTTGGCCACCATGGCGTGTGCCAGAGGACACCCACCTGCTCCTTAAGGAGA
11H GCGTTCAGGTTGGCCACCATGGCGTGTGCCAGAGGACACCCACCTGCTCCTTAAGGAGA
12D GCGTTCAGGTTGGCCACCATGGCGTGTGCCAGAGGACACCCACCTGCTCCTTAAGGAGA
13C GCGTTCAGGTTGGCCACCATGGCGTGTGCCAGAGGACACCCACCTGCTCCTTAAGGAGA
HSA GCGTTCAGGTTGGCCACCATGGCGTGTGCCAGAGGACACCCACCTGCTCCTTAAGGAGA

-301 -278

00A GGACAAGTTGGGTGGTATCTCTGGCCGACACTCTGTGCACAGCCCTCACAATGCTGGTGA
00B GGACAAGTTGGGTGGTATCTCTGGCCGACACTCTGTGCACAGCCCTCACAATGCTGGTGA
00G GGACAAGTTGGGTGGTATCTCTGGCCGACACTCTGTGCACAGCCCTCACAATGCTGGTGA

01F GGACAAGTTGGGTGGTATCTCTGGCCGACACTCTGTGCACAGCCCTCACAATGCTGGTGA
02D GGACAAGTTGGGTGGTATCTCTGGCCGACACTCTGTGCACAGCCCTCACAATGCTGGTGA
04A GGACAAGTTGGGTGGTATCTCTGGCCGACACTCTGTGCACAGCCCTCACAATGCTGGTGA
04F GGACAAGTTGGGTGGTATCTCTGGCCGACACTCTGTGCACAGCCCTCACAATGCTGGTGA
05C GGACAAGTTGGGTGGTATCTCTGGCCGACACTCTGTGCACAGCCCTCACAATGCTGGTGA
05D GGACAAGTTGGGTGGTATCTCTGGCCGACACTCTGTGCACAGCCCTCACAATGCTGGTGA
05E GGACAAGTTGGGTGGTATCTCTGGCCGACACTCTGTGCACAGCCCTCACAATGCTGGTGA
06A GGACAAGTTGGGTGGTATCTCTGGCCGACACTCTGTGCACAGCCCTCACAATGCTGGTGA
06D GGACAAGTTGGGTGGTATCTCTGGCCGACACTCTGTGCACAGCCCTCACAATGCTGGTGA
06E GGACAAGTTGGGTGGTATCTCTGGCCGACACTCTGTGCACAGCCCTCACAATGCTGGTGA
06G GGACAAGTTGGGTGGTATCTCTGGCCGACACTCTGTGCACAGCCCTCACAATGCTGGTGA
07A GGACAAGTTGGGTGGTATCTCTGGCCGACACTCTGTGCACAGCCCTCACAATGCTGGTGA
07C GGACAAGTTGGGTGGTATCTCTGGCCGACACTCTGTGCACAGCCCTCACAATGCTGGTGA
07F GGACAAGTTGGGTGGTATCTCTGGCCGACACTCTGTGCACAGCCCTCACAATGCTGGTGA
08A GGACAAGTTGGGTGGTATCTCTGGCCGACACTCTGTGCACAGCCCTCACAATGCTGGTGA
08F GGACAAGTTGGGTGGTATCTCTGGCCGACACTCTGTGCACAGCCCTCACAATGCTGGTGA
08G GGACAAGTTGGGTGGTATCTCTGGCCGACACTCTGTGCACAGCCCTCACAATGCTGGTGA
09G GGACAAGTTGGGTGGTATCTCTGGCCGACACTCTGTGCACAGCCCTCACAATGCTGGTGA
10G GGACAAGTTGGGTGGTATCTCTGGCCGACACTCTGTGCACAGCCCTCACAATGCTGGTGA
11H GGACAAGTTGGGTGGTATCTCTGGCCGACACTCTGTGCACAGCCCTCACAATGCTGGTGA
12D GGACAAGTTGGGTGGTATCTCTGGCCGACACTCTGTGCACAGCCCTCACAATGCTGGTGA
13C GGACAAGTTGGGTGGTATCTCTGGCCGACACTCTGTGCACAGCCCTCACAATGCTGGTGA
HSA GGACAAGTTGGGTGGTATCTCTGGCTGACACTCTGTGCACAACCCTCACAACACTGGTGA

00A CGGTGAGAAGGGAAAAGATGACAAGCCAGGGGGCATGATCCCAGCATGTGTGGGAGGAGCT
00B CGGTGAGAAGGGAAAAGATGACAAGCCAGGGGGCATGATCCCAGCATGTGTGGGAGGAGCT
00G CGGTGAGAAGGGAAAAGATGACAAGCCAGGGGGCATGATCCCAGCATGTGTGGGAGGAGCT
01F CGGTGAGAAGGGAAAAGATGACAAGCCAGGGGGCATGATCCCAGCATGTGTGGGAGGAGCT
02D CGGTGAGAAGGGAAAAGATGACAAGCCAGGGGGCATGATCCCAGCATGTGTGGGAGGAGCT
04A CGGTGAGAAGGGAAAAGATGACAAGCCAGGGGGCATGATCCCAGCATGTGTGGGAGGAGCT
04F CGGTGAGAAGGGAAAAGATGACAAGCCAGGGGGCATGATCCCAGCATGTGTGGGAGGAGCT
05C CGGTGAGAAGGGAAAAGATGACAAGCCAGGGGGCATGATCCCAGCATGTGTGGGAGGAGCT
05D CGGTGAGAAGGGAAAAGATGACAAGCCAGGGGGCATGATCCCAGCATGTGTGGGAGGAGCT
05E CGGTGAGAAGGGAAAAGATGACAAGCCAGGGGGCATGATCCCAGCATGTGTGGGAGGAGCT
06A CGGTGAGAAGGGAAAAGATGACAAGCCAGGGGGCATGATCCCAGCATGTGTGGGAGGAGCT
06D CGGTGAGAAGGGAAAAGATGACAAGCCAGGGGGCATGATCCCAGCATGTGTGGGAGGAGCT
06E CGGTGAGAAGGGAAAAGATGACAAGCCAGGGGGCATGATCCCAGCATGTGTGGGAGGAGCT
06G CGGTGAGAAGGGAAAAGATGACAAGCCAGGGGGCATGATCCCAGCATGTGTGGGAGGAGCT
07A CGGTGAGAAGGGAAAAGATGACAAGCCAGGGGGCATGATCCCAGCATGTGTGGGAGGAGCT
07C CGGTGAGAAGGGAAAAGATGACAAGCCAGGGGGCATGATCCCAGCATGTGTGGGAGGAGCT
07F CGGTGAGAAGGGAAAAGATGACAAGCCAGGGGGCATGATCCCAGCATGTGTGGGAGGAGCT
08A CGGTGAGAAGGGAAAAGATGACAAGCCAGGGGGCATGATCCCAGCATGTGTGGGAGGAGCT
08F CGGTGAGAAGGGAAAAGATGACAAGCCAGGGGGCATGATCCCAGCATGTGTGGGAGGAGCT
08G CGGTGAGAAGGGAAAAGATGACAAGCCAGGGGGCATGATCCCAGCATGTGTGGGAGGAGCT
09G CGGTGAGAAGGGAAAAGATGACAAGCCAGGGGGCATGATCCCAGCATGTGTGGGAGGAGCT
10G CGGTGAGAAGGGAAAAGATGACAAGCCAGGGGGCATGATCCCAGCATGTGTGGGAGGAGCT
11H CGGTGAGAAGGGAAAAGATGACAAGCCAGGGGGCATGATCCCAGCATGTGTGGGAGGAGCT
12D CGGTGAGAAGGGAAAAGATGACAAGCCAGGGGGCATGATCCCAGCATGTGTGGGAGGAGCT
13C CGGTGAGAAGGGAAAAGATGACAAGCCAGGGGGCATGATCCCAGCATGTGTGGGAGGAGCT
HSA CGGTGGGAAGGGAAAAGATGACAAGCCAGGGGGCATGATCCCAGCATGTGTGGGAGGAGCT

00A TCTCAATTATCCATTAGCACAGGCACGTCAGTAGCCCCATGCATAAATGTACAGAGAAAC
00B TCTCAATTATCCATTAGCACAGGCACGTCAGTAGCCCCATGCATAAATGTACAGAGAAAC
00G TCTCAATTATCCATTAGCACAGGCACGTCAGTAGCCCCATGCATAAATGTACAGAGAAAC
01F TCTCAATTATCCATTAGCACAGGCACGTCAGTAGCCCCATGCATAAATGTACAGAGAAAC
02D TCTCAATTATCCATTAGCACAGGCACGTCAGTAGCCCCATGCATAAATGTACAGAGAAAC

04A TCTCAATTATCCATTAGCACAGGCACGTCAGTAGCCCCATGCATAAAATGTACAGAGAAAC
04F TCTCAATTATCCATTAGCACAGGCACGTCAGTAGCCCCATGCATAAAATGTACAGAGAAAC
05C TCTCAATTATCCATTAGCACAGGCACGTCAGTAGCCCCATGCATAAAATGTACAGAGAAAC
05D TCTCAATTATCCATTAGCACAGGCACGTCAGTAGCCCCATGCATAAAATGTACAGAGAAAC
05E TCTCAATTATCCATTAGCACAGGCACGTCAGTAGCCCCATGCATAAAATGTACAGAGAAAC
06A TCTCAATTATCCATTAGCACAGGCACGTCAGTAGCCCCATGCATAAAATGTACAGAGAAAC
06D TCTCAATTATCCATTAGCACAGGCACGTCAGTAGCCCCATGCATAAAATGTACAGAGAAAC
06E TCTCAATTATCCATTAGCACAGGCACGTCAGTAGCCCCATGCATAAAATGTACAGAGAAAC
06G TCTCAATTATCCATTAGCACAGGCACGTCAGTAGCCCCATGCATAAAATGTACAGAGAAAC
07A TCTCAATTATCCATTAGCACAGGCACGTCAGTAGCCCCATGCATAAAATGTACAGAGAAAC
07C TCTCAATTATCCATTAGCACAGGCACGTCAGTAGCCCCATGCATAAAATGTACAGAGAAAC
07F TCTCAATTATCCATTAGCACAGGCACGTCAGTAGCCCCATGCATAAAATGTACAGAGAAAC
08A TCTCAATTATCCATTAGCACAGGCACGTCAGTAGCCCCATGCATAAAATGTACAGAGAAAC
08F TCTCAATTATCCATTAGCACAGGCACGTCAGTAGCCCCATGCATAAAATGTACAGAGAAAC
08G TCTCAATTATCCATTAGCACAGGCACGTCAGTAGCCCCATGCATAAAATGTACAGAGAAAC
09G TCTCAATTATCCATTAGCACAGGCACGTCAGTAGCCCCATGCATAAAATGTACAGAGAAAC
10G TCTCAATTATCCATTAGCACAGGCACGTCAGTAGCCCCATGCATAAAATGTACAGAGAAAC
11H TCTCAATTATCCATTAGCACAGGCACGTCAGTAGCCCCATGCATAAAATGTACAGAGAAAC
12D TCTCAATTATCCATTAGCACAGGCACGTCAGTAGCCCCATGCATAAAATGTACAGAGAAAC
13C TCTCAATTATCCATTAGCACAGGCACGTCAGTAGCCCCATGCATAAAATGTACAGAGAAAC
HSA TCTAAATTATCCATTAGCACAAAGCCCGTCAGTGGCCCCATGCATAAAATGTACAGAGAAAC
*** ***** ** ***** *****

00A AGGTGGGGTCACGCAGCGAGAGATAAGGGGCCAGGGTATAAAAAGGGGCCACAAGAGACC
00B AGGTGGGGTCACGCAGCGAGAGATAAGGGGGCCAGGGTATAAAAAGGGGCCACAAGAGACC
00G AGGTGGGGTCACGCAGCGAGAGATAAGGGGGCCAGGGTATAAAAAGGGGCCACAAGAGACC
01F AGGTGGGGTCACGCAGCGAGAGATAAGGGGGCCAGGGTATAAAAAGGGGCCACAAGAGACC
02D AGGTGGGGTCACGCAGCGAGAGATAAGGGGGCCAGGGTATAAAAAGGGGCCACAAGAGACC
04A AGGTGGGGTCACGCAGCGAGAGATAAGGGGGCCAGGGTATAAAAAGGGGCCACAAGAGACC
04F AGGTGGGGTCACGCAGCGAGAGATAAGGGGGCCAGGGTATAAAAAGGGGCCACAAGAGACC
05C AGGTGGGGTCACGCAGCGAGAGATAAGGGGGCCAGGGTATAAAAAGGGGCCACAAGAGACC
05D AGGTGGGGTCACGCAGCGAGAGATAAGGGGGCCAGGGTATAAAAAGGGGCCACAAGAGACC
05E AGGTGGGGTCACGCAGCGAGAGATAAGGGGGCCAGGGTATAAAAAGGGGCCACAAGAGACC
06A AGGTGGGGTCACGCAGCGAGAGATAAGGGGGCCAGGGTATAAAAAGGGGCCACAAGAGACC
06D AGGTGGGGTCACGCAGCGAGAGATAAGGGGGCCAGGGTATAAAAAGGGGCCACAAGAGACC
06E AGGTGGGGTCACGCAGCGAGAGATAAGGGGGCCAGGGTATAAAAAGGGGCCACAAGAGACC
06G AGGTGGGGTCACGCAGCGAGAGATAAGGGGGCCAGGGTATAAAAAGGGGCCACAAGAGACC
07A AGGTGGGGTCACGCAGCGAGAGATAAGGGGGCCAGGGTATAAAAAGGGGCCACAAGAGACC
07C AGGTGGGGTCACGCAGCGAGAGATAAGGGGGCCAGGGTATAAAAAGGGGCCACAAGAGACC
07F AGGTGGGGTCACGCAGCGAGAGATAAGGGGGCCAGGGTATAAAAAGGGGCCACAAGAGACC
08A AGGTGGGGTCACGCAGCGAGAGATAAGGGGGCCAGGGTATAAAAAGGGGCCACAAGAGACC
08F AGGTGGGGTCACGCAGCGAGAGATAAGGGGGCCAGGGTATAAAAAGGGGCCACAAGAGACC
08G AGGTGGGGTCACGCAGCGAGAGATAAGGGGGCCAGGGTATAAAAAGGGGCCACAAGAGACC
09G AGGTGGGGTCACGCAGCGAGAGATAAGGGGGCCAGGGTATAAAAAGGGGCCACAAGAGACC
10G AGGTGGGGTCACGCAGCGAGAGATAAGGGGGCCAGGGTATAAAAAGGGGCCACAAGAGACC
11H AGGTGGGGTCACGCAGCGAGAGATAAGGGGGCCAGGGTATAAAAAGGGGCCACAAGAGACC
12D AGGTGGGGTCACGCAGCGAGAGATAAGGGGGCCAGGGTATAAAAAGGGGCCACAAGAGACC
13C AGGTGGGGTCACGCAGCGAGAGATAAGGGGGCCAGGGTATAAAAAGGGGCCACAAGAGACC
HSA AGGTGGGGTCA-ACAGTGGGAGAGAAGGGGGCCAGGGTATAAAAAGGGGCCACAAGAGACC
***** ** * ** *****

00A AGCTCCAGGATTCCAAGGCCCAACTCCCCGAACCACTCAGGGTCTGTGGACAGCCCACC
00B AGCTCCAGGATTCCAAGGCCCAACTCCCCGAACCACTCAGGGTCTGTGGACAGCCCACC
00G AGCTCCAGGATTCCAAGGCCCAACTCCCCGAACCACTCAGGGTCTGTGGACAGCCCACC
01F AGCTCCAGGATTCCAAGGCCCAACTCCCCGAACCACTCAGGGTCTGTGGACAGCCCACC
02D AGCTCCAGGATTCCAAGGCCCAACTCCCCGAACCACTCAGGGTCTGTGGACAGCCCACC
04A AGCTCCAGGATTCCAAGGCCCAACTCCCCGAACCACTCAGGGTCTGTGGACAGCCCACC

```

04F      AGCTCCAGGATTCCAAGGCCCAACTCCCCGAACCACTCAGGGTCCTGTGGACAGCCCACC
05C      AGCTCCAGGATTCCAAGGCCCAACTCCCCGAACCACTCAGGGTCCTGTGGACAGCCCACC
05D      AGCTCCAGGATTCCAAGGCCCAACTCCCCGAACCACTCAGGGTCCTGTGGACAGCCCACC
05E      AGCTCCAGGATTCCAAGGCCCAACTCCCCGAACCACTCAGGGTCCTGTGGACAGCCCACC
06A      AGCTCCAGGATTCCAAGGCCCAACTCCCCGAACCACTCAGGGTCCTGTGGACAGCCCACC
06D      AGCTCCAGGATTCCAAGGCCCAACTCCCCGAACCACTCAGGGTCCTGTGGACAGCCCACC
06E      AGCTCCAGGATTCCAAGGCCCAACTCCCCGAACCACTCAGGGTCCTGTGGACAGCCCACC
06G      AGCTCCAGGATTCCAAGGCCCAACTCCCCGAACCACTCAGGGTCCTGTGGACAGCCCACC
07A      AGCTCCAGGATTCCAAGGCCCAACTCCCCGAACCACTCAGGGTCCTGTGGACAGCCCACC
07C      AGCTCCAGGATTCCAAGGCCCAACTCCCCGAACCACTCAGGGTCCTGTGGACAGCCCACC
07F      AGCTCCAGGATTCCAAGGCCCAACTCCCCGAACCACTCAGGGTCCTGTGGACAGCCCACC
08A      AGCTCCAGGATTCCAAGGCCCAACTCCCCGAACCACTCAGGGTCCTGTGGACAGCCCACC
08F      AGCTCCAGGATTCCAAGGCCCAACTCCCCGAACCACTCAGGGTCCTGTGGACAGCCCACC
08G      AGCTCCAGGATTCCAAGGCCCAACTCCCCGAACCACTCAGGGTCCTGTGGACAGCCCACC
09G      AGCTCCAGGATTCCAAGGCCCAACTCCCCGAACCACTCAGGGTCCTGTGGACAGCCCACC
10G      AGCTCCAGGATTCCAAGGCCCAACTCCCCGAACCACTCAGGGTCCTGTGGACAGCCCACC
11H      AGCTCCAGGATTCCAAGGCCCAACTCCCCGAACCACTCAGGGTCCTGTGGACAGCCCACC
12D      AGCTCCAGGATTCCAAGGCCCAACTCCCCGAACCACTCAGGGTCCTGTGGACAGCCCACC
13C      AGCTCCAGGATTCCAAGGCCCAACTCCCCGAACCACTCAGGGTCCTGTGGACAGCCCACC
HSA      AGCTCCAGGATTCCAAGGCCCAACTCCCCGAACCACTCAGGGTCCTGTGGACAGCTCACC
          *****
          -6  -1  +3                +16      +25

```

```

00A      TAGCTGCAATGGCTGCAGGTAA-----
00B      TAGCTGCAATGGCTGCAGGTAAGTGCCCCATAAAATCCCTTTCCGC-----
00G      TAGCTGCAATGGCTGCAGGTAA-----
01F      TAGCTGCAATGGCTGCAGGTAAGTGCCCCATAAAATCCCTTTCCGCATGA-----
02D      TAGCTGCAATGGCTGCAGGT-----
04A      TAGCTGCAATGGCTGCAGGTAAGTGCCCCATAAAATCCCTT-----
04F      TAGCTGCAATGGCTGCAGGTAA-----
05C      TAGCTGCAATGGCTGCAGGT-----
05D      TAGCTGCAATGGCTGCAGGTAAGTGCCCCATAAAATCCCTTTCCGCATGA-----
05E      TAGCTGCAATGGCTGCAGGT-----
06A      TAGCTGCAATGGCTGCAGGT-----
06D      TAGCTGCAATGGCTGCAGGTAAGTGCCCCATAAAATCCCTTTCCGCA-----
06E      TAGCTGCAATGGCTGCAGGTAAGTGCCCCATAAAATCCC-----
06G      TAGCTGCAATGGCTGCAGGTAA-----
07A      TAGCTGCAATGGCTGCAGGTAA-----
07C      TAGCTGCAATGGCTGCAGGTAAGTGCCCC-----
07F      TAGCTGCAATGGCTGCAGGTA-----
08A      TAGCTGCAATGGCTGCAGGT-----
08F      TAGCTGCAATGGCTGCAGGTAA-----
08G      TAGCTGCAATGGCTGCAG-----
09G      TAGCTGCAATGGCTGCAGGTAAGTGCCCC-----
10G      TAGCTGCAATGGCTGCAGGTAAGTGCCCCATAAAATC-----
11H      TAGCTGCAATGGCTGCAGGTAAGTGCCCCATAAAATCCCTTTCCGCATG-----
12D      TAGCTGCAATGGCTGCAGGTAA-----
13C      TAGCTGCAATGGCTGCAGGTAA-----
HSA      TAGCTGCAATGGCTACAGGTAAGCGCCCCATAAAATCCCTTTGGGCACAATGTGTCCTGAG
          *****
          +59

```

Danksagung

An erster Stelle danke ich PD Dr. H. Kehrer-Sawatski für die hervorragende Betreuung. Danke an Helene für die umfassende labortechnische Einweisung und dass sie mir immer mit Rat und Tat beiseite stand. Darüber hinaus möchte ich mich bei allen meinen Kollegen bedanken, besonders bei Kathi und Jessica, für ihre fachliche und moralische Unterstützung.

Nicht zuletzt gilt mein Dank auch meiner Familie und meinen Freunden und Kommilitonen für ihre Unterstützung und Geduld.