

# **Modellierung der Tertiärstrukturen lysosomaler Cysteinproteasen, Charakterisierung der katalytisch aktiven Bindungszentren und strukturbasiertes Ligandendesign**



## **Dissertation**

Zur Erlangung des akademischen Grades  
doctor rerum naturalium (Dr. rer. nat.)

vorgelegt der

Mathematisch-Naturwissenschaftlich-Technischen Fakultät  
(mathematisch-naturwissenschaftlicher Bereich)  
der Martin-Luther-Universität Halle-Wittenberg

von Frau Dipl.-Chem. Annett Fengler

geb. am 05.09.1970 in Halle (Saale)

Gutachter:

1. Herr PD Dr. W. Brandt
2. Herr Prof. Dr. D. Brömme
3. Herr Prof. Dr. D. Schomburg

Halle (Saale), den 03.05.2000

Die Arbeiten der vorliegenden Dissertation wurden im Zeitraum von Oktober 1996 bis Dezember 1999 am Institut für Biochemie in der Arbeitsgruppe Molecular Modelling – Theoretische Wirkstoffbiochemie der Martin-Luther-Universität Halle-Wittenberg angefertigt.

Herrn PD Dr. habil. W. Brandt bin ich für die interessante Themenstellung, das fördernde Interesse am Fortgang meiner Arbeit, die stets gewährte Unterstützung und das in mich gesetzte Vertrauen zu besonderem Dank verpflichtet.

Mein spezieller Dank gilt Herrn Prof. Dr. D. Brömme sowie Herrn Prof. Dr. E. Mehler für ihre hilfreiche Unterstützung und die inspirierenden Diskussionen während meines Forschungsaufenthaltes am Department of Physiology and Biophysics der Mount Sinai School of Medicine in New York, USA.

Herrn Dr. F. Bühling danke ich herzlichst für die durchgeführten enzymkinetischen Untersuchungen, für die Unterstützung bei Fragestellung hinsichtlich dieser Untersuchungen sowie für zahlreiche anregende und kritische Hinweise.

Herrn Dr. Faust möchte ich für die Synthese und der Bereitstellung der untersuchten Verbindungen danken.

Allen Mitgliedern der Arbeitsgruppe danke ich für das sehr angenehme Arbeitsklima und insbesondere möchte ich Herrn O. Ludwig und Frau Dr. I. Thondorf für die vielen Tips und den wertvollen Hinweisen bei der Anfertigung dieser Arbeit danken.

Besonderer Dank gilt an dieser Stelle meinen Eltern und meinem Lebensgefährten, die durch ihre Unterstützung großen Anteil an der vorliegenden Arbeit haben und mir stets ein großer Rückhalt waren.

---

# INHALTSVERZEICHNIS

<b>1</b>	<b>EINLEITUNG</b>	<b>1</b>
1.1	Die Cysteinprotease Cathepsin B	5
1.2	Die Cysteinprotease Cathepsin H	6
1.3	Die Cysteinprotease Cathepsin L	7
1.4	Die Cysteinprotease Cathepsin S	7
1.5	Die Cysteinprotease Cathepsin K	8
1.6	Die Cysteinprotease Cathepsin F	9
<b>2</b>	<b>AUFGABENSTELLUNG</b>	<b>10</b>
<b>3</b>	<b>METHODEN</b>	<b>12</b>
3.1	Modellierung von Proteinstrukturen	12
3.1.1	Sequenzanalyse	12
3.1.2	Sekundärstrukturvorhersage	14
3.1.3	Tertiärstrukturvorhersage	15
3.1.3.1	COMPOSER	16
3.1.3.2	MODELLER	19
3.1.4	Optimierung der Strukturmodelle	20
3.1.5	Genauigkeit und Bewertung der Modelle	21
3.1.5.1	Abschätzung der Fehler in Proteinstrukturmodellen	21
3.1.5.1.1	Bestimmung von Mißfaltungen und Fehlern in lokalen Regionen	22
3.1.5.1.2	Bewertung der stereochemischen Qualität der Tertiärstrukturmodelle	23
3.1.6	Protein-Ligand-Docking	24
3.1.6.1	Beurteilung der Affinität des Liganden zum Enzym	25
3.2	Design von Proteinliganden	27
3.2.1	Aufbau und Analyse der Substratbindungsregionen von Cathepsinen	27
3.2.2	Entwicklung von neuen Liganden	29
<b>4</b>	<b>ERGEBNISSE UND DISKUSSION</b>	<b>32</b>
4.1	Sequenzanalyse der Papain-ähnlichen Cysteinproteasen	32
4.2	Modellierung der Tertiärstrukturen der Cathepsine K, S, H und F	35
4.2.1	Ein Modell der dreidimensionalen Struktur von humanem Cathepsin K	35
4.2.1.1	Die Bindungsspalte des Strukturmodells von Cathepsin K	40
4.2.2	Ein Tertiärstrukturmodell von humanem Cathepsin S	42
4.2.2.1	Die Bindungsspalte des Strukturmodells von Cathepsin S	47
4.2.2.2	Docking von Liganden in die katalytisch aktiven Bindungszentren der Strukturmodelle der Cathepsine K und S	48
4.2.3	Ein Tertiärstrukturmodell des humanen Cathepsin H	52
4.2.3.1	Lage der Minikette in der Bindungsspalte von Cathepsin H	57

---

4.2.3.2	Untersuchung der Bindungsspalte des Strukturmodells von Cathepsin H	59
4.2.3.3	Vergleichende Untersuchungen zu Cathepsin H aus Schweineleber	60
4.2.3.4	Substratspezifität von Cathepsin H im Vergleich zu Cathepsin L	61
4.2.4	Ein dreidimensionales Strukturmodell von humanem Cathepsin F	63
4.2.4.1	Untersuchung der Bindungsspalte von Cathepsin F	67
4.2.4.2	Weitere Überprüfungen des Strukturmodells von Cathepsin F	68
4.3	Design neuer Liganden der Cathepsine L, K und F	71
4.3.1	Vergleichende Studien an lysosomalen Cysteinproteasen	71
4.3.1.1	Die elektrostatischen Potentiale der molekularen Oberfläche der Cathepsine L, B, K, H, S und F	74
4.3.1.2	Bestimmung bevorzugter Wechselwirkungen zwischen Ligand und Enzym	79
4.3.2	Erstellung neuer Peptidsequenzen für hochaffine Cathepsin L-Substrate	84
4.3.2.1	Enzymkinetische Untersuchungen	89
4.3.2.2	Verifizierungen der Synthesevorschläge	89
4.3.3	Erstellung von Peptidsequenzen für hochaffine Cathepsin K-Substrate	91
4.3.3.1	Überarbeitung der Substratvorschläge für Cathepsin K	95
4.3.3.2	Enzymkinetische Untersuchungen	98
4.3.4	Erstellung neuer Peptidsequenzen für effiziente Cathepsin F-Inhibitoren	99
<b>5</b>	<b>ZUSAMMENFASSUNG UND AUSBLICK</b>	<b>104</b>
<b>6</b>	<b>LITERATUR</b>	<b>108</b>
<b>7</b>	<b>ANHANG</b>	<b>A-1</b>
7.1	Tabellen	A-1
7.2	Schemata	A-5
7.3	Abbildungen	A-7
7.4	Synthese der Quenchingsubstrate	A-15
7.5	Enzymkinetische Untersuchungen	A-15

## ABKÜRZUNGSVERZEICHNIS UND ERLÄUTERUNGEN

In der Arbeit werden folgende Abkürzungen, Symbole und Bezeichnungen verwendet (alphabetisch geordnet):

1-Nal	3-(1'-Naphthyl)alanin
3D	dreidimensional
aa	amino acid (Aminosäure)
Abu	2-Aminobuttersäure
Abz	2-Aminobenzoessäure (Anthranilsäure)
Ac	N-terminale Acetylgruppe
BANA	Benzyloxycarbonyl-L-arginyl-2-naphthylamid
BLAST	Basic Local Alignment Search Tool (Programm zur Suche nach homologen Proteinen)
Boc	N-terminale t-Butyloxycarbonylgruppe
Bz	N-terminale Benzoylgruppe
CA030	Ethylester von Epoxysuccinyl-L-Ile-L-Pro-OH
CA074	Propylamid von Epoxysuccinyl-L-Ile-L-Pro-OH
CLUSTAL W	Programm zum multiplen Sequenzalignment
COMPOSER	Programm zur Generierung von Proteinstrukturmodellen
DSSP	Database of Secondary Structure in Proteins (Programm zur Ermittlung von Sekundärstrukturelementen in Proteinen anhand des Vergleiches mit Röntgenkristallstrukturen, die in der PDB eingetragen sind)
DTT	1,4-Dithio-DL-threit
E-64	1-[L-N-(trans-Epoxysuccinyl)leucyl]amino-4-guanidinobutan
E-64c	1-[L-N-(trans-Epoxysuccinyl)leucyl]isoamylamid
E.C.	Enzyme Commission Number (Enzymverfügungsnummer)
EDTA	Ethylendiamintetraacetat
EMBL	Datenbank für Sequenzen von Nukleinsäuren
EP	Elektrostatisches Potential
FLEXIDOCK	Dockingprogramm
Fmoc	N-(9-Fluorenyl)methoxycarbonyl
GA	Genetischer Algorithmus
GRID	Molecular Modelling Software zur Berechnung von Wechselwirkungsfeldern
HBTU	2-(1H-Benzotriazol-1-yl)-1,1,3,3,-tetramethyluroniumhexafluorophosphat
hCatB	humanes Cathepsin B
hCatF	humanes Cathepsin F

---

hCatH	humanes Cathepsin H
hCatK	humanes Cathepsin K
hCatL	humanes Cathepsin L
hCatS	humanes Cathepsin S
Hph	Homophenylalanin
$k_{\text{cat}}$	Geschwindigkeitskonstante enzymatischer Reaktionen
$K_{\text{d}}$	Dissoziationskonstante [M]
$K_{\text{i}}$	Inhibitorkonstante, Dissoziationskonstante des Enzym-Inhibitor-Komplexes [M]
$K_{\text{m}}$	Michaelis-Konstante [M]
MCA	7-Amino-4-methyl-cumarin
MD	Moleküldynamik
MetO <sub>2</sub>	Methioninsulfon
MHC-II	Major histocompatibility complex II
MODELLER	Programm zur Modellierung von dreidimensionalen Proteinstrukturen
MOLCAD	MOlecular Computer Aided Design (Programm zur Berechnung verschiedener Potentiale der molekularen Oberflächen von Molekülen)
Mu	4-Morpholincarboxyl
NHMec	7-Amino-4-methyl-cumarin
NHNap	2-Naphthylamid
NMec	7-Amino-4-methyl-cumarin
NMR	Nuclear Magnetic Resonance (Kernspinresonanz)
pCatH	Cathepsin H, isoliert aus Schweineleber
PDB	Brookhaven Protein DataBase
pdfs	Probability density functions (Wahrscheinlichkeitsdichtefunktionen)
Ph	Phenyl
PHD	Profile Network from Heidelberg (Programm zur Sekundärstrukturvorhersage von Proteinen)
Phe(4-NO <sub>2</sub> )	4-Nitrophenylalanin
Pn	Bindungsposition in der Substratsequenz
pNA	p-Nitranilin
PROSA	PROtein Structure Analysis (Programm zur Strukturanalyse von Proteinen)
RMS	Root Mean Square (gemittelte Abweichung von miteinander verglichenen Atomkoordinaten)
SCORE	Programm zur Bestimmung der Bindungsaffinität eines Protein-Ligand-Komplexes
SCR	Structurally Conserved Region (strukturell konservierte Region)
Sn	Bezeichnung der Substratbindungsstellen im Enzym
Suc	N-terminale Succinylgruppe

---

SVR	Structurally Variable Region (strukturell variable Region)
Swiss-Prot	Proteinsequenzdatenbank
SYBYL	Molecular Modelling Software zur Molekülgraphik
TrEMBL	Datenbank, welche die übersetzten Proteinsequenzen der EMBL-Datenbank enthält
TrEMBLNEW	Ergänzung der TrEMBL-Datenbank
Tyr(3-NO <sub>2</sub> )	3-Nitrotyrosin
Z	N-terminale Benzyloxycarbonylgruppe

Wenn es in den aufgeführten farbigen Molekülgraphiken nicht explizit angegeben ist, wurden für die verschiedenen Atomtypen folgende Farben verwendet:

Kohlenstoff : grau

Sauerstoff : rot

Stickstoff : blau

Schwefel : gelb

In den Abbildungen der Protein-Ligand-Komplexe sind nur die Wasserstoffatome dargestellt, die an Wasserstoffbrückenbindungen beteiligt sind. Wasserstoffbrücken werden generell als gestrichelte Linie gekennzeichnet (orange gefärbt).

# 1 EINLEITUNG

Proteasen haben die Eigenschaft, die Hydrolyse von Amidbindungen zu katalysieren. Obwohl Proteine vielen reversiblen posttranslatorischen Modifizierungen während ihrer Lebensdauer unterliegen können, wie z. B. den Phosphorylierungen und den allosterischen Übertragungen, ist die Proteolyse weitestgehend irreversibel. Basierend auf der Natur der Proteolyse, ist es nicht überraschend, daß proteolytische Enzyme an Prozessen beteiligt sind, die häufig irreversibel verlaufen: Koagulation (Gerinnung), Verdauung, Reifung von Zytokinen und Prohormonen, Apoptose und Degradation von intrazellulären Proteinen [1]. Alle Proteasen zeigen einen allgemein vergleichbaren Mechanismus im nukleophilen Angriff auf das Carbonyl-Kohlenstoffatom einer Amidbindung. Im Verlauf eines säure- und basekatalysierten Prozesses wird die kovalente Bindung gespalten [2]. Unterschiedliche Proteasen nutzen zur Generierung der nukleophilen Eigenschaften sowie zur Übertragung dieser Eigenschaften auf die zu spaltende Bindung verschiedene Vorgehensweisen. Diese Differenzierung liefert ein gutes Klassifizierungsschema, wodurch die Proteasen in vier Hauptklassen eingeteilt werden können: Serin-, Cystein-, Aspartat- und Metalloproteasen. Die Cysteinproteasen umfassen eine große und mannigfaltige Gruppe von Enzymen, die von *Rawlings & Barrett* in Clans und Familien klassifiziert wurden [3]. Dabei gehören die Enzyme zu einer Familie, die eine evolutionäre Beziehung untereinander aufweisen. Im Unterschied dazu setzt sich ein Clan aus Gruppen von Familien zusammen, die eine Beziehung aufweisen, ungeachtet dessen, daß signifikante Sequenzähnlichkeiten fehlen könnten. Die größte Familie der Cysteinproteasen, die aufgrund der Sequenzhomologie identifiziert werden konnte, ist die Papain-Superfamilie, deren Mitglieder eine große Enzymbreite von Prokaryoten und Eurokaryoten, einschließlich Pflanzen, Bakterien, wirbellosen Tieren und Wirbeltieren beinhaltet [4]. Papain-ähnliche Cysteinproteasen von Säugern spielen eine Schlüsselrolle in der lysosomalen Proteindegredation, sind aber auch in der extrazellulären Matrix vorhanden [5]. In dieser Arbeit werden die Betrachtungen auf die Enzyme der Papain-Superfamilie beschränkt.

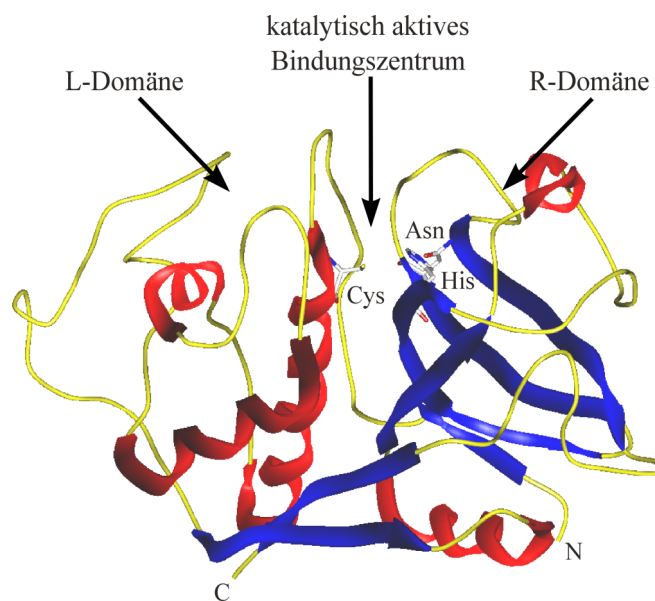
Obwohl der Name "Cathepsin" auch andere Proteasen beinhaltet, wie z. B. Cathepsin G (Serinprotease) und Cathepsin D (Aspartatprotease), sind die meisten Enzyme, die diese Bezeichnung tragen, Cysteinproteasen. Man sieht daher sehr häufig, daß der Name Cathepsin als allgemeiner Begriff für lysosomale Cysteinproteasen benutzt wird. Zum ersten Mal wurden die Cathepsine 1929 beschrieben [6]. Das Wort Cathepsin kommt aus der griechischen Sprache und bedeutet Verdauung.

Funktionelle Analysen dieser Proteasen haben gezeigt, daß sie zusätzlich zum Proteinabbau auch in andere Prozesse involviert sind, so z. B. in die Antigenpräsentation [7], in den Knochenumbau [8] und in die Prohormonaktivierung [9]. Außerdem wird angenommen, daß Cysteinproteasen an einer Reihe von Krankheitsprozessen beteiligt sind, wie z. B. Lungenemphysem [10], Osteoporose [11], Alzheimer Krankheit [12], rheumatische Arthritis [13], Metastasen und Krebs [14]. Deshalb stellt die Entwicklung von Inhibitoren, die die unkontrollierte Aktivität dieser Enzyme unter pathologischen Bedingungen blockieren könnten, einen wichtigen Schwerpunkt in der Therapeutikaentwicklung dar.



Die Röntgenkristallstrukturanalysen von mehreren Mitgliedern der Papain-Superfamilie lieferten Informationen über den strukturellen Aufbau dieser Proteine sowie detaillierte Strukturen der Umgebung der katalytisch aktiven Bindungszentren. Die Bedeutung der Aminosäuren, die signifikant für den katalytischen Mechanismus sind, konnte aufgeklärt werden [15]. Die Bildung eines Thiolatanions ist essentiell für die Katalyse. Dabei ist die Bildung dieses Anions abhängig von der Ionenpaarbildung zwischen dem Cystein des aktiven Zentrums und der benachbarten Aminosäure Histidin [16]. Der Katalysemechanismus der Papain-ähnlichen Cysteinproteasen, wie er von *Storer & Menard* beschrieben wurde, ist im Anhang dargestellt (Schema A-1) [17]. In saurer Umgebung katalysieren die reifen Enzyme die Hydrolyse von zahlreichen Peptidsubstraten und sind an vielen biologischen Prozessen beteiligt.

Struktur- und Sequenzanalysen von Cathepsinen haben gezeigt, daß sie eine sehr ähnliche Aminosäuresequenz sowie ein vergleichbares Faltungsmuster besitzen [18]. Die reifen Enzyme bestehen aus zwei Domänen, wobei die Form an ein leicht geöffnetes Buch mit einem Spalt in der Mitte erinnert. Die Domänen werden durch eine "V"-förmige Spalte getrennt (Abb. 1).



**Abb. 1** Tertiärstruktur von Papain (PDB-Eintrag: 9PAP [19]). Die Aminosäuren der katalytischen Triade Cystein, Histidin und Asparagin sind mit dargestellt. Die Orientierung von Papain wurde so gewählt, wie die Eigenschaften des Enzyms im Text beschrieben werden. Die L-Domäne nimmt die linke Seite der Abbildung ein und die R-Domäne die rechte Seite. N- und C-Terminus des Enzyms sind durch ein N bzw. C gekennzeichnet.

In der Mitte dieser Spalte befinden sich die Aminosäuren Cystein, Histidin und Asparagin, die das katalytisch aktive Zentrum dieser Enzyme bilden. Ein sehr markantes Merkmal für die L-Domäne ist eine zentrale  $\alpha$ -Helix, welche aus ca. 20-30 Aminosäuren besteht. Die R-Domäne enthält neben mehreren  $\beta$ -Faltblattsegmenten auch ein kurzes  $\alpha$ -Helixmotiv [18]. Die Wechselwirkungen zwischen diesen Domänen haben sowohl hydrophoben als auch hydrophilen Charakter und sind spezifisch für jedes einzelne Enzym. Die Aminosäuren Cystein und Histidin können ein Thiolat-Imidazolium-

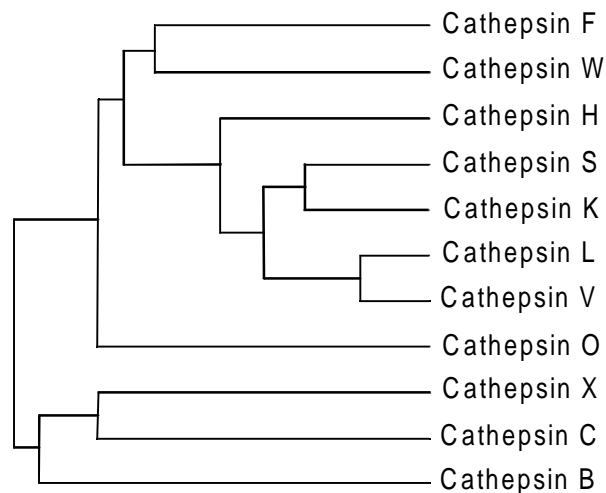
Ionenpaar ausbilden, das wesentlich für die enzymatische Aktivität der Cysteinproteasen ist (Schema A-1). Die Existenz eines Thiolatanions als Bestandteil dieses Ionenpaars wurde über Spektrophotometrie nachgewiesen [20]. Die Protonierung des katalytisch aktiven Histidins wurde mittels NMR-Untersuchungen und anderer Studien bestätigt [20]. Das Asparagin der katalytischen Triade richtet über Wasserstoffbrückenbindung den Imidazolring des Histidins aus. Eine weitere Aminosäure, die essentiell für die Katalyse ist, ist ein Glutamin, dessen Seitenkette zur Stabilisierung des während der Katalyse gebildeten Oxyanions beiträgt [21].

Es ist erwiesen, daß die Papain-ähnlichen Cysteinproteasen als Präproenzyme vorliegen, die zu den korrespondierenden Proenzymen umgewandelt werden. Für die Cathepsine hat sich gezeigt, daß das Propeptid als potenter, pH-abhängiger Inhibitor wirkt [22]. Das Propeptid ist außerdem essentiell für die korrekte Faltung und Stabilisierung von Cysteinproteasen [23,24]. Die Abspaltung des Propeptides mit gleichzeitiger Aktivierung des Enzyms basiert auf der Autokatalyse unter sauren Bedingungen [25]. *In vitro*-Experimente über die Reifung lysosomaler Cysteinproteasen haben gezeigt, daß die Abspaltung des Propeptides auch durch Cathepsin D, Pepsin, Serinproteasen, Metalloproteasen sowie durch Cysteinproteasen katalysiert werden kann [26].

Trotz ihrer großen Ähnlichkeit unterscheiden sich die Papain-ähnlichen Cysteinproteasen in der Zellverteilung und in einigen enzymatischen Eigenschaften, insbesondere der Substratspezifität und pH-Stabilität. In der Regel wurden sie als Endopeptidasen charakterisiert. Allerdings haben Untersuchungen ergeben, daß Cathepsin B zusätzlich die Eigenschaften einer Dipeptidylcarboxypeptidase und Cathepsin H die einer Aminopeptidase besitzen [27]. Es ist deshalb von großem Interesse, für diese Cathepsine die Rolle der Aminosäuresubstitutionen im katalytisch aktiven Bindungszentrum zu untersuchen, die diese Divergenzen der proteolytischen Aktivitäten hervorrufen.

In den letzten Jahren ist die Anzahl der beschriebenen humanen Cysteinproteasen, die zur Papain-Superfamilie C1 gehören, drastisch gestiegen. Insgesamt 11 Familienmitglieder wurden bezüglich ihrer Aminosäuresequenz charakterisiert. Zu diesen Cysteinproteasen gehören Cathepsin B [28], Cathepsin L [29,30], Cathepsin H [31,32], Cathepsin S [33,34], Cathepsin C [34,35], Cathepsin O [36], Cathepsin K [37,38], Cathepsin W [39], Cathepsin F [40,41], Cathepsin V [42] und Cathepsin X [43]. Diese Enzyme konnten aus zahlreichen Zelltypen isoliert werden, obwohl einige von ihnen eine selektive (Cathepsin S) [44] bzw. spezifische (Cathepsin K) [8] Zellverteilung haben. Die Cathepsine, die bisher am besten charakterisiert wurden, sind die Cathepsine B und L. Resultierend aus den Ergebnissen der Sequenzanalysen und Strukturcharakteristika der Cathepsine konnte festgestellt werden, daß diese Proteine in 2 Unterfamilien eingeteilt werden können [45]. Die Cathepsine L, K, V, S, H, O, W und F gehören zu der größeren Cathepsin L-Unterfamilie. Die Cathepsine B, C und X bilden die Cathepsin B-Unterfamilie. Zusätzlich zu einigen Unterschieden in der Enzymsequenz gibt es gravierende Divergenzen innerhalb dieser beiden Unterfamilien in der Sequenz und der Länge der jeweiligen Proregionen. Cathepsin B-ähnliche Enzyme besitzen Proregionen, die sich aus ca. 60

Aminosäuren zusammensetzen, wohingegen die Proregionen von Cathepsin L-ähnlichen Proteasen ca. 100 Aminosäuren enthalten [46]. Basierend auf diesen Studien läßt sich ein "genetischer Stammbaum" zur Bestimmung der evolutionären Beziehungen zwischen den einzelnen Cathepsinen konstruieren (Abb. 2) [40]. Anhand dieser Abbildung läßt sich die Einteilung der Cathepsine in die beiden Unterfamilien erkennen. Außerdem ist zu sehen, daß die beiden erst kürzlich beschriebenen Cathepsine F und W eine weitere neue Untergruppe der Cathepsin L-Unterfamilie bilden.



**Abb. 2** Schematische Darstellung der evolutionären Beziehungen zwischen den humanen Papain-ähnlichen Cysteinproteasen [40].

Die Kenntnis der Substratbindungsregionen ( $S_n$ ) der Cysteinproteasen basiert auf kinetischen Untersuchungen und Röntgenkristallstrukturen von substratähnlichen Inhibitoren, die im katalytisch aktiven Zentrum der Enzyme gebunden sind [47]. *Schechter & Berger* haben beschrieben, daß die Substratbindungsregion der lysosomalen Cysteinproteasen fähig ist, bis zu sieben Reste eines Substrates zu erkennen [47]. Von den sieben entsprechenden Bindungspositionen ( $S1-S4$  und  $S1'-S3'$ ) konnten fünf ( $S1-S3$  und  $S1'-S2'$ ) strukturell charakterisiert werden [18].

Ein Großteil der Tertiärstrukturen von Cathepsin-Komplexen mit synthetischen Inhibitoren, die über Röntgenkristallstrukturanalyse ermittelt werden konnten, ist zur Aufklärung der Eigenschaften der Bindungspositionen geeignet. Die Inhibitoren sind kovalent an das Cystein der katalytischen Triade gebunden. Kristallstrukturen mit Substratanaloga, z. B. den Chlormethylketonen im Komplex mit Papain gaben aufschlußreiche Informationen über die Bindungspositionen  $S1$  und  $S2$  sowie über die Lage der  $P1$ - und  $P2$ -Reste auf der Proteinoberfläche [47,48]. Die  $S1'$ - und  $S2'$ -Bindungsstellen der Cathepsine wurden auf der Basis der Kristallstruktur von Cathepsin B komplexiert mit CA030, einem Inhibitor mit einer reaktiven Epoxysuccinylgruppe, näher definiert [49]. Die Strukturen zeigen, daß die Peptide entlang der aktiven Bindungsspalte in einer gestreckten Konformation binden. Es wurde dabei gefunden, daß die Substratbindungspositionen auf der linken und rechten Seite der Bindungsspalte lokalisiert sind (Abb. A-1).

Bisher konnten die dreidimensionalen Proteinstrukturen von mehreren Cysteinproteasen, die zur Papain-Superfamilie C1 gehören, mittels Röntgenkristallstrukturanalyse aufgeklärt werden. Ihre strukturellen Daten sind über die Brookhaven Proteindatenbank (PDB) [50,51] zugänglich (Tab. A-1). Darüber hinaus werden zahlreiche Kristallstrukturen von Enzym-Inhibitor-Komplexen der Cathepsine B und K sowie für Papain beschrieben (Tab. A-1).

## 1.1 Die Cysteinprotease Cathepsin B

Cathepsin B (E.C. 3.4.22.1) ist eine der am meisten untersuchten lysosomalen Cysteinprotease. Cathepsin B wird von verschiedenen Zelltypen, einschließlich Makrophagen, Fibroblasten und Osteoklasten sezerniert. Cathepsin B wurde zum ersten Mal aus der Rindermilz isoliert [52]. Die sekretierten Formen sind hauptsächlich Proenzyme, die sich von der reifen Form unterscheiden. Die Präproform des humanen Cathepsin B (hCatB) besteht aus 339 Aminosäuren, entsprechend einer Molmasse von 39 kDa (Tab. A-2). Auf das N-terminale Signalpeptid folgt eine Prosequenz. Das Propeptid von Cathepsin B ist unter leicht sauren Bedingungen ein potenter langsam bindender Inhibitor für dieses Enzym (pH-Wert = 6.0) [22]. Katalytisch aktives Cathepsin B wird durch spezielle Zellen und Zellkulturen bei einem pH-Wert unter 7.0 ausgeschieden. Die reife Form des humanen Cathepsin B besteht aus 254 Aminosäuren. Cathepsin B hat eine breite Substratspezifität und spaltet zahlreiche synthetische Substrate bei einem pH-Wertbereich von 5-6 [26]. Im Gegensatz zu anderen Papain-ähnlichen Cathepsinen zeigt Cathepsin B sowohl Exo- als auch Endopeptidaseaktivität. Zusätzlich zur Hydrolyse von Peptidsubstraten kann dieses Enzym darüber hinaus Dipeptideinheiten vom C-Terminus abspalten (Dipeptidylaktivität) [5].

*Musil et al.* gelang es, die Kristallstruktur des humanen Cathepsin B mit 2.15 Å Auflösung zu bestimmen (PDB-Eintrag: 1HUC) [53]. Anhand der Strukturdaten stellten sie fest, daß Cathepsin B einen "Einschubloop", bestehend aus ca. 20 Aminosäuren enthält, der die Bindungsspalte in der S2'-Bindungsposition blockiert. Für diese im Vergleich zu den anderen Cathepsinen zusätzliche Aminosäuresequenz wurde gezeigt, daß sie über zwei Histidinreste als Anker für die C-terminale Carboxylgruppe eines Substrates fungiert, wodurch dieses Cathepsin die Eigenschaften einer Carboxypeptidase besitzt. Erst kürzlich wurde durch *Illy et al.* diese Rolle des Cathepsin B-Einschubloops durch Mutationsstudien bestätigt [54]. Allerdings ist darauf hinzuweisen, daß dieser zusätzliche Loop eine Reihe von unterschiedlichen Konformationen einnehmen kann. Wird die Anordnung dieses Einschubloops zum Enzym über eine ortsgerichtete Punktmutation so verändert, daß eine Interaktion dieses Loops mit dem Enzym unterbunden wird, kann für Cathepsin B Endopeptidaseaktivität beobachtet werden [55].

Neben der Röntgenkristallstruktur des reifen Cathepsin B konnten auch die Proteinstrukturen des humanen Procathepsin B (PDB-Eintrag: 1PBH, 2PBH, 3PBH [56]) sowie des Procathepsin B, isoliert aus Rattenleber (PDB-Eintrag: 1MIR [57]), bestimmt werden.

## 1.2 Die Cysteinprotease Cathepsin H

Cathepsin H (E.C. 3.4.22.16) wurde erstmalig aus Rattenleber isoliert [27]. Wie gezeigt werden konnte, sind Cathepsin H bzw. Cathepsin H-ähnliche Enzyme an Krankheitsstadien wie Brustkrebs, Hautkrebs und Tumormetastasierung beteiligt [58]. Weitere Namen für Cathepsin H sind Cathepsin I und BANA-Hydrolase. Dieses Enzym kommt ubiquitär in Zellen von Säugern vor. Cathepsin H zeigt nicht nur eine hohe Verwandtschaft zu den Cathepsinen L und S, sondern auch zu Proteasen, die in Zellen von Schimmelpilzen auftreten [26]. Die Präproform des humanen Cathepsin H besteht aus 335 Aminosäuren und hat eine Molmasse von 41 kDa (Tab. A-2). Die reife Form des Enzyms ist irreversibel inaktiv bei einem pH-Wert  $> 7$ , wobei der Präkursor bei physiologischem pH-Wert stabil ist [26]. Im Gegensatz zu seiner hohen Korrelation zu anderen Papain-ähnlichen Cathepsinen ist die Spezifität von Cathepsin H etwas andersartig. Durch *Fruton et al.* wurde wahrscheinlich zum ersten Mal die Aminopeptidaseaktivität dieses Enzyms untersucht und als thiolabhängige "Leucinaminopeptidase" deklariert [59]. *Kirschke et al.* beschrieben Cathepsin H als eine "Endoaminopeptidase", da dieses Protein Substrate sowohl als Aminopeptidase, aber auch als Endopeptidase hydrolysieren kann [27]. Diese Befunde resultieren aus Studien über die Spaltung von Substraten mit einer freien  $\alpha$ -Aminogruppe, wie z. B. Arg-NH<sub>2</sub> und Thr-NH<sub>2</sub> sowie der Hydrolyse von N-substituierten Aminosäurederivaten, wie z. B. Bz-Arg-NH<sub>2</sub> und Bz-Arg-2-Nitroanilid [27]. Das pH-Optimum von Cathepsin H zur Hydrolyse der meisten Substrate liegt in einem Bereich von 6.5-6.8 [26]. Spezifische Substrate für dieses Cathepsin sind N-terminal ungeschützte Derivate, wie z. B. Arg-NH<sub>2</sub> und Arg-NH<sub>2</sub> [26].

Es wurde für dieses Protein angenommen und auch kürzlich nachgewiesen, daß ein Fragment des Cathepsin H-Propeptides, die sogenannte Minikette, in das katalytisch aktive Zentrum des Enzyms bindet, woraus die Aminopeptidaseaktivität dieses Cathepsins resultiert [60,61]. Die Befunde von *Guncar et al.* zeigten, daß die aus acht Aminosäuren bestehende Minikette (Glu-Pro-Gln-Asn-Cys-Ser-Ala-Thr) über eine Disulfidbrücke an das Enzym gebunden ist [61]. Sie hat die Orientierung eines in der Bindungsspalte gebundenen Substrates. Der negativ geladene C-Terminus des Oktapeptides kann mit dem positiv geladenen N-Terminus eines Substrates interagieren. Der C-Terminus bindet an der Stelle, wo sich in vergleichbaren Enzymen die S<sub>2</sub>-Bindungsstelle befindet und imitiert dadurch einen P<sub>2</sub>-Rest.

Cathepsin H zeigt ein ähnliches Verhalten wie Cathepsin B, indem es neben seiner Endopeptidaseaktivität auch Exopeptidaseaktivität zeigt. Es können so durch dieses Enzym Substrate mit blockierter  $\alpha$ -Aminogruppe hydrolysiert werden [27,62].

Die Tertiärstruktur des Cathepsin H, isoliert aus Schweineleber, wurde erst kürzlich über Röntgenkristallstrukturanalyse aufgeklärt (PDB-Eintrag: 8PCH [61]). Im Gegensatz dazu konnte die Struktur des humanen Cathepsin H noch nicht ermittelt werden.

### 1.3 Die Cysteinprotease Cathepsin L

Cathepsin L (EC 3.4.22.27) wurde erstmals über Lysosomenpräparation aus einer Rattenleber isoliert [27,63]. Dieses Cathepsin stellt unter den lysosomalen Cysteinproteasen ein Enzym mit der höchsten Aktivität in Hinblick auf seine Fähigkeit zur Hydrolyse von Azocasein [62], Elastin [10] und Kollagen dar [64]. Neben Cathepsin B konnte Cathepsin L bisher am genauesten charakterisiert werden. Cathepsin L spielt nicht nur eine sehr wichtige physiologische Rolle in der intrazellulären Proteindegradation [65], es dient auch der Aktivierung des Proplasminogenaktivators [26]. Wegen seiner direkten Beteiligung an der Knochenresorption [66], der Myelindegeneration [67] und der Tumormetastasierung [68] ist dieses Enzym ein interessantes Zielmolekül, um Wirkstoffe zu entwickeln, die die Aktivität dieses Proteins beeinflussen.

Cathepsin L kommt als ein inaktives Präproenzym vor, das zur reifen Form prozessiert wird [69]. Die Präproform des humanen Enzyms enthält 333 Aminosäuren und hat eine Molmasse von 38 kDa (Tab. A-2). Procathepsin L ist stabil bei neutralem bzw. basischem pH-Wert [70]. Das Propeptid, das 96 Aminosäuren enthält, ist zwischen der Signalsequenz und dem N-Terminus des reifen Enzyms lokalisiert. Die reife Form des humanen Cathepsin L besteht aus 220 Aminosäuren. Katalytisch aktiv ist Cathepsin L bei pH-Werten zwischen 3.0-6.5. Das Enzym ist irreversibel inaktiv bei einem pH-Wert  $> 7$ , wobei der Präkursor bei physiologischem pH-Wert stabil ist. Cathepsin L hat eine begrenzte Spezifität in der Hydrolyse von synthetischen Substraten [26]. Das Enzym spaltet Substrate mit Arginin in P1-Position und hydrophoben Resten in P2- und P3-Position bei einem pH von 5.5. Allerdings sind diese Substrate nicht spezifisch unter den lysosomalen Cysteinproteasen, da sie auch von den Cathepsinen B und S hydrolysiert werden [71].

### 1.4 Die Cysteinprotease Cathepsin S

Cathepsin S (E.C. 3.4.22.27) ist eine hochaktive lysosomale Cysteinprotease, die zum ersten Mal aus den Lymphknoten des Rindes isoliert werden konnte [72]. Aufgrund seiner Sequenzhomologie gehört es zur Papain-Superfamilie. Cathepsin S ist im Vergleich zu den Cathepsinen B, H und L in seiner Zellverteilung begrenzt, wodurch eine umfangreiche Isolierung des Enzyms schwierig ist. Es wurde hauptsächlich in Lymphknoten, Milz und Makrophagen gefunden [44]. Im Gegensatz dazu konnte es nur auf sehr geringem Niveau aus Niere, Leber, Herz und Bauchspeicheldrüse expremiert werden [73,74]. Studien von *Riese et al.* zeigten, daß Cathepsin S bei der MHC-II vermittelnden Antigenpräsentation eine Schlüsselrolle spielt [7]. Durch die hohe elastinolytische Aktivität dieses Proteins in Lungenmakrophagen bei neutralem pH kann man eine Beteiligung bei der Pathogenese von Emphysemen annehmen [75].

Die reife Form des humanen Cathepsin S besteht aus 217 Aminosäuren und hat eine Molmasse von 24 kDa (Tab. A-2). Es unterscheidet sich jedoch von den anderen vergleichbaren Proteasen durch

seine hohe Stabilität bei neutralem bzw. leicht basischem pH [44]. Bisher konnte noch kein Peptidsubstrat beschrieben werden, das bei einem sauren pH-Wert spezifisch für Cathepsin S ist [76]. Die Substratspezifität dieses Enzyms scheint sehr ähnlich zu Cathepsin L zu sein. Jedoch erlaubt eine Erhöhung des pH-Wertes auf 7.5 und die Verwendung von Substraten, wie z. B. Bz-Phe-Val-Arg-NHMec oder Z-Val-Arg-Arg-NHMec eine spezifische Bestimmung von Cathepsin S [76]. Eine Röntgenkristallstruktur von Cathepsin S wird zwar in der Literatur beschrieben, jedoch sind die Strukturdaten nicht zugänglich [77].

## 1.5 Die Cysteinprotease Cathepsin K

Die cDNA-Sequenz des Cathepsin K (EC 3.4.22.38) wurde unabhängig durch mehrere Arbeitsgruppen beschrieben [37,38,78,79]. Neben den Cathepsinen B, L, H und S ist dieses Cathepsin an der Resorption der Knochenmatrix beteiligt [80,81]. Es konnte gezeigt werden, daß Cathepsin K Osteoklasten-spezifisch ist und im Vergleich zu den anderen Cysteinproteasen am häufigsten in diesen Zellen vorkommt [82]. Seine potentielle Funktion im Knochenumbau ist durch sein reichliches Vorkommen in Osteoklasten und Osteoklasten-ähnlichen Multikernzellen begründet [38]. Der Knochenumbau ist ein dynamischer Prozeß, der die Knochenresorption und -bildung beinhaltet. Die Resorptionsphase wird durch Osteoklasten hervorgerufen, die an der Knochenoberfläche haften. Es resultiert so die Bildung eines extrazellulären Fragmentes, dem Resorptionsloch, in das die Osteoklasten die proteolytischen Enzyme absondern. In saurer Umgebung wird die Mineralschicht dieses Bereiches zerstört und es kann die Degradation der Proteinmatrix durch die proteolytischen Enzyme erfolgen [38].

Für Cathepsin K wurde eine hohe kollagenolytische Aktivität gegenüber Kollagenen der Typen I, II und IV sowie eine starke gelatinolytische Aktivität beschrieben [83]. Der wichtigste Beweis der besonderen Bedeutung dieser Protease liegt in genetischen Studien, die zeigen, daß Mutationen ein nichtfunktionelles Enzym hervorrufen, das in einer seltenen Erbkrankheit, der Pycnodysostose, eine Rolle spielt [38,84]. Weiterhin konnte gezeigt werden, daß Cathepsin K in Bronchialepithelzellen expremiert wird [85]. Andere Untersuchungen deuten auf eine sehr differenzierte Expression des Enzyms im embryonalen Lungengewebe hin [86].

Cathepsin K liegt als inaktives Präproenzym vor. Nach der Umwandlung in die Proform wird das Enzym zu seiner reifen Form durch proteolytische Spaltung des aus 99 Aminosäuren bestehenden Propeptides konvertiert. Die Prozessierung zum reifen Cathepsin K *in vitro* erfolgt autokatalytisch bei einem pH-Wert von 4.0 [87]. Die Umgebung des Knochenresorptionsloches hat einen niedrigen pH-Wert [88]. Das ist konsistent mit der Beobachtung, daß Cathepsin K sehr große Bedeutung bei dem Knochenumbau hat und führt zu der Annahme, daß die Aktivierung von Procathepsin K *in vivo* in diesem Resorptionsloch auftreten kann.

Das Präproprotein von humanem Cathepsin K enthält 329 Aminosäuren und hat eine Molmasse von 37 kDa (Tab. A-2). Die reife Form des Enzyms besteht aus 215 Aminosäuren. Sein pH-Optimum zur Hydrolyse der meisten Substrate liegt zwischen 6.0 und 6.5 und ist vergleichbar mit den Werten von Cathepsin S [38].

## 1.6 Die Cysteinprotease Cathepsin F

Cathepsin F (TrEMBLNEW: AAC78838), ein neues Mitglied der Papain-Superfamilie, kommt ubiquitär in den meisten humanen Geweben und Krebszellen vor [41]. Das weitverbreitete Auftreten von Cathepsin F korreliert mit den Cathepsinen B, L und H. Es ist allerdings nicht vergleichbar mit den Cathepsinen K und S, die meist gewebespezifisch vorkommen. Die Funktion dieses Enzyms konnte bisher noch nicht genau aufgeklärt werden. Aufgrund der großen Verbreitung dieses Enzyms in den meisten Geweben wird jedoch angenommen, daß dieses Cathepsin eine allgemeine Bedeutung im lysosomalen Proteinkatabolismus in allen Zelltypen hat [40]. Diese neue Cysteinprotease weist im Vergleich zu den anderen Cathepsinen in seiner Aminosäuresequenz einige spezielle Merkmale auf. Während der Teil der Sequenz, aus dem sich das reife Enzym zusammensetzt, in hoher Konformität zu den anderen Mitgliedern der Papainfamilie steht, ist die Proregion einzigartig in ihrer Länge und Sequenz. Es ist die längste Proregion, die bisher für Cysteinproteasen dieser Familie beobachtet wurde. Sie besteht aus 251 Aminosäuren und enthält ein N-terminales Segment, das strukturelle Ähnlichkeiten zu den Cystatinen, einer Gruppe von Cysteinprotease-Inhibitoren, aufweist [89]. Die Präproform von humanem Cathepsin F besteht aus 489 Aminosäuren und hat eine Molmasse von 53 kDa (Tab. A-2).

Rekombinantes Cathepsin F hat zur Hydrolyse von Substraten ein pH-Optimum zwischen 5.2 und 6.8 [41].

Zwischen den Cathepsinen F und W wurde eine hohe Sequenzhomologie gefunden [40,90]. Das läßt darauf schließen, daß dieses Enzym in engerer Beziehung zu Cathepsin W als zu den anderen Mitgliedern der Papainfamilie steht. Diese Befunde deuten darauf hin, daß diese beiden Cathepsine eine neue Untergruppe der Cathepsin L-Unterfamilie bilden (Abb. 2) [40,90].



## 2 AUFGABENSTELLUNG

Lysosomale Cysteinproteasen besitzen eine entscheidende Bedeutung bei dem Umbau intrazellulärer und extrazellulärer Proteine. Diese Proteasen spielen eine wichtige Rolle in pathologischen Prozessen, die mit einer Gewebedestruktion verbunden sind. Papain-ähnliche Cathepsine besitzen eine relativ ähnliche Substratspezifität, die durch die Strukturhomologie der Enzyme untereinander bedingt wird. Aufgrund dieser Befunde ist es von großem Interesse, Wirkstoffe zu entwickeln, die gezielt die Aktivität eines Enzyms beeinflussen. Mit enzymologischen Methoden gestaltet sich aufgrund der geringen Substratspezifität innerhalb der Cathepsine eine genaue Differenzierung sehr schwierig. Mit Hilfe artifizierter enzymspezifischer fluorogener Substrate sollte die enzymatische Aktivität bestimmt werden, um die Cathepsine im Organismus möglichst selektiv detektieren zu können. Daher stand das Design fluorogener Substrate für Cathepsine im Vordergrund dieser Arbeit. Darüber hinaus bildete die Erstellung neuer Peptidsequenzen für spezifische Inhibitoren für Cathepsin F eine weitere Zielstellung. Über den Vergleich der Papain-ähnlichen Cathepsine B, F, H, K, L und S in ihrer Proteinstruktur und der detaillierten Charakterisierung der katalytisch aktiven Zentren sollte versucht werden, neue hochaffine Liganden für die Cathepsine K, L und F zu entwickeln. Zur Realisierung dieser Aufgabenstellung ist die Kenntnis der Proteinstrukturen dieser Cathepsine von wesentlicher Bedeutung. Resultierend aus diesen Problemstellungen gliedert sich die vorliegende Arbeit in zwei Schwerpunkte.

### 1. *Bestimmung der Tertiärstrukturen von catheptischen Cysteinproteasen*

Die Kenntnis der Tertiärstruktur eines Enzyms kann nicht nur zum Verständnis der biologischen Funktion und zur Untersuchung seiner Eigenschaften beitragen, sie stellt auch eine wesentliche Voraussetzung zur Entwicklung neuer Inhibitoren oder Substrate mit spezifischem Wirkstoffprofil dar. Bis zum heutigen Zeitpunkt sind die Sequenzen von über 80.000 Proteinen (Swiss-Prot-Datenbank [91,92]) bekannt und zugänglich. Im Gegensatz dazu konnten bisher von nur ca. 10.000 Proteinen bzw. Peptiden die 3D-Strukturen mittels Röntgenkristallstrukturanalyse bzw. NMR-Spektroskopie bestimmt werden [50], was darauf hindeutet, wie schwierig sich die Aufklärung einer Proteinstruktur gestaltet. Aus diesem Grund stellt die Generierung von Tertiärstrukturmodellen eine wichtige Alternative zu den experimentellen Methoden dar, um essentielle Strukturinformationen über das jeweilige Protein zu erhalten [93,94].

Zu Beginn der hier dargestellten Untersuchungen an den lysosomalen Cysteinproteasen waren die Röntgenkristallstrukturen der Cathepsine B und L sowie von Papain und Actinidin bekannt und deren 3D-Koordinaten über die PDB zugänglich. Basierend auf der hohen Homologie innerhalb der Mitglieder der Papain-Superfamilie sollten 3D-Strukturmodelle der Cathepsine K, S, H und F über verschiedene Verfahren der Proteinmodellierung generiert und hinsichtlich ihrer Qualität und

Molekülgeometrie untersucht werden. Ein besonderes Augenmerk mußte hierbei auf Cathepsin H gelegt werden, da das aktive Enzym im Gegensatz zu den anderen zu untersuchenden Cathepsinen eine Minikette enthält und die Position dieses Oktapeptides in der Bindungsspalte des Cathepsins bestimmt werden mußte.

Zur Verifizierung der modellierten Tertiärstrukturen sollten diese in ihrer Substrat- und Inhibitorspezifität anhand von experimentell bestimmten kinetischen Daten, die in der Literatur beschrieben sind, untersucht werden.

Ausgehend von den bekannten Röntgenkristallstrukturen und den modellierten Proteinstrukturen sollten die Substratbindungsregionen der Cathepsine genauer untersucht werden. Größe und Eigenschaften der einzelnen Bindungspositionen im Enzym sollen im Detail analysiert werden, um so zusätzliche Informationen über charakteristische Merkmale und die Spezifität der Bindungsspalte zu erhalten.

## 2. *Entwicklung von neuen Liganden für die Cathepsine L, K und F*

Die Entwicklung möglichst spezifischer Liganden für die lysosomalen Cysteinproteasen kann zur weiteren Aufklärung der Funktion dieser Enzyme beitragen und von potentiellen therapeutischen Interesse sein.

Anhand der Strukturmodelle der Cathepsine und der Charakterisierung der Substratbindungspositionen in den katalytisch aktiven Zentren dieser Enzyme ist es mit Hilfe der Methoden des "Molecular Modelling" möglich, Vorschläge für gezielte Synthesen spezifischer Effektoren abzuleiten. Über zahlreiche Verfahren, wie z. B. der Analyse der sterischen Verhältnisse in den Substratbindungsregionen, der molekularen elektrostatischen und lipophilen Oberflächenpotentiale im Protein sowie der Ermittlung von möglichen Interaktionen zwischen Substratresten und den Aminosäuren der einzelnen Bindungspositionen des Enzyms sollten die Bindungsspalten der Cathepsine intensiv studiert und charakterisiert werden. Basierend auf den resultierenden Ergebnissen sollten Liganden für die Cathepsine K, L und F zur Synthese vorgeschlagen werden, die möglichst spezifisch und hochaffin nur an jeweils eines der Cathepsine binden.

## 3 METHODEN

### 3.1 Modellierung von Proteinstrukturen

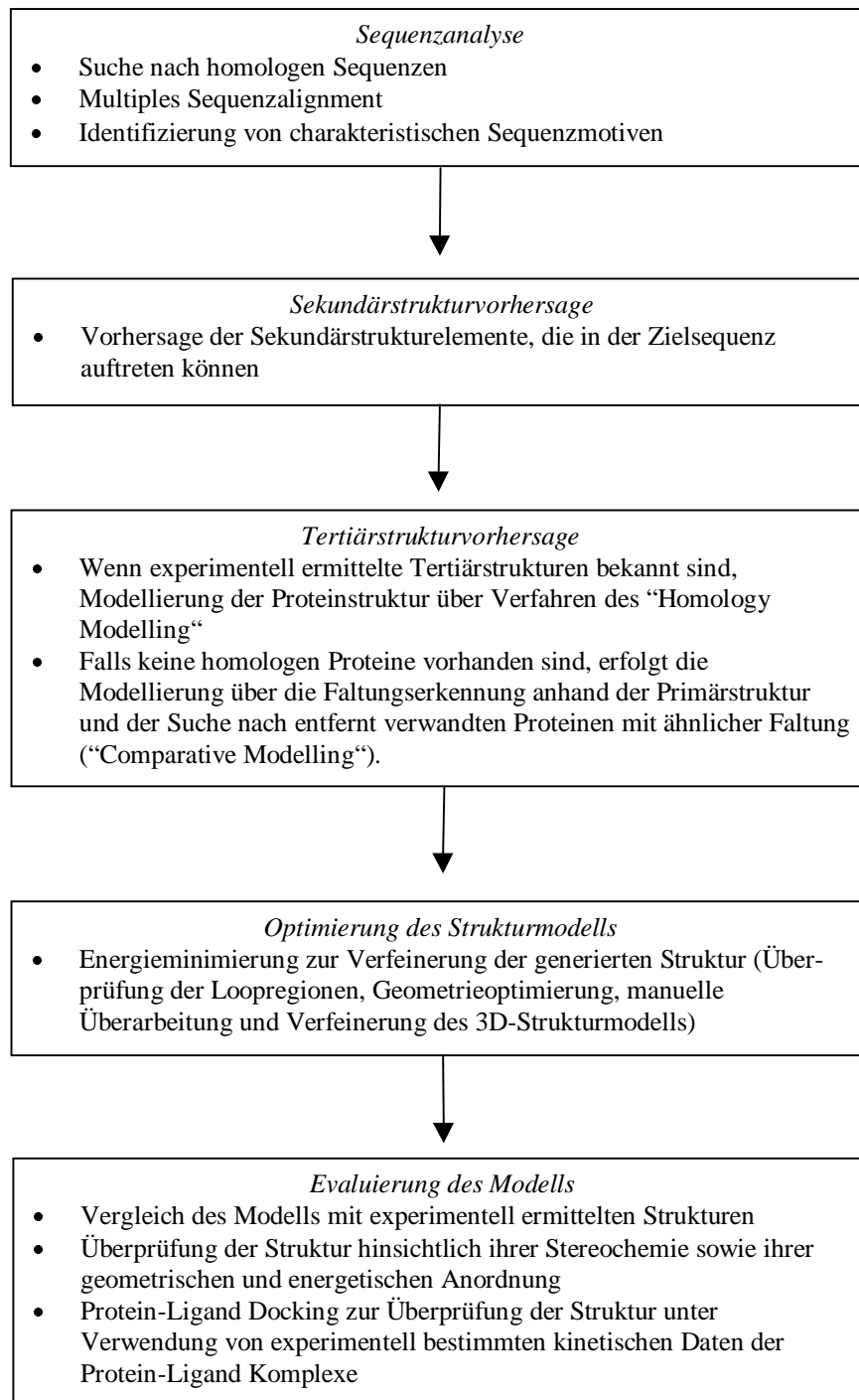
Das steigende Interesse in den Bereichen der Medizin, Molekularbiologie, Pharmazie, Biotechnologie an Proteinstrukturen unterstreicht die Bedeutung der Bestimmung bzw. Vorhersage von dreidimensionalen Strukturen von Proteinen. Neben den experimentellen Techniken spielt die Modellierung von Proteinstrukturen eine wesentliche Rolle im Design von neuen Proteinen und Liganden.

Die allgemeine Vorgehensweise zur Modellierung von Proteinstrukturen ist im Schema 1 dargestellt. Jedoch erweckt dieses Fließschema den Eindruck, daß dabei eine Reihe von Algorithmen verwendet werden, wobei die Ausgabe des einen Schrittes die Eingabe des nächsten Schrittes ist. Die beste Methode zur Generierung von Proteinstrukturen besteht in der Kombination der Nutzung von Algorithmen sowie von manuellen Interventionen und Erfahrung.

#### 3.1.1 Sequenzanalyse

##### *Suche nach homologen Proteinen*

Generell ist der erste Schritt zur Generierung einer Proteinstruktur die Untersuchung, ob die Zielsequenz homolog zu anderen Sequenzen ist. Es ist jedoch dabei zu beachten, daß sich viele resultierende "Treffer" in einer sogenannten "Zwielichtzone" befinden, d. h. es ergeben sich paarweise Sequenzübereinstimmungen, bei denen bei genauer Überprüfung der separaten Sequenzen nicht zwischen einer zufälligen Korrelation oder strukturellen Gemeinsamkeiten unterschieden werden kann. Das Auftreten dieser Zone kann eine Verfälschung der Strukturinformationen in einer lokalen Sequenz bewirken. Außerdem können Proteinfragmente, die eine Identität von 25% zu anderen Proteinen aufweisen, trotzdem verschiedene Konformationen in nichtverwandten Proteinen besitzen [95]. *Sander & Schneider* entwickelten ein relativ einfaches Verfahren zur Bestimmung der prozentualen Identität in der lokalen Sequenzübereinstimmung bei einer gegebenen Sequenzlänge [96].



**Schema 1** Vorgehensweise für die Generierung von Proteinstrukturen.

### *Multiples Sequenzalignment*

Konnten Homologe zur Zielsequenz gefunden werden, ist der nächste Schritt die Durchführung eines multiplen Sequenzalignments der Zielsequenz mit den Sequenzen der homologen Proteine. Ein

Alignment ist besonders hilfreich zur Identifizierung von konservierten Regionen, die von wesentlicher funktioneller und struktureller Bedeutung sind. Multiple Sequenzalignments erhöhen die Genauigkeit von Sekundärstrukturvorhersage und anschließender Tertiärstrukturmodellierung. Die Sequenzalignments wurden unter Verwendung von CLUSTAL W durchgeführt [97]. Eine aminosäurespezifische "Gap-Strafe" wird bei der Bildung von Gaps in potentiellen Loopregionen angewendet, so daß die Ausbildung von Gaps in regulären Sekundärstrukturen reduziert wird. Die Gaps, die sich in vorherigen Sequenzalignments ausgebildet haben, werden mit einer geringeren Gap-Strafe behandelt, um die Entstehung von neuen Gaps zu verhindern. Weiterhin wird die Gap-Strafe verwendet, wenn bei einem Sequenzalignment unterschiedliche Sequenzlängen auftreten und dadurch ein Gap geöffnet werden muß.

In den durchgeführten Sequenzalignments mit CLUSTAL W wurde die Blosom62-Matrix verwendet und der Parameter für die Gap-Strafe auf 10 gesetzt [97]. Für die weiteren Parameter (zusätzliche Gap-Strafen für einen erweiterten Gap, separate Gap-Strafen) wurden die Standardeinstellungen von CLUSTAL W verwendet.

#### *Bestimmung von Sequenzmotiven*

Falls für die Gesamtsequenz keine homologen Sequenzen bestimmt werden konnten, können Anhaltspunkte über die mögliche Faltung des Zielproteins über die Bestimmung von vorher charakterisierten lokalen Sequenz- und Faltungsmotiven erhalten werden [98]. Es hat sich gezeigt, daß dieses Verfahren ein geeigneter Weg ist, wobei die ermittelten Motive eine bemerkenswert hohe Genauigkeit besitzen [98].

Aufgrund der hohen Homologie der untersuchten Proteine und des Vorhandenseins von Röntgenkristallstrukturen von verwandten Enzymen wurde keine Bestimmung von Sequenzmotiven durchgeführt. Der Vollständigkeit halber sollte diese Verfahrensweise zur Modellierung der Tertiärstrukturen von Proteinen nicht unerwähnt bleiben.

### **3.1.2 Sekundärstrukturvorhersage**

Die Vorhersage der Sekundärstruktur von Proteinen ist für die Erkennung der entsprechenden Proteinfaltung von großer Bedeutung und stellt eine relativ einfache Methode dar, um Strukturinformationen über eine neu bestimmte Sequenz zu erhalten. Die Sekundärstrukturvorhersage ist vor allem zur Lösung einer Reihe von Fragestellungen bezüglich der Struktur von Proteinen geeignet:

1. Man erhält erste Informationen über das mögliche Auftreten von  $\alpha$ -Helices und  $\beta$ -Faltblattsegmenten in der zu modellierenden Proteinstruktur.

2. Die Zuordnung von Sekundärstrukturelementen kann bei der Bestimmung der Struktur-Funktions-Beziehung zwischen Proteinen helfen, auch wenn nur eine geringe Sequenzähnlichkeit auftritt.
3. Die Vorhersage von lokalen Sekundärstrukturelementen kann darüber hinaus bei der Strukturbestimmung über 2D-NMR-Spektroskopie hilfreich sein.

In der Literatur sind zahlreiche Methoden zur Sekundärstrukturvorhersage von Proteinen aus ihrer Primärstruktur beschrieben [99]. In der vorliegenden Arbeit wurde das Programm PHD ("Profile Network from Heidelberg") verwendet [100,101]. Vergleicht man die Ergebnisse, die aus der Sekundärstrukturvorhersage mit PHD resultieren, mit den Sekundärstrukturelementen von zugänglichen Proteinstrukturen, zeigt sich, daß dieses Programm die Regionen der Sekundärstrukturen mit einer ausreichenden Genauigkeit wiedergibt. Dieses Verfahren beruht auf einem neuronalen Netzwerk-Algorithmus. Die Genauigkeit der vorhergesagten Sekundärstrukturelemente in Proteinen unter Verwendung dieses Programms liegt über 70% und sogar über 80%, wenn zusätzliche Strukturinformationen berücksichtigt werden (z. B. ermittelt über ein multiples Sequenzalignment) [102].

### 3.1.3 Tertiärstrukturvorhersage

Die Generierung einer dreidimensionalen Proteinstruktur anhand seiner Aminosäuresequenz ist am zuverlässigsten, wenn die Struktur von einem oder mehreren Sequenzhomologen bekannt und zugänglich ist. Die Sequenzhomologie ist als Maß der Ähnlichkeit von Proteinen bzw. identischer (ähnlicher) Bausteinsequenzen in mehr oder weniger ausgedehnten Teilbereichen definiert. Sind mehrere Aminosäuresequenzen zueinander homolog, bedeutet das, daß sie eine Ähnlichkeit zueinander aufweisen. Die Strukturdaten der homologen Proteine können auf die Zielsequenz extrapoliert werden und es läßt sich so ein 3D-Modell ableiten.

Die Verfahren, die zur Modellierung von Proteinstrukturen am häufigsten verwendet werden, sind "Comparative Modelling" und "Homology Modelling". Comparative Modelling basiert auf der Strukturhomologie zwischen Ziel- und Vorlagesequenz. Im Gegensatz dazu nutzen die Verfahren des Homology Modelling die Sequenzhomologie zwischen Ziel- und Vorlagesequenz aus. Es ist allerdings darauf hinzuweisen, daß oftmals in der Literatur keine Differenzierung zwischen beiden Methoden getroffen wird. Eine grundlegende Voraussetzung für die Generierung eines zufriedenstellenden 3D-Strukturmodells ist, daß die Sequenzidentität zwischen der Vorlagesequenz und den verwendeten Zielsequenzen nicht unter 30% liegt.

Zur Entwicklung der Strukturmodelle der Cathepsine K, S, H und F wurden die Programme COMPOSER [103] und MODELLER [104] verwendet. Aus vergleichenden Analysen von einer Reihe mit COMPOSER bestimmten Modellen mit den korrespondierenden Röntgenkristallstrukturen resultieren

RMS-Werte (über alle C $\alpha$ -Atome), die in einem Bereich von 0.7 Å-1.7 Å liegen, was für eine hohe Genauigkeit der modellierten 3D-Strukturen spricht [105].

### 3.1.3.1 COMPOSER

*Blundell et al.* entwickelten zur Generierung von 3D-Proteinstrukturen die wissensbasierte Methode COMPOSER, die in das Softwarepaket SYBYL6.5 integriert wurde [103,106,107].

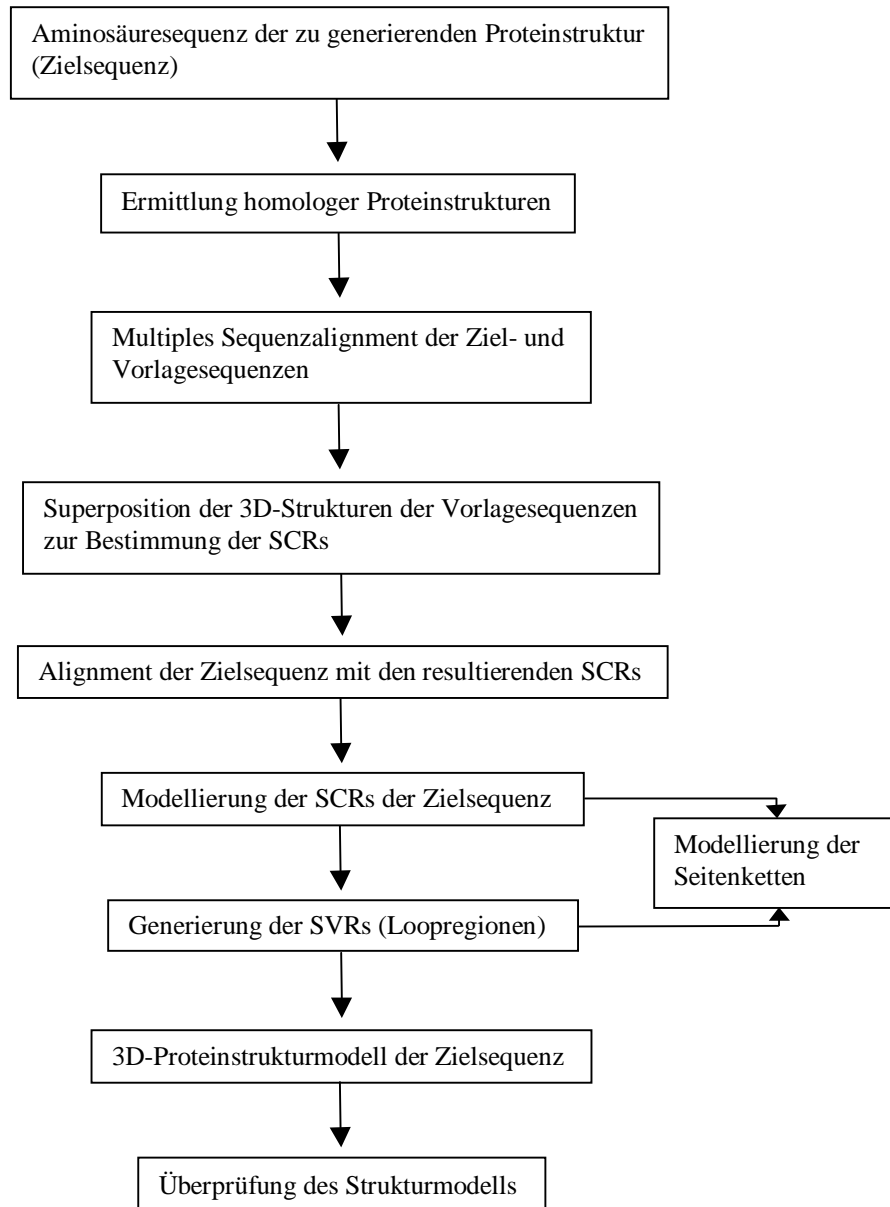
Die Vorgehensweise ist in einem Fließschema dargestellt (Schema 2). Das Programm ist weitestgehend automatisiert, erlaubt aber auch eine Reihe von manuellen Eingaben, z. B. bei der Definition der strukturell konservierten Regionen (SCR) und bei der Auswahl der strukturell variablen Regionen (SVR, Loopregion).

#### *Ermittlung von homologen Proteinsequenzen*

Gerade in der letzten Zeit wurden von zahlreichen Enzym-Inhibitor-Komplexen lysosomaler Cysteinproteasen die Tertiärstruktur mittels Röntgenkristallstrukturanalyse bestimmt (Tab. A-1). Deshalb muß noch einmal explizit darauf hingewiesen werden, daß zum Zeitpunkt der Modellierung der Proteinstrukturen der Cathepsine K, S, H und F keine Röntgenkristallstrukturen für diese Enzyme zugänglich waren.

Zur Durchführung der Tertiärstrukturmodellierung von Proteinen mit COMPOSER wurde zunächst unter Verwendung von BLAST nach Proteinen mit zugänglicher Röntgenkristallstruktur, die eine hohe Homologie zur Zielsequenz besitzen, gesucht [108,109].

Um die Aminosäuresequenzidentität zwischen Ziel- und Vorlagesequenz zu bestimmen, wurde ein multiples Sequenzalignment zwischen der Sequenz des zu generierenden Proteins und den Aminosäuresequenzen der Vorlageproteine unter Verwendung des Needleman-Wunsch-Algorithmus durchgeführt [110]. Das multiple Sequenzalignment wurde mit der pmutation-Homologiematrix des SYBYL-Programms durchgeführt [106]. Die längenabhängige Gap-Strafe wurde auf 8 gesetzt. Die Auswahl der Proteine, die als Vorlagestrukturen dienten, wurde durch den Identitätsscore begrenzt [103]. Es wurden nur solche Proteine ausgewählt, bei denen mindestens 30% der Aminosäuren identisch zu der Aminosäuresequenz des Zielproteins sind. Allerdings mußte beachtet werden, daß in einigen Proteinen häufig bestimmte Aminosäurereste auftreten, die dazu führen können, daß diese Proteine ähnlicher zur Zielsequenz zu sein scheinen, als sie es wirklich sind. Um diesen Effekt zu berücksichtigen und um besser zwischen den Homologen differenzieren zu können, wurde neben dem Identitätswert ein weiteres Bewertungskriterium eingeführt, der Signifikanzscore, der auf den Wert 4 gesetzt wurde [106].



**Schema 2** Allgemeines Fließschema für das Programm COMPOSER.

### *Aufbau eines Gerüsts der strukturell konservierten Regionen*

Bevor ein Strukturalignment zwischen den ermittelten homologen Proteinen durchgeführt werden kann, muß ein Initialset definiert werden, der aus mindestens drei Aminosäuren mit strukturell gleichen Positionen (z. B. katalytische Triade) besteht. Durch diese Festlegung wird die Qualität des zu bestimmenden Modells deutlich erhöht. COMPOSER verwendet diesen Initialset, um ein optimales Strukturalignment der homologen Proteine zu generieren. Es werden die strukturell konservierten Regionen (SCR) bestimmt und daraus eine Durchschnitts- bzw. Gerüststruktur abgeleitet. Dieses vorläufige Gerüst besteht somit aus den konservierten Regionen der homologen Proteine [111]. Die erhaltene Gerüststruktur wird über die Superposition der homologen Strukturen sowie über den



Abstands-cut-off zwischen den strukturell äquivalenten  $C\alpha$ -Atomen definiert [103]. Dieser cut-off Wert beträgt 3.5 Å.

Basierend auf den resultierenden SCRs der Vorlageproteine werden anschließend die Bereiche in der Zielsequenz bestimmt, die zu diesen SCRs korrespondieren. Dazu wird die Zielsequenz mit den Sequenzen der SCRs übereinandergelegt, wobei die Proteinfragmente ausgewählt werden, die zur Generierung der SCR-Rückgratstruktur des Modells dienen. Dieses Alignment wird unter Anwendung des Needleman-Wunsch-Algorithmus und der pmutation-Homologiematrix durchgeführt. Die Gap-Strafe beträgt 8.

Die resultierende Gerüststruktur stellt eine 3D-Struktur der konservierten Regionen des Proteins dar. Jedoch fehlt noch jegliche reale Proteingeometrie, da es sich hierbei um eine Durchschnittsstruktur handelt. In einem nächsten Schritt wird das Rückgrat der einzelnen SCRs im Modell konstruiert, indem eine Region der Gerüststruktur auf das entsprechende Fragment des homologen Proteins angepaßt wird. Anschließend werden unter Anwendung einer wissensbasierten Methode die Seitenkettenkonformation der jeweiligen Aminosäurereste bestimmt (Seite 19).

#### *Modellierung der Loopstrukturen*

Um die einzelnen Fragmente des Gerüsts zu verbinden, müssen die Loopregionen konstruiert werden. Es wird versucht, für jeden Loopbereich Fragmente von zugänglichen Strukturen zu finden, die kompatibel mit dem Rest des Modells sind. Für die Suche nach geeigneten Loopfragmenten werden die Koordinaten der  $C\alpha$ -Atome der Aminosäuren, welche die Loopregion begrenzen, verwendet, um die Suche einzuschränken. Diese Aminosäurereste werden auch als Ankerregionen bezeichnet. Im allgemeinen werden drei Ankerreste pro Loop genutzt, um die strukturelle Umgebung des Loops mit zu berücksichtigen. Für jede zu modellierende Loopregion wurden folgende Schritte ausgeführt:

1. Suche nach Homologen für die zu bestimmende Loopregion mit der geeigneten Anzahl der Aminosäuren.
2. Alle Fragmente von ähnlicher Länge, die in der Datenbank enthalten sind, werden auf folgende Kriterien untersucht und ausgewählt:
  - a) Auswahl aller Teilstücke, deren End-to-end-Abstand in einem Toleranzbereich von 1.5 Å ist.
  - b) Für jedes Fragment werden alle Abstände zwischen dem  $C\alpha$ -Atom in der N-terminalen Ankerregion und dem  $C\alpha$ -Atom der C-terminalen Ankerregion berechnet. Diese Abstände werden mit den korrespondierenden Abständen im Modell verglichen und der RMS-Wert berechnet. Es werden nur die Abschnitte ausgewählt, deren RMS-Wert unter 1.0 Å liegt.
3. Anpassen der Koordinaten des ausgewählten Fragmentes auf die Reste der Ankerregion,
4. Generierung der Seitenketten (siehe unten),
5. Berechnung des Sequenzhomologiescores zwischen jeder Teilsequenz und der Sequenz im Modell. Die Loops werden nach der Homologie und dem berechneten RMS-Wert zwischen den

Aminosäuren der Vorlagesequenz und der Ankerregion sortiert. Der Loop mit der höchsten Homologie wird in das Modell eingefügt.

Das generierte Modell wird anschließend darauf überprüft, daß keine sterischen Überlappungen auftreten, d. h. es wird untersucht, ob die C $\alpha$ -Atome der Loopregionen zu geringe Abstände zu C $\alpha$ -Atomen aus der Nichtloopregion des Modells einnehmen.

#### *Generierung der Seitenketten*

Die richtige Seitenkettenkonformation von Aminosäuren im Protein ist bedeutend für die Packung der Aminosäuren. Bei der Modellierung der Seitenketten müssen bevorzugte Torsionswinkel der Seitenketten, Disulfidbrücken, Wasserstoffbrückenbindungen, Ionenwechselwirkungen und andere elektrostatische Wechselwirkungen berücksichtigt werden.

COMPOSER benutzt einen schnellen und effektiven Weg zur Modellierung der Seitenketten. Die Seitenketten werden unter Verwendung von bestimmten Richtlinien modelliert, wodurch die wahrscheinlichste Orientierung der Seitenketten im Vergleich zu äquivalenten Positionen in homologen Proteinen bestimmt wird. Es werden bevorzugte Konformationen der Seitenketten, wie sie in Sekundärstrukturelementen auftreten, mit einbezogen [103,106]. Wenn die Seitenkettenorientierung nicht als konserviert in einer topographisch äquivalenten Position in homologen Proteinen angenommen werden kann, werden die allgemeinen Richtlinien, die aus den Konformationen der Seitenketten in Helices und Faltblattsegmenten resultieren, angewendet [103].

Außerdem wurde eine Analyse der Seitenkettenanordnung in gleichen Positionen für repräsentative Vertreter mit  $\alpha$ -helikalen Proteinen und Proteinen mit  $\beta$ -Faltblattstrukturen durchgeführt. Da es sich gezeigt hat, daß die beste Seitenkettenkonformation nicht nur von der Sekundärstruktur abhängt, sondern auch von zahlreichen Richtlinien (insgesamt 1200) [111]. Das Verfahren zur Seitenkettengenerierung kann als potentielle paarweise Substitution innerhalb der strukturell gleichen Reste angesehen werden. Die Aminosäuren werden in Bezug auf die N-, C $\alpha$ - und C' $\alpha$ -Atome verglichen und die mittlere Position der C $\beta$ - und C $\gamma$ -Atome der Seitenkette bestimmt.

### **3.1.3.2 MODELLER**

*Sali & Blundell* entwickelten das Programm MODELLER, mit dem man ein 3D-Modell durch optimale Ausnutzung von bekannten Strukturen, die homolog zur Aminosäuresequenz des zu generierenden Proteins sind, erhalten kann [104]. Es werden "Restrains" definiert, die über Dichtefunktionen (pdfs) von Merkmalen, die in ihren Freiheitsgraden eingeschränkt werden sollen, ausgedrückt werden. Solche Merkmale sind Struktureigenschaften für bestimmte Aminosäurepositionen und Beziehungen zwischen den Aminosäureresten. Sie schließen die Zugänglichkeit des Lösungsmittels sowie mögliche

Sekundärstrukturelemente und Wasserstoffbrücken ein. Es werden z. B. die Möglichkeiten der Rückgratkonformation einer äquivalenten Aminosäure in homologen Proteinen als Ähnlichkeitsfunktion zwischen zwei Sequenzen behandelt. Mehrere solcher pdfs können aus der Korrelation von Strukturmerkmalen von Familien mit homologen Proteinen erhalten werden, die auf der Basis ihrer 3D-Strukturen verglichen und angepaßt werden. Die pdfs schränken die Variabilität der  $C\alpha$ - $C\alpha$ -Abstände, der N-O-Abstände der Hauptkette, die Torsionswinkel des Proteinrückgrates und der Seitenketten ein. Das 3D-Modell wird über die Optimierung dieser molekularen pdfs erhalten. Der pdf kann als eine Kombination von pdfs-Restraints individueller räumlicher Merkmale des gesamten Moleküls betrachtet werden. Das Optimierungsverfahren ist eine alternierende Methode, die einen "Conjugate Gradient"-Algorithmus auf alle Atome anwendet (außer Wasserstoffatome) [104].

Generell unterscheidet sich diese Methode nicht sehr gravierend von COMPOSER, außer daß zur Identifizierung der SCRs und SVRs und bei der Generierung der Seitenketten die beschriebenen Restraints angewendet werden [105]. Der grundlegende Schritt ist auch hier ein multiples Sequenzalignment der Zielsequenz mit homologen Proteinsequenzen. Dieses Alignment wurde mit CLUSTAL W durchgeführt (siehe Abschnitt 3.1.1) [97]. Mit dem Programm MODELLER werden aus diesem Sequenzalignment die Abstands- und Torsionswinkelrestraints von den bekannten 3D-Vorlagestrukturen abgeleitet und diese auf die Sequenz des Zielenzym übertragen [104]. Als Restraints der Torsionswinkel  $\Phi$ ,  $\Psi$  und  $\chi_i$  des Proteinrückgrates und der Seitenketten der Zielsequenz sind zufällige Werte in einem Bereich von  $0^\circ$ - $360^\circ$  zulässig. Diese räumlichen Restraints werden unter Verwendung von Moleküldynamik-Methoden ("Simulated Annealing") optimiert [104,112]. Die Stärke der abgeleiteten Restraints ist wesentlich von der Sequenzidentität abhängig. Je höher die Identität ist, um so größer ist die Stärke der verwendeten Restraints, wodurch wiederum ein genaueres 3D-Modell der unbekannt Proteinstruktur bedingt wird.

### 3.1.4 Optimierung der Strukturmodelle

Energieoptimierungen eines neu generierten Modells sind notwendig, um sterisch ungünstige Kontakte bzw. eine schlechte Geometrie, die innerhalb des Modells auftreten können, aufzuheben. Mit Hilfe der Energieoptimierung des Strukturmodells lassen sich solche Fehler korrigieren, wobei allerdings die Optimierung keine wesentlichen Strukturveränderungen hervorrufen sollte. Bevor eine Energieoptimierung durchgeführt wird, muß jedoch die Struktur hinsichtlich ihrer Geometrie untersucht werden. Treten hierbei Probleme auf (z. B. Überlappung von Aminosäureseitenketten), müssen diese manuell beseitigt werden, ohne daß dabei grundlegende Veränderungen innerhalb des Modells auftreten.

Die modellierten Proteinstrukturen wurden mit dem TRIPOS-Kraftfeld in ihrer Energie minimiert. Verwendet wurde das POWELL-Optimierungsverfahren, welches in SYBYL/MAXIMIN2 enthalten ist

[106]. Während des gesamten Prozesses beruhten die elektrostatischen Wechselwirkungen auf der Gasteiger-Partialladungsverteilungen. Es wurde eine abstandsabhängige Dielektrizitätskonstante von  $\epsilon = 4r$  und ein cut-off-Wert von  $8 \text{ \AA}$  bei der Berechnung der nichtbindenden Wechselwirkungen zwischen den Atomen verwendet [113,114]. Das Energieminimum wurde erreicht, wenn die Veränderung der generierten Strukturmodelle in ihren Gradienten nicht größer als  $0.01 \text{ kcal}/(\text{mol}\cdot\text{\AA})$  war. Da die Strukturen eine sehr hohe Korrelation zu den homologen Proteinen aufwiesen, wurden die Koordinaten des Proteinrückgrates bei einer ersten Minimierung eingefroren, so daß es nur den Seitenketten der Aminosäuren erlaubt war, ihre Geometrie während des Minimierungsprozesses zu ändern. Es wurden so die energetisch ungünstigen Kontakte zwischen den Seitenketten aufgehoben. Wurden die auftretenden Probleme in der Stereochemie berichtigt und die Seitenkettenpositionen energetisch verfeinert, konnte die Energieminimierung auf alle Atome des Proteins ausgedehnt werden.

### 3.1.5 Genauigkeit und Bewertung der Modelle

Die Genauigkeit und Qualität der Proteinmodelle läßt sich über den Vergleich mit ihren eventuell später zugänglichen Röntgenkristallstrukturen bzw. über den Vergleich mit verwandten Proteinen analysieren. Die Berechnung des RMS-Wertes zwischen dem Modell und dem bekannten Protein über alle  $C\alpha$ -Atome sowie über alle Atome ist ein Bewertungskriterium für die Qualität des Strukturmodells.

Der RMS-Wert wird über folgende Gleichung berechnet:

$$RMS = \sqrt{\frac{\sum d^2}{n}}$$

$d$  = Abstand zwischen den Atomen, die miteinander verglichen werden sollen  
 $n$  = Anzahl der zu vergleichenden Atome

Dieser Wert repräsentiert die gemittelte Abweichung von miteinander verglichenen Atomkoordinaten.

#### 3.1.5.1 Abschätzung der Fehler in Proteinstrukturmodellen

Eine Abschätzung auf mögliche Fehler ist für jedes Modell notwendig und kann auf verschiedenen Niveaus bezüglich des Strukturaufbaus durchgeführt werden:

- Überprüfen der Richtigkeit der gesamten Faltung des Proteins,
- Auffinden von Fehlern einer bestimmten Region in einer ansonsten korrekt gefalteten Sequenz,

- Ermittlung der Parameter, wie z. B. Bindungslängen, -winkel, Torsionswinkel und Identifizierung von sterisch ungünstigen Kontakten.

### 3.1.5.1.1 Bestimmung von Mißfaltungen und Fehlern in lokalen Regionen

*Sippl et al.* entwickelten das Programm PROSA, mit dem die Konformationsenergie von verschiedenen Faltungen in einer gegebenen Aminosäuresequenz berechnet werden kann [115]. Dieses Verfahren zur Bestimmung der mittleren Energie von Proteinen basiert auf der inversen Boltzmann-Statistik:

$$E_{ijk} = -kT \ln(f_{ijk})$$

$E_{ijk}$  = Energiefunktion

$f_{ijk}$  = relative Wahrscheinlichkeit

$k$  = Boltzmann-Konstante

$T$  = Temperatur

Die Energiefunktion  $E_{ijk}$  wird als das Potential der mittleren Energie bezeichnet ("Potential of mean force"). Die Energiezustände  $i, j, k$  werden von der relativen Wahrscheinlichkeit  $f_{ijk}$  abgeleitet.

Werden experimentell bestimmte 3D-Proteinstrukturen bezüglich ihres Aufbaus analysiert, erhält man grundlegende Informationen über die Kräfte, die zur Stabilisierung nativ gefalteter Strukturen beitragen. Unter Verwendung der inversen Boltzmann-Gleichung wurden diese Kräfte zahlreicher Proteine, die in der PDB enthalten sind, zur Erstellung der mittleren Energiepotentiale von nativ gefalteten Proteinstrukturen herausgefiltert. Dazu wurde die statistische Verteilung der  $C\alpha$ - $C\alpha$ - und  $C\beta$ - $C\beta$ -Wechselwirkungspotentiale sowie zur Berücksichtigung von Solvationseffekten die statistische Verteilung der Oberflächenpotentiale der  $C\alpha$ - und  $C\beta$ -Atome analysiert [116]. Dadurch kann die Energieverteilung eines bestimmten Proteins mit bekannter oder unbekannter Struktur durch die Kombination dieser Potentiale als Funktion seiner Aminosäuresequenz ermittelt werden. Mit Hilfe von PROSA wird ein Energiegraph berechnet, der die energetische Architektur der Proteinfaltung als eine Funktion der Aminosäuresequenzlänge darstellt. Dieser Graph gibt Aufschluß über die Potentiale der mittleren Energieverteilung im Protein in Abhängigkeit zur Aminosäuresequenz. In diesem Diagramm weisen Werte mit positiver Energie auf gespannte Bereiche im Protein und negative Werte korrelieren mit stabilen Teilen im Protein. Es ist daher möglich zwischen korrekt und mißgefalteten Strukturen zu unterscheiden.

Außerdem kann unter Anwendung von PROSA der sogenannte Z-Score berechnet werden [115]. Dieser Wert, der von der betrachteten Aminosäuresequenzlänge abhängt, ist ein zusätzlicher Hinweis auf die Qualität der zu untersuchenden Proteinstruktur. Ein Polyprotein wurde zur Berechnung des Z-Scores verwendet. Dieses Polyprotein ist aus Proteineinheiten aufgebaut, die miteinander verbunden sind. Es besteht aus 230 Proteinen mit bekannter Struktur mit einer Gesamtlänge von ca. 50000

Aminosäureresten. Die Konformationen des Polyproteins nehmen eine ideale Stereochemie sowie zahlreiche weitere Merkmale von nativ gefalteten Proteinen ein [115].

Die modellierten Strukturen wurden mit dem Polyprotein kombiniert und deren Energien berechnet. Der Z-Score ( $Z_p$ ) geht aus der resultierenden Energieverteilung hervor und berechnet sich über die folgende Gleichung:

$$Z_p = \frac{E_p - \bar{E}}{\sigma}$$

$E_p$  = Gesamtenergie eines Proteins in Abhängigkeit zu seiner Sequenz und Konformation

$\bar{E}$  = Durchschnittsenergie aller Fragmente, abgeleitet vom Polyprotein

$\sigma$  = Standardabweichung

Für einige Vertreter der Papain-Superfamilie (Papain, Actinidin, Cathepsin B, Cathepsin L) mit zugänglicher Röntgenstruktur (PDB-Eintrag: 9PAP [19], 1AEC [117], 1HUC [53], 1CJL [118]) wurde der Z-Score unter Berücksichtigung der  $C\alpha$ - $C\alpha$ -,  $C\beta$ - $C\beta$ -Wechselwirkung sowie der Oberflächenpotentiale der  $C\alpha$ - und  $C\beta$ -Atome berechnet. Daraus resultiert für den Z-Score dieser Proteasen ein Bereich von -7.41 bis -8.85.

### 3.1.5.1.2 Bewertung der stereochemischen Qualität der Tertiärstrukturmodelle

Bindungslängen, -winkel und Torsionswinkel müssen für eine zuverlässige Verwendung von Proteinmodellen so exakt wie möglich sein. *Labanowski et al.* analysierten die ( $\Phi$ ,  $\Psi$ )-Torsionswinkel, die Peptidplanarität, Bindungslängen, -winkel, die Wasserstoffbrückengeometrie sowie die Seitenkettenkonformationen von bekannten Proteinstrukturen als Funktion zu ihrer kristallographischen Auflösung [119]. Resultierend aus diesen Daten leiteten sie Werte abhängig von ihrer Auflösung (für Röntgenkristallstrukturen) der 3D-Strukturen ab. Das Programm PROCHECK gibt visuelle und quantitative Informationen über diese Merkmale des Strukturmodells wieder (wie z. B. Verteilung der ( $\Phi$ ,  $\Psi$ )-Torsionswinkel, Bindungswinkel und -länge, Qualität der Peptidbindungen, Seitenkettenkonformation) [119]. Bezugnehmend auf diese Kriterien wurden die modellierten Proteinstrukturen in ihrer Geometrie analysiert, wobei diese Berechnungen auf die Werte einer Röntgenstruktur mit einer Auflösung von 2.0 Å extrapoliert wurden. Der resultierende Ramachandran-Plot erlaubt eine Einteilung der ( $\Phi$ ,  $\Psi$ )-Torsionswinkel des Proteins in favorisierte, erlaubte und nichterlaubte Konformationen der untersuchten Struktur. Dabei bezeichnet man den Torsionswinkel um die Bindung zwischen dem N-Atom und dem  $C\alpha$ -Atom als  $\Phi$  ( $C'-C\alpha-N-C'$ ) und den um die  $C\alpha$ - $C'$ -Bindung als  $\Psi$  ( $N-C'-C\alpha-N$ ) [120].

### 3.1.6 Protein-Ligand-Docking

Die Bindung eines Liganden an seinen hochmolekularen Bindungspartner ist ein komplexer Vorgang, bei dem energetische und entropische Faktoren eine Rolle spielen. Dabei erfolgt eine (partielle) Desolvatation des Liganden und der Bindungsstelle des Enzyms und es kommt zu Konformations- und Entropieänderungen. Ein der Realität nahekommendes Eindocken des Liganden, bei dem alle Effekte Berücksichtigung finden, ist mit Hilfe von Computerprogrammen bis heute nicht möglich. Klassische Dockingprogramme, wie z. B. DOCK oder AUTODOCK, verwenden die sterische Komplementarität des Liganden zum Protein als Auswahlkriterium für den Dockingvorgang in Kombination mit einem Kraftfeld [121,122]. Dabei wird die konformelle Flexibilität des Proteins vernachlässigt. Der Nachteil dieser Verfahren liegt in der Annahme, daß die Bindung primär durch die Komplementarität der Gestalt bestimmt wird und das Einpassen des Liganden die Proteingestalt nicht beeinflusst. Wenn die Flexibilität des Bindungspartners berücksichtigt wird, erhöht sich die Anzahl möglicher Dockingpositionen und Orientierungen exponentiell. Um diesen großen Suchraum effizient abzutasten, ist der Einsatz genetischer Algorithmen von Vorteil. Das bei den nachfolgenden Berechnungen benutzte Programm FLEXIDOCK verwendet zur optimalen Positionierung des Liganden im Bindungszentrum des Enzyms einen genetischen Algorithmus [106,123].

Zur Überprüfung der Genauigkeit der Struktur der Umgebung der Substratbindungsregion wurden Substrate und Inhibitoren von Cathepsinen, deren kinetische Daten in der Literatur beschrieben sind, in die Bindungsspalte der zu untersuchenden Cysteinproteasen unter Verwendung des Programms FLEXIDOCK gedockt. Dazu wurden die Liganden in einem ersten Schritt manuell in die Substratbindungsregion des Enzyms eingepaßt. Das Programm FLEXIDOCK erlaubt dem Liganden sowie den Aminosäureseitenketten der Bindungsregion des Enzyms während der Dockingprozedur ihre Anordnung zu verändern, so daß ein "induced fit"-Mechanismus simuliert werden kann. Das Programm besteht aus zwei Hauptkomponenten:

1. dem genetischen Algorithmus (GA) zur Veränderung der Konformation des Liganden und der Proteinbindungsstellen und ihre relative Anpassung zueinander,
2. der Energieberechnung zur Bewertung der resultierenden Wechselwirkungen zwischen Ligand und Enzym [123].

Die Liganden sollten in jedem Fall in der Bindungsregion manuell vorpositioniert sein, um eine optimale Anordnung zu erhalten. Während der Prozedur ist das Proteinrückgrat fixiert, es werden jedoch Variationen der Seitenkettenanordnungen zugelassen. Der Ligand ist mobil (Rotation und Translation werden angewendet) und enthält zahlreiche flexible Bindungen. Folgende Kriterien wurden bei der Auswahl der rotierbaren Bindungen verwendet: Einfachbindungen, die weder Bestandteil von Ringen bzw. von Peptidbindungen sind sowie die Bindungen der Aminosäureseitenketten, die zu den Bindungsstellen des Enzyms gehören. Der cut-off-Abstand der nichtbindenden Wechselwirkungen zwischen den Aminosäureresten wurde auf 16 Å gesetzt.

Gasteiger-Partialladungsverteilungen wurden den Liganden und auch dem Enzym zugeordnet. Es wurde eine abstandsabhängige Dielektrizitätskonstante von  $\epsilon = 4r$  verwendet. Es hat sich gezeigt, daß wenn die Anzahl der Generationszyklen auf 10000 gesetzt wird, man davon ausgehen kann, daß die Anordnungsmöglichkeiten der Liganden weitestgehend vollständig ausgeschöpft wurden. Eine sukzessive Generierung der Anordnungen wurden über einen "steady-state" GA realisiert. Neue Möglichkeiten in der Anordnung des Liganden wurden durch Überkreuzungen gebildet, wobei die ursprünglichen Konformationen zufällig ausgewählt wurden. Geeignete Anordnungen wurden mittels einer Bewertungsmethode klassifiziert. Die Anzahl der Inseln (für die Überkreuzungen) wurde auf 8 gesetzt. Gleiche Mitglieder einer Insel wurden zusammengefaßt. Die so erhaltenen Protein-Ligand-Komplexe der Cathepsine wurden anschließend in ihrer Energie mit dem TRIPOS-Kraftfeld optimiert (siehe Abschnitt 3.1.4).

Um die konformelle Stabilität der in das Enzym gedockten Liganden zu überprüfen, wurden die isolierten Strukturen dieser Liganden in ihren energetischen Verhalten untersucht. Zusammenfassend konnte dabei festgestellt werden, daß aus den Energieoptimierungen keine gravierenden Veränderungen im strukturellen Aufbau der Verbindungen resultierten. Daher kann man davon ausgehen, daß die Liganden in der Konformation, wie sie im Enzym vorliegen, nahe eines lokalen Energieminimums liegen.

### 3.1.6.1 Beurteilung der Affinität des Liganden zum Enzym

Über die Berechnungen der Bindungsaffinität in Form der Wechselwirkungsenergie des Liganden im Enzym erhält man eine erste Information über das Wechselwirkungsverhalten dieses Liganden mit dem Enzym. Diese Energie wurde mit dem TRIPOS-Kraftfeld berechnet [106]. Allerdings ist diese Vorgehensweise nur eine Abschätzung der realen Wechselwirkungen zwischen dem Liganden und dem Enzym. Es werden hierbei keinerlei Effekte der Solvation bzw. Desolvation des Enzym-Ligand-Komplexes berücksichtigt. Die Änderungen der Entropie oder Enthalpie des Komplexes auf die Energiewerte können nicht mit einbezogen werden. Berechnet wird die Energie aus der Differenz des Komplexes und seinen einzelnen Komponenten.

Als weitere Methode zur Berechnung der Affinität des Liganden zum Enzym wurde LEAPFROG genutzt [106]. Dieses Programm wurde sowohl zum Ligandendesign als auch zur Berechnung der resultierenden Wechselwirkungsenergien zwischen Ligand und Protein getestet. Allerdings konnte zwischen den berechneten Energiewerten und den experimentellen Daten keine bzw. nur eine sehr geringe Korrelation festgestellt werden. Aus diesem Grund wurde dieses Programm für weiterführende Untersuchungen nicht verwendet.

Daher wurde zur weiteren Bestimmung der Affinität des Liganden zum Enzym das Programm SCORE ausgewählt [124]. Mit SCORE läßt sich die Bindungsaffinität von Protein-Ligand-Komplexen



berechnen, deren 3D-Strukturen bekannt sind. Dabei wird eine empirische Bewertungsfunktion verwendet, um die freie Bindungsenergie zu beschreiben. Diese enthält verschiedene Terme für van der Waals-Kontakte, Metall-Ligand-Bindungen, Wasserstoffbrückenbindungen, Desolvationseffekte sowie für die Deformation, die während des Bindungsprozesses auftreten kann. Die Koeffizienten für jeden Term wurden über eine Regressionsanalyse eines Trainingsatzes, bestehend aus 170 Protein-Ligand-Komplexen, erhalten [124]. Es wird angenommen, daß die freien Energieänderungen während des Bindungsprozesses in grundlegende Komponenten aufgegliedert werden können. Die Bewertungsfunktion von SCORE nimmt die nachfolgend beschriebene Form ein:

$$pK_d = K_0 + K_{vdw} + K_{Metall} + K_{H-Brücke} + K_{Desolvation} + K_{Deformation}$$

wobei  $K_{vdw}$  den Beitrag der van der Waals-Wechselwirkungen zwischen Protein und seinem Liganden darstellt,  $K_{metall}$  den der Metall-Ligand-Bindung und  $K_{H-Brücke}$  den der Bildung von Wasserstoffbrückenbindungen.  $K_{Desolvation}$  beschreibt den Desolvationseffekt und  $K_{Deformation}$  die Deformation, die während des Bindungsprozesses auftreten kann.  $K_0$  ist die Regressionskonstante, welche die Entropieänderung des Bindungsprozesses enthält [124]. Auf der Grundlage dieser Bewertungsfunktion resultiert bei der Berechnung der Komplexe mit SCORE der negative dekadische Logarithmus der Dissoziationskonstante  $pK_d$  für den gegebenen Protein-Ligand-Komplex. Außerdem erhält man detaillierte Informationen über den Bindungswert für jedes einzelne Atom des Liganden, was mit dem SYBYL-Programm graphisch dargestellt werden kann. Man kann eine genaue Abstufung der Bindungsaffinität jedes Atoms im Liganden zum Protein ableiten. Die berechnete Standardabweichung für SCORE beträgt 1.50 kcal/mol.

Die berechneten Werte stellen die Bindungsaffinität eines Liganden zum Enzym dar. Da die  $K_m$ -Werte eines Enzym-Substrat-Komplexes nicht nur von dieser Affinität sondern auch von der Hydrolyse des Substrates abhängen, können keine genauen und vollständigen Aussagen über  $K_m$ -Werte unter Verwendung der hier aufgeführten Methoden des "Molecular Modelling" getroffen werden. Um einen Zusammenhang zwischen den Wechselwirkungsenergien sowie der mit SCORE berechneten Dissoziationskonstanten und den experimentell bestimmten kinetischen Daten erstellen zu können, müssen mehrere Vereinfachungen getroffen werden (siehe Seite A-6).

## 3.2 Design von Peptidliganden

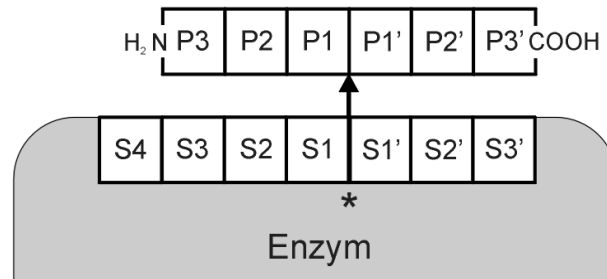
### 3.2.1 Aufbau und Analyse der Substratbindungsregionen von Cathepsinen

Um für die Cathepsine neue Liganden zu modellieren, müssen die Eigenschaften der Substratbindungsregionen der Enzyme charakterisiert werden. Das katalytisch aktive Bindungszentrum eines Proteins hat eine doppelte Funktion: die Bindung des Substrates sowie die Katalyse der Reaktion. Die Effizienz dieser Interaktionen bestimmt die Spezifität des Substrates zum Enzym.

Die Cysteinproteasen besitzen in der Umgebung der katalytischen Triade eine Bindungsregion für Substrate bzw. Inhibitoren. Die Bindungsregion besteht aus mehreren Bindungspositionen ( $S_n$ ). Durch eine spezifische Anordnung der Enzymhauptkette und der Aminosäureseitenketten werden die Bindungsstellen gebildet, die eine bestimmte Form und Größe besitzen und durch elektrostatische sowie lipophile Eigenschaften charakterisiert sind. Die räumliche Anordnungen der einzelnen Bindungspositionen haben ebenfalls Einfluß auf die Substratspezifität.

Eine enzymatische Reaktion ist in verschiedene nacheinander ablaufende Schritte zu untergliedern. Zunächst nähert sich der Ligand der Bindungsregion des Enzyms. Durch nichtkovalente Wechselwirkungen wird der Ligand fixiert und optimal für eine Reaktion in das katalytisch aktive Zentrum des Enzyms gebunden. Nach Abschluß der Reaktion verlassen die Reaktionsprodukte die Bindungsregion wieder. Schon *Emil Fischer* hat erkannt, daß für eine energetisch bevorzugte Bindung von Liganden am Enzym komplementäre Eigenschaften der Liganden zu den Bindungsstellen des Enzyms von Vorteil sind [125]. Es konnte aber auch gezeigt werden, daß diese Bindungstaschen keine starren Gebilde, sondern in gewissen Grenzen variabel sind. Auch die Liganden können ihre Konformation während des Bindungsprozesses ändern. Zum Teil nehmen diese im gebundenen Zustand am Enzym eine andere Konformation ein als im Lösungsmittel. Man spricht dann vom "soft lock and key"-Konzept [126].

Die Spezifität der Cysteinproteasen läßt sich unter Verwendung eines Modells beschreiben, in dem das aktive Zentrum so betrachtet wird, daß es von einer oder beiden Seiten von Bindungsstellen flankiert wird (Abb. 3) [47]. Die Bindungsstellen besitzen die Fähigkeit, mit einer Aminosäure zu interagieren. Diese Bindungspositionen sind ausgehend von dem katalytischen Zentrum mit  $S_1 \dots S_n$  zum N-Terminus der Struktur sowie mit  $S_1' \dots S_n'$  in Richtung zum C-Terminus numeriert. Die Reste des Liganden, welche die jeweiligen Bindungsstellen belegen, sind entsprechend numeriert:  $P_1 \dots P_n$  und  $P_1' \dots P_n'$  (Abb. 3) [47].

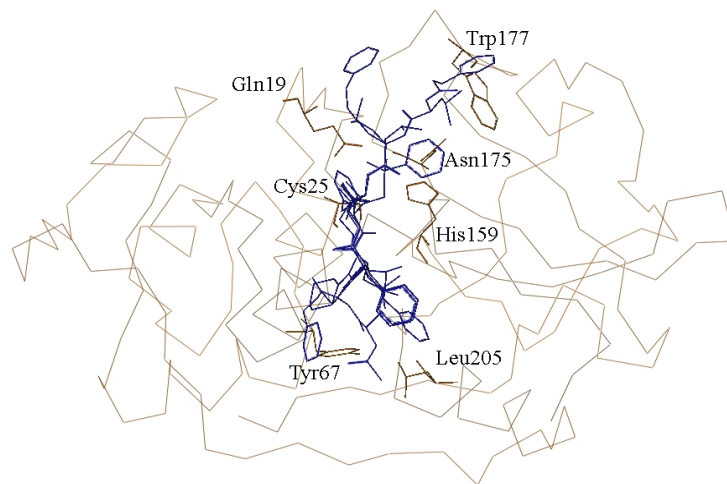


**Abb. 3** Schematische Darstellung eines Enzym-Substrat-Modells mit einem Hexapeptid [47]. Das katalytisch aktive Bindungszentrum des Enzyms besteht aus sieben Bindungspositionen. Das katalytische Zentrum ist mit einem Stern gekennzeichnet. Die Positionen  $P_n$  des Substrates werden vom Spaltungspunkt in Richtung des N-Terminus, die  $P_n'$ -Positionen in Richtung C-Terminus gezählt und haben die gleiche Numerierung wie die Bindungsstellen, die durch sie belegt werden.

Ein Ziel der vorliegenden Arbeit war es, Form, Größe und Eigenschaften der Bindungsstellen der Cathepsine zu untersuchen, um daraus Schlußfolgerungen über die Spezifität dieser Enzyme ziehen zu können. Ausgangspunkt dieser Studien waren die zugänglichen Röntgenkristallstrukturen bzw. modellierten Proteinstrukturen der Cathepsine. Um neu entwickelte Liganden optimal in den einzelnen Bindungspositionen zu orientieren, mußten zunächst die Bindungsstellen genauer charakterisiert werden. Zum einen wurde dazu das Programm GRID von *Goodford* verwendet, um Einblicke in die Wechselwirkungen zwischen Ligand und Enzym zu gewinnen [127]. Dieses Programm arbeitet nach folgendem Prinzip: es wird ein Gitter um das zu untersuchende Molekül gelegt. An jedem Gitterpunkt wird die Wechselwirkung zwischen dem Molekül und einer "Probe" berechnet. Man erhält als Ergebnis Bereiche im Raum, in denen günstige Wechselwirkungen zwischen dieser Probe und dem Molekül zu erwarten sind. Auf diese Weise kann man sterische und elektrostatische Eigenschaften von Molekülen abtasten. Als Proben können einzelne Atome oder funktionelle Gruppe ausgewählt werden. In diesem Fall wurden die Methylgruppe, eine positiv geladene Ammoniumgruppe sowie eine negativ geladene Carboxylatgruppe verwendet. Die graphische Veranschaulichung dieser GRID-Felder wurde mit dem SYBYL-Programm durchgeführt. Die Konformation der Substrate kann dann so variiert werden, daß die entsprechenden funktionellen Gruppen zu den jeweiligen GRID-Feldern positioniert sind. Weiterhin wurden die Eigenschaften der elektrostatischen Potentiale der molekularen Oberfläche mit dem Programm MOLCAD untersucht [128]. Für eine Bewertung der elektrostatischen Eigenschaften auf der Enzymoberfläche der Cathepsine wurden mit MOLCAD die elektrostatischen Potentiale der Substratbindungsregion des Enzyms basierend auf Gasteiger-Ladungen berechnet [113]. Für jedes Cathepsin wurde dazu eine vom Lösungsmittel zugängliche Connolly-Oberfläche kreiert [129]. Der Probenradius wurde auf 1.4 Å gesetzt. Als Endgruppen der untersuchten Enzymfragmente wurde die Acetylgruppe (N-Terminus) und das N-Methylamid (C-Terminus) verwendet.

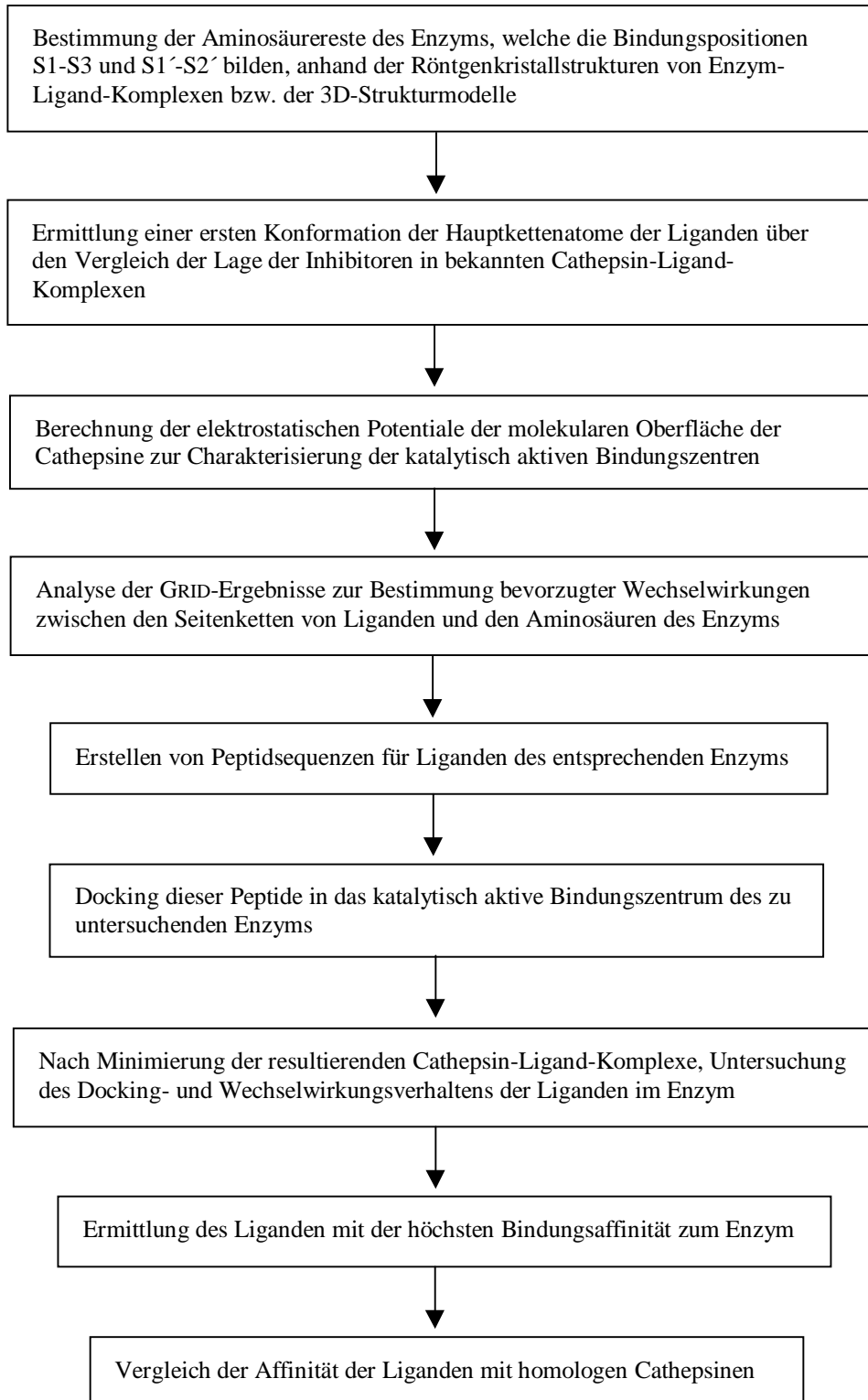
### 3.2.2 Entwicklung von neuen Liganden

Über Röntgenkristallstrukturanalyse konnten die Proteinstrukturen von mehreren Mitgliedern der Papain-Superfamilie ermittelt bzw. generiert werden (Tab. A-1). Eine Vielzahl dieser 3D-Strukturen sind Protein-Ligand-Komplexe dieser Enzymklasse. Basierend auf diesen Proteinstrukturen erhält man erste Informationen über mögliche Struktur-Funktions-Beziehungen der Liganden im Enzym. Vergleicht man Lage und Konformation der Hauptkettenatome dieser Liganden wird ersichtlich, daß große Ähnlichkeiten auftreten (Abb. 4).



**Abb. 4** Vergleich der Inhibitorstrukturen (blau) der Enzym-Inhibitor-Komplexe von 1AU0 [130], 1BP4 [131], 1MEM [132], 1PAD [48] und 5PAD [48] zur Struktur eines Cathepsin K-Komplexes (1MEM [132]). Die Rückgratstruktur, gebildet über die C $\alpha$ -Atome des Cathepsin K, sowie einige wesentliche Aminosäuren des Enzyms sind im Hintergrund dargestellt (braun).

Zum Design von neuen Liganden für die Cathepsine L, K, und F wurde die Vorgehensweise angewendet, wie sie im Schema 3 dargestellt ist.



**Schema 3** Verwendete Vorgehensweise zum Design von neuen Liganden der Cathepsine L, K und F.

Auf der Basis der bekannten Röntgenkristallstrukturen der Enzym-Inhibitor-Komplexe von Papain sowie der Cathepsine B und K (Tab. A-1) wurden die Rückgratkonformationen der neu erstellten Peptidsequenzen der Liganden ermittelt. Die Seitenketten der Inhibitorreste wurden so modifiziert,

---

daß sie der Sequenz des zu untersuchenden Liganden entsprachen, ohne die Koordinaten der Hauptkettenatome stark zu verändern. Zur Positionierung der erstellten Peptidsequenz im Enzym wurden die Eigenschaften der Substratbindungsregionen untersucht. Die auf diese Weise aufgebauten Enzym-Ligand-Komplexe wurden als Startstrukturen zur Durchführung von FLEXIDOCK verwendet (siehe Abschnitt 3.1.6). Es resultierten zahlreiche Anordnungsmöglichkeiten der Liganden im Enzym, deren Energie mit dem TRIPOS-Kraftfeld minimiert wurden (siehe Abschnitt 3.1.4). Die erhaltenen Komplexe wurden hinsichtlich der Affinität des Liganden zum Enzym analysiert (Interaktionen zwischen Ligand und Enzym, Tendenz der Ausbildung von Wasserstoffbrücken zwischen Ligand und Enzym). Die Wechselwirkungsenergien bzw. die  $K_d$ -Werte der Enzym-Ligand-Komplexe wurden in der Weise bestimmt, wie es in Abschnitt 3.1.6.1 beschrieben wurde. Um nun eine Differenzierung zwischen den einzelnen Cathepsinen ableiten zu können, wurden die Liganden, die z. B. für das Cathepsin L entwickelt wurden, zum Vergleich in andere catheptische Cysteinproteasen gedockt. Die oben beschriebene Verfahrensweise wurde für diese Komplexe wiederholt. Die Liganden, welche sich aufgrund dieser Untersuchungen als hochaffin für das jeweilige Cathepsin erwiesen, wurden zur Synthese vorgeschlagen (AK Prof. K. Neubert, Universität Halle-Wittenberg; Dr. R. Singh, Rigel Inc., San Francisco, USA) und an den Cathepsinen getestet (Dr. F. Bühling, Universität Magdeburg, AK Prof. D. Brömme, Mount Sinai School of Medicine, New York, USA).

## 4 ERGEBNISSE UND DISKUSSION

### 4.1 Sequenzanalyse der Papain-ähnlichen Cysteinproteasen

Mitglieder der Papain-Superfamilie werden als inaktive Präkursoren beschrieben, die sich zu ihrer katalytisch aktiven Form umsetzen [33,133]. Diese Enzyme werden in ihrer Präproform expremiert, die eine kurze N-terminale Signalsequenz von durchschnittlich 12-16 Aminosäuren enthält. Daran schließt sich eine Sequenz an, die als Propeptid bezeichnet wird und spezifisch für das jeweilige Cathepsin ist (siehe Abschnitt 1).

Zur Untersuchung der Sequenzähnlichkeiten zwischen den lysosomalen Cysteinproteasen wurde ein Vergleich innerhalb der Sequenzen der katalytisch aktiven Formen der humanen Cathepsine K, L, H, S, F und B sowie für Papain und Actinidin durchgeführt (Tab. A-3). Ein Primärsequenzalignment von Cathepsin K zu den genannten Enzymen ist in Abb. 5 zu sehen (Seite 33). Die Werte der Proteinsequenzhomologie wurden unter Verwendung von BLAST ermittelt [108]. Die höchste Sequenzhomologie tritt zwischen den Cathepsinen K und L auf (74% Homologie, siehe Tab. 1). Eine stufenweise reduzierte Homologie wurde zu Cathepsin S (71%), Cathepsin H (61%), Actinidin (59%), Papain (57%), Cathepsin F (56%) und Cathepsin B (21%) gefunden.

Anhand des Sequenzalignments ist zu erkennen, daß in allen Enzymen an der zweiten Position ein konservierter Prolinrest auftritt. Außerdem erlaubt dieses Alignment die Identifizierung des katalytisch aktiven Cysteins sowie der anderen Aminosäuren, die an der katalytischen Triade beteiligt sind (Abb. 5). Die Bereiche, die das aktive Zentrum umgeben, sind innerhalb der catheptischen Cysteinproteasen konserviert. Die N-terminale Region enthält ein Glutamin, dessen Seitenkette zur Stabilisierung des während der Katalyse gebildeten Oxyanions beiträgt sowie ein konserviertes Tryptophan und weitere hydrophobe Bereiche, die zum katalytisch aktiven Cystein unmittelbar benachbart sind [134,118]. Weiterhin beinhaltet die C-terminale Region eine Reihe konservierter aromatischer Aminosäuren und Glycinreste, die um das Histidin und das Asparagin des aktiven Zentrums lokalisiert sind.

Neben dem Cystein der katalytischen Triade besitzen die Papain-ähnlichen Proteasen sechs zusätzliche Cysteinreste, die in ihrer Position konserviert sind und Disulfidbrücken ausbilden können. Ausnahmen bilden die Cathepsine S und B, welche eine bzw. drei zusätzliche Disulfidbrücken enthalten.

		10	20	30	40	50	60
				*			
hCatK	APDSVDYRKK	G----	YVTPV	KN <b>QGQCGSCW</b>	AFSSVGALEG	QLKKKTGKLL	NLSP--QNLV
hCatL	APRSVDWREK	G----	YVTPV	KN <b>QGQCGSCW</b>	AFSATGALEG	QMFRTGRLI	SLSE--QNLV
hCatF	APPEWDWRSK	G----	AVTKV	KD <b>QGMCGSCW</b>	AFSVTGNVEG	QWFLNQGTLL	SLSE--QELL
hCatS	LPDSVDWREK	G----	CVTEV	KY <b>QGSCGACW</b>	AFSAVGALEA	QLKLTGKLV	TLSA--QNLV
hCatH	YPPSVDWRKK	GN---	FVSPV	KN <b>QGACGSCW</b>	TFSTTGALES	AIAIATGKML	SLAE--QQLV
hCatB	LPASFDAREQ	WPQCPTIKEI		RD <b>QGSCGSCW</b>	AFGAVEAISD	RICIHTNAHV	SVEVSAEDLL
Papain	IPEYVDWRQK	G----	AVTPV	KN <b>QGSCGSCW</b>	AFSAVVTIEG	IIKIRTGNLN	EYSE--QELL
Actinidin	LPSYVDWRSA	G----	AVVDI	KS <b>QGECCGSCW</b>	AFSAIATVEG	INKIIVTGVLI	SLSE--QELI
		70	80	90	100	110	120
hCatK	DCVSE---ND	G <b>CGGGY</b> MTNA	FQYVQKNRGI	DS-----	-----	EDA	YPYVGQEESC
hCatL	DCSGPQG-NE	G <b>CNGGL</b> MDYA	FQYVQDNGLL	DS-----	-----	EES	YPYEATEESC
hCatF	DCDK---MDK	A <b>CMGGL</b> PSNA	YSAIKNLGGL	ET-----	-----	EDD	YSYQGHMQSC
hCatS	DCSTEKYGNK	G <b>CNGGF</b> MTTA	FQYI IDNKGI	DS-----	-----	DAS	YPYKAMDQKC
hCatH	DCAQD-FNNY	G <b>CQGG</b> LPSQA	FEYILYNKGI	MG-----	-----	EDT	YPYQKDGVC
hCatB	TCCGS-MCGD	G <b>CNGG</b> YPAEA	WNFWTRKGLV	SGGLYESHVG	CRPYSIPPCE	HHVNGSRPPC	
Papain	DCDR---RSY	G <b>CNNG</b> YPWSA	LQLVAQYG-I	HY-----	-----	RNT	YPYEGVQRYC
Actinidin	DCGRTQ-NTR	G <b>CNNG</b> YITDG	FQFI INNGGI	NT-----	-----	EEN	YPYTAQDGEC
		130	140	150	160	170	180
hCatK	MYNPTG-KAA	KCRGYREIPE	G-----	-----	<b>NEKA</b>	LKRAVARVGP	VSVAIDASLT
hCatL	KYNPKY-SVA	NDTGFVDIPK	-----	-----	<b>QEKA</b>	LMKAVATVGP	ISVAIDAGHE
hCatF	NFSAEK-AKV	YINDSVELSQ	-----	-----	<b>NEQK</b>	LAAWLAKRGP	ISVAINAFG-
hCatS	QYDSKY-RAA	TCSKYTELPY	G-----	-----	<b>REDV</b>	LKEAVANKGP	VSVGVDARHP
hCatH	KFQPGK-AIG	FVKDVANITI	Y-----	-----	<b>DEEA</b>	MVEAVALYNP	VSFAFEVT-Q
hCatB	TGEGDTPKCS	KICEPGYSPT	YKQDKHYGYN	SYSVSN <b>SEKD</b>	IMAEIYKNGP	VEGAFSVYS-	
Papain	RSREKGPYAA	KTDGVRQVQP	Y-----	-----	<b>NEGA</b>	LLYSIAN-QP	VSVVLEAAGK
Actinidin	NLDDLQNEKYV	TIDTYENVPY	N-----	-----	<b>NEWA</b>	LQTAVTY-QP	VSVVALDAAGD
		190	200	210	220	230	240
				*		*	
hCatK	SFQFY <b>SKGVY</b>	--YDESCNS-	-DNLN <b>HAVLA</b>	V <b>GYGIQ</b> ---	KGNKH <b>WIKN</b>	<b>SWGENWGNKG</b>	
hCatL	SFLFY <b>KEGIY</b>	--FEPDCSS-	-EDMD <b>HGVLV</b>	V <b>GYGFESTES</b>	DNNKY <b>WLVKN</b>	<b>SWGEEWGMGG</b>	
hCatF	-MQFY <b>RHGIS</b>	RPLRPLCSP-	-WLID <b>HAVLL</b>	V <b>GYGNRS</b> ---	-DVPF <b>WAIKN</b>	<b>SWGTDWGEKG</b>	
hCatS	SFFLY <b>RSGVY</b>	--YEPSCT--	-QNVN <b>HGVLV</b>	V <b>GYGDLN</b> ---	-GKEY <b>WLVKN</b>	<b>SWGHNFGGEEG</b>	
hCatH	DFMMY <b>RTGIY</b>	--SSTSCHKT	PDKVN <b>HAVLA</b>	V <b>GYGEKN</b> ---	-GIPY <b>WIVKN</b>	<b>SWGPOWGMNG</b>	
hCatB	DFLLY <b>KSGVY</b>	---QHVTG--	EMMG <b>HAIRI</b>	L <b>GWGVEN</b> ---	-GTPY <b>WLVAN</b>	<b>SWNTDWDGNG</b>	
Papain	DFQLY <b>RGGIF</b>	--VGP-CG--	-NKVD <b>HAVAA</b>	V <b>GYGPN</b> ---	----Y <b>ILIKN</b>	<b>SWGTTWGGENG</b>	
Actinidin	AFKH <b>YSSGIF</b>	--TGP-CG--	-TAID <b>HAVTI</b>	V <b>GYGTEG</b> ---	-GIDY <b>WIVKN</b>	<b>SWDTTWGGEEG</b>	
		250	260	270	280	290	300
hCatK	YIILMARNKN-	--N <b>ACG</b> IANL	ASFPKM----	-----	-----	-----	-----
hCatL	YVKMAKDRR-	--N <b>HCG</b> IASA	ASYPTV----	-----	-----	-----	-----
hCatF	YYLHRGSG-	---A <b>CGV</b> NTM	ASSAVVD---	-----	-----	-----	-----
hCatS	YIRMARNKG-	--N <b>HCG</b> IASF	PSYPEI----	-----	-----	-----	-----
hCatH	YFLIERGKN-	---M <b>CGL</b> AAC	ASYIPLV--	-----	-----	-----	-----
hCatB	FFKILRGQD-	---H <b>CGI</b> ESE	VVAGIPRTD-	-----	-----	-----	-----
Papain	YIRIKRGTGN	SYG <b>VCGL</b> YTS	SFY <b>PKN</b> ---	-----	-----	-----	-----
Actinidin	YMRILRNVG-	GAG <b>TCGI</b> ATM	PSY <b>PKYNNQ</b>	NHP <b>KPYSSLI</b>	NPP <b>AFS</b> MSKD	GPV <b>GVD</b> DGQR	
		310					
hCatK	---						
hCatL	---						
hCatF	---						
hCatS	---						
hCatH	---						
hCatB	---						
Papain	---						
Actinidin	YSA						



**Abb. 5** Multiples Sequenzalignment von Cathepsin K mit den Cathepsinen L, F, S, H, B sowie Papain und Actinidin. Die Aminosäuren der katalytischen Triade sind mit einem Stern gekennzeichnet. Die Aminosäuren, die konserviert in allen Proteinsequenzen auftreten, wurden fett hervorgehoben. Gaps sind mit einem Bindestrich markiert. Die Proteinsequenzen wurden aus der Swiss-Prot-Datenbank und der TrEMBL-Datenbank entnommen. Die Zugriffsnummern der Cysteinproteasen sind P43235 (hCatK), P07711 (hCatL), O95240 (hCatF), P25774 (hCatS), P09668 (hCatH), P07858 (hCatB), P00784 (Papain) und P00785 (Actinidin).

**Tab. 1** Proteinsequenzhomologie und -identität von Cathepsin K zu den Cathepsinen L, F, S, H, B sowie zu Papain und Actinidin. Die Identitätswerte resultieren aus dem Sequenzalignment, das mit CLUSTAL W [97] durchgeführt wurde. Die Sequenzhomologie wurde unter Verwendung von BLAST ermittelt [108].

Enzym	Homologie [%]	Identität [%]
hCatL	74	58
hCatF	56	39
hCatS	71	57
hCatH	61	46
hCatB	43	21
Papain	57	41
Actinidin	59	38

## 4.2 Modellierung der Tertiärstrukturen der Cathepsine K, S, H und F

Zu Beginn der Untersuchungen an lysosomalen Cysteinproteasen bildete die Modellierung der Proteinstrukturen der Cathepsine K und S den ersten Schwerpunkt. Da im Rahmen der vorliegenden Arbeit zahlreiche vergleichende Betrachtungen an diesen beiden Cathepsinen durchgeführt wurden, insbesondere bei der Verifizierung der Strukturmodelle hinsichtlich ihrer Substratspezifität, werden als erstes diese beiden Enzyme in den nachfolgenden Kapiteln behandelt.

### 4.2.1 Ein Modell der dreidimensionalen Struktur von humanem Cathepsin K

Ein Proteinstrukturmodell des humanen Cathepsin K (hCatK) wurde basierend auf seiner Aminosäuresequenz und seiner Homologie zu verwandten Proteinen mit zugänglichen Röntgenkristallstrukturen unter Verwendung des Programms COMPOSER generiert (siehe Abschnitt 3.1.3.1). Resultierend aus der Sequenzidentität von hCatK zu seinen homologen Proteinen wurden die Koordinaten der Enzyme zur Tertiärstrukturmodellierung verwendet, die in Tab. 2 aufgelistet sind. Die Proteinsequenzidentität von diesen Enzymen zu Cathepsin K liegt generell über 40%.

**Tab. 2** Sequenzidentität von Cathepsin K zu den Proteinen, die zur Strukturmodellierung verwendet wurden. Diese Werte wurden über COMPOSER ermittelt [103].

Name des Proteins	PDB-Eintrag	Identität [%]
Procathepsin L	1CJL	60
Chymopapain	1YAL	46
Protease $\Omega$	1PPO	45
Glycylendopeptidase	1GEC	44
Papain	1PPN	43
Papain	9PAP	43
2-Hydroxyethylthiopapain	1PPD	43
Actinidin	1AEC	43

Ein multiples Sequenzalignment zwischen der Zielsequenz und den Vorlagesequenzen mit der Kennzeichnung der strukturell konservierten Regionen (SCR) ist in Abb. 6 dargestellt. Die Röntgenkristallstruktur von Cathepsin L (PDB-Eintrag: 1CJL [118]) enthält auch die ermittelten Koordinaten des Propeptides dieses Enzyms. Da dieser Bereich zur Modellierung der Tertiärstruktur von hCatK keine Rolle gespielt hat, wurde die Aminosäuresequenz des Propeptides in dem Sequenzalignment nicht mit abgebildet. Außerdem enthält die Sequenz des reifen Cathepsin L zwei Mutationen (Cys25Ser und Thr110Ala), die allerdings keine Auswirkungen auf die Modellierung der Proteinstruktur von Cathepsin K haben.

	10	20	30	40	50	60
			*			
hCatK	APDSVDYRKK	GYVTPVKNQG	QCGSCWAFSS	VGALEGQLKK	KTGKLLNLSL	QNLVDCVS-E
1CJL	APRSVDWREK	GYVTPVKNQG	QCGSSWAFSA	TGALEGQMFR	KTGRLISLSE	QNLVDCSGPE
1YAL	YPQSIDWRAK	GAVTPVKNQG	ACGS-WAFST	IATVEGINKI	VTGNLLELSE	QELVDCDKH-
1PPO	LPENVDWRKK	GAVTPVRHQG	SCGSCWAFSA	VATVEGINKI	RTGKLVLESE	QELVDCERR-
1GEC	LPESVDWRAK	GAVTPVKHQG	YCESCWAFST	VATVEGINKI	KTGNLLELSE	QELVDCDLQ-
1PPN	IPEYVDWRQK	GAVTPVKNQG	SCGSCWAFSA	VVTIEGI IKI	RTGNLNEYSE	QELLDCCR-
9PAP	IPEYVDWRQK	GAVTPVKNQG	SCGSCWAFSA	VVTIEGI IKI	RTGNLNQYSE	QELLDCCR-
1PPD	IPEYVDWRQK	GAVTPVKNQG	SCGSCWAFSA	VVTIEGI IKI	RTGNLNQYSE	QELLDCCR-
1AEC	LPSYVDWRSR	GAVVDIKSQG	ECGGCWAFSA	IATVEGINKI	VTGVLISLSE	QELIDCGR--
	----HHHH--	----E-----	--HHHHHHHH	HHHHH-EEE-	----EEEE--	HHHH-----
	70	80	90	100	110	120
hCatK	-ND--GCGG	GYMTNAFQYV	QKNRGIDSED	AYPYVGQEEES	CMYNPTGK-A	AKCRGYREIP
1CJL	GNE--GCNG	GLMDYAFQYV	QDNNGLDSEE	SYPYEATEES	CKYNPKYS-V	ANDAGFVDIP
1YAL	-SY--GCKG	GYQTTSLQYV	ANN-GVHTSK	VYPYQAKQYK	CRATDKPGPK	VKITGYKRVV
1PPO	-SH--GCKG	GYPYALEYV	AKN-GIHLRS	KYPYKAKQGT	CRAKQVGGPI	VKTSGVGRVQ
1GEC	-SY--GCNR	GYQSTSLQYV	AQN-GIHLRA	KYPYIAKQQT	CRANQVGGPK	VKTNGVGRVQ
1PPN	-SY--GCNG	GYPWSALQLV	AQY-GIHYRN	TYPYEGVQRY	CRSREKGPYA	AKTDGVRQVQ
9PAP	-SY--GCNG	GYPWSALQLV	AQY-GIHYRN	TYPYEGVQRY	CRSREKGPYA	AKTDGVRQVQ
1PPD	-SY--GCNG	GYPWSALQLV	AQY-GIHYRN	TYPYEGVQRY	CRSREKGPYA	AKTDGVRQVQ
1AEC	-TQNTRGCG	GYITDGFQFI	INNGGINTEE	NYPYTAQDGE	CNVDLQNEKY	VTIDTYENVP
	-----	HHHHHHHHHH	HHH-----	-----	-----	EEE---EE--
	130	140	150	160	170	180
					*	
hCatK	EGNEKALKRA	VARVGPVSVV	IDASLTSFQF	YKSGVYYDES	CNSDNLNHAV	LAVGYGIQKG
1CJL	KQ-EKALMKA	VATVGPISVA	IDAGHESFLF	YKEGIYFEPD	CSSDMDHGV	LVVGYGFESN
1YAL	S-NETSFLGA	LA-NQPLSVL	VEAGGKPFQL	YKSGVFDGP-	CGT-KLDHAV	TAVGYGTS DG
1PPO	PNNEGNLLNA	IA-KQPVSVV	VESKGRPFQL	YKGGIFEGP-	CGT-KVDHAV	TAVGYGKSGG
1GEC	SNNEGSLLNA	IA-HQPVSVV	VESAGRDFQN	YKGGIFEGS-	CGT-KVDHAV	TAVGYGKSGG
1PPN	PYNEGALLYS	IA-NQPVSVV	LEAAGKDFQL	YRGGIFVGP-	CGN-KVDHAV	AAVGYG----
9PAP	PYNQGALLYS	IA-NQPVSVV	LQAAGKDFQL	YRGGIFVGP-	CGN-KVDHAV	AAVGYG----
1PPD	PYNQGALLYS	IA-NQPVSVV	LQAAGKDFQL	YRGGIFVGP-	CGN-KVDHAV	AAVGYG----
1AEC	YNNWALQTA	VT-YQPVSVV	LDAAGDAFKQ	YSSGIFTGP-	CGT-AIDHAV	TIVGYGTEGG
	--HHHHHHHH	HHH---EEEE	EE---HHHEE	EE---EEE---	-----EE	EEEE-----
	190	200	210	220	230	240
	*					
hCatK	NKHWIIKNSW	GENWGNKGYI	LMARNK-N--	NACGIANLAS	FPKM---	
1CJL	-KYWLKNSW	GEEWGMGGYV	KMAKDRR---	NHCGIASAAS	YPTV---	
1YAL	KNYIIKNSW	GNWGEKGYM	RLKRQSGNSQ	GTCGVYKSSY	YPFKGFA	
1PPO	KGYILIKNSW	GTAWGEKGYI	RIKRAPGNSP	GVCGLYKSSY	YPTKN--	
1GEC	KGYILIKNSW	GPGWGNGYI	RIRRASGNSP	GVCGVYRSSY	YPIKN--	
1PPN	PNYILIKNSW	GTGWGNGYI	RIKRGTGNSY	GVCGLYTSSF	YPVKN--	
9PAP	PNYILIKNSW	GTGWGNGYI	RIKRGTGNSY	GVCGLYTSSF	YPVKN--	
1PPD	PNYILIKNSW	GTGWGNGYI	RIKRGTGNSY	GVCGLYTSSF	YPVKN--	
1AEC	IDYWIVKNSW	DTTWGEEGYM	RILRNVGGA-	GTCGIATMPS	YPVK-Y-	
	-----	-----EE	EEEE-----	--E--EE-----		

**Abb. 6** Multiples Sequenzalignment von hCatK mit seinen Vorlagesequenzen (durchgeführt mit COMPOSER). Die SCRs sind grau unterlegt und die Aminosäuren der katalytischen Triade sind mit einem Stern gekennzeichnet. Die letzte Zeile enthält die mit PHD berechneten Sekundärstrukturelemente (H = Helix, E =  $\beta$ -Faltblattstrang).

Anhand dieses Alignments ist zu sehen, daß sich die Bereiche der SCRs über große Abschnitte innerhalb der Primärstrukturen erstrecken. Das zur katalytischen Triade gehörende Cystein ist jedoch nicht in einem SCR enthalten. An dieser Position tritt in der Sequenz der Röntgenkristallstruktur von Chymopapain ein Gap auf. Zusätzlich enthält diese Abbildung neben den Primärsequenzen der

Enzyme die ermittelten Sekundärstrukturelemente für hCatK, um erste Informationen über die Anordnung dieser Elemente in der Proteinstruktur von hCatK zu erhalten.

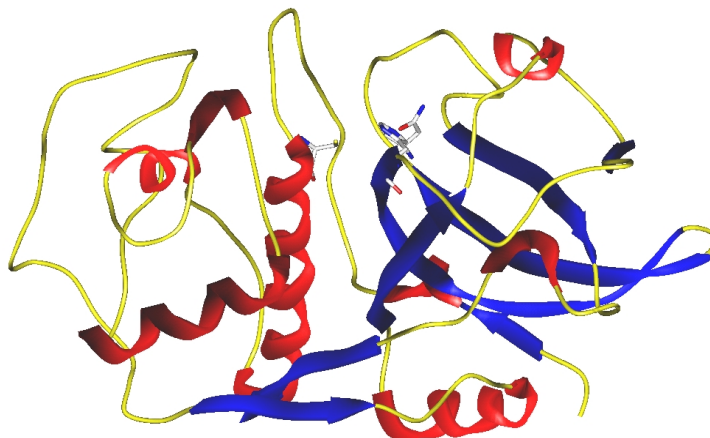
Das Modell von hCatK enthält 11 SCRs, bestehend aus 9-47 Aminosäuren sowie neun strukturell variable Regionen (SVR), die sich aus 1-10 Aminosäuren zusammensetzen (Tab. 3).

**Tab. 3** SCRs innerhalb der Proteinstruktur von hCatK, die von den homologen Cysteinproteasen bestimmt wurden und die resultierenden SVRs dieses Enzyms

SCR/ SVR	Länge der Sequenz	Aminosäuresequenz- nummer von hCatK	PDB-Eintrag des Vorlageproteins	erste Aminosäure des Vorlageproteins	Identität [%]
<b>SCR</b>					
1	24	1-24	1CJL	1	88
2	22	26-58	1CJL	26	67
3	15	62-76	1CJL	62	71
4	18	80-97	1CJL	82	61
5	7	109-115	1YAL	109	43
6	9	118-126	1CJL	119	89
7	25	128-152	1CJL	129	60
8	12	159-170	1CJL	160	58
9	23	176-198	1CJL	181	61
10	14	202-215	1CJL	207	57
<b>SVR</b>					
					Homologie der Loopregion [%]
1	1	25	1AEC	25	100
2	1	59	1GEC	59	82
3	3	77-79	1AEC	79	50
4	11	98-108	1CJL	100	59
5	2	116-117	1GEC	116	72
6	1	127	1CJL	128	30
7	6	153-158	1CJL	154	84
8	5	171-175	1YAL	168	66
9	3	199-201	1CJL	204	82

Die aus den homologen Proteinsequenzen generierten SCRs weisen zur Sequenz von hCatK eine Identität von mehr als 55% auf. Einzige Ausnahme bilden die Aminosäuren 109-115 (Tab. 3). Basierend auf dem multiplen Sequenzalignment konnten Disulfidbrücken zwischen den Cysteinresten Cys22-Cys63, Cys56-Cys96 und Cys155-Cys204 addiert werden. Vergleicht man das berechnete Strukturmodell von hCatK mit den SCRs der verwendeten homologen Proteine, so ergibt sich eine RMS-Differenz von 0.36 Å (berechnet über die C $\alpha$ -Atome).

Eine Minimierung des Proteinmodells mit dem TRIPOS-Kraftfeld wurde nach der Ergänzung der Wasserstoffatome sowie manueller Überarbeitung des Modells bezüglich auftretender sterischer Überlappungen durchgeführt (siehe Abschnitt 3.1.4). Das resultierende Tertiärstrukturmodell des humanen Cathepsin K ist in Abb. 7 zu sehen.



**Abb. 7** 3D-Strukturmodell von hCatK. Die Sekundärstrukturelemente wurden mit DSSP berechnet. Helices sind rot eingefärbt und  $\beta$ -Faltblattstrukturen blau. Die Aminosäurereste der katalytischen Triade Cys25, His162 und Asn182 wurden zur besseren Orientierung mit dargestellt.

Die Sekundärstrukturelemente des Modells von hCatK wurden mit DSSP berechnet und sind in Abb. A-5 aufgelistet. Daran ist zu sehen, daß die über DSSP ermittelten Strukturelemente mit den Ergebnissen der PHD-Sekundärstrukturvorhersage in Korrelation stehen.

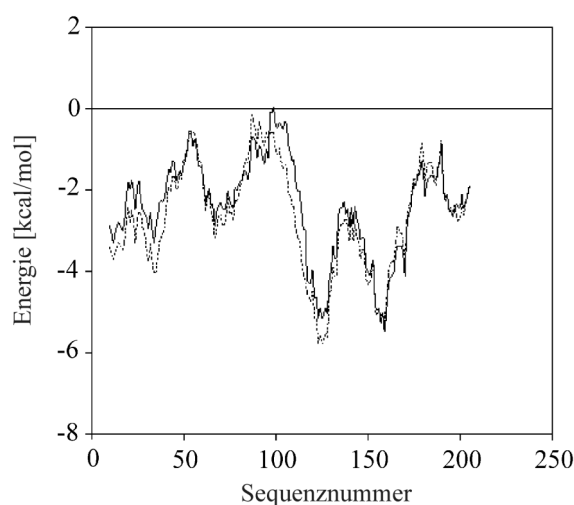
Die 3D-Struktur von hCatK enthält alle Faltungsmerkmale, die für Papain-ähnliche Enzyme typisch sind (siehe Abschnitt 1). Das Proteinmodell besteht aus zwei Domänen, die durch eine V-förmige Bindungsspalte getrennt werden. In der Mitte dieser Spalte sind die Aminosäuren Cys25, His162 und Asn182 lokalisiert, welche die katalytische Triade des Enzyms bilden. Die Seitenkette von Cys25 liegt als Thiolatanion und der Imidazolring von His162 in protonierter Form vor. Die Anordnung dieser Aminosäuren korreliert mit der, wie sie in den homologen Proteinen gefunden wird. Der Abstand zwischen Cys25S $\gamma$  und His162N $\delta$ 1 ist 3.75 Å, so daß sich das Thiolat-Imidazolium-Ionenpaar ausbilden kann. Diese Werte liegen in dem Bereich der homologen Cysteinproteasen mit bekannter Röntgenkristallstruktur (Tab. A-4). Eine Wasserstoffbrücke kann sich zwischen Asn182O $\delta$ 1 und His162N $\epsilon$ 2H ausbilden (2.26 Å).

Die modellierte Proteinstruktur von Cathepsin K wurde anschließend in ihrer Qualität und Geometrie unter Verwendung der Programme PROSA und PROCHECK überprüft. Das Modell von Cathepsin K erfüllt alle Kriterien, die in PROCHECK implementiert sind, in der gleichen Weise, wie Röntgenkristallstrukturen mit einer Auflösung von 2.0 Å. Der Ramachandran-Plot der Proteinstruktur von Cathepsin K spiegelt eine nahezu optimale Stereochemie innerhalb der Struktur des Modells wider (Abb. A-10). Die ( $\Phi$ ,  $\Psi$ )-Torsionswinkel von 81% der Aminosäurereste nehmen Werte ein, die für optimale Proteinstrukturen am stärksten begünstigt sind, die restlichen Torsionswinkel liegen in zusätzlich erlaubten Bereichen.

Da in der Zwischenzeit die Röntgenkristallstruktur von Cathepsin K bestimmt wurde und die Koordinaten zugänglich sind (PDB-Eintrag: 1MEM [132]), konnten beide 3D-Strukturen miteinander

verglichen werden. Somit war es möglich, qualitative Aussagen über das Modell der Tertiärstruktur von hCatK sowie über die verwendete Methode zu treffen.

Das Diagramm in Abb. 8 zeigt die Energiegraphen, die unter Verwendung des Programms PROSA von der Röntgenkristallstruktur sowie dem Proteinmodell von Cathepsin K berechnet wurden. Es ist zu sehen, daß die Energieprofile von den untersuchten Sequenzstrukturpaaren beider Strukturen negative Werte einnehmen. Dies korrespondiert mit energetisch stabilen Proteinfragmenten dieser Moleküle. Weiterhin kann man eine hohe Korrelation im Verlauf dieser beiden Graphen beobachten. Nur geringfügige Unterschiede treten im N-terminalen Bereich Ala1-Gly20 sowie in den Bereichen Cys25-Lys40, Asn70-Ser95, und Lys102-Gly109 auf (Abb. 8).



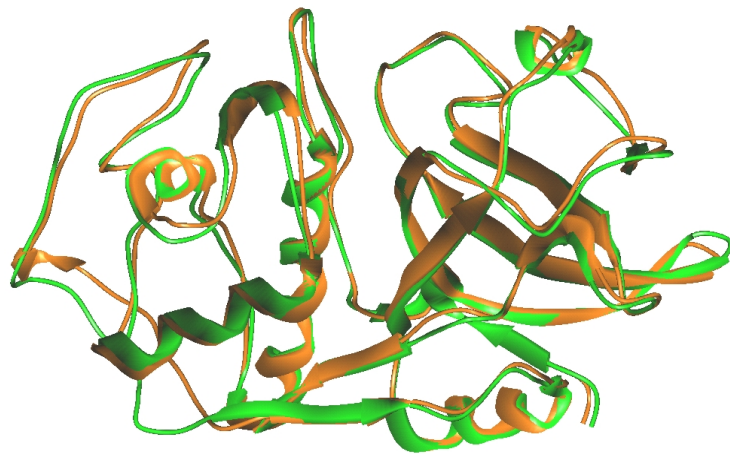
**Abb. 8** Energiegraphen des Proteinmodells von Cathepsin K (durchgezogene Linie) und der Röntgenkristallstruktur von diesem Enzym (dünn gepunktete Linie). Es wurden die durchschnittlichen Energiewerte der Proteinsequenz über einen Bereich von 20 Aminosäuren abgebildet.

Diese Differenzen lassen sich über folgende Eigenschaften innerhalb der Proteinstruktur von hCatK erklären. Der N-terminale Bereich Ala1-Gly20 beinhaltet große Loopregionen, die an der Oberfläche des Enzyms liegen. In diesem Bereich sind keine definierten Sekundärstrukturelemente enthalten. Die dadurch bedingte Flexibilität dieser Region ist wahrscheinlich die Ursache für die energetische Abweichung, die für die Struktur des Proteinmodells von hCatK beobachtet wurde. Trotzdem zeigt sich eine hohe Ähnlichkeit in der Lage der Hauptkettenatome (RMS-Wert = 0.70 Å, berechnet über alle  $\alpha$ -Atome). Der Abschnitt zwischen Cys25 und Lys40 enthält eine lange  $\alpha$ -Helix. Aus diesem Grund ist die RMS-Abweichung zwischen beiden Strukturfragmenten wesentlich geringer (0.21 Å). Die Unterschiede der Energieverläufe können hier durch eine geringfügig veränderte Anordnung des Proteinrückgrates hervorgerufen werden. Die Region von Asn70-Ser95 beinhaltet neben einer  $\alpha$ -Helix eine große Loopregion. Die Positionen der Hauptkettenatome der  $\alpha$ -Helix ist mit denen der entsprechenden Atome der Röntgenkristallstruktur vergleichbar (RMS-Wert = 0.29 Å). Durch die unterschiedliche Orientierung der Loopregion in diesem Abschnitt können diese Divergenzen in der Architektur des Proteins erklärt werden. Die gleiche Aussage kann für das letzte Fragment

Lys102-Gly109 getroffen werden, da sich hier ebenfalls ein Loop befindet, der an der Enzymoberfläche liegt.

Die sehr gute Übereinstimmung von beiden 3D-Strukturen wird zusätzlich anhand der geringen Differenz zwischen dem Z-Score des Modells (-9.34) und dem der Röntgenkristallstruktur (-9.59) deutlich.

Der RMS-Wert der SCRs des Strukturmodells zu den korrespondierenden Regionen in der experimentell bestimmten Struktur beträgt 0.54 Å (berechnet über die Hauptkettenatome). Wird die Berechnung für die Seitenkettenatome der Aminosäuren durchgeführt, beträgt der RMS-Wert 1.44 Å. Die entsprechenden Werte für die SVRs sind 1.12 Å (Hauptkettenatome) und 2.83 Å (Seitenkettenatome). Wird bei diesem Vergleich das ganze Protein mit einbezogen, ist die RMS-Abweichung zur Röntgenkristallstruktur von Cathepsin K 1.14 Å (Berechnung über alle Atome, außer den Wasserstoffatomen). Diese Werte verdeutlichen die hohe Korrelation zwischen beiden Molekülstrukturen und sie unterstreichen die grundlegende Richtigkeit des mit COMPOSER bestimmten Proteinmodells. Anhand dieser Befunde konnte gezeigt werden, daß die theoretisch ermittelte Struktur von hCatK und die Röntgenkristallstruktur dieses Cathepsins weitestgehend übereinstimmen (siehe auch Abb. A-5). Zur besseren Veranschaulichung sind in Abb. 9 beide Tertiärstrukturen dargestellt.



**Abb. 9** Darstellung der Tertiärstrukturen vom Strukturmodell von Cathepsin K (grün) im Vergleich zu seiner Röntgenkristallstruktur (orange).

#### 4.2.1.1 Die Bindungsspalte des Strukturmodells von Cathepsin K

So wie das Enzym in der Abb. 9 orientiert ist, befindet sich die Bindungsspalte des Cathepsins in der Mitte des Enzyms. In Cathepsin K werden die Begrenzungen dieser Spalte durch das in die zentrale  $\alpha$ -Helix übergehende Segment Gly20-Trp26 sowie durch das darüber liegende langgestreckte Proteinsegment von Gly64-Tyr67 gebildet. Diese beiden Abschnitte sind in der L-Domäne des Enzyms zu finden. In der R-Domäne wird die Bindungsspalte durch die Proteinsegmente

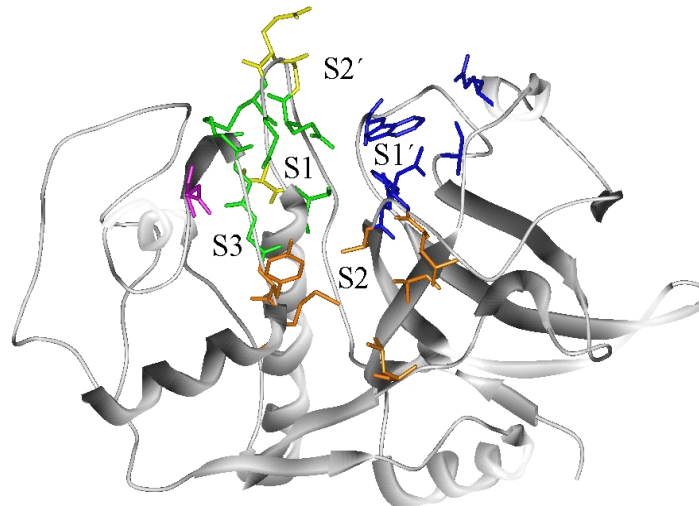
Val132-Ser134, Phe142-Phe144, Leu160-Ala163 sowie Trp184-Asn187 begrenzt. Die Spalte wird hauptsächlich durch folgende Aminosäureseitenketten gebildet: Gln19, Cys25, Met68, Ala134, Phe142, His162, Ala163, Trp184 und Leu209.

#### *Beschreibung der Bindungsstellen im Modell von Cathepsin K*

Bisherige Studien haben ergeben, daß sich die Substratbindungsregionen der Papain-ähnlichen Cysteinproteasen aus bis zu sieben Bindungsstellen S1-S4 und S1'-S3' zusammensetzen, von denen S1-S3 sowie S1'-S2' näher charakterisiert werden konnten [18,47] (siehe auch Abschnitt 3.2.1).

Die S1-Bindungsstelle befindet sich vollständig in der L-Domäne (Abb. 10). Diese Bindungsposition kann man als "Graben" in der Begrenzung der Bindungsspalte in der L-Domäne ansehen. Sie wird aus Aminosäuren gebildet, die in catheptischen Cysteinproteasen konserviert vorliegen (Gln19, Cys22, Gly23, Cys63, Gly64, Gly65, Gly66) (Abb. 5). Diese Aminosäuren gehören zu der zentralen  $\alpha$ -Helix sowie zu den umgebenden Loopregionen. Die S2-Bindungstasche ist wahrscheinlich die am besten definierte Bindungsstelle der Cathepsine und auch diejenige, die als einzige die Form einer "Tasche" besitzt. Diese Bindungsstelle wird meist aus hydrophoben Aminosäuren gebildet, die sich in der L- und R-Domäne des Enzyms befinden (Tyr67, Met68, Ala134, Leu160, Ala163, Leu209) (Abb. 10). Die Seitenketten dieser Aminosäuren liegen vorwiegend innerhalb der Bindungstasche, so daß diese überwiegend hydrophoben Charakter besitzt. Die Aminosäure Leu209 von hCatK bildet eine Begrenzung der Bindungsspalte. Im Gegensatz dazu sind experimentelle Untersuchungen zur Bindungsposition S3 wesentlich seltener [18]. Sie wird aus den Aminosäuren Tyr67 und Asp61 gebildet (L-Domäne), wobei nur die Seitenketten einen Einfluß auf diese Bindungsstelle im Enzym besitzen. Die S1'-Bindungsstelle wird aus Aminosäureresten gebildet (Ala137, Gln143, Asn161, His162, Trp184), die sich in der R-Domäne des Enzyms befinden, wobei die Seitenketten so ausgerichtet sind, daß sie in die Substratbindungsposition weisen. Die Aminosäuren der Bindungsposition S2' (Gln19, Gly20, Gln21, Cys22, Gly23, Ser24) liegen überwiegend in der L-Domäne. Allerdings existieren bis jetzt nur sehr wenige Studien über diese Bindungsstelle, so daß sie noch nicht exakt definierbar ist [18].





**Abb. 10** Darstellung der Bindungsstellen S1-S3 und S1'-S2' im Proteinmodell von hCatK. Die Aminosäuren, welche die jeweiligen Bindungspositionen bilden, sind verschieden farbig (S1-grün, S2-orange, S3-magenta, S1'-blau, S2'-gelb) dargestellt. Die einzelnen Bindungsstellen sind gekennzeichnet.

#### 4.2.2 Ein Tertiärstrukturmodell von humanem Cathepsin S

Zur Charakterisierung der Umgebung des katalytisch aktiven Zentrums sowie zur Untersuchung der Spezifität der Papain-ähnlichen Cysteinproteasen ist eine 3D-Struktur der jeweiligen Cathepsine essentiell. Zwar wurde in der Literatur eine Röntgenkristallstruktur des humanen Cathepsin S (hCatS) beschrieben, jedoch sind die Koordinaten nicht zugänglich [75,77]. Aus diesem Grund wurde ein 3D-Modell der Tertiärstruktur von hCatS entwickelt, wobei analog zu hCatK vorgegangen wurde. Das Modell für hCatS basiert auf der Sequenzidentität der Proteine, die in der Tab. 4 angegeben sind. Die Identitäten zwischen der Zielsequenz und den Vorlagesequenzen liegen über 40% (Tab. 4).

**Tab. 4** Proteinsequenzidentitäten von Cathepsin S zu seinen homologen Proteinen. Berechnet wurden diese Werte über COMPOSER [103].

Name des Proteins	PDB-Eintrag	Identität [%]
Procathepsin L	1CJL	56
Actinidin	1AEC	47
Glycylendopeptidase	1GEC	45
Actinidin	2ACT	45
Protease $\Omega$	1PPO	45
Papain	1PPN	44
Papain	9PAP	44
2-Hydroxyethylthiopapain	1PPD	44

In Abb. 11 ist das multiple Sequenzalignment zwischen den Primärstrukturen des zu bestimmenden Proteins und den Vorlagesequenzen dargestellt. Neben den SCRs, die in Tab. 5 detailliert aufgelistet

sind und in dieser Abbildung hervorgehoben wurden, sind in der letzten Zeile die berechneten Sekundärstrukturelemente für die Zielsequenz von hCatS aufgeführt.

	10	20	30	40	50	60
hCatS	LPDSVDWREK	GCVTEVKYQG	SCGACWAFSA	VGALEAQLKL	KTGKLVTLSA	QNLVDCSTEK
1CJL	APRSVDWREK	GYVTPVKNQG	QCGSSWAFSA	TGALEGQMFR	KTGRLISLSE	QNLVDCSGPE
1AEC	LPSYVDWRSA	GAVVDIKSQG	ECGGCWAFSA	IATVEGINKI	VTGVLISLSE	QELIDCGRTQ
1GEC	LPESVDWRAK	GAVTPVKHQG	YCESCWAFST	VATVEGINKI	KTGNLVELSE	QELVDCDLQS
2ACT	LPSYVDWRSA	GAVVDIKSQG	ECGGCWAFSA	IATVEGINKI	TSGLISLSE	QELIDCGRTQ
1PPO	LPENVDWRKK	GAVTPVRHQG	SCGSCWAFSA	VATVEGINKI	RTGKLVSE	QELVDCERRS
1PPN	IPEYVDWRQK	GAVTPVKNQG	SCGSCWAFSA	VVTIEGIIKI	RTGNLNEYSE	QELLDCCRRS
9PAP	IPEYVDWRQK	GAVTPVKNQG	SCGSCWAFSA	VVTIEGIIKI	RTGNLNQYSE	QELLDCCRRS
1PPD	IPEYVDWRQK	GAVTPVKNQG	SCGSCWAFSA	VVTIEGIIKI	RTGNLNQYSE	QELLDCCRRS
	-----	-----	---HHHHHHH	HHHHHHEEEE	----EEEE--	HHHH-----
	70	80	90	100	110	120
hCatS	YGNKGCNGGF	MTTAFQYIID	NKGIDSDASY	PYKAMDQKCQ	YDSKYR-AAT	CSKYTELPG
1CJL	-GNEGCNGGL	MDYAFQYVQD	NGGLDSEESY	PYEATEESCK	YNPKYS-VAN	DAGFVDIPK-
1AEC	-NTRGCNGGY	ITDGFQFIIN	NGGINTEENY	PYTAQDGECN	VDLQNEKYVT	IDTYENVPYN
1GEC	---YGCNRY	QSTSLQYVAQ	N-GIHLRAKY	PYIAKQQTCT	ANQVGGPKVK	TNGVGRVQSN
2ACT	-NTRGCDGGY	ITDGFQFIIN	DGINTEENY	PYTAQDGDCC	VALQDQKYVT	IDTYENVPYN
1PPO	---HGCKGGY	PPYALEYVAK	N-GIHLRSKY	PYKAKQGTCT	AKQVGGPIVK	TSGVGRVQPN
1PPN	---YGCNGGY	PWSALQLVAQ	Y-GIHYRNTY	PYEGVQRYCR	SREKGPYAAK	TDGVRQVQPY
9PAP	---YGCNGGY	PWSALQLVAQ	Y-GIHYRNTY	PYEGVQRYCR	SREKGPYAAK	TDGVRQVQPY
1PPD	---YGCNGGY	PWSALQLVAQ	Y-GIHYRNTY	PYEGVQRYCR	SREKGPYAAK	TDGVRQVQPY
	-----HH	HHHHHHHHHH	H-----	-----	-----EEE	EE-EEE----
	130	140	150	160	170	180
hCatS	REDVLKEAVA	NKGPVSVGVD	ARHPSFFLYR	SGVYYEPSCT	-QNVNHGVLV	VGYGDLNGKE
1CJL	QEKALMKAVA	TVGPISVAID	AGHESFLFYK	EGIYFEPDCS	SEDMDHGVLV	VGYGFESN-K
1AEC	NEWALQTAVT	-YQPVSVALD	AAGDAFKQYS	SGIFTGPCGT	--AIDHAVTI	VGYGTEGGID
1GEC	NEGSLNNAIA	-HQPVSVVVE	SAGRDFQNYK	GGIFEGSCGT	-K-VDHAVTA	VGYGKSGGK
2ACT	NEWALQTAVT	-YQPVSVALD	AAGDAFKQYA	SGIFTGPCGT	--AVDHAIIV	VGYGTEGGVD
1PPO	NEGNLLNAIA	-KQPVSVVLE	SKGRPFQLYK	GGIFEGPCGT	-K-VDHAVTA	VGYGKSGGK
1PPN	NEGALLYSIA	-NQPVSVVLE	AAGKDFQLYR	GGIFVGPCCN	-K-VDHAVAA	VGYG---P-N
9PAP	NQGALLYSIA	-NQPVSVVLE	AAGKDFQLYR	GGIFVGPCCN	-K-VDHAVAA	VGYG---P-N
1PPD	NQGALLYSIA	-NQPVSVVLE	AAGKDFQLYR	GGIFVGPCCN	-K-VDHAVAA	VGYG---P-N
	-HHHHHHHHH	H---EEEEEE	---HHHEEEE	--EEE-----	-----EEEE	EEE-----
	190	200	210	220	230	240
hCatS	YWLKNSWGH	NFGEEGYIRM	ARNKG---NH	CGIASFPSYP	EI-	
1CJL	YWLKNSWGE	EWGMGGYVKM	AKDRR---NH	CGIASAASYP	TV-	
1AEC	YWIVKNSWDT	TWGEEGYMRI	LRNVGGA-GT	CGIATMPSYP	VKY	
1GEC	YILIKNSWGP	GWGENGYIRI	RRASGNSPGV	CGVYRSSYYP	IKN	
2ACT	YWIVKNSWDT	TWGEEGYMRI	LRNVGGA-GT	CGIATMPSYP	VKY	
1PPO	YILIKNSWGT	AWGEKGYIRI	KRAPGNSPGV	CGLYKSSYYP	TKN	
1PPN	YILIKNSWGT	GWGENGYIRI	KRGTGNSYGV	CGLYTSSFYP	VKN	
9PAP	YILIKNSWGT	GWGENGYIRI	KRGTGNSYGV	CGLYTSSFYP	VKN	
1PPD	YILIKNSWGT	GWGENGYIRI	KRGTGNSYGV	CGLYTSSFYP	VKN	
	EEEE-----	-----EEEE	EE-----	-EE-----	--	

**Abb. 11** Multiples Sequenzalignment von hCatS mit seinen Vorlagesequenzen. Das Alignment wurde mit COMPOSER durchgeführt. Die SCRs sind grau unterlegt und Aminosäuren der katalytischen Triade sind mit einem Stern gekennzeichnet. Die Sequenz des Propeptides von Cathepsin L (PDB-Eintrag: 1CJL [118]) wurde nicht mit abgebildet. Die letzte Zeile enthält die Ergebnisse der Sekundärstrukturvorhersage, ermittelt mit PHD (H = Helix, E =  $\beta$ -Faltblattstrang).

Da bei der Generierung dieses Proteinmodells die Koordinaten von Chymopapain nicht verwendet wurden, erstreckt sich die Region des ersten SCR über einen wesentlich größeren Bereich und schließt das Cystein der katalytischen Triade mit ein (Abb. 11).

Das Strukturmodell von hCatS setzt sich aus neun SCRs, die 7-59 Aminosäuren enthalten sowie aus acht SVRs, bestehend aus 1-11 Aminosäureresten, zusammen (Tab. 5).

Die RMS-Abweichung des ermittelten Strukturmodells zu den SCRs der verwendeten homologen Proteine ist geringer als 0.40 Å. Der Vergleich der Zielsequenz von hCatS mit den SCRs, die aus den Vorlageenzymen generiert wurden, liefert Übereinstimmungswerte von  $\geq 50\%$  (Tab. 5). Einzige Ausnahme bildet die Region der Sequenz 112-118, wo nur ein Wert von 43% erreicht wird.

Zwischen den Cysteinresten 12-110, 22-66, 56-99 und 158-206 bilden sich vier Disulfidbrücken in hCatS aus.

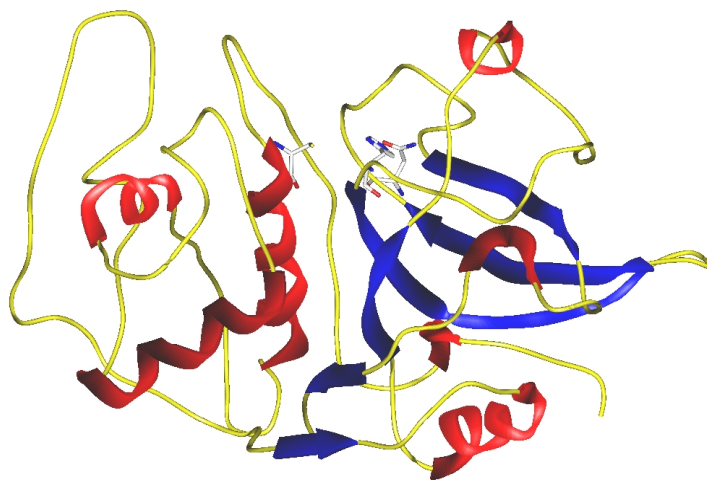
**Tab. 5** SCRs und SVRs der Proteinstruktur von hCatS, die aus den Cysteinproteasen resultieren, welche als Vorlagesequenzen verwendet wurden

SCR / SVR	Länge der Sequenz	Aminosäuresequenznummer von hCatS	PDB-Eintrag des Vorlageproteins	erste Aminosäure des Vorlageproteins	Identität [%]
<b>SCR</b>					
1	58	1-58	1PPO	1	63
2	16	64-79	1AEC	63	63
3	18	83-100	1CJL	82	50
4	7	112-118	1AEC	112	43
5	10	120-129	1CJL	118	50
6	28	131-158	1CJL	129	54
7	11	162-172	1CJL	161	82
8	23	178-200	1CJL	181	61
9	14	204-217	1CJL	207	71
<b>SVR</b>					
					Homologie der Loopregion [%]
1	4	60-63	1CJL	204	37
2	3	80-82	1CJL	79	74
3	11	101-111	1CJL	100	72
4	1	119	1AEC	119	58
5	1	130	1CJL	128	67
6	3	159-161	1CJL	158	83
7	5	173-177	1PPO	168	71
8	3	201-203	1CJL	204	64

Nach der Ergänzung der Wasserstoffatome und manueller Überarbeitung des Modells in Bezug auf mögliche sterische Überlappungen, wurde das Proteinmodell mit dem TRIPOS-Kraftfeld minimiert (siehe Abschnitt 3.1.4). Unter Verwendung von DSSP wurden die Sekundärstrukturelemente des Strukturmodells von Cathepsin S berechnet (Abb. A-6).

Aus dem Vergleich der konservierten Aminosäuren von hCatL und hCatS (Abb. 5) errechnet sich ein RMS-Wert von 1.16 Å (nur die Wasserstoffatome wurden bei dieser Berechnung ausgelassen). Diese geringe Abweichung unterstreicht Genauigkeit des ermittelten Modells von hCatS.

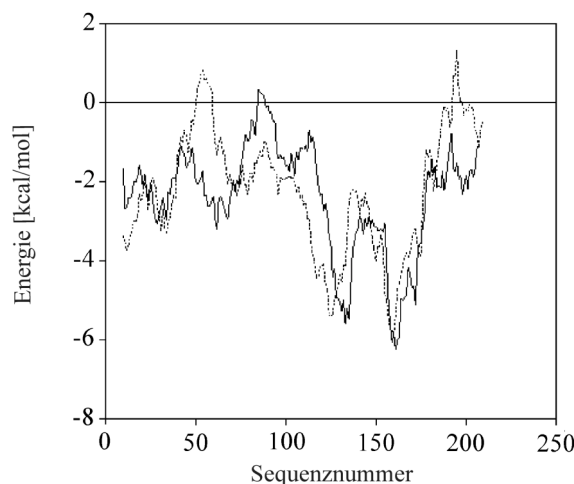
Analog zum Tertiärstrukturmodell von hCatK zeigt das hier diskutierte 3D-Strukturmodell von hCatS die typischen Merkmale im Faltungsmuster catheptischer Cysteinproteasen (Abb. 12). Die Aminosäuren der katalytischen Triade Cystein, Histidin und Asparagin, die sich in der Mitte der Bindungspalte des Cathepsins befinden, sind so ausgerichtet, daß sich Wasserstoffbrücken zwischen Cys25S $\gamma$  und His164N $\delta$ 1H sowie zwischen Asn184O $\delta$ 1 und His164N $\epsilon$ 2H ausbilden können (Tab. A-4).



**Abb. 12** Das Tertiärstrukturmodell von hCatS. Die Aminosäuren der katalytischen Triade sind eingezeichnet. Die in diesem Modell enthaltenen Sekundärstrukturelemente wurden mit DSSP berechnet und sind farbig gekennzeichnet (Helices sind rot und  $\beta$ -Faltblattstrukturen sind blau gefärbt).

Das Proteinmodell von hCatS wurde mit PROCHECK und PROSA analysiert. Von den Aminosäureresten nehmen 64% der ( $\Phi$ ,  $\Psi$ )-Torsionswinkel Werte ein, die für eine optimale Proteinstruktur am meisten begünstigt werden (Abb. A-11). Die restlichen ( $\Phi$ ,  $\Psi$ )-Torsionswinkel liegen in zusätzlich erlaubten Bereichen.

Abb. 13 zeigt den Energiegraphen des hCatS-Modells sowie den berechneten Energieverlauf der Röntgenkristallstruktur von hCatL (PDB-Eintrag: 1CJL [118]), die aus den Berechnungen mit PROSA resultieren. Vergleicht man beide Energieprofile miteinander, stellt man fest, daß Unterschiede für die N-terminale Region und für die Bereiche Ala50-Phe70, Asp80-Asn130 sowie Ser185-Ile217 beobachtet werden können (hCatS-Numerierung).



**Abb. 13** Energiegraphen des Proteinmodells von hCatS (durchgezogene Linie) sowie der Röntgenkristallstruktur von hCatL (dünn gepunktete Linie). Abgebildet sind die durchschnittlichen Energiewerte der Proteinsequenz über einen Bereich von 20 Aminosäuren.

Für die N-terminale Region von hCatS und hCatL (Leu1-Gly20, hCatS-Numerierung) errechnet sich eine RMS-Abweichung von 1.14 Å (berechnet über die C $\alpha$ -Atome beider Enzyme). Trotz dieses geringen Wertes treten in dieser Region stärkere Unterschiede im Energieverlauf beider Cathepsine auf (Abb. 13). Dieser Abschnitt enthält keine definierten Sekundärstrukturelemente. Daher ist eine genaue Bestimmung der Lage dieses Bereiches sehr schwierig. Das Proteinfragment zwischen den Aminosäureresten Ala50 und Phe70 von hCatS weist im Energieplot im Vergleich zu hCatL Unterschiede auf. Das spiegelt sich auch in der RMS-Abweichung wider (2.54 Å). Auf eine Modifizierung dieses Fragmentes wurde verzichtet, da die Architektur von hCatS energetisch günstiger ist als die von hCatL. Der Abschnitt, der sich aus den Aminosäuren Asp80-Asn130 zusammensetzt, enthält neben einer großen Loopregion auch definierte Sekundärstrukturelemente. Die Loopregion befindet sich an der Oberfläche des Enzyms und kann so in ihrer Position relativ variabel sein. Ausdruck für diese mögliche Flexibilität der Loopregion ist die berechnete RMS-Differenz von 2.89 Å. Für die anschließenden Fragmente, die Sekundärstrukturen enthalten, verringert sich dieser Wert (1.89 Å). Für den Bereich von Ser185 und Ile217 konnte festgestellt werden, daß die räumliche Anordnung der Aminosäuren in hCatS energetisch stabiler ist als in Cathepsin L. Neben Loopregionen enthält dieses Proteinfragment  $\beta$ -Faltblattstrukturen sowie den C-terminalen Bereich des Cathepsins. Die RMS-Abweichung, die für diesen Bereich berechnet wurde, beträgt 1.54 Å. Der Z-Score für das Modell von hCatS beträgt -8.86 und liegt in dem Bereich für nativ gefaltete Proteine [115]. Dieser Wert ist in hoher Übereinstimmung mit dem Z-Score von Cathepsin L (Vorlagesequenz mit der höchsten Homologie zu Cathepsin S), der -8.85 beträgt. Die aus der Analyse des hCatS-Modells resultierenden Ergebnisse zeigen, daß diese Proteinstruktur gut für weiterführende Studien an den Cathepsinen geeignet ist.

### 4.2.2.1 Die Bindungsspalte des Strukturmodells von Cathepsin S

In Analogie zu hCatK wird die Bindungsspalte in hCatS von mehreren Proteinsegmenten der L- und R-Domäne begrenzt. In der L-Domäne handelt es sich dabei um die Abschnitte Gly20-Trp26 sowie Asn67-Phe70. In der R-Domäne sind es die Enzymabschnitte Ser135-Gly137, Phe145-Leu147, Val162-Gly165 und Trp186-Asn189. Die Eigenschaften der Bindungsspalte werden überwiegend durch die Seitenketten der Aminosäurereste Gln19, Cys25, Met71, Phe145, His146, Val162, Trp186 und Phe211 bestimmt.

#### *Beschreibung der Bindungsstellen im Modell von Cathepsin S*

Die Aminosäurereste, welche die Bindungspositionen S1-S3 und S1'-S2' des katalytisch aktiven Bindungszentrums von hCatS bilden, sind in Tab. 6 aufgelistet.

**Tab. 6** Bindungsstellen von Cathepsin S, die aus dem Tertiärstrukturmodell dieses Enzyms resultieren

Bindungsstellen	S1	S2	S3	S1'	S2'
Aminosäurereste	Gln19	Phe70	Lys64	Ala140	Gln19
	Cys22	Met71	Phe70	Phe146	Gly20
	Gly23	Gly137		Asn163	Ser21
	Cys25	Val162		His164	Cys22
	Cys66	Gly165		Trp186	Gly23
	Asn67	Phe211			Ala24
	Gly68				
	Gly69				

Die S1-Bindungsposition in hCatS setzt sich, analog zu hCatK, aus konservierten Aminosäuren zusammen. Die S2-Bindungsstelle dieses Cathepsins ist im Vergleich zu hCatK in ihrer Größe und Tiefe erweitert. Dies ist auf eine andere Ausrichtung der Seitenketten der Aminosäuren, welche diese Bindungstasche bilden, zurückzuführen. Diese Aminosäuren befinden sich sowohl in der R- als auch in der L-Domäne. Im Gegensatz zu hCatK (Asp61, Tyr67) wird die S3-Bindungsposition in hCatS aus den Aminosäureresten Lys64 und Phe70 gebildet. Aus diesem Grund besitzt diese Bindungsstelle hinsichtlich ihrer Eigenschaften signifikante Unterschiede auf. Derartige Differenzen weist auch die S1'-Bindungsstelle von hCatS (Tab. 6).

#### 4.2.2.2 Docking von Liganden in die katalytisch aktiven Bindungszentren der Strukturmodelle der Cathepsine K und S

Zur abschließenden Überprüfung der 3D-Strukturmodelle der Cathepsine K und S wurde die Substrat- und Inhibitorspezifität dieser Cathepsine genauer untersucht. Dazu wurde Boc-Ala-Gly-Pro-Arg-MCA (1) als Substrat verwendet, da Aibe *et al.* zeigen konnten, daß Cathepsin K dieses chromogene Substrat selektiv hydrolysieren kann [135]. Dagegen ist aufgrund des Prolins in P2-Position die Bindungsaffinität dieses Substrates zu hCatS sowie auch zu den Cathepsinen B und L drastisch reduziert [71]. Für weitere Untersuchungen zum Docking- und Wechselwirkungsverhalten zwischen den Cathepsinen und seinen Liganden wurden mehrere selektive Peptidylvinylsulfone (Mu-X-Hph-Ψ(CH=CH-SO<sub>2</sub>-Ph)) mit unterschiedlichen Substitutionen in der P2-Position (X = Leucin, Methioninsulfon, Phenylalanin, Alanin) als Inhibitoren der Cysteinproteasen genutzt [136]. Die Strukturformeln dieser Verbindungen sind im Anhang dargestellt (Abb. A-14a, b).

Das Substrat und die Inhibitoren wurden unter Verwendung von FLEXIDOCK in das katalytisch aktive Bindungszentrum der erstellten 3D-Modelle von hCatK und hCatS gedockt (siehe auch Abb. A-2). Dabei resultierten mehrere Komplexe, die ein unterschiedliches strukturelles und energetisches Verhalten aufweisen. Nach der Energieoptimierung der Enzym-Ligand-Komplexe wurden diese Komplexe hinsichtlich hydrophober Wechselwirkungen und der Ausbildung von Wasserstoffbrückenbindungen zwischen Enzym und Ligand untersucht. Basierend auf den Interaktionen des Substrates in den katalytisch aktiven Bindungszentren der Cathepsine K und S war es möglich, die unterschiedliche Substratspezifität dieser Enzyme zu erklären. Die berechneten nichtbindenden Wechselwirkungsenergien und Dissoziationskonstanten ( $K_d$ ) dieser Komplexe sind in Tab. 7 aufgelistet.

**Tab. 7** Bindungsaffinität der Liganden zu den Cathepsinen K und S

Ligand	Cathepsin K			Cathepsin S		
	$K_m$ bzw. $K_i$ [ $\mu$ M] <sup>a</sup>	$\Delta E$ [kcal/mol] <sup>b</sup>	$K_d$ [ $\mu$ M] <sup>c</sup>	$K_m$ bzw. $K_i$ [ $\mu$ M] <sup>a</sup>	$\Delta E$ [kcal/mol] <sup>b</sup>	$K_d$ [ $\mu$ M] <sup>c</sup>
(1)	13	-61.65	3.55	- <sup>d</sup>	-58.39	8.51
(2)	0.11±0.01	-53.95	0.48	0.013±0.006	-61.23	0.19
(3)	14±5	-46.37	11.75	0.031±0.011	-61.10	0.24
(4)	36±5	-43.85	19.27	0.018±0.009	-61.44	0.19
(5)	23±2	-44.88	12.14	7±2	-45.54	11.86

(1) Boc-Ala-Gly-Pro-Arg-MCA; (2) Mu-Leu-Hph-Ψ(CH=CH-SO<sub>2</sub>-Ph);  
 (3) Mu-MetO<sub>2</sub>-Hph-Ψ(CH=CH-SO<sub>2</sub>-Ph), (4) Mu-Phe-Hph-Ψ(CH=CH-SO<sub>2</sub>-Ph);  
 (5) Mu-Ala-Hph-Ψ(CH=CH-SO<sub>2</sub>-Ph)

<sup>a</sup> kinetische Daten, beschrieben in [135,136]

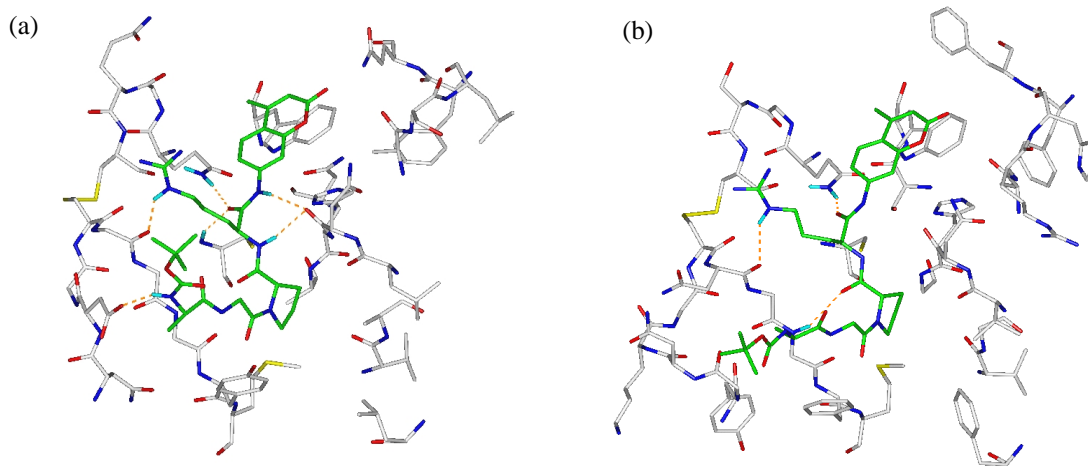
<sup>b</sup> nichtbindende Wechselwirkungsenergien

<sup>c</sup> Dissoziationskonstanten, berechnet mit SCORE

<sup>d</sup>  $K_m$  wurde nicht bestimmt [135]

Die berechneten Werte der Wechselwirkungsenergien stellen eine Abschätzung der freien Wechselwirkungsenthalpie dar. Entropieeffekte sowie Einflüsse der Solvation bzw. Desolvation, die bei der Komplexbildung auftreten können, wurden nicht berücksichtigt. Bei den sich bildenden Komplexen ist ein kleinerer (negativerer) Wert mit einer höheren Affinität der Liganden zu den Cathepsinen gleichzusetzen. Zur Vermeidung von eventuell auftretenden Fehlern bei der Bestimmung der Affinitätswerte wurden die  $K_d$ -Werte unter Verwendung von SCORE berechnet (siehe Abschnitt 3.1.6.1 und Tab. 7). Diese Werte korrelieren gut mit den experimentell bestimmten kinetischen Daten. Weiterhin ist zu sehen, daß zwischen den berechneten Wechselwirkungsenergien und den mit SCORE berechneten Werten ein enger Zusammenhang besteht. Anhand dieser Übereinstimmung kann man davon ausgehen, daß die Methoden qualitativ und quantitativ geeignet sind, die Bindungsaffinität eines Liganden im Enzym zu bestimmen.

Lage und Konformation des Substrates (1), gedockt in die Cathepsine K und S, zeigen nur geringe Unterschiede. In hCatS ist die Boc-Gruppe verglichen mit dem hCatK-Komplex, um  $90^\circ$  gedreht wurden (Abb. 14). In den Enzym-Substrat-Komplexen der beiden Cathepsine wurden hydrophobe Wechselwirkungen zwischen der fluorogenen MCA-Gruppe und der aromatischen Seitenkette von Trp184 der S1-Bindungsposition beobachtet (hCatK-Numerierung). In Cathepsin K bildet sich eine Wasserstoffbrücke zwischen dem Amidproton der MCA-Gruppe und dem Carbonyl-Sauerstoffatom der Hauptkette von Asn161 aus.



**Abb. 14** Darstellung der Enzym-Substrat-Komplexe von hCatK (a) und hCatS (b). Von den Cathepsinen sind jeweils nur die Aminosäuren zu sehen, die in der Bindungsspalte des Enzyms liegen. Die Kohlenstoffatome des Substrates sind grün eingefärbt.

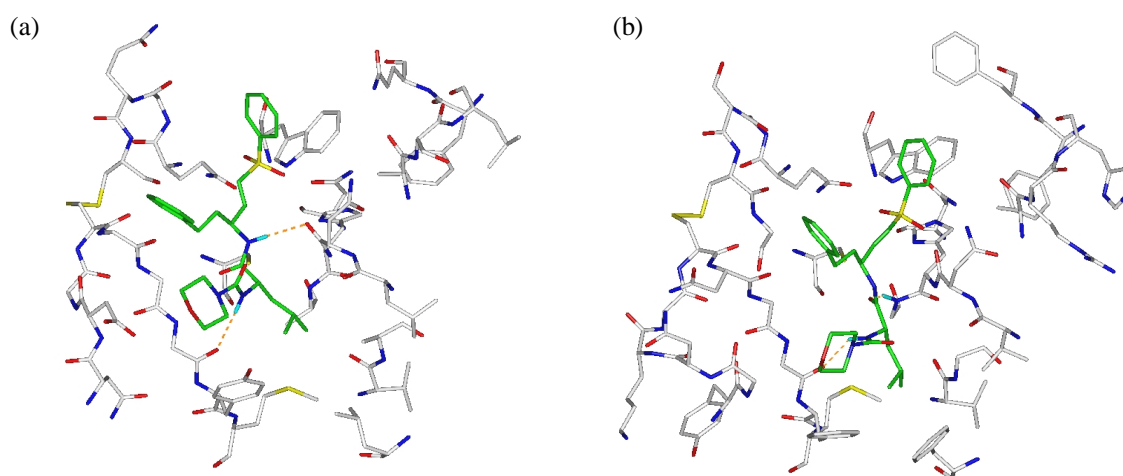
Durch die günstige Anordnung des Arginins des Substrates zu den Hauptkettenatomen von Gly23, Gly64 und Gly65 in hCatK können sich Wasserstoffbrückenbindungen ausbilden. Diese werden sowohl für ein Proton der Guanidinogruppe der Argininseitenkette zum Carbonyl-Sauerstoffatom von Gly64 als auch zwischen dem Carbonyl-Sauerstoffatom des Arginins und den Aminosäuren des Enzyms (Gln19, Cys25, Asn161) beobachtet (Abb. 14a). Im Fall des hCatS-Komplexes ist die



Seitenkette des Arginins vom Substrat zwar ähnlich ausgerichtet, allerdings ist die Tendenz der Ausbildung von Wasserstoffbrückenbindungen reduziert (Abb. 14b), da im Vergleich zu hCatK die Konformation des Proteinrückgrates von hCatS leicht verändert ist. Aufgrund der unterschiedlichen Eigenschaften und der Größe der S2-Bindungsposition der Cathepsine K und S zeigt das in P2-Position befindliche Prolin in seinem Dockingverhalten zu beiden Enzymen größere Differenzen. Es bildet sich eine große Anzahl von Wechselwirkungen zwischen dem Prolin und den Aminosäuren der S2-Bindungstasche von hCatK aus. Die Seitenketten der Aminosäuren Tyr67, Met68, Ala134 und Ala163 in hCatK sind so ausgerichtet, daß hier hydrophobe Wechselwirkungen auftreten können. Für den hCatS-Komplex konnte dieser Befund nicht beobachtet werden. Hier sind die Wechselwirkungen zwischen dem Prolin in P2-Position und den Aminosäuren der S2-Bindungstasche geschwächt und wurden nur zu Met71 beobachtet. Vergleichbare Wechselwirkungen zu beiden Cathepsinen treten für die Aminosäuren Alanin und Glycin in den Positionen P3 bzw. P4 des Substrates auf. Da die S3-Bindungsposition von hCatK Asp61 enthält, kann sich zwischen dem Amidproton des Alanins des Substrates und der Seitenkette von Asp61 eine Wasserstoffbrücke ausbilden.

Die niedrigere Affinität des Substrates zu Cathepsin S wird auch durch die reduzierten Energiewerte im Vergleich zum hCatK-Komplex deutlich sowie durch die Werte, die aus den Berechnungen mit SCORE resultieren (Tab. 7).

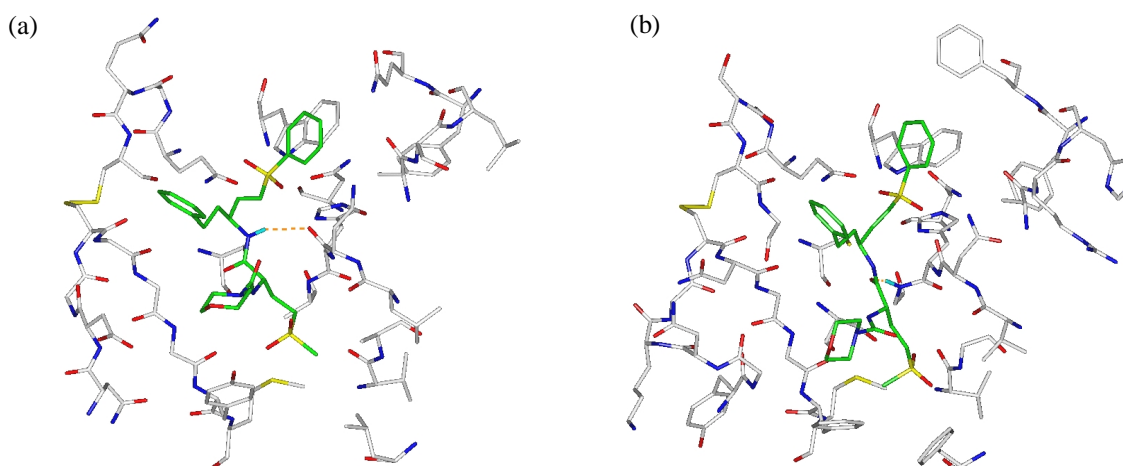
Im Gegensatz zu dem beschriebenen Substrat nehmen die Inhibitoren eine unterschiedliche Anordnung in der Bindungspalte der Cathepsine K und S ein. Die berechneten Werte für die Bindungsaffinität sind in der Tab. 7 aufgeführt. Im folgenden wird das Wechselwirkungsverhalten der Inhibitoren (Mu-X-Hph- $\Psi$ (CH=CH-SO<sub>2</sub>-Ph)) detaillierter beschrieben, die in P2-Position ein Leucin (2) bzw. ein Methioninsulfon (3) enthalten (X = Leucin, Methioninsulfon) (siehe Abb. 15 und Abb. 16).



**Abb. 15** Enzym-Inhibitor-Komplex von hCatK (a) sowie von hCatS (b). Die Kohlenstoffatome des Inhibitors Mu-Leu-Hph- $\Psi$ (CH=CH-SO<sub>2</sub>-Ph) (2) sind grün eingefärbt.

In beiden Cathepsinen treten hydrophobe Wechselwirkungen zwischen dem Phenylring des Inhibitors in P1'-Position und der Seitenkette von Trp184 (hCatK-Numerierung) sowie zum Histidin der katalytischen Triade auf. Für die S1-Bindungsstelle wurden ähnliche nichtkovalente Wechselwirkungen für hCatK und hCatS beobachtet. Die Seitenkette von Homophenylalanin (Hph) in der Position P1 richtet sich zu den Hauptkettenatomen von Gly23, Gly64 und Gly65 (hCatK-Numerierung) aus. Im hCatK-Komplex kann sich zwischen dem Amidproton des Homophenylalanins und Asn161 eine Wasserstoffbrücke ausbilden. Für die Interaktionen des Leucins in Position P2 mit den Aminosäuren der S2-Bindungstaschen treten starke Unterschiede zwischen beiden Cathepsinen auf. Verglichen zu hCatK ist die S2-Bindungstasche von hCatS in ihrer Größe weiter und tiefer. Es konnten für den hCatK-Komplex Wechselwirkungen zwischen Leucin in P2-Position und Tyr67, Met68 und Ala133 von hCatK beobachtet werden. Der Abstand der Leucinseitenkette zu Leu209 von hCatK ist zu groß, als dass sich stabile hydrophobe Wechselwirkungen ausbilden können. Der kürzeste Abstand von  $\text{LeuC}_{\text{Inh}}^{\delta}$  zu  $\text{Leu209C}^{\delta}$  beträgt 4.94 Å. Der entsprechende Abstand in hCatS zwischen  $\text{LeuC}_{\text{Inh}}^{\delta}$  und dem Mittelpunkt des Phenylringes von Phe211 ist 3.82 Å. Weiterhin ist die Seitenkette von Leu160 in hCatK so ausgerichtet, daß sich keine Wechselwirkungen zwischen dem Liganden und diesem Aminosäurerest ausbilden können (Abb. 15a). In der S2-Bindungstasche von hCatS treten hydrophobe Wechselwirkungen zwischen Leucin in P2-Position und Phe70, Met71 und Gly137 auf (Abb. 15b). Außerdem können sich stabilisierende Wechselwirkungen zu den Seitenketten von Val162 und Phe211 ausbilden. Der Inhibitor liegt im Vergleich zu hCatK tiefer in der Bindungsspalte von hCatS. Zwischen dem Carbonyl-Sauerstoffatom des Leucins und dem Amidproton von Gly165 bildet sich eine Wasserstoffbrücke aus, die für den hCatK-Komplex nicht beobachtet wurde. Aufgrund der Lage des Inhibitors in der Bindungsspalte von hCatS nehmen der Sechsring der 4-Morpholincarboxylgruppe (Mu) und Phe70 von hCatS eine energetisch günstige Anordnung zueinander ein (Abb. 15b). Diese Anordnung wirkt sich stabilisierend auf die Bildung des Enzym-Inhibitor-Komplexes aus.

Ersetzt man die Aminosäure in P2-Position durch ein Methioninsulfon ( $\text{MetO}_2$ ), wird das unterschiedliche Wechselwirkungsverhalten des Inhibitors mit beiden Enzymen noch deutlicher hervorgehoben (Abb. 16). Die Oxidation der Seitenkette des Methionins bewirkt eine drastische Erhöhung des sterischen Anspruchs dieses Restes in P2-Position. Für den hCatK-Komplex kommt es dadurch zu einer Reduzierung der Wechselwirkungsenergie dieses Peptidylvinylsulfons und demzufolge auch zu einer Abnahme der Inhibitoraktivität dieses Liganden (Tab. 7). Aufgrund der Größe von  $\text{MetO}_2$  wird die Seitenkette von diesem Aminosäurerest geringfügig aus der S2-Bindungstasche von hCatK heraus gedrückt (Abb. 16a). Dadurch werden die stabilisierenden Interaktionen zwischen dem Methioninsulfon in P2-Position und den Aminosäuren der korrespondierenden Bindungsstelle abgeschwächt. Im Gegensatz dazu ist es für hCatS möglich, diesen Rest in die S2-Bindungstasche aufzunehmen und stabilisierende Wechselwirkungen mit Met71, Val162, Phe70 und Phe211 auszubilden (Abb. 16b).



**Abb. 16** Darstellung der katalytisch aktiven Bindungszentren von hCatK (a) und hCatS (b) mit dem Inhibitor Mu-MetO<sub>2</sub>-Hph-Ψ(CH=CH-SO<sub>2</sub>-Ph) (3). Die Kohlenstoffatome des Peptidylvinylsulfons sind grün gefärbt.

Die reduzierte Bindungsaffinität der Inhibitoren gegenüber hCatK im Vergleich zu hCatS kann anhand der unterschiedlichen stabilisierenden Wechselwirkungen der beiden Cathepsin-Inhibitor-Komplexe erklärt werden. Die berechneten Werte der Bindungsaffinität der Liganden zu den Enzymen weisen die gleiche Tendenz auf, wie die experimentell bestimmten Daten (Tab. 7). Darin zeigt sich eine weitere Bestätigung für die Genauigkeit der modellierten Tertiärstrukturen von hCatK und hCatS.

Werden die Werte der Bindungsaffinität untereinander verglichen, ist eine gute Korrelation zwischen den Werten der nichtbindenden Wechselwirkungsenergien und den berechneten Dissoziationskonstanten  $K_d$  festzustellen (Tab. 7). Aufgrund der Übereinstimmung der theoretisch bestimmten Werte untereinander, aber auch anhand der Übereinstimmung dieser Werte mit den kinetischen Daten ist davon auszugehen, daß sich die verwendeten Methoden zur Bestimmung der Bindungsaffinität eines Liganden zum Enzym als geeignet erweisen.

### 4.2.3 Ein Tertiärstrukturmodell des humanen Cathepsin H

Die Röntgenkristallstruktur des humanen Cathepsin H (hCatH) konnte bis jetzt noch nicht aufgeklärt werden. Um vergleichende Studien an den lysosomalen Cathepsinen durchführen zu können, wurde ein 3D-Strukturmodell für die Tertiärstruktur von hCatH entwickelt. Erst kürzlich wurde in der Literatur eine hochaufgelöste Röntgenkristallstruktur von Cathepsin H, isoliert aus Schweineleber, beschrieben [61]. Allerdings waren die Koordinaten zum Zeitpunkt der Modellierung von hCatH nicht verfügbar.

Die 3D-Struktur von hCatH wurde basierend auf seiner Primärstruktur und seiner Homologie zu ähnlichen Proteinen unter Verwendung des Programms COMPOSER modelliert. Zur Generierung der

Tertiärstruktur von hCatH wurden Enzyme mit größtmöglicher Sequenzidentität verwendet. Die ausgewählten Enzyme, die als Vorlagestrukturen verwendet wurden, sind in Tab. 8 aufgelistet.

**Tab. 8** Primärsequenzidentität von Cathepsin H zu seinen homologen Proteinen. Berechnet wurden diese Werte über COMPOSER [103].

Name des Proteins	PDB-Eintrag	Identität [%]
Procathepsin L	1CJL	44
Glycylendopeptidase	1GEC	41
Papain	1PPN	41
Papain	9PAP	41
2-Hydroxyethylthiopapain	1PPD	41
Actinidin	1AEC	40
Actinidin	2ACT	39

Das multiple Sequenzalignment zwischen der Ziel- und den Vorlagesequenzen wurde mit COMPOSER durchgeführt. Es ist einschließlich der resultierenden SCRs in Abb. 17 dargestellt. Zusätzlich enthält diese Abbildung auch die Sekundärstrukturelemente, die aus den Ergebnissen der Vorhersage dieser Elemente mit PHD resultieren.

	10	20	30	40	50	60
hCatH	YPPSVDWRKK	GNFVSPVKNQ	GACGSCWTF <sup>*</sup> S	TTGALES <sup>*</sup> AI	IATGKMLSLA	EQQLVDCAQD
1CJL	APRSVDWREK	G-YVTPVKNQ	GQCGSSWAFS	ATGALEGQMF	RKTGRLISLS	EQNLVDCSGP
1GEC	LPESVDWRAK	G-AVTPVKHQ	GYCESCWAFS	TVATVEGINK	IKTGNLVELS	EQELVDCDLQ
1PPN	IPEYVDWRQK	G-AVTPVKNQ	GSCGSCWAFS	AVVTIEGIIK	IRTGNLNEYS	EQELLDCCR
9PAP	IPEYVDWRQK	G-AVTPVKNQ	GSCGSCWAFS	AVVTIEGIIK	IRTGNLNQYS	EQELLDCCR
1PPD	IPEYVDWRQK	G-AVTPVKNQ	GSCGSCWAFS	AVVTIEGIIK	IRTGNLNQYS	EQELLDCCR
1AEC	LPSYVDW <sup>*</sup> RSA	G-AVVDIKSQ	GECGGCWAFS	AIATVEGINK	IVTGVLSLS	EQELIDCGRT
2ACT	LPSYVDW <sup>*</sup> RSA	G-AVVDIKSQ	GECGGCWAFS	AIATVEGINK	ITSGLISLS	EQELIDCGRT
	---HHH---	-----	---HHHHHH	HHHHHHHHH	HH---EEEE-	-HHHH----
	70	80	90	100	110	120
hCatH	FNNYGCQGG	PSQAF <sup>*</sup> EYILY	NKGIMGEDTY	PYQ <sup>*</sup> GKDG <sup>*</sup> YCK	FQPGKAIGFV	-KDVANITTY
1CJL	EGNEGCNGGL	MDYAFQYVQD	NGGLDSEESY	PYEATEESCK	YNPKYSVAND	-AGFVDIPK-
1GEC	--SYGCNRGY	QSTSLQYVAQ	N-GIHLRAKY	PYIAKQQT <sup>*</sup> CR	ANQVGGPKVK	TNGVGRVQSN
1PPN	--SYGCNGGY	PWSALQLVAQ	Y-GIHYRNTY	PYEGVQRYCR	SREKGPYAAK	TDGVRQVQPY
9PAP	--SYGCNGGY	PWSALQLVAQ	Y-GIHYRNTY	PYEGVQRYCR	SREKGPYAAK	TDGVRQVQPY
1PPD	--SYGCNGGY	PWSALQLVAQ	Y-GIHYRNTY	PYEGVQRYCR	SREKGPYAAK	TDGVRQVQPY
1AEC	QNTRGCNGGY	ITDGFQFIIN	NGGINTEENY	PYTAQDGECN	VDLQNEKYVT	IDTYENVPYN
2ACT	QNTRGCNGGY	ITDGFQFIIN	DGGINTEENY	PYTAQDGD <sup>*</sup> CD	VALQDQKYVT	IDTYENVPYN
	-----HH	HHHHHHHHH	H-----	-----	-----EEE	-E--EE----
	130	140	150	160	170	180
hCatH	DEEAMVEAVA	LYNPVSFAFE	VT-QDFM <sup>*</sup> MYR	TGIYSST <sup>*</sup> SCH	KTPDKVNHAV	LAVGYGKNG
1CJL	QEKALMKAVA	TVGPISVAID	AGHESFLFYK	EGIYFEPDCS	SE--DMDHGV	LVVGYGFESN
1GEC	NEGSLN <sup>*</sup> IAIA	-HQPVS <sup>*</sup> VVE	SAGRDFQNYK	GGIFEGSCGT	---KVDHAV	TAVGYGKSGG
1PPN	NEGALLYSIA	-NQPVSVVLE	AAGKDFQLYR	GGIFVGP <sup>*</sup> CGN	---KVDHAV	AAVGYGPN--
9PAP	NQGALLYSIA	-NQPVSVVLQ	AAGKDFQLYR	GGIFVGP <sup>*</sup> CGN	---KVDHAV	AAVGYGPN--
1PPD	NQGALLYSIA	-NQPVSVVLQ	AAGKDFQLYR	GGIFVGP <sup>*</sup> CGN	---KVDHAV	AAVGYGPN--
1AEC	NEWALQTAVT	-YQPVSVALD	AAGDAFKQYS	SGIFTG <sup>*</sup> PCGT	---AIDHAV	TIVGYGTEGG
2ACT	NEWALQTAVT	-YQPVSVALD	AAGDAFKQYA	SGIFTG <sup>*</sup> PCGT	---AVDHAI	VIVGYGTEGG
	HHHHHHHHH	H---EEEE-	---HHHHEEE	---EEEE---	-----EEE	EEEE-----
	190	200	210	220	230	240
hCatH	IPYWIVKNSW	GPQWGMNGYF	LIERGK----	NMCGLAACAS	YPIPLV	
1CJL	K-YWLVKNSW	GEEWGMGGYV	KMAKDRR---	NHCGIASAAS	YPTV--	
1GEC	KGYILIKNSW	GPGWGNGYI	RIRRASGNSP	GVC <sup>*</sup> GVYRSSY	YPIKN-	
1PPN	--YILIKNSW	GTGWGNGYI	RIKRGTGNSY	GVCGLYTSS <sup>*</sup> F	YPVKN-	
9PAP	--YILIKNSW	GTGWGNGYI	RIKRGTGNSY	GVCGLYTSS <sup>*</sup> F	YPVKN-	
1PPD	--YILIKNSW	GTGWGNGYI	RIKRGTGNSY	GVCGLYTSS <sup>*</sup> F	YPVKN-	
1AEC	IDYWIVKNSW	DTTWGEEGYM	RILRNVGGA-	GTCGIATMPS	YPVKY-	
2ACT	VDYWIVKNSW	DTTWGEEGYM	RILRNVGGA-	GTCGIATMPS	YPVKY-	
	-----EE	-----EE	EEE-----	-----EEE	-----	

**Abb. 17** Multiples Sequenzalignment von hCatH mit seinen Vorlagesequenzen (COMPOSER). Die SCR's sind grau unterlegt und Aminosäuren der katalytischen Triade sind mit einem Stern hervorgehoben. Die letzte Zeile enthält die Sekundärstrukturelemente, die aus der Vorhersage mit PHD resultieren (H = Helix, E =  $\beta$ -Faltblattstrang).

Das Modell von hCatH setzt sich aus 11 SCR's mit 9-47 Aminosäuren und aus neun SVR's zusammen, die 1-10 Aminosäuren enthalten (Tab. 9).

Die RMS-Differenz des Strukturmodells von hCatH zu den SCR's der verwendeten Proteine ist kleiner als 0.40 Å (berechnet über die C $\alpha$ -Atome). Die Übereinstimmung der SCR's mit der Aminosäuresequenz von Cathepsin H beträgt über 45% (Tab. 9). Mit einem Wert von 36% bildet die Region der Aminosäuren 111-119 die einzige Ausnahme. In Tab. 9 sind diese Werte für jeden einzelnen SCR aufgeführt. Aus der berechneten Proteinstruktur von hCatH resultieren drei Disulfidbrücken zwischen Cys23-Cys66, Cys57-Cys99 und Cys157-Cys207.

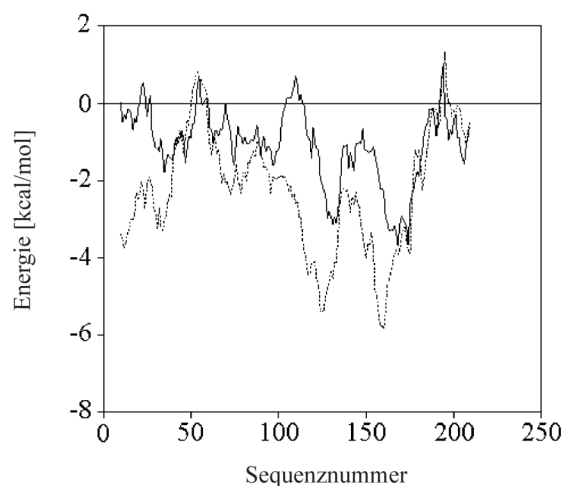
**Tab. 9** SCRs innerhalb der Proteinstruktur von hCatH, die von den homologen Cysteinproteasen bestimmt wurden und die resultierenden SVRs dieses Enzyms

SCR / SVR	Länge der Sequenz	Aminosäuresequenznummer von hCatH	PDB-Eintrag des Vorlageproteins	erste Aminosäure des Vorlageproteins	Identität [%]
SCR					
1	11	1-11	1CJL	1	50
2	47	13-59	1CJL	12	52
3	19	63-81	1CJL	62	53
4	18	83-100	1PPN	79	50
5	8	111-118	1GEC	108	36
6	10	120-129	1CJL	118	50
7	11	131-141	1AEC	130	46
8	13	142-154	1PPN	139	46
9	13	163-175	1GEC	156	77
10	22	181-202	1CJL	182	59
11	14	205-218	1CJL	198	57
SVR					
					Homologie der Loopregion [%]
1	1	12	-	-	-
2	3	60-62	1AEC	56	44
3	1	82	1AEC	78	30
4	10	101-110	1CJL	97	59
5	1	130	9PAP	113	100
6	8	155-162	1CJL	125	33
7	5	176-180	1FIE	124	59
8	2	203-204	1AEC	169	72
9	2	219-220	1HUC	233	82

Nach Ergänzung der Wasserstoffatome wurde das Strukturmodell mit dem TRIPOS-Kraftfeld minimiert und anschließend analysiert. Unter Verwendung des Programms DSSP wurden die Sekundärstrukturelemente der Proteinstruktur bestimmt (Abb. A-7). Die berechneten Sekundärstrukturelemente kommen in den SCRs vor. Von den Aminosäureresten, die SCRs ausbilden, nehmen 38% dieser Reste Konformationen von definierten Sekundärstrukturen ein.

Der berechnete RMS-Wert über die konservierten Aminosäuren von hCatL und hCatH (Abb. 5) beträgt 2.24 Å (nur die Wasserstoffatome wurden bei dieser Berechnung ausgelassen). Durch diesen geringen Wert wird die Qualität des ermittelten Strukturmodells von hCatH unterstrichen.

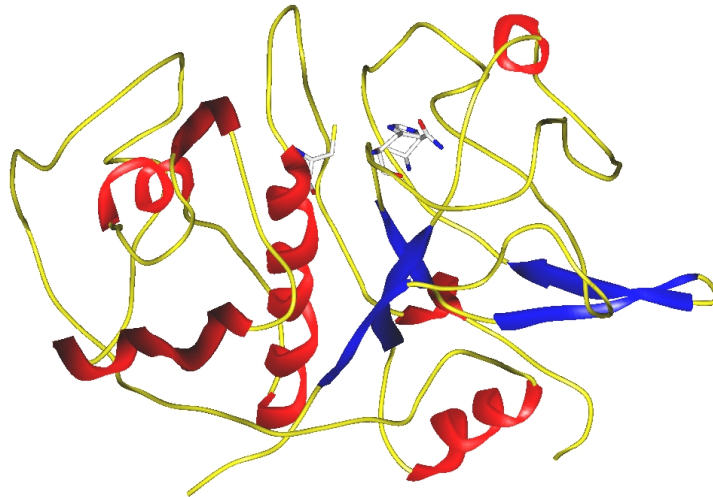
Die ( $\Phi$ ,  $\Psi$ )-Torsionswinkel von 69% der Aminosäuren nehmen Werte ein, die am stärksten begünstigt werden (Abb. A-12). Außerdem wurde das erhaltene Strukturmodell von hCatH mit dem Programm PROSA überprüft. Die Energieverläufe der modellierten Struktur und der Röntgenkristallstruktur von Cathepsin L, zu dem die höchste Sequenzidentität gefunden wurde (Tab. 8), sind in Abb. 18 dargestellt.



**Abb. 18** Energiegraphen des Strukturmodells von Cathepsin H (durchgezogene Linie) sowie der Röntgenkristallstruktur von Cathepsin L (dünn gepunktete Linie), berechnet mit PROSA. Es wurden die durchschnittlichen Energiewerte der Proteinsequenz über einen Bereich von 20 Aminosäuren abgebildet.

Die reduzierte Sequenzhomologie zwischen hCatH und Cathepsin L (44%) im Vergleich zu den entsprechenden Werten der Cathepsine K (66%) und S (56%) führt zu der Annahme, daß auch die Ähnlichkeit der Proteinstrukturen geringer ist. Diese Vermutung wird durch die erhöhte RMS-Abweichung, welche über die konservierten Aminosäuren beider Cathepsine ( $2.24 \text{ \AA}$ ) berechnet wurde, bestätigt. Aus diesem Grund treten auch größere Unterschiede für den Verlauf der Energiegraphen von hCatH und hCatL auf (Abb. 18). Sehr hohe Korrelationen zwischen beiden Proteinstrukturen werden nur für zwei Sequenzabschnitte (Ala40-Leu70 und Ala170-Ile220) beobachtet. Der Z-Score dieser Struktur beträgt  $-7.79$  und nimmt somit einen Wert für nativ gefaltete Proteinstrukturen ein. Für Cathepsin L wurde ein Z-Score von  $-8.85$  berechnet.

Aufgrund der Ergebnisse der PROSA- und PROCHECK-Analysen kann von einer grundlegenden Richtigkeit des Modells von hCatH ausgegangen werden. Abb. 19 zeigt die Tertiärstruktur des Proteinmodells von hCatH und verdeutlicht die Übereinstimmung des Faltungsmusters dieses Enzyms im Vergleich zu anderen catheptischen Cysteinproteasen.



**Abb. 19** Tertiärstruktur des Modells von hCatH. Die Sekundärstrukturelemente wurden mit Dssp berechnet. Die Aminosäurereste der katalytischen Triade Cys26, His166 und Asn186 sind dargestellt. Helices sind rot gefärbt und  $\beta$ -Faltblattstrukturen blau.

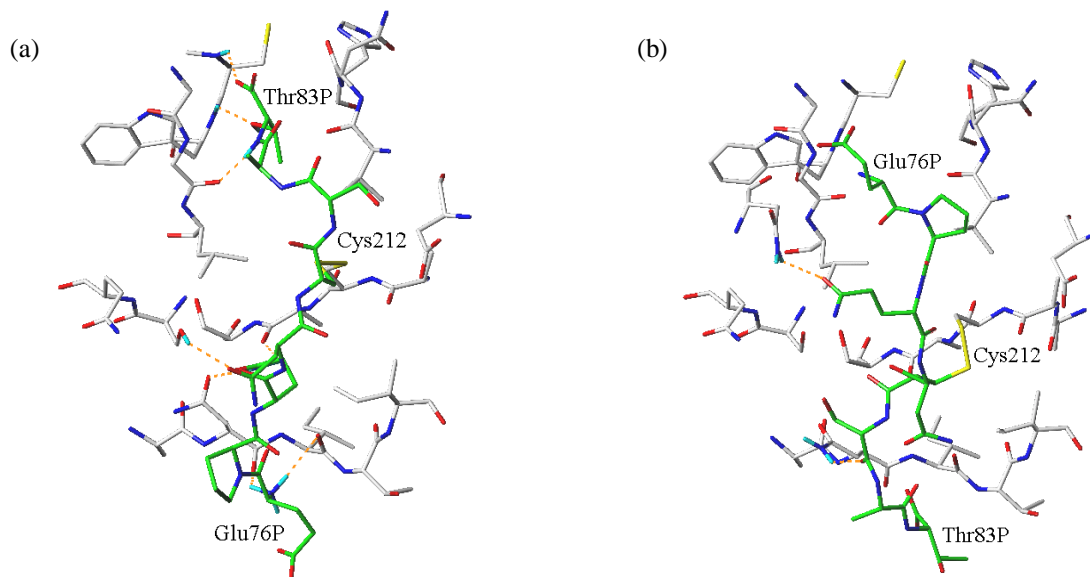
#### 4.2.3.1 Die Lage der Minikette in der Bindungsspalte von Cathepsin H

Über die Aminosäuresequenzen von Cathepsin H, isoliert aus Ratte [137], Maus [32] und Mensch [138] wurde nachgewiesen, daß eine Minikette (Glu-Pro-Gln-Asn-Cys-Ser-Ala-Thr), die vom Propeptid des Cathepsins abgespalten wird, zwischen den Propeptidresten Glu76P-Thr83P lokalisiert ist (Propeptid-Numerierung). Die Strukturen der Procathepsine B, K und L zeigen, daß die Propeptide der Cathepsine zum Teil in der Bindungsspalte dieser Enzyme liegen. Dabei ist die Orientierung dieser Propeptide entgegengesetzt zur Richtung eines gebundenen Substrates [56,139,118]. Für Cathepsin H wird jedoch in der Literatur beschrieben, daß die Minikette dieses Proteins bei der Bindung in das aktive Zentrum des Enzyms die Lage eines gebundenen Substrates einnimmt [61]. Erst kürzlich konnte gezeigt werden, daß diese Minikette über eine Disulfidbrücke an das aktive Enzym gebunden ist [60,61]. Dieses Peptid hat eine definierte Rolle bei der Substraterkennung. Die Aminopeptidaseaktivität des Enzyms wird durch die Lage der C-terminalen Carboxylgruppe der Minikette in der Bindungsspalte von Cathepsin H hervorgerufen. Studien an Substraten lassen die Schlußfolgerung zu, daß der negativ geladene C-Terminus dieses Peptides als Anker für die positiv geladene N-terminale Aminogruppe des Substrates agiert [61].

Das Oktapeptid Glu-Pro-Gln-Asn-Cys-Ser-Ala-Thr wurde unter Verwendung des Programms FLEXIDOCK in die Bindungsspalte des berechneten Modells von hCatH gedockt. Ausgehend von den Studien von *Guncar et al.* wurden Position und Orientierung dieses Peptides im katalytisch aktiven Bindungszentrum von hCatH untersucht, um die Verwendung einer inkorrekten Struktur dieses Enzyms für weitere Untersuchungen zu verhindern [61]. Der negativ geladene C-Terminus dieses Oktapeptides (Thr83P) bindet an der Stelle, wo in vergleichbaren Enzymen die S2-Bindungstasche



vorhanden ist (Orientierung (1), Abb. 20a). Er imitiert so einen Rest in Position P2. Zur Überprüfung dieser Befunde wurde die Minikette in die Bindungsspalte des Enzyms gedockt, in dem sie um 180° gedreht wurde. Dadurch belegt Glu76P die S2-Bindungsstelle (Orientierung (2), Abb. 20b).



**Abb. 20** Illustration des 3D-Modells von hCatH mit der Minikette: (a) die Minikette liegt im katalytisch aktiven Bindungszentrum in Richtung eines gebundenen Substrates (Orientierung (1)). (b) Die Minikette liegt invers in der Bindungsspalte von hCatH (Orientierung (2)).

In beiden Anordnungsmöglichkeiten wird das Peptid über eine Disulfidbrücke an das Cathepsin gebunden. Während der Dockingprozedur bildeten sich für beide Orientierungen mehrere Komplexe mit unterschiedlicher Struktur und Energieinhalt. Nach der Energieoptimierung dieser Komplexe mit dem TRIPOS-Kraftfeld konnte ein differenziertes Docking- und Wechselwirkungsverhalten der Minikette in der Bindungsspalte von hCatH beobachtet werden. Nach diesen Dockingstudien sind die Komplexe von hCatH, in denen Thr83P die Bindungstasche S2 belegt (Orientierung (1)) über 8 kcal/mol energetisch bevorzugt zu den Komplexen mit einer inversen Position der Minikette (Orientierung (2)). Die erhaltenen nichtbindenden Wechselwirkungsenergien der energetisch günstigsten Komplexe betragen -42.87 kcal/mol für Orientierung (1) und -34.49 kcal/mol für Orientierung (2). Basierend auf den Wechselwirkungen der Minikette mit hCatH kann versucht werden, die Divergenzen in der Bindungsaffinität des Oktapeptides mit dem Enzym zu erklären, wenn die Lage dieses Moleküls in der Bindungsspalte des Cathepsins verändert wird.

Die Minikette, in der Orientierung (1), bindet an hCatH über nichtkovalente Wechselwirkungen mit den Seitenketten Gln78P, Cys80P, Ser81P und Thr83P (Abb. 20a). Anziehende Wechselwirkungen können sich zwischen den Methylengruppen der Gln78P-Seitenkette zu Ile118 von hCatH sowie zwischen der Thr83P-Seitenkette und Val164 ausbilden (Cathepsin H-Numerierung). Zusätzlich tragen mehrere Wasserstoffbrückenbindungen, die sich zwischen den Resten der Minikette Glu76P, Gln78P, dem negativ geladenen C-Terminus Thr83P und den Aminosäureresten des hCatH-Enzyms

ausbilden, wesentlich zu einer stabilen Anordnung der Minikette in der Bindungsspalte des Cathepsins bei. Die Seitenketten der Aminosäuren Ala82P, Asn79P und Pro77P sind so positioniert, daß sie von der Bindungsspalte weg weisen. Für eine bessere Übersicht ist in Abb. A-3 die gesamte Proteinstruktur von hCatH einschließlich seiner Minikette in der Orientierung (1) dargestellt.

Das Wechselwirkungsverhalten zum katalytisch aktiven Bindungszentrum ist abgeschwächt, wenn die Minikette um 180° gedreht wird (Abb. 20b). Der Pyrrolidinring von Pro77P richtet sich so aus, daß sich zwischen dem Prolin der Minikette und Val164 von hCatH hydrophobe Wechselwirkungen ausbilden. Die Tendenz zur Bildung von Wasserstoffbrücken zwischen dem Cathepsin und dem Oktapeptid ist reduziert. Es wurde nur eine Wasserstoffbrücke zwischen dem Carbonyl-Sauerstoffatom von Ser81P und der Seitenkette von Asn115 gefunden.

Anhand des hier diskutierten Wechselwirkungsverhaltens der Minikette in hCatH ist zu sehen, daß die Affinität dieser Minikette zum Enzym in der Orientierung (1) im Vergleich zur anderen Anordnung deutlich erhöht ist. Diese Aussage wird durch die berechneten Wechselwirkungsenergien unterstrichen. Man kann daher davon ausgehen, daß das Oktapeptid Glu-Pro-Gln-Asn-Cys-Ser-Ala-Thr so in der Bindungsspalte von hCatH liegt, daß der negativ geladene C-Terminus Thr83P die S2-Bindungsposition des Enzyms belegt (Orientierung (1)).

#### **4.2.3.2 Untersuchung der Bindungsspalte des Strukturmodells von Cathepsin H**

Im Vergleich zu den anderen catheptischen Proteinen weist die Bindungsspalte von hCatH signifikante Unterschiede auf, da die Größe der Spalte durch die Minikette des Enzyms stark eingeschränkt wird. In der L-Domäne des Enzyms wird die Begrenzung der Spalte aus den Proteinabschnitten Gly21-Cys26 und Gln67-Leu70 gebildet. In der R-Domäne begrenzen die Proteinsegmente Phe144-Met146, Asn165-Ala167 und Trp188-Gln191 die Bindungsspalte. Die Eigenschaften innerhalb der Spalte werden im wesentlichen durch die Seitenketten der Aminosäuren Gln20, Cys26, Met145, His166 und Trp188 bestimmt.

##### *Beschreibung der Bindungsstellen im Modell von Cathepsin H*

Die Aminosäuren, aus denen die Bindungspositionen S1 sowie S1'-S2' gebildet werden, sind in Tab. 10 zusammengestellt.

**Tab. 10** Bindungspositionen (Sn) von Cathepsin H, die aus dem Tertiärstrukturmodell dieses Enzyms resultieren

Bindungsstellen Sn	S1	S1'	S2'
Aminosäurereste	Gln20	Val140	Gln20
	Cys23	Met145	Gly21
	Gly24	Asn165	Ala22
	Cys26	His166	Cys23
	Cys66	Trp188	Gly24
	Gln67		Ser25
	Gly68		
	Gly69		

hCatH bildet keine S2- bzw. S3-Substratbindungsstellen, da diese Regionen durch die Minikette belegt werden. Dagegen können die Bindungsstellen S1, S1' und S2' und die Positionierung von entsprechenden P1-, P1'- und P2'-Substratresten durch den Vergleich mit den Röntgenkristallstrukturen verwandter Enzyme beschrieben werden. Ein Substratrest in Position P1 kann ionische Wechselwirkungen mit dem negativ geladenen C-Terminus der Minikette eingehen. Das vorgeschlagene Strukturmodell von hCatH zeigt, daß sich die Seitenkette eines P1-Restes entlang der Begrenzung der Bindungsspalte ausrichten würde. Im Vergleich zu den korrespondierenden Bindungsstellen von Enzymen der Papain-Superfamilie treten Differenzen bei den Aminosäuren auf, welche die S1'-Bindungsstelle ausbilden. Ein Rest in P1'-Position kann zahlreiche stabilisierende Wechselwirkungen zu den Seitenketten der Aminosäuren, die diese Bindungsstelle bilden, eingehen. Die S2'-Bindungsseite ist relativ breit und zum Lösungsmittel ausgerichtet. Es zeigt sich jedoch, daß sie im Vergleich zu anderen catheptischen Cysteinproteasen keine besonderen Merkmale aufweist.

#### 4.2.3.3 Vergleichende Untersuchungen zu Cathepsin H aus Schweineleber

Während der Studien am humanen Cathepsin H (hCatH) waren die 3D-Koordinaten der Röntgenkristallstruktur von Cathepsin H aus Schweineleber (pCatH) nicht zugänglich. Diese Daten sind erst seit kurzer Zeit über die PDB verfügbar (PDB-Eintrag: 8PCH [61]) (Tab. A-1). Die Sequenzhomologie zwischen hCatH und pCatH beträgt 91%. Innerhalb der Sequenzen treten 19 unterschiedliche Aminosäuren auf. Ein Sequenzalignment der beiden Primärstrukturen ist im Anhang abgebildet (Abb. A-8). Die unterschiedlichen Aminosäuren sind überwiegend an der Enzymoberfläche zu finden und haben keinen wesentlichen Einfluß auf die Eigenschaften der Bindungsspalte.

Die Hauptkettenatome der katalytischen Triade von pCatH befinden sich in Positionen, wie sie für Papain-ähnlichen Enzymen definiert wurden. Es tritt jedoch ein signifikanter Unterschied zwischen der Röntgenkristallstruktur von pCatH und der modellierten 3D-Struktur von hCatH auf. Der Imidazolring des aktiven Histidins (pCatH) kann mit dem Thiolatom des aktiven Cysteins kein Thiolat-Imidazolium-Ionenpaar ausbilden. In der Röntgenkristallstruktur von pCatH bildet das

N $\epsilon$ 2H-Proton des Histidins der katalytischen Triade mit dem negativ geladenen C-Terminus eines benachbarten Moleküls im Kristall eine Salzbrücke aus [61]. Der Torsionswinkel  $\chi_1$  des Histidins der katalytischen Triade in hCatH ist im Vergleich zu pCatH um ca. 90° gedreht. Diese Orientierung der Aminosäurereste der katalytischen Triade im hCatH-Modell ist ideal, um das Thiolat-Imidazolium-Ionenpaar zu bilden. Die Abstände Cys26S $\gamma$ ...His166N $\delta$ 1 von 3.78 Å und His166N $\epsilon$ 2...Asn186O $\delta$ 1 von 3.37 Å im Modell stehen in Korrelation zu den entsprechenden Abständen der Röntgenkristallstrukturen von Enzymen der Papain-Superfamilie (Tab. A-4). Eine weitere Diskrepanz tritt in der S1'-Bindungsposition auf. Die Aminosäuren Trp188, Val140, Met145, His166 und Asn165 bilden diese Bindungsstelle von hCatH. In pCatH ist Met145 durch Leu145 substituiert, wodurch die Hydrophobizität dieser Bindungsposition verändert wird. Im Vergleich zu Met145 (hCatH) ist die Seitenkette von Leu145 (pCatH) weniger in die S1'-Bindungsposition gerichtet, was auf eine variierte Seitenkettenkonformation zurückzuführen ist. Daher kann man annehmen, daß in pCatH nichtkovalente Wechselwirkungen zu einem Liganden reduziert werden.

Insgesamt kann festgestellt werden, daß eine hohe Übereinstimmung zwischen der modellierten Proteinstruktur von hCatH und der Röntgenkristallstruktur von pCatH auftritt, was durch den niedrigen RMS-Wert von 2.24 Å bestätigt wird. Dieser Wert wurde über alle Hauptkettenatome von beiden Enzymen berechnet.

#### 4.2.3.4 Substratspezifität von Cathepsin H im Vergleich zu Cathepsin L

Zur weiteren Überprüfung der Qualität des Strukturmodells von hCatH wurde die Substratspezifität dieses Enzyms im Vergleich zu Cathepsin L untersucht. Dazu wurden kinetische Daten, die in der Literatur beschrieben sind, verwendet [62]. Als Substrat für hCatH wurde Arg-NMec (6) genutzt, da *Barrett & Kirschke* zeigen konnten, daß Cathepsin H dieses Substrat wesentlich effizienter hydrolysiert als Cathepsin L [62]. Unter Verwendung von FLEXIDOCK wurde dieses Molekül in das katalytisch aktive Bindungszentrum von hCatH gedockt. Für die vergleichenden Betrachtungen mit Cathepsin L wurde seine Röntgenkristallstruktur genutzt (PDB-Eintrag: 1CJL [118]). Diese Röntgenkristallstruktur enthält jedoch auch die Koordinaten des Propeptides, die für die weiteren Berechnungen entfernt wurden. Die im Vergleich zur Sequenz des reifen Enzyms auftretenden Mutationen (Cys25Ser und Thr110Ala) wurden durch den Austausch dieser Reste mit den entsprechenden Aminosäuren aufgehoben. Die zugehörigen Seitenkettenkonformationen wurden über den Vergleich der Anordnung dieser Atome in der ursprünglichen Röntgenkristallstruktur sowie Energieoptimierung dieser Aminosäurereste ermittelt. Ein Ausschnitt aus einer Loopregion (Thr175-Asn179, Cathepsin L-Numerierung) konnte mittels der Röntgenkristallstrukturanalyse nicht bestimmt werden. Dieser Loop wurde unter Anwendung der "LOOP-SEARCH"-Option des SYBYL-Programms generiert. Für weitere Betrachtungen wurde die energetisch günstigste Loopkonformation verwendet.

Nach Optimierung der resultierenden Enzym-Ligand-Komplexe wurden diese hinsichtlich ihres Wechselwirkungsverhaltens zwischen Protein und Substrat untersucht. Die berechneten Wechselwirkungsenergien und theoretisch bestimmten Bindungskonstanten sind in Tab. 11 aufgelistet.

**Tab. 11** Affinität des Substrates Arg-NMec (6) gegenüber den Cathepsinen H und L

Enzym	$K_m$ [mM] <sup>a</sup>	$\Delta E$ [kcal/mol] <sup>b</sup>	$K_d$ [mM] <sup>c</sup>
Cathepsin H	0.15	-52.03	0.59
Cathepsin L	- <sup>d</sup>	-39.19	46.35

<sup>a</sup> kinetische Daten [62]

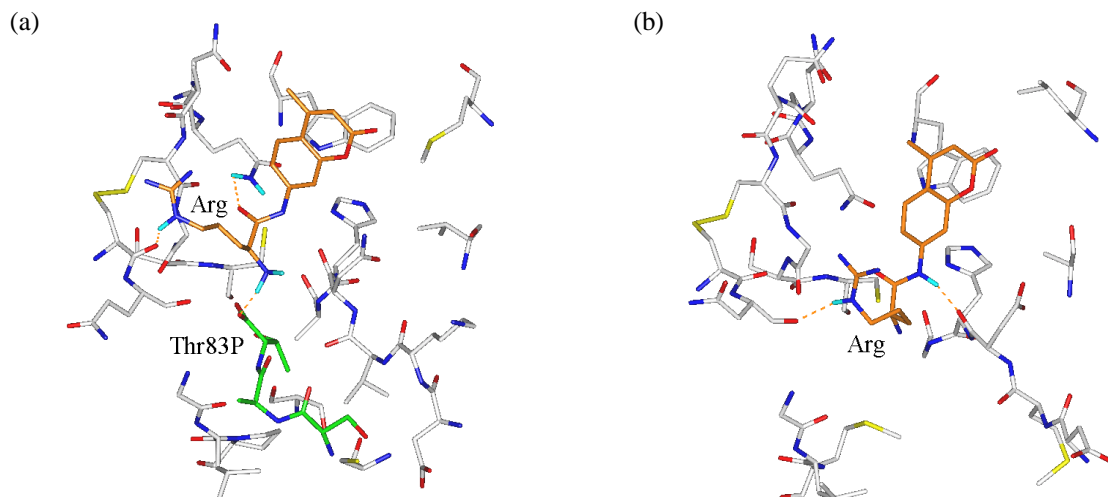
<sup>b</sup> nichtbindende Wechselwirkungsenergien

<sup>c</sup> berechnete Werte, unter Verwendung von SCORE

<sup>d</sup>  $K_m$  wurde nicht bestimmt [62]

Im Vergleich zu dem sich bildenden hCatL-Komplex ist der Enzym-Substrat-Komplex von hCatH deutlich energetisch bevorzugt. Der Affinitätsverlust des Liganden zu Cathepsin L wird durch die niedrigere Wechselwirkungsenergie dieses Komplexes ausgedrückt (ca. 13 kcal/mol).

Die Lage und Konformation des Substrates, gedockt in beide Cathepsine, weist einige Differenzen auf. Das Arginin in P1-Position entwickelt ein variiertes Dockingverhalten in beiden Cathepsinen. Für Cathepsin H konnte eine Reihe von Wechselwirkungen zur S1-Bindungsstelle (Gly68, Gly69 und Gly24) beobachtet werden (Abb. 21a). Wasserstoffbrückenbindungen bilden sich zwischen der Argininseitenkette und dem Carbonyl-Sauerstoffatom von Cys66 sowie zwischen dem Carbonyl-Sauerstoffatom von Arginin und der Seitenkette von Gln20 aus. Die Ausbildung einer Salzbrücke zwischen dem negativ geladenen C-Terminus der Minikette und dem positiv geladenen N-Terminus dieses Substrates trägt entscheidend zur Stabilisierung des hCatH-Substrat-Komplexes bei (Abb. 21a). Im Gegensatz dazu sind die anziehenden Wechselwirkungen zwischen Enzym und Substrat im korrespondierenden Komplex des Cathepsin L abgeschwächt (Abb. 21b). Aus diesen beschriebenen Bindungsunterschieden resultiert eine unterschiedliche Bindungsaffinität des Substrates zu hCatH bzw. hCatL. Die Tendenz der berechneten Wechselwirkungsenergien stimmt mit den kinetischen Daten überein. Es zeigt sich darin eine weitere Bestätigung für die Qualität der modellierten Tertiärstruktur von hCatH.



**Abb. 21** Enzym-Substrat-Komplex von hCatH (a) und hCatL (b). Von den Enzymen sind nur die Aminosäuren abgebildet, die an den Interaktionen zwischen dem Cathepsin und dem Substrat beteiligt sind sowie solche, die unmittelbar benachbart sind. Die Kohlenstoffatome des Substrates sind orangefarben. Thr83P, Ala82P und Ser81P der Minikette sind mit eingeblendet (Kohlenstoffatome sind grün eingefärbt).

#### 4.2.4 Ein dreidimensionales Strukturmodell von humanem Cathepsin F

Humanes Cathepsin F (hCatF) wurde erstmalig durch *Wang et al.* in der Literatur beschrieben [41]. Eine Röntgenkristallstrukturanalyse dieses Enzyms wurde bisher noch nicht durchgeführt. Da während des Zeitraums der Tertiärstrukturmodellierung dieser Protease das Programm COMPOSER nicht zur Verfügung stand, wurde das Programm MODELLER zur Generierung eines 3D-Strukturmodells von hCatF verwendet. Zunächst wurde ein multiples Sequenzalignment zwischen der Primärstruktur von hCatF und den verwandten Enzymen der Papain-Superfamilie, deren Koordinaten über die PDB zugänglich waren, durchgeführt. Basierend auf der Proteinsequenzhomologie von Cathepsin F zu den Cathepsinen B, K und L sowie Papain und Actinidin, wurden die Cathepsine L und K sowie Papain und Actinidin als optimales Set von Vorlagestrukturen zur Generierung des Strukturmodells von hCatF ermittelt (Tab. 12). Das Sequenzalignment zwischen hCatF und seinen Vorlagesequenzen ist in Abb. 22 dargestellt. Da bei dieser Strukturmodellierung weniger Vorlagestrukturen als bei den Cathepsinen K, S und H verwendet wurden, waren einige Modifizierungen in der Röntgenkristallstruktur von Cathepsin L (PDB-Eintrag: 1CJL [118]) notwendig (siehe Abschnitt 4.2.3.4).

**Tab. 12** Proteinsequenzhomologie und Identitätswerte von Cathepsin F zu den Cathepsinen L, K, B, Papain und Actinidin. Die Identitätswerte resultieren aus dem mit CLUSTAL W [97] durchgeführten Sequenzalignment. Die Homologiewerte wurden mit BLAST bestimmt [108].

Enzym	PDB-Eintrag	Homologie [%]	Identität [%]
Cathepsin L	1CJL	57	40
Cathepsin K	1MEM	56	39
Cathepsin B	1HUC	42	19
Papain	9PAP	51	34
Actinidin	1AEC	52	38

	10	20	30	40	50	60
hCatF	APPEWDWRSK	GAVTKVKDQG	MCGSCWAFSV	TGNVEGQWFL	NQGTLLSLSE	QELLDCDK--
1CJL	APRSVDWREK	GYVTPVKNQG	QCGSCWAFSA	TGALEGMFR	KTGRLISLSE	QNLVDCSGPQ
1MEM	APDSVDYRKK	GYVTPVKNQG	QCGSCWAFSS	VGALBGLK	KTGKLLNLS	QNLVDCVSE-
9PAP	IPFYVDWRQK	GAVTPVKNQG	SCGSCWAFSA	VVTIEGIKI	RTGNLNEYSE	QELLDCDR--
1AEC	LPSYVDWRSR	GAVVDIKSQG	ECGSCWAFSA	IATVEGINKI	VTGVLISLSE	QELIDCGRTQ
	-----	-----	---HHHHHHHH	HHHHHHHHHH	----EEEE-H	HHHH-----
	70	80	90	100	110	120
hCatF	MDKACMGGLP	SNAYSAIKNL	GGLETEDDYS	YQGHMQSCNF	SAEKAKVYIN	DSVELSQ--N
1CJL	GNEGCGGLM	DYAFQYVQDN	GGLDSEESYP	YEATEESCKY	NPY-SVAND	TGFVDIPK-Q
1MEM	-NDGCGGGM	TNAFQYVQKN	RGIDSEDAYP	YVQEEESCMY	NPTG-KAAKC	RGYREIPEGN
9PAP	RSYCGNGGYP	WSALQLVAQY	G-IHYRNTYP	YEGVQRYCRS	REKGPYAAK	DGVRQVQPYN
1AEC	NTRGCGGGYI	TDGFQFIINN	GGINTEENYP	YTAQDGCNCL	DLQNEKYVTI	DTYENVYPYN
	-----HHH	HHHHHHHHHH	-----	-----	-----EEEE	-EEE-----H
	130	140	150	160	170	180
hCatF	EQKLAAWLAK	RGPISVAINA	FG--MQFYRH	GISRPLRPLC	SPWLIDHAVL	LVGYGNRS--
1CJL	EKALMKAVAT	VGPISVDA	GHESEFLFYKE	GIY--FEPDC	SSEDMDHGVL	VVGYGFESTE
1MEM	EKALKRAVAR	VGPVSVDA	SLTSFQFYSK	GVY--YDESC	NSDNLNHAVL	AVGYGIQ---
9PAP	EGALLYSIAN	Q-PVSVVLEA	AGKDFQLYRG	GIF--VGP-C	GN-KVDHAVA	AVGYGPN---
1AEC	EWALQTAVTY	Q-PVSVVALDA	AGDAFKHYSS	GIF--TGP-C	GT-AIDHAVT	IVGYGTEG--
	HHHHHHHHHH	---EEEEEE-	---EEEEEE	EEEE-----	-----EEEE	EEEE-----
	190	200	210	220	230	240
hCatF	--DVPFWAIK	NSWGTDWGK	GYYYLHRGSG	----ACGVNT	MASSAVVD	
1CJL	SDNNKYWLVK	NSWGEWGMG	GYVKMAKDRR	---NHCGIAS	AASYPTV-	
1MEM	-KGNKHIIK	NSWGENWGNK	GYILMARNKN	---NACGIAN	LASFPM-	
9PAP	-----YILIK	NSWGTGWGEN	GYIRIKRGTG	NSYGVCGLYT	SSFYPVKN	
1AEC	--GIDYWIVK	NSWDTTWEE	GYMRILRNVG	-GAGTCGIAT	MPSYPVKY	
	---EEEEEE	-----	-EEEEEE-	-----EEEE	-----	

**Abb. 22** Multiples Sequenzalignment von hCatF mit Cathepsin L (1CJL [118]), Cathepsin K (1MEM [132]), Papain (9PAP [19]) und Actinidin (1AEC [117]) unter Verwendung von CLUSTAL W [97]. Identische Aminosäuren wurden fett hervorgehoben. Die Aminosäuren der katalytischen Triade sind mit einem Stern gekennzeichnet.

Das berechnete Modell erfüllt alle Restraints, die zur Erzeugung des Strukturmodells verwendet werden (siehe Abschnitt 3.1.3.2). Die RMS-Abweichungen zu den idealen Werten für Bindungslänge, Bindungswinkel und Torsionswinkel sind 0.005 Å, 1.981° und 1.329°. Neben dem Cystein der

katalytischen Triade enthält hCatF sechs weitere Cysteinreste, welche Disulfidbrücken ausbilden (Cys22-Cys63, Cys56-96 und Cys154-Cys202).

Die Wasserstoffatome wurden ergänzt und eine Energieminimierung der Proteinstruktur mit dem TRIPOS-Kraftfeld durchgeführt. Das resultierende Tertiärstrukturmodell dieses Enzyms ist in der Abb. 23 zu sehen (siehe auch Abb. A-9). Analog zu anderen catheptischen Cysteinproteasen besteht die Enzymstruktur dieses Proteins aus zwei Domänen. Im Zentrum der Spalte befinden sich die Aminosäuren Cys25, His161 und Asn181, die das katalytisch aktive Zentrum des Enzyms bilden. Die Seitenkette von Cys25 liegt als Thiolatanion und der Imidazolring von His161 in protonierter Form vor.



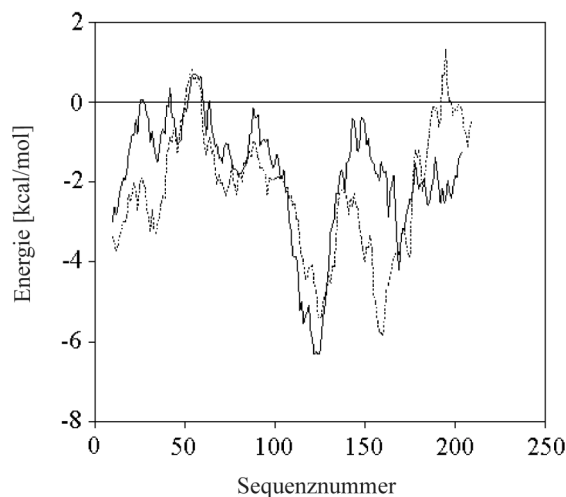
**Abb. 23.** Tertiärstruktur des humanen Cathepsin F. Die Sekundärstrukturelemente wurden mit DSSP berechnet. Helices sind rot gefärbt und  $\beta$ -Faltblattstrukturen blau. Mit dargestellt sind die Aminosäuren der katalytischen Triade Cys25, His161 und Asn181.

Cathepsin F weist im Vergleich zu anderen catheptischen Cysteinproteasen gemeinsame katalytische und inhibitorische Merkmale auf, die auf das Thiolat-Imidazolium-Ionenpaar des aktiven Zentrums des Enzyms zurückzuführen sind. Obwohl die Homologie zwischen den Aminosäuresequenzen des gesamten Proteins in einem Bereich von 40%-50% liegt, ist die Homologie der Umgebung der katalytischen Triade größer 90%. Aus diesem Grund kann man annehmen, daß die strukturellen Charakteristika der Region des aktiven Zentrums sehr gut über die Prozedur des "Homology Modelling" wiedergegeben werden. Der Abstand zwischen den Ionenpaar von Cys25S $\gamma$  und His161N $\delta$ 1 beträgt 3.41 Å. Die Seitenkette von Asn181 bildet zum His161 eine Wasserstoffbrücke aus (His161N $\epsilon$ 2H $\cdots$ Asn181O $\delta$ 1 = 1.94 Å). Ähnliche Werte entsprechender Abstände findet man in homologen Cysteinproteasen (Tab. A-4).

Eine Analyse des Strukturmodells von hCatF wurde mit PROCHECK durchgeführt. Der Ramachandran-Plot, der von dem Strukturmodell von hCatF berechnet wurde, spiegelt die hohe Qualität des Modells wider (Abb. A-13). Die ( $\Phi$ ,  $\Psi$ )-Torsionswinkel von 81% der Aminosäurereste nehmen Werte ein, die für eine nativ gefaltete Struktur sprechen.



Eine Analyse der Proteinfaltung des Modells von hCatF kann unter Anwendung des PROSA-Programms durchgeführt werden. Das Energieprofil für hCatF wird in Abb. 24 gezeigt. In den nachfolgenden Betrachtungen wird der Energieverlauf des hCatF-Modells mit dem Energiegraphen, der von der Röntgenkristallstruktur von Cathepsin L berechnet wurde, verglichen. Genauere Untersuchungen ergeben, daß die Energiegraphen von beiden Cathepsinen Unterschiede in den Sequenzabschnitten von hCatF Ala1-Leu48, Gln140-Asn170 und Pro175-Asp214 aufweisen (Cathepsin F-Numerierung) (Abb. 24).



**Abb. 24** Energieprofile von Cathepsin F (durchgezogene Linie) und Cathepsin L (gepunktete Linie). Es wurden die durchschnittlichen Energiewerte der Proteinsequenz über einen Bereich von 20 Aminosäureresten abgebildet.

Die Lage der Hauptkettenatome innerhalb des ersten Segmentes ist in beiden Cathepsinen ähnlich. Die RMS-Wert beträgt  $0.42 \text{ \AA}$ , berechnet über die  $C\alpha$ -Atome dieses Bereiches. Die energetischen Unterschiede der Struktur von hCatF im Vergleich zu Cathepsin L kann durch die unterschiedliche Ausrichtung von Aminosäureseitenketten sowie durch benachbarte Aminosäuren innerhalb des Enzyms verursacht werden. Der Bereich zwischen den Aminosäuren Gln140 und Asn170, für den weitere Unterschiede zwischen den Energiegraphen beobachtet werden, enthält eine lange Loopregion. Diese Loopregion besitzt im Vergleich zum Cathepsin L zwei zusätzliche Aminosäuren, woraus eine veränderte Anordnung dieser Region resultiert. Eine weitere Erklärung für die auftretenden Differenzen könnte sein, daß sich dieser Bereich an der Oberfläche des Enzyms befindet. Diese Gründe können die Ausbildung einer unterschiedlichen Architektur in der Proteinstruktur bedingen. Demzufolge wurde zwischen beiden Cathepsinen für diesen Bereich eine erhöhte RMS-Abweichung von  $0.75 \text{ \AA}$  ermittelt. Wie der Verlauf der Energiegraphen zeigt, ist die Struktur des letzten Abschnitts (Pro175-Asp214) für hCatF stärker begünstigt als die entsprechende Region von Cathepsin L. In diesem Abschnitt befinden sich sowohl drei  $\beta$ -Faltblattsegmente, eine helikale Struktur als auch Loopregionen (Abb. A-9). Die Positionen der Hauptkettenatome der definierten Sekundärstrukturelemente stimmen mit der Lage der korrespondierenden Atome von Cathepsin L

überein (RMS-Wert  $\leq 0.39 \text{ \AA}$ ). Daher kann man annehmen, daß die Unterschiede im Energieprofil zwischen beiden Cathepsinen durch eine veränderte Orientierung der Loopregionen hervorgerufen werden. Der berechnete Z-Score des hCatF-Modells beträgt  $-7.90$  und liegt somit in dem Bereich für nativ gefaltete Proteine.

#### 4.2.4.1 Untersuchung der Bindungsspalte von Cathepsin F

Die Bindungsspalte des Modells von Cathepsin F beinhaltet große Ähnlichkeiten zu homologen Cysteinproteasen. Die Spalte wird analog zu diesen verwandten Papain-ähnlichen Enzymen in der L-Domäne durch die Proteinsegmente Gly20-Trp26 und Met64-Leu67 begrenzt. In der R-Domäne wird diese Begrenzung der Spalte durch die Bereiche Ser131-Ala133, Met139-Phe141, Ile159-Ala162 sowie Trp183-Asp186 gebildet. Die Seitenketten der Aminosäurereste Gln19, Cys25, Pro68, Ala133, Gln140, Ile159, Trp183 und Met207 bestimmen durch ihre Anordnung in der Bindungsspalte von hCatF die Eigenschaften dieser Region.

##### *Beschreibung der Bindungsstellen im Modell von Cathepsin F*

Anhand des Strukturmodells von hCatF und der auftretenden Übereinstimmung des Faltungsmusters zu verwandten Cysteinproteasen kann das katalytisch aktive Bindungszentrum dieses Enzyms mit seinen Bindungsstellen S1-S3 und S1'-S2' charakterisiert werden. Die Aminosäurereste, die diese Bindungsstellen ausbilden können, sind in Tab. 13 aufgeführt.

**Tab. 13** Bindungsstellen (Sn) von Cathepsin F, die aus dem Tertiärstrukturmodell dieses Enzyms resultieren

Bindungsstellen	S1	S2	S3	S1'	S2'
Aminosäurereste	Cys22	Ala162	Lys61	Trp183	Gln19
	Gly23	Met207	Leu67	Ala136	Gly20
	Cys63	Leu67		Gln140	Met21
	Gly65	Pro68		Asp160	Cys22
	Gly66	Ala133		His161	Gly23
	Cys25	Ile159			Ser24
	Gln19				

Die S1-Bindungsstelle befindet sich vollständig in der L-Domäne. Sie wird aus Aminosäuren gebildet, die in homologen Cysteinproteasen konserviert vorliegen. Diese Aminosäuren gehören zu einer zentralen  $\alpha$ -Helix sowie zu den umgebenden Loopregionen. Die S2-Bindungsstelle setzt sich im Vergleich zu homologen Cysteinproteasen aus unterschiedlichen Aminosäuren zusammen, die einen überwiegend hydrophoben Charakter aufweisen. Die Aminosäuren der S2-Bindungstasche richten ihre Seitenketten so aus, daß sie in die Bindungsspalte zeigen. Begrenzt wird diese Bindungstasche durch

Met207. Die S3-Bindungsstelle wird aus den Aminosäuren Lys61 und Leu67 gebildet (L-Domäne), wobei für die Wechselwirkungen zwischen den Aminosäuren dieser Bindungsstelle und einem Substrat nur die Seitenketten von Bedeutung sind. Die Eigenschaften dieser Bindungsposition sind im Vergleich zu den anderen catheptischen Enzymen verschieden, wodurch diese S3-Bindungsstelle von großen Interesse für das Design von neuen spezifischen Liganden für die Cathepsine sein kann. Weitere Unterschiede lassen sich für die Aminosäuren der S1'-Bindungsstelle beobachten. Vergleicht man den Aufbau dieser Bindungsposition mit anderen Enzymen der Papain-Superfamilie, wird ersichtlich, daß die S1'-Bindungsstelle Aminosäuren enthält, die sich in ihren Eigenschaften deutlich zu den anderen Proteinen unterscheiden (Gln140, Asp160).

Es ist anzunehmen, daß die S2'-Bindungsstelle keinen essentiellen Einfluß auf die Substratspezifität für hCatF ausübt, da die Seitenketten der Aminosäuren dieser Bindungsstelle nicht zur Bindungsspalte ausgerichtet sind.

#### 4.2.4.2 Weitere Überprüfungen des Strukturmodells von Cathepsin F

Durch *Brömme et al.* wurde die Bindungsaffinität von Inhibitoren in der Bindungsspalte von Cathepsin F im Vergleich zu Cathepsin L und Calpain untersucht [140]. Für eine zusätzliche Abschätzung der Qualität des 3D-Proteinmodells von hCatF wurden die Inhibitoren Z-Leu-Abu-CONH-CH<sub>2</sub>-2-furan (7) und Z-Leu-Phe-CONH-(CH<sub>2</sub>)<sub>2</sub>-Ph (8) in die katalytisch aktiven Bindungszentren der Cathepsine F und L gedockt und die Affinität sowie das Wechselwirkungsverhalten zwischen Inhibitor und Enzym genauer untersucht. Eine Abbildung der Strukturformeln dieser Inhibitoren befindet sich im Anhang (Abb. A-14d,e). Die Dockingprozedur wurde unter Verwendung von FLEXIDOCK durchgeführt.

Aufgrund der Größe des katalytisch aktiven Bindungszentrums von hCatL können sich die Inhibitoren besser in die Bindungsspalte dieses Enzyms einpassen, als in hCatF. Daraus resultieren stärkere nichtkovalente Interaktionen zwischen dem Inhibitor und Cathepsin L, im Vergleich zu dem Enzym-Inhibitor-Komplex von Cathepsin F. Diese Befunde werden auch über die K<sub>d</sub>-Werte und die berechneten Energien für die Wechselwirkungen im Enzym-Inhibitor-Komplex wiedergegeben (Tab. 14).

**Tab. 14** Affinität der Inhibitoren zu den Cathepsinen F und L

Nr.	Inhibitor	Cathepsin F			Cathepsin L		
		$K_i$ [nM] <sup>a</sup>	$\Delta E$ [kcal/mol] <sup>b</sup>	$K_d$ [nM] <sup>c</sup>	$K_i$ [nM] <sup>a</sup>	$\Delta E$ [kcal/mol] <sup>b</sup>	$K_d$ [nM] <sup>c</sup>
(7)	Z-Leu-Abu-CH <sub>2</sub> -2-furan	7.2±2	-40.76	8.51	1.5±0.4	-49.62	1.89
(8)	Z-Leu-Phe-(CH <sub>2</sub> ) <sub>2</sub> -Ph	0.5±0.1	-54.69	0.39	0.7±0.2	-55.18	0.97

<sup>a</sup> bestimmt durch *Brömme et al.* [140]

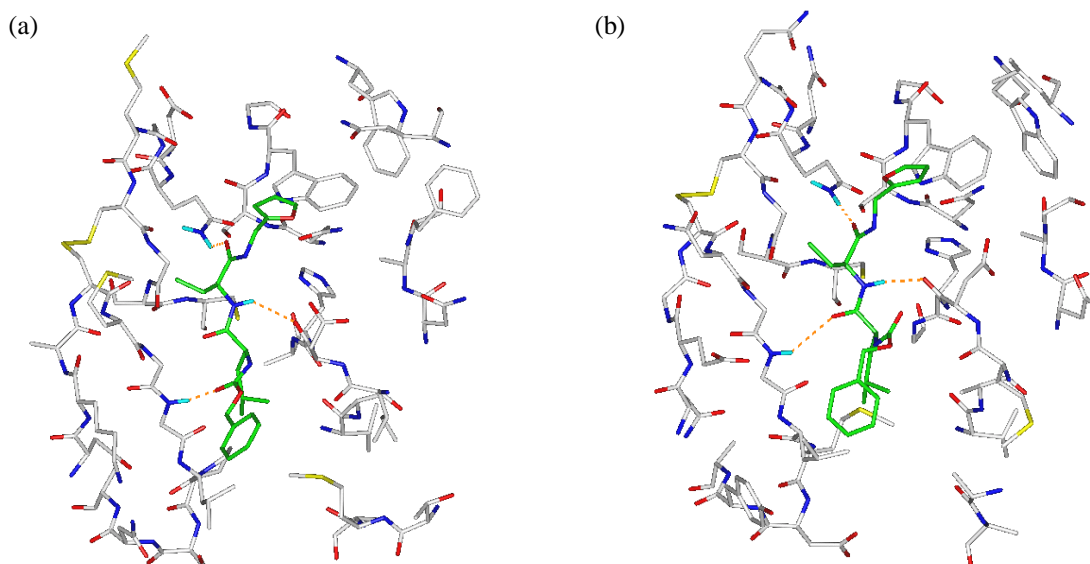
<sup>b</sup> berechnete Wechselwirkungsenergien

<sup>c</sup> Dissoziationskonstanten, ermittelt unter Verwendung von SCORE

Die unterschiedliche Affinität der Inhibitoren gegenüber den Cathepsinen soll anhand des Inhibitors Z-Leu-Abu-CH<sub>2</sub>-2-furan (7) detaillierter beschrieben werden.

Wird dieser Ligand in die Bindungsspalte von Cathepsin L gedockt, nehmen der Furanring des Inhibitors und die Seitenkette von Trp189 des Enzyms eine solche Anordnung zueinander ein, die stabile Wechselwirkungen zwischen diesen Resten ermöglicht. Aufgrund der veränderten Lage des Inhibitors im katalytisch aktiven Bindungszentrum von hCatF werden diese Wechselwirkungen abgeschwächt (Abb. 25). In beiden Cathepsinen ordnet sich die Seitenkette der 2-Aminobuttersäure des Inhibitors zwischen den Proteinsegmenten Gly23-Cys25 und Met64-Met65 an (Cathepsin F-Numerierung). Unterschiedliche Wechselwirkungen werden zwischen Leucin in P2-Position und den Aminosäuren der S2-Bindungstasche beobachtet. Da der Inhibitor tiefer in der Bindungsspalte von Cathepsin L liegt, bilden sich anziehende Wechselwirkungen zwischen dem P2-Rest und den Seitenketten von Leu69, Met70 und Ala135 aus. Zusätzlich richtet sich der Phenylring der Z-Gruppe so zur Seitenkette von Leu69 aus, daß weitere stabile Interaktionen zwischen dem Liganden und Cathepsin L auftreten können (Abb. 25b). Im hCatL-Komplex bilden sich Wasserstoffbrückenbindungen zwischen dem Amidproton von 2-Aminobuttersäure und dem Carbonyl-Sauerstoffatom der Hauptkette von Asp162, vom Carbonyl-Sauerstoffatom des Abu-Restes zum Amidproton der Gln19-Seitenkette sowie zwischen dem Carbonyl-Sauerstoffatom des Leucins und dem Amidproton von Gly68 aus (Abb. 25b).

Im Fall des hCatF-Komplexes werden diese Wechselwirkungen zwischen Leucin in Position P2 sowie der Z-Gruppe und dem Enzym eingeschränkt (Abb. 25a). Wasserstoffbrücken können sich zwischen dem Amidproton von Abu und dem Carbonyl-Sauerstoffatom von Asp160 sowie vom Carbonyl-Sauerstoffatom der Z-Gruppe zum Amidproton von Gly66 ausbilden (Abb. 25a).



**Abb. 25** Die katalytisch aktiven Bindungszentren von Cathepsin F (a) und Cathepsin L (b) mit dem Inhibitor Z-Leu-Abu-CH<sub>2</sub>-2-furan (7). Die Kohlenstoffatome des Inhibitors sind grün.

Anhand dieser Ergebnisse kann man annehmen, daß das Proteinmodell von Cathepsin F von grundlegender Richtigkeit ist und für weitere Studien verwendet werden kann.

Zusammenfassend kann gesagt werden, daß es möglich war, die 3D-Proteinstrukturen der Cathepsine K, S, H und F basierend auf ihrer Sequenzhomologie zu anderen Papain-ähnlichen Cysteinproteasen zu generieren. Dabei erwiesen sich die verwendeten Methoden als sehr gut geeignet. Die Ergebnisse der Überprüfungen dieser Strukturmodelle, die über verschiedene Methoden durchgeführt wurden, unterstreichen die hohe Relevanz dieser Proteinstrukturen. Wie der Vergleich der 3D-Strukturmodelle mit den Röntgenkristallstrukturen homologer Cathepsine zeigt, enthalten die generierten Proteinstrukturen das für die Papain-Superfamilie charakteristische Faltungsmuster. Aus diesem Grund können die Tertiärstrukturmodelle der Cathepsine K, S, H und F als ein hilfreiches Instrument zur Untersuchung der Substratspezifität dieser Enzyme verstanden werden.

Die Substratspezifität der modellierten Proteinstrukturen wurde anhand von in der Literatur beschriebenen kinetischen Daten untersucht. Dabei konnte festgestellt werden, daß die berechneten Werte der Bindungsaffinität der Liganden zum Enzym in guter Übereinstimmung mit den experimentell bestimmten Werten stehen. Es kann somit davon ausgegangen werden, daß die aufgezeigten Methoden zur Berechnung der Wechselwirkungsenergie sowie der Dissoziationskonstanten der Enzym-Ligand-Komplexe ein geeigneter Weg ist, um Aussagen über die Bindungsaffinität der untersuchten Liganden zu treffen.

### 4.3 Design neuer Liganden der Cathepsine L, K und F

#### 4.3.1 Vergleichende Studien an lysosomalen Cysteinproteasen

Zur Entwicklung von spezifischen Liganden für catheptische Cysteinproteasen wurden vergleichende Untersuchungen an den katalytisch aktiven Bindungszentren der Cathepsine B, F, H, K, L und S durchgeführt. Basierend auf den nun vorliegenden Tertiärstrukturen dieser Cathepsine konnten die Aminosäuren, welche die einzelnen Bindungspositionen S1-S3 sowie S1'-S2' bilden, ermittelt werden (Tab. 15). Die Beschreibung der Bindungspositionen soll dazu dienen, das bisherige Wissen über die strukturelle Organisation der Cathepsine zusammenzufassen. Es muß allerdings berücksichtigt werden, daß die resultierende Substrat- bzw. Inhibitoraffinität nicht die einfache Summierung der Beiträge aller Fragmente darstellt, sondern ein Ergebnis der gesamten Interaktionen mit den Aminosäuren des Enzyms ist.

Die bisherige Kenntnis der Substratbindungsregion basiert auf kinetischen Daten sowie auf den Röntgenkristallstrukturen von Enzym-Inhibitor-Komplexen. Ein Großteil der Strukturen dieser Komplexe liefert Erkenntnisse über die S<sub>n</sub>-Bindungspositionen. Die Inhibitoren sind durch Reaktion des Cysteins der katalytischen Triade mit einem Chlormethylketon (1PAD, 5PAD, 6PAD [48]), einem Fluormethylketon (1AIM [141], 2AIM [142]), einem Diazomethan (1GEC [143]) oder einem Aldehyd (1POP [144]) kovalent mit dem Enzym verknüpft. Mittlerweile konnten auch Röntgenkristallstrukturen von Enzym-Inhibitor-Komplexen bestimmt werden, durch die aufschlußreiche Informationen neben den S<sub>n</sub>-Bindungspositionen auch Informationen über die S<sub>n</sub>'-Bindungsstellen zugänglich sind (1THE [145], 1MEM [132], 1ATK [146], 1AU0 [130], 1AU2 [130], 1AU3 [130], 1AU4 [130], 1AYU [147], 1AYV [147], 1AYW [147], 1BGO [148]). Aus den Strukturen von CA030 im Komplex mit humanem Cathepsin B (1CSB [49]) und dem fast identischen Molekül CA074 mit Cathepsin B, isoliert aus der Rindermilz (1QDQ [149]), lassen sich Rückschlüsse über die S<sub>2</sub>'-Bindungsstelle ziehen.

Die Strukturen zeigen, daß der Inhibitor in der aktiven Bindungsspalte der Cathepsine in einer gestreckten Konformation bindet. Die Amidprotonen (Hauptkette) der Reste in Positionen P1-P2 und P1'-P2' können Wasserstoffbrückenbindungen mit konservierten Aminosäuren der Cathepsine ausbilden, wobei sich die Seitenketten der Inhibitoren überwiegend in die Bindungsstellen S1-S2 und S1'-S2' ausrichten.

**Tab. 15** Aminosäuren der Bindungspositionen S1-S3 und S1'-S2' der untersuchten Cathepsine B, L, H, S, K und F

Bindungsposition	hCatB	hCatL	hCatH	hCatS	hCatK	hCatF
S1	<i>Gln23</i>	<i>Gln19</i>	<i>Gln20</i>	<i>Gln19</i>	<i>Gln19</i>	<i>Gln19</i>
	<i>Cys26</i>	<i>Cys22</i>	<i>Cys23</i>	<i>Cys22</i>	<i>Cys22</i>	<i>Cys22</i>
	<i>Gly27</i>	<i>Gly23</i>	<i>Gly24</i>	<i>Gly23</i>	<i>Gly23</i>	<i>Gly23</i>
	<b>Cys29</b>	<b>Cys25</b>	<b>Cys26</b>	<b>Cys25</b>	<b>Cys25</b>	<b>Cys25</b>
	<i>Cys71</i>	<i>Cys65</i>	<i>Cys66</i>	<i>Cys66</i>	<i>Cys63</i>	<i>Cys63</i>
	<i>Asn72</i>	<i>Asn66</i>	<i>Gln67</i>	<i>Asn67</i>	<i>Gly64</i>	<i>Met64</i>
	<i>Gly73</i>	<i>Gly67</i>	<i>Gly68</i>	<i>Gly68</i>	<i>Gly65</i>	<i>Gly65</i>
	<i>Gly74</i>	<i>Gly68</i>	<i>Gly69</i>	<i>Gly69</i>	<i>Gly66</i>	<i>Gly66</i>
S2	<b>Tyr75</b>	<b>Leu69</b>		<b>Phe70</b>	<b>Tyr67</b>	<b>Leu67</b>
	<b>Pro76</b>	<b>Met70</b>		<b>Met71</b>	<b>Met68</b>	<b>Pro68</b>
	<i>Ala173</i>	<i>Ala135</i>		<i>Gly137</i>	<i>Ala134</i>	<i>Ala133</i>
	-	<b>Met161</b>		<b>Val162</b>	<b>Leu160</b>	<b>Ile159</b>
	<i>Ala200</i>	<i>Gly164</i>		<i>Gly165</i>	<i>Ala163</i>	<i>Ala162</i>
	<b>Glu245</b>	<b>Ala214</b>		<b>Phe211</b>	<b>Leu209</b>	<b>Met207</b>
S3	<b>Asp69</b>	<b>Glu63</b>		<b>Lys64</b>	<b>Asp61</b>	<b>Lys61</b>
	<b>Tyr75</b>	<b>Leu69</b>		<b>Phe70</b>	<b>Tyr67</b>	<b>Leu67</b>
S1'	<b>Val176</b>	<b>Ala138</b>	<b>Val140</b>	<b>Ala140</b>	<b>Ala137</b>	<b>Ala136</b>
	<b>Leu181</b>	<b>Leu144</b>	<b>Met145</b>	<b>Phe146</b>	<b>Gln143</b>	<b>Gln140</b>
	<i>Gly198</i>	<b>Asp162</b>	<b>Asn165</b>	<b>Asn163</b>	<b>Asn161</b>	<b>Asp160</b>
	<b>His199</b>	<b>His163</b>	<b>His166</b>	<b>His164</b>	<b>His162</b>	<b>His161</b>
	<b>Trp221</b>	<b>Trp189</b>	<b>Trp188</b>	<b>Trp186</b>	<b>Trp184</b>	<b>Trp183</b>
S2'	<i>Gln23</i>	<i>Gln19</i>	<i>Gln20</i>	<i>Gln19</i>	<i>Gln19</i>	<i>Gln19</i>
	<i>Gly24</i>	<i>Gly20</i>	<i>Gly21</i>	<i>Gly20</i>	<i>Gly20</i>	<i>Gly20</i>
	<i>Ser25</i>	<i>Gln21</i>	<i>Ala22</i>	<i>Ser21</i>	<i>Gln21</i>	<i>Met21</i>
	<i>Cys26</i>	<i>Cys22</i>	<i>Cys23</i>	<i>Cys22</i>	<i>Cys22</i>	<i>Cys22</i>
	<i>Gly27</i>	<i>Gly23</i>	<i>Gly24</i>	<i>Gly23</i>	<i>Gly23</i>	<i>Gly23</i>
	<i>Ser28</i>	<i>Ser24</i>	<i>Ser25</i>	<i>Ala24</i>	<i>Ser24</i>	<i>Ser24</i>
	<b>His110</b>					
	<b>His111</b>					

Kennzeichnung der Aminosäuren:

kursiv: die Aminosäuren sind konserviert bzw. weisen keine großen Unterschiede untereinander auf,

fett: diese Aminosäuren besitzen einen Einfluß auf die Substratspezifität dieser Enzyme,

kursiv und fett: die Aminosäuren sind konserviert und die Seitenketten dieser Aminosäuren liegen in der Bindungsspalte.

Die Numerierung der Aminosäuren der Bindungsstellen bezieht sich in den nachfolgenden Abschnitt immer auf Cathepsin L, sofern es nicht anders angegeben wird.

Die Aminosäuren der Bindungsstelle S1 liegen in der L-Domäne der Cathepsine. Die Bindungsstelle für die Seitenkette des P1-Restes wird aus einer Loopregion (Cys22, Gly23) gebildet, die sich in die zentrale  $\alpha$ -Helix fortsetzt. Ein weiterer Bereich der S1-Bindungsstelle besteht aus den Aminosäuren Cys65, Asn66, Gly67 und Gly68 (Tab. 15). Beide Ketten des Enzyms werden über eine Disulfidbrücke Cys22-Cys65 verbunden. Vergleicht man die Cathepsine B, F, H, K, L und S untereinander, wird ersichtlich, daß sich die Bindungsstelle S1 meist aus Aminosäuren zusammensetzt, die in den Papain-ähnlichen Enzymen konserviert vorliegen.

Die S2-Bindungstasche wird aus Aminosäuren der R- und L-Domäne gebildet. Im Gegensatz zu den anderen untersuchten Cathepsinen enthält Cathepsin B in dieser Bindungsstelle ein Glutamat. Diese Substitution bewirkt, daß ein basischer Aminosäurerest in der P2-Position (Arginin, Lysin) effizienter gebunden wird als ein hydrophober Rest. Untersuchungen haben bestätigt, daß Glu245 (Cathepsin B-Numerierung) eine wesentliche Bedeutung für die Substratspezifität dieses Enzyms besitzt [150]. Die Cathepsine L, K, F und S enthalten in dieser Position Ala214, Leu209, Met207 bzw. Phe211 (Tab. 15). Die unterschiedlichen Eigenschaften dieser Aminosäurereste hinsichtlich ihrer Hydrophobizität und ihres sterischen Anspruchs können zur Ermittlung eines Restes in P2-Position dienen, der ideale Wechselwirkungen zum Enzym ausbilden kann. Die Seitenketten der Aminosäuren der S2-Bindungstasche weisen meist in die Bindungsspalte der Cathepsine. Es ist darauf hinzuweisen, daß Cathepsin H keine S2-Bindungstasche enthält, da diese durch die Minikette belegt wird.

Im Gegensatz zu den gut charakterisierten Bindungspositionen S1 und S2 ist die S3-Bindungsstelle nicht genau lokalisierbar. Für diese Bindungsposition können Wechselwirkungen zwischen der Enzymoberfläche und dem Liganden beobachtet werden. Diese Interaktionen sind auf die Seitenketten der Aminosäuren, welche diese Bindungsstelle bilden, beschränkt. Anhand Tab. 15 ist zu sehen, daß sich die Aminosäurereste, die Einfluß auf die Eigenschaften dieser Bindungsstelle besitzen, unterscheiden. Die Cathepsine F und S enthalten einen basischen Aminosäurerest (Lysin), während die Cathepsine B, L und K eine Aminosäure mit einer sauren Seitenkette aufweisen (Aspartat bzw. Glutamat). Die Aminosäuren an den Positionen 63 und 69 der S3-Bindungsstelle sind nicht in den Enzymen dieser Familie konserviert. Sie richten ihre Seitenkette zum Lösungsmittel aus. Die beschriebenen unterschiedlichen Eigenschaften der S3-Bindungsstelle können für ein rationales Ligandendesign ausgenutzt werden.

Die S1'-Bindungsstelle der lysosomalen Cysteinproteasen setzt sich überwiegend aus Aminosäuren der R-Domäne zusammen. Sie enthält neben einem Tryptophan, das für die Enzyme der Papain-Superfamilie konserviert ist und dem Histidin der katalytischen Triade, Aminosäuren mit unterschiedlichen Eigenschaften (Tab. 15), die zum effektiven Design von neuen Liganden ausgenutzt werden können. Vor allem ist dabei auf das Asp162 in den Cathepsinen L und F hinzuweisen. Im Gegensatz dazu enthalten die Cathepsine B, H, K und S an dieser Position ein Glycin bzw. Asparagin (Tab. 15). Die Seitenketten der Aminosäuren, die diese Bindungsposition bilden, ragen in die Bindungsspalte (Tab. 15).

Die Struktur des CA030-Komplexes von Cathepsin B deutet darauf hin, daß die Seitenkette der P2'-Position in einer Loopregion oberhalb von Gln23-Ser28 bindet (Cathepsin B-Numerierung). Aufgrund der Aminosäuren, die diese Bindungsstelle bilden, besitzen die Cathepsine F, H, K, L und S ein ähnliches Bindungsprofil (Tab. 15). Allerdings werden die Eigenschaften dieser Bindungsstelle in Cathepsin B durch seinen Einschubloop, der die beiden Histidinreste His110 und His111 enthält, beeinflusst.



### 4.3.1.1 Die elektrostatischen Potentiale der molekularen Oberfläche der Cathepsine L, B, K, H, S und F

Eine weitere Möglichkeit zur Charakterisierung des katalytisch aktiven Bindungszentrums ist die Berechnung der elektrostatischen Potentiale der molekularen Oberfläche des Enzyms. Im einfachsten Fall bedient man sich dazu eines Punktladungsmodells. Jedem Atom des Modells wird eine Partialladung zugeordnet. Das durch die Konformation des Moleküls bestimmte Zusammenspiel dieser Ladungen kann auf der Connolly-Oberfläche in Form der Wechselwirkung mit einer positiven Einheitsladung abgebildet werden und stellt das elektrostatische Potential dar.

Die elektrostatischen Oberflächenpotentiale (EP) der Bindungsspalte und seiner Umgebung der Cathepsine B, L, K, S, H und F wurden mit dem im SYBYL implementierten Programm MOLCAD berechnet (Abb. 26a-f, Seite 77 f.). Die Bereiche mit einem positiven EP sind in dieser Abbildung rot gefärbt und violett gekennzeichnete Bereiche stellen negative EP dar. Zum Vergleich wurden jeweils die Aminosäurereste der Enzymausschnitte mit abgebildet, über die diese Potentiale berechnet wurden (siehe auch Abb. A-4).

Die für die Papain-ähnlichen Cathepsine charakteristische Bindungsspalte ist anhand der Darstellung der EP gut zu erkennen. Die Begrenzungen dieser Spalte treten deutlich hervor. Eine Ausnahme bildet die Abbildung des Potentials von Cathepsin H, da die Bindungsspalte teilweise durch die Minikette belegt wird (Abb. 26d). Anhand dieser Visualisierung wird ersichtlich, daß die Größe der Substratbindungsregionen von hCatL, hCatB, hCatS und hCatF ähnlich ist. Aufgrund der Ausrichtung einiger Aminosäureseitenketten innerhalb der Proteinstruktur von hCatK ist diese Bindungsregion in ihrer Größe reduziert.

Im Gegensatz zu der hohen Sequenzhomologie dieser Enzyme zeigen die Eigenschaften der berechneten EP gravierende Unterschiede. Vergleicht man diese Oberflächenpotentiale der Cathepsine untereinander, kann man folgende Abstufung vornehmen: hCatL besitzt ein sehr ausgeprägtes elektronegatives Potential, das zu hCatB und hCatK stark abnimmt (vgl. Abb. 26a-c). hCatF besitzt ein schwach elektronegatives bzw. neutrales EP (Abb. 26f). Im Gegensatz dazu ist die molekulare Oberfläche von hCatS elektropositiv (Abb. 26e).

#### *Cathepsin L*

Zur Berechnung des elektrostatischen Potentials von hCatL wurde seine Röntgenkristallstruktur verwendet (PDB-Eintrag: 1CJL [118]). Im Vergleich zu den anderen hier aufgeführten Cathepsinen ist das elektrostatische Potential von hCatL über ausgedehnte Bereiche elektronegativ (Abb. 26a). Einige stark elektronegative Regionen befinden sich auf der molekularen Oberfläche dieses Ausschnittes des Enzyms, die sich durch Aminosäuren mit negativ geladenen Seitenketten ausbilden. Hervorzuheben ist der Bereich um Glu63 sowie von Asp162, da diese Aminosäuren Bestandteil der S3- und S1'-

Bindungspositionen sind. Diese Merkmale lassen die Vermutung zu, daß elektrostatische Interaktionen an diesen Positionen zwischen einem Substrat und hCatL eine wesentliche Rolle spielen. Weitere stark elektronegative Bereiche bilden sich durch die Ausrichtung der Seitenketten von Asp71 und Asp117 sowie von Asp160 zur Oberfläche des Enzyms aus. Allerdings sind diese Aminosäuren für die Substratspezifität von hCatL ohne Relevanz.

### *Cathepsin B*

Das Volumen des dargestellten Potentials von hCatB, das von seiner Röntgenkristallstruktur berechnet wurde (PDB-Eintrag: 1HUC [53]), ist im Vergleich zu den anderen Cathepsinen größer, da der Einschubloop dieses Enzyms mit abgebildet wurde (Abb. 26b). Dieser Loop legt sich über die S2'-Bindungsposition des Cathepsins und beeinflusst dadurch die Eigenschaften dieses Bereiches. Im Vergleich zu hCatL ist die Bindungsspalte von hCatB weniger elektronegat. Stärkere negative Potentialbereiche auf der Moleküloberfläche von hCatB, die für die Substratspezifität dieses Enzyms von Bedeutung sein könnten, lassen sich in der L-Domäne der Enzymoberfläche sowie in der R-Domäne lokalisieren. Hervorgerufen werden diese elektronegativen Bereiche durch die Seitenkette von Asp69 (S3-Bindungsposition) und durch Glu245 der S2-Bindungsposition von hCatB. Ein weiterer kleiner negativer Bereich bildet sich durch die Carbonylgruppe von Gly74 aus. Diese Gruppe richtet sich zur Bindungsspalte von hCatB aus.

### *Cathepsin K*

Das Potential der Bindungsspalte von hCatK (Abb. 26c) ist in seinen Eigenschaften ähnlich zu hCatB. Die Einschränkungen der S2-Bindungsposition dieses Enzyms in ihrer Größe lassen sich sehr gut erkennen (vgl. Abb. 26a-f). Die Gestalt dieser Bindungstasche wird durch die Ausrichtung des Phenylringes von Tyr67 sowie durch die Seitenkette von Leu209 beeinflusst. Stark elektronegative Potentialbereiche entstehen durch die Seitenketten von Asp61 (S3-Bindungsstelle) und dem benachbarten Glu59. Ein weiterer elektronegativer Bereich um Glu112 und Glu115 hat für die Charakteristiken der Bindungsregion von hCatK keine Bedeutung.

### *Cathepsin H*

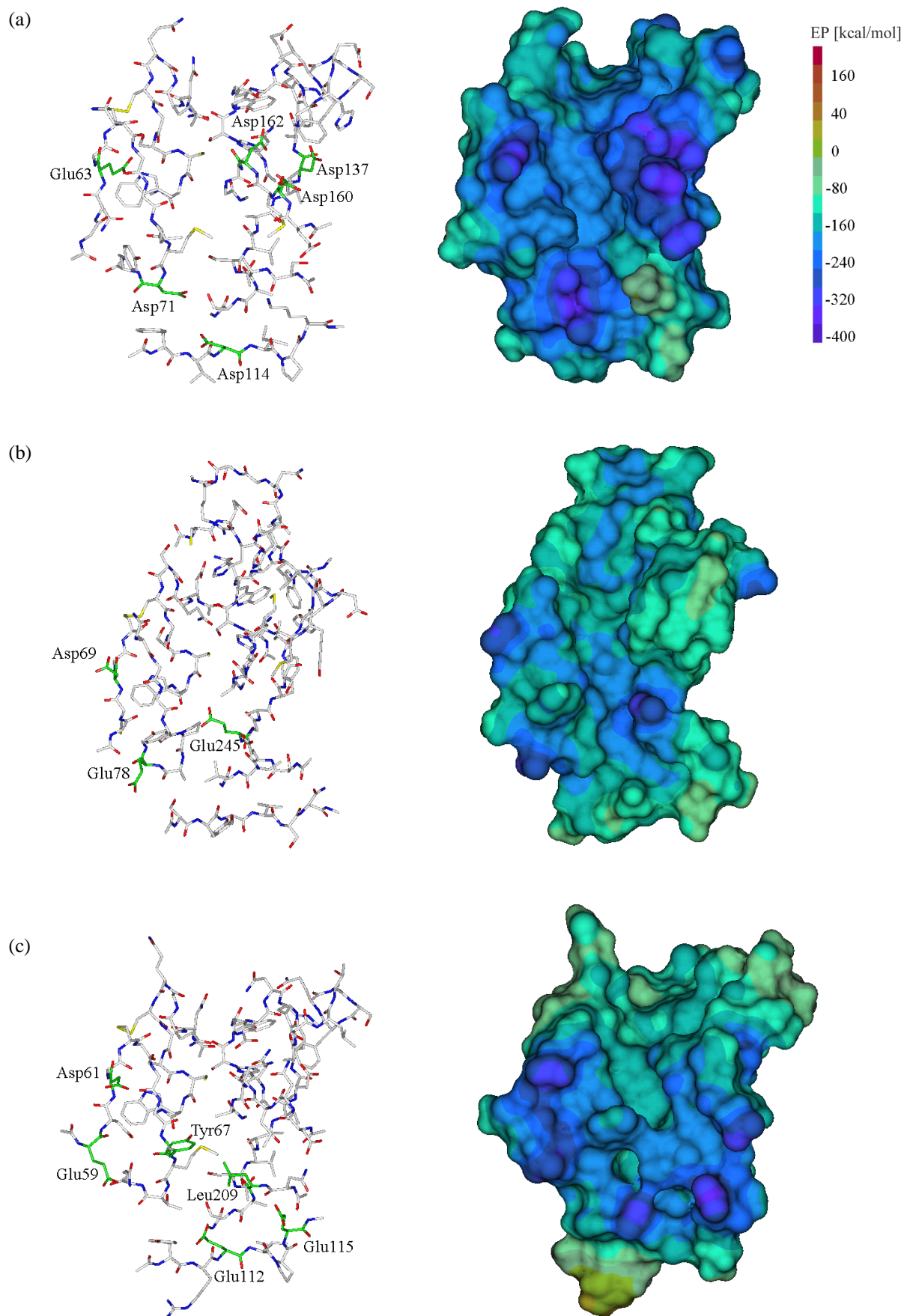
Das sichtbare elektrostatische Oberflächenpotential von hCatH, das nicht durch seine Minikette bedeckt wird, ist geringfügig elektronegat (Abb. 26d). Es zeichnet sich nicht durch besondere Merkmale, wie sie beispielsweise bei hCatL oder hCatK beobachtet werden, aus.

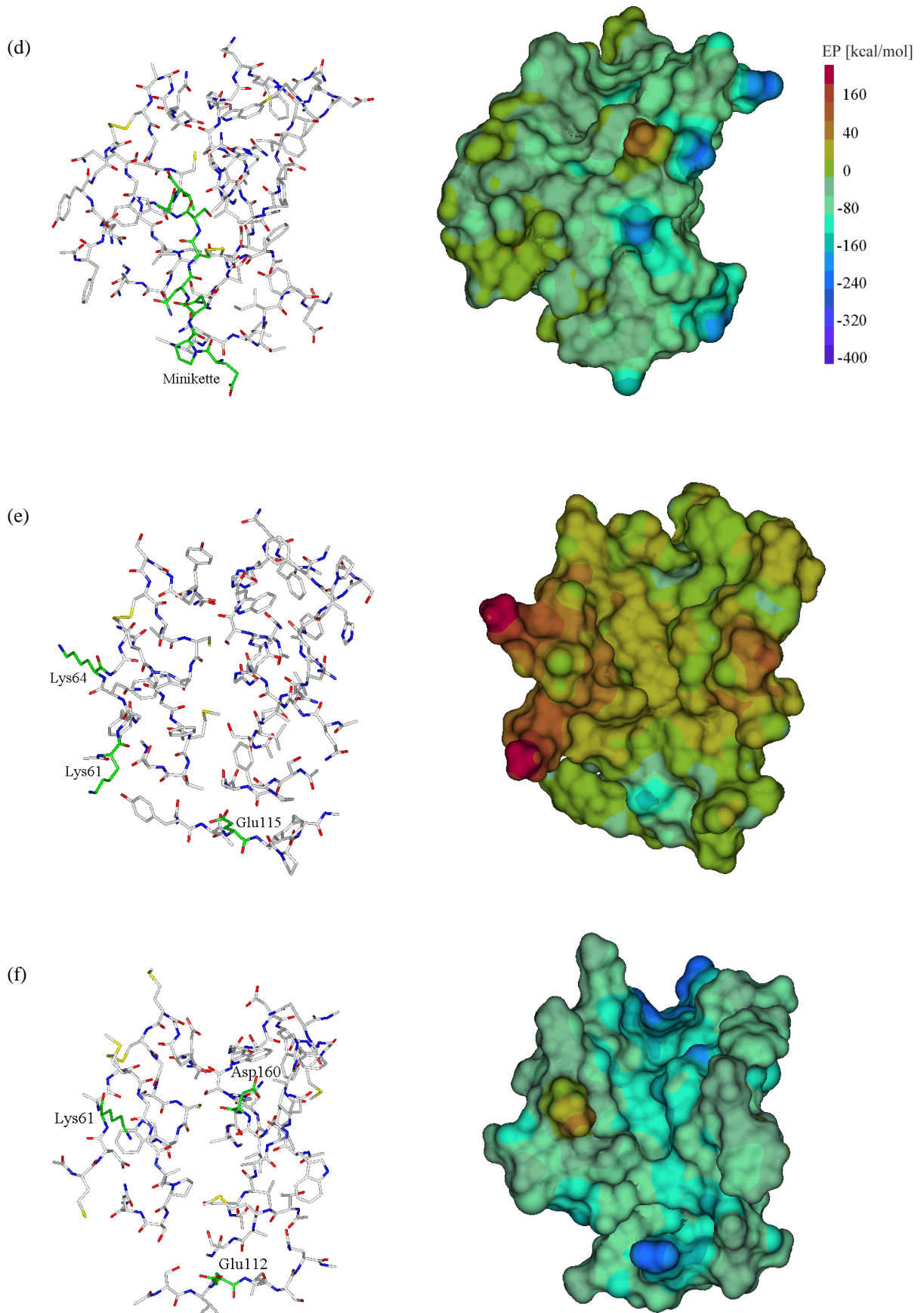
*Cathepsin S*

Im Gegensatz zu den bisher beschriebenen EP ist das Oberflächenpotential von hCatS elektropositiv. Wie anhand Abb. 26e zu sehen ist, beinhaltet die L-Domäne des hier dargestellten Potentials von hCatS mehrere positive Bereiche (Seitenketten von Lys60 und Lys64). Lys64 ist Bestandteil der Bindungsposition S3 von hCatS und es ist anzunehmen, daß diese Bindungsposition einen großen Einfluß auf die Substratspezifität dieses Proteins besitzt. Im unteren Bereich des hier abgebildeten Potentials tritt eine schwach elektronegative Stelle auf, die sich durch Glu115 ausbildet. Allerdings befindet sich diese Aminosäure außerhalb des katalytisch aktiven Bindungszentrums von hCatS.

*Cathepsin F*

Das elektrostatische Oberflächenpotential von hCatF ist schwach elektronegativ (Abb. 26f). Jedoch kann innerhalb der L-Domäne des sichtbaren Potentials auch ein elektropositiver Bereich beobachtet werden, welcher durch Lys61 der S3-Bindungsposition hervorgerufen wird. Ein stärkeres negatives Potential entsteht durch die Seitenkette Asp160 (S1'-Bindungsstelle), das durch die benachbarten Seitenketten teilweise verdeckt wird. Durch Glu112 bildet sich ein weiteres elektronegatives Potential von hCatF aus, das jedoch keine Bedeutung für das katalytisch aktive Bindungszentrum des Enzyms besitzt.





**Abb. 26** Molekülstrukturen und elektrostatische Potentiale der katalytisch aktiven Bindungszentren und deren Umgebung von hCatL (a), hCatB (b), hCatK (c), hCatH (d), hCatS (e) und hCatF (f). Die Lage der Strukturen korrespondiert mit ihren EP. Die für diese Potentiale entsprechende Farbskala zeigt die Abstufung der EP.

#### 4.3.1.2 Bestimmung bevorzugter Wechselwirkungen zwischen Ligand und Enzym

Um Aussagen über Wechselwirkungen zwischen einem möglichen Liganden und den Cathepsinen L, B, K, H, S und F zu erhalten, wurden Berechnungen mit dem Programm GRID durchgeführt [127] (siehe Abschnitt 3.2.1). Als Proben wurden die Methylgruppe, eine positiv geladene Ammoniumgruppe sowie eine negativ geladene Carboxylatgruppe verwendet. Um die berechneten GRID-Felder und die Anzahl der Interaktionspunkte der verwendeten Proben innerhalb der Cathepsine vergleichen zu können, wurden folgende Energiegrenzwerte verwendet: -10 kcal/mol für die Ammoniumprobe, -3.0 kcal/mol für die Carboxylatprobe und -0.7 kcal/mol für die Methylgruppenprobe. Das bedeutet, daß Energiewerte bzw. GRID-Felder oberhalb der angegebenen Werte nicht dargestellt wurden. Auf diese Weise wurden die Substratbindungsregionen der Cathepsine L, B, K, H, S und F untersucht. Vergleicht man die berechneten elektrostatischen Oberflächenpotentiale dieser Cathepsine (Abb. 26a-f) mit den Ergebnissen der GRID-Berechnungen (Abb. 27a-f, Seite 83), lassen sich viele Parallelen aufzeigen. Vor allem werden die unterschiedlichen Eigenschaften der Cathepsine bezüglich ihrer Potentialverteilung gut wiedergegeben. Die Cathepsine L, B, K und H sind in der Lage mit einer Ammoniumprobe zu interagieren (Abstufung der Stärke dieser Wechselwirkung:  $hCatL \approx hCatB > hCatK \approx hCatH$ ) (vgl. Abb. 27a-d). Diese Eigenschaften konnten für die Cathepsine S und F nicht beobachtet werden (Abb. 27e,f). Dahingegen wurden für die Bindungsspalte von hCatS ausgedehnte Interaktionen zu einer Carboxylatprobe beobachtet, was für einen elektropositiven Charakter dieser Region spricht (Abb. 27e).

Wird bei den GRID-Berechnungen als Probe eine Methylgruppe verwendet, kann die Hydrophobizität dieser Enzyme untersucht werden. Von großer Relevanz ist hierbei die S2-Bindungsposition dieser Cathepsine, da diese Region eine wesentliche Bedeutung für die Substratspezifität dieser Proteasen besitzt. Anhand der Darstellung der berechneten hydrophoben GRID-Felder (Methylgruppenprobe) ist zu sehen, daß sich nur für die Cathepsine L, S und F größere Bereiche hydrophober Wechselwirkungen ausbilden können, die im Vergleich zu den Cathepsinen K und B weiter in die Bindungsspalte reinragen (Abb. 27a-c,e,f). Die Ursachen hierfür sollen später näher erläutert werden.

Die berechneten Wechselwirkungsfelder sowie die lokalen Energieminima zwischen den Aminosäuren des Enzyms und der Probe können dazu dienen, bevorzugte Wechselwirkungen zwischen dem Enzym und einem möglichen Liganden aufzufinden. Aus diesem Grund sollen die resultierenden Wechselwirkungen zwischen den unterschiedlichen Proben und den Cathepsinen L, B, K, S, H und F im folgenden diskutiert werden.

##### *Cathepsin L*

In Abb. 27a ist das katalytisch aktive Bindungszentrum von hCatL und die GRID-Felder, welche aus den Berechnungen an diesem Enzym mit einer Methylgruppenprobe und einer Ammoniumprobe

resultieren, dargestellt. Lokale Energieminima zur hydrophoben Gruppe werden zu den Seitenketten von Trp189 und Trp193 gefunden. Wesentlich sind die Punkte und die sich daraus bildenden Bereiche zu den Aminosäuren der S2-Bindungsposition von hCatL (Leu69, Met70, Ala214), da diese Bindungsstelle für die Spezifität eines Liganden gegenüber hCatL von Bedeutung ist. Ein weiteres Feld bildet sich zwischen den Seitenketten von Leu69 und Tyr72 aus. Betrachtet man die Wechselwirkungen zwischen der Ammoniumprobe und hCatL, ist zu sehen, daß sich zu den sauren Seitenketten von Asp162 (S1'-Bindungsposition) und Glu63 (S3-Bindungsposition) stabile Wechselwirkungen ausbilden, die zum Ligandendesign für dieses Cathepsin ausgenutzt werden können.

Anhand dieser Ergebnisse ist es möglich, die Seitenketten eines Liganden im Enzym so anzuordnen, daß diese die GRID-Felder belegen und so optimal mit den Aminosäuren des Enzyms interagieren können.

### *Cathepsin B*

Im Vergleich zu Cathepsin L treten in der S2-Bindungsposition von hCatB weniger Wechselwirkungspunkte zu der Methylgruppenprobe auf (Abb. 27b). Die Felder, die sich zu den Aromaten von Tyr75 ausbilden, werden deutlich durch die Seitenkette von Glu245 begrenzt. Für Glu245 resultieren lokale Minima mit der Ammoniumprobe, die sich um die Carboxylatgruppe der Seitenkette von Glu245 orientieren. Es ist zu erwarten, daß für die Bindungsstelle elektrostatische Wechselwirkungen favorisiert werden. Weitere Interaktionen zwischen der Ammoniumprobe und hCatB können durch die Ausrichtung des Carbonyl-Sauerstoffatoms von Gly27, der Seitenkette von Asn72 (S1-Bindungsposition) sowie zu Asp69 (S3-Bindungsstelle) beobachtet werden.

Zu einigen Aminosäuren der S1'-Bindungsposition (Trp221, Leu181, Val176) können sich hydrophobe Wechselwirkungen ausbilden. Allerdings wird die hydrophobe Fläche der Seitenkette von Trp221 durch His110 und His111 (Einschubloop) teilweise verdeckt, so daß diese Wechselwirkung eingeschränkt wird. Im Gegensatz dazu treten zwischen den beiden Imidazolringen der Histidinseitenketten Wechselwirkungspunkte zu der Carboxylatprobe auf. Zu der Seitenkette von Gln23 werden weitere Wechselwirkungen zu der negativ geladenen Probe beobachtet.

### *Cathepsin K*

Abb. 27c zeigt die Ergebnisse der GRID-Berechnungen an hCatK. Vergleicht man die hydrophoben Felder in der S2-Bindungsposition mit denen von hCatL, ist eine Verkleinerung der Bereiche zu erkennen, die unmittelbar in dieser Bindungsstelle liegen. Ein größeres Feld wird nur im unteren Bereich der Bindungsspalte beobachtet. Es bildet sich durch die Seitenketten von Tyr67 und Leu209 aus. Da die S2-Bindungsposition allerdings durch die Ausrichtung anderer Aminosäuren eingeschränkt wird (Abschnitt 4.3.1.1), ist es ungewiß, ob diese hydrophoben Eigenschaften für

Liganden von Cathepsin K ausgenutzt werden können. Im Gegensatz dazu sind die Wechselwirkungspunkte, die sich unter Verwendung der Ammoniumprobe zu Asp61, Asn60 und Glu69 ausbilden, von größerer Bedeutung, da diese Merkmale zur Entwicklung von neuen Liganden für hCatK ausgenutzt werden können. Weiterhin spielen auch die berechneten GRID-Felder zwischen der Carboxylatprobe und hCatK eine wichtige Rolle. Unter Verwendung dieser Gruppe treten Wechselwirkungen vor allem zu Aminosäuren der S1'-Bindungsposition auf (Gln143), die mit Ausnahme von hCatF für die anderen Cathepsine nicht beobachtet werden konnten (vgl. Abb. 27a-f). Unter Verwendung der hier beschriebenen bevorzugten Wechselwirkungen sollte es möglich sein, neue Peptidsequenzen zu erstellen, die im Vergleich zu verwandten catheptischen Cysteinproteasen eine hohe Selektivität gegenüber hCatK aufweisen.

### *Cathepsin H*

Wie anhand der Felder und der Wechselwirkungspunkte zu sehen ist, bilden sich hydrophobe Wechselwirkungen zwischen der Methylgruppenprobe und hCatH zu den Seitenketten von Trp188 und Met145 aus (Abb. 27d). Die Ammoniumprobe kann mit dem negativ geladenen C-Terminus der Minikette (Thr83P) sowie mit den Seitenketten von Gln20 und Asn165 interagieren. Basierend auf diesen Ergebnissen ist daher zu erwarten, daß überwiegend elektrostatische Wechselwirkungen für die Selektivität eines Liganden von hCatH von Bedeutung sind.

### *Cathepsin S*

In Abb. 27e ist das katalytisch aktive Bindungszentrum von hCatS mit den GRID-Feldern einer Methylgruppenprobe und einer Carboxylatprobe dargestellt. Lokale Energieminima zur hydrophoben Gruppe werden zu den Seitenketten von Trp186 und Phe146 (S1'-Bindungsposition) gefunden. Für ein effektives Ligandendesign sind vor allem die Punkte und die sich daraus bildenden Bereiche zu den Aminosäuren der S2-Bindungsposition von hCatS von großer Relevanz (Phe70, Val162, Phe211). Die Größe und Lage dieser hydrophoben Regionen sind ähnlich mit denen von hCatL (Abb. 27a und e).

Die berechneten Wechselwirkungsfelder von hCatS mit einer Carboxylatprobe unterstreichen die Aussagen über das elektropositive Potential dieser Protease (Abschnitt 4.3.1.1). Über ausgedehnte Bereiche können Wechselwirkungen zwischen dieser Probe und hCatS beobachtet werden.

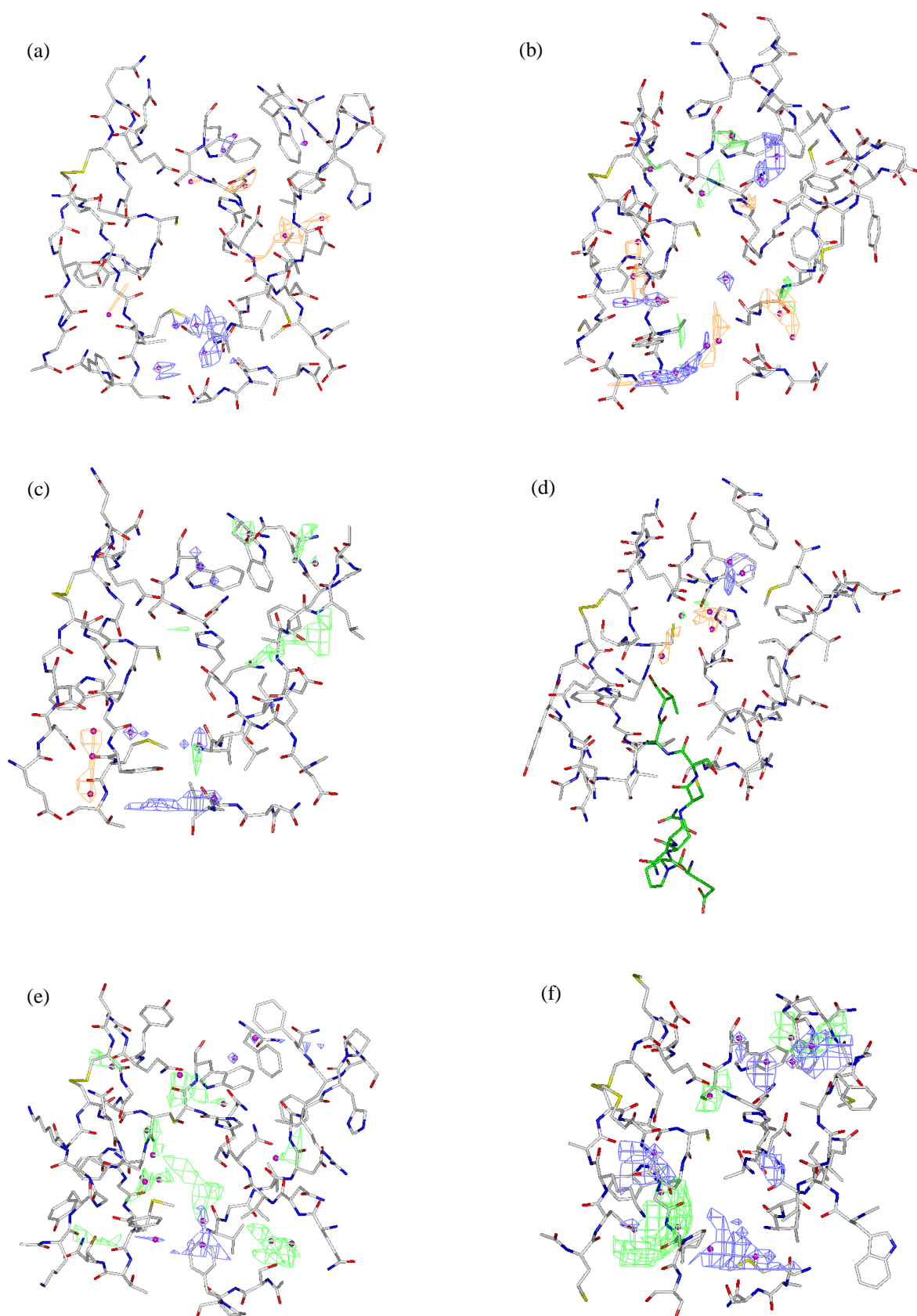
### *Cathepsin F*

Die bevorzugten Wechselwirkungen zwischen einer Methylgruppenprobe bzw. der Carboxylatprobe sind in Abb. 27f dargestellt. Anhand dieser Abbildung ist zu sehen, daß hCatF im Vergleich zu den anderen Cathepsinen ein leicht differenziertes Wechselwirkungsverhalten zu diesen Gruppen aufweist.



Die sich ausbildenden hydrophoben Wechselwirkungen zu Aminosäuren der S2-Bindungsposition (Leu67, Pro68, Ile159, Met207) erstrecken sich im Vergleich zu den Cathepsinen L, K und S über größere Bereiche. Aufgrund der größeren hydrophoben Bindungstasche ist zu erwarten, daß die S2-Bindungsstelle von hCatF sterisch anspruchsvollere Aminosäuren aufnehmen kann. Weitere Wechselwirkungspunkte zwischen der Methylgruppe und hCatF bilden sich zu Trp183 und Ala136. Ein ausgedehntes GRID-Feld liegt zwischen diesen beiden Aminosäuren. Wird für diese Berechnungen die Carboxylatprobe verwendet, treten mehrere lokalen Minima für die sich bildenden Wechselwirkungen auf (Seitenkette von Gln19, Gln140). Hervorzuheben sind die Bereiche um Lys61 (S3-Bindungsposition). Die Protonen der Aminogruppe der Seitenkette von Lys61 werden von den GRID-Feldern umschlossen (Abb. 27f). Zwischen den Methylgruppen dieser Seitenkette und der Methylgruppe bildet sich ein weiteres Feld hydrophober Wechselwirkungen aus.

Basierend auf den Ergebnissen, die unter Verwendung des GRID-Programms erhalten wurden, ist es möglich, günstige Wechselwirkungspartner für die Aminosäuren der einzelnen Substratbindungsregionen zu ermitteln. Die Konformation der Aminosäureseitenketten der Liganden kann so variiert werden, daß diese Seitenketten zu den jeweiligen GRID-Feldern positioniert sind.



**Abb. 27** Darstellung der bevorzugten Wechselwirkungen zwischen einer Methylgruppenprobe (blau), Carboxylatprobe (grün), Ammoniumprobe (orange) und den katalytisch aktiven Bindungszentren der Cathepsine L (a), B (b), K (c), H (d), S (e) und F (f). Die magentafarbenen Punkte stellen jeweils die lokalen Minima der Wechselwirkungen zwischen der Probe und dem Enzym dar. Für Cathepsin H wurde seine Minikette (grün) mit abgebildet.

### 4.3.2 Erstellung neuer Peptidsequenzen für hochaffine Cathepsin L-Substrate

Zum strukturbasierten Design von neuen Substraten für Cathepsin L wurde die Röntgenkristallstruktur dieses Enzyms verwendet (PDB-Eintrag: 1CJL [118]). An dieser Proteinstruktur wurden eine Reihe von Modifizierungen vorgenommen, die im Abschnitt 4.2.3.4 beschrieben sind. Nach der Untersuchung des katalytisch aktiven Bindungszentrums von Cathepsin L (MOLCAD, GRID) und anhand des Vergleiches mit verwandten Enzym-Inhibitor-Komplexen, ergibt sich für die einzelnen Pn-Positionen eines Substrates folgende Einteilung. In Analogie zu verwandten Papain-ähnlichen Cysteinproteasen bevorzugt ein Substrat in P1-Position ein Arginin bzw. Lysin. Betrachtet man das berechnete EP von hCatL (Abb. 26a), wird ersichtlich, daß sich elektrostatische Wechselwirkungen zwischen dem P1-Rest und den Aminosäuren der S1-Bindungsposition ausbilden können. Anhand Tab. 15 sowie Abb. 27a ist zu sehen, daß die S2-Bindungstasche von hCatL hydrophoben Charakter aufweist. Aufgrund dieser Eigenschaften und der Größe der S2-Bindungstasche von hCatL, werden in P2-Position eines Liganden Aminosäuren mit sterisch anspruchsvollen Resten in der hier angegebenen Reihenfolge Phenylalanin > Leucin > Valin bevorzugt [38,150]. Die S3-Bindungsposition von hCatL enthält unter anderem eine hydrophobe Aminosäure (Leu69). Daher wurde in P3-Position eines Substrates Leucin und Phenylalanin verwendet. In der S1'-Bindungsposition von hCatL werden Aminosäuren mit kurzen Seitenketten (Alanin, Serin) bzw. mit neutralen, hydrophilen Seitenketten (Asparagin, Glutamin) bevorzugt [18]. Hervorzuheben ist, daß hCatL neben hCatF als einziges Cathepsin dieser Enzymklasse in dieser Bindungsstelle ein Aspartat (Asp162) enthält (Abb. 27a). Es ist daher anzunehmen, daß für die P1'-Position ein Lysin bzw. ein Arginin sehr gut geeignet ist. Verwendet wurde Lysin, da Lysinderivate leichter synthetisch zugänglich sind als die entsprechenden Argininderivate.

Es wurden deshalb zur Erstellung von Peptidsequenzen folgende Aminosäuren für die jeweiligen Pn-Positionen ausgewählt:

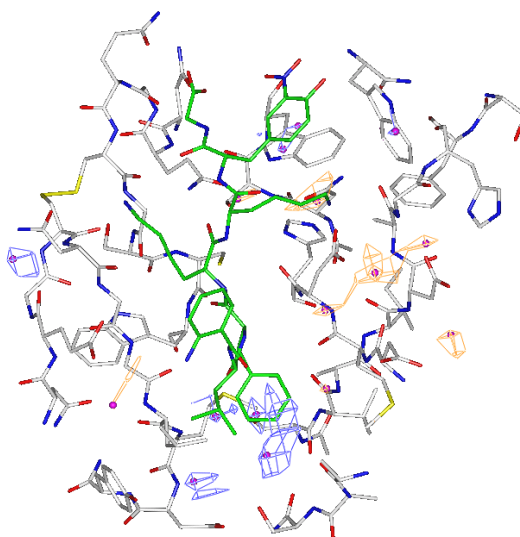
<b>P3</b>	-	<b>P2</b>	-	<b>P1</b>	-	<b>P1'</b>		<b>P2</b>	-	<b>P1</b>	-	<b>P1'</b>
Leu		Phe		Arg		Lys		Leu		Arg		Lys
Phe		Leu		Lys		Ser		Phe		Lys		Ser

Daraus resultieren zahlreiche Peptide, die in ihrer Bindungsaffinität zu hCatL untersucht werden sollten (Tab. 16). Zum Nachweis der enzymatischen Aktivität dieser Liganden wurden 2-Aminobenzoesäure (Abz) und 3-Nitrotyrosin (Tyr(3-NO<sub>2</sub>)) als Endgruppen verwendet. Es entstehen dadurch Quenchingsubstrate, wobei Abz den fluoreszenten Donor darstellt und 3-Nitrotyrosin als Akzeptor (Quencher) fungiert [151,152].

**Tab. 16** Substrate für hCatL

Nr.	Hexapeptide	Nr.	Pentapeptide
	P3 - P2 - P1 - P1'		P2 - P1 - P1'
(9)	Abz-Leu-Phe-Lys-Lys-Tyr(3-NO <sub>2</sub> )-Gly	(17)	Abz-Phe-Arg-Lys-Tyr(3-NO <sub>2</sub> )-Gly
(10)	Abz-Leu-Phe-Arg-Lys-Tyr(3-NO <sub>2</sub> )-Gly	(18)	Abz-Phe-Lys-Lys-Tyr(3-NO <sub>2</sub> )-Gly
(11)	Abz-Phe-Phe-Lys-Lys-Tyr(3-NO <sub>2</sub> )-Gly	(19)	Abz-Leu-Lys-Lys-Tyr(3-NO <sub>2</sub> )-Gly
(12)	Abz-Phe-Phe-Arg-Lys-Tyr(3-NO <sub>2</sub> )-Gly	(20)	Abz-Phe-Arg-Ser-Tyr(3-NO <sub>2</sub> )-Gly
(13)	Abz-Leu-Leu-Arg-Lys-Tyr(3-NO <sub>2</sub> )-Gly		
(14)	Abz-Phe-Leu-Arg-Lys-Tyr(3-NO <sub>2</sub> )-Gly		
(15)	Abz-Leu-Phe-Arg-Ser-Tyr(3-NO <sub>2</sub> )-Gly		
(16)	Abz-Phe-Leu-Arg-Ser-Tyr(3-NO <sub>2</sub> )-Gly		

Diese Peptide wurden in die Bindungsspalte von hCatL eingepaßt. Die Lage der Hauptkettenatome wurde an die Position der Atome von Liganden in bekannten Enzym-Inhibitor-Komplexen angeglichen (Abschnitt 3.2.2). Um die terminalen Gruppen sowie die Aminosäureseitenketten optimal in die einzelnen Bindungspositionen zu orientieren, wurde das katalytisch aktive Bindungszentrum von hCatL bezüglich seiner elektrostatischen und hydrophoben Eigenschaften unter Verwendung des Programms GRID abgetastet (Abschnitt 4.3.1.2). Die Anordnung der Seitenketten des Liganden wurde so variiert, daß sie optimal in den entsprechenden GRID-Feldern positioniert waren (Abb. 28). Über diese Vorgehensweise wurden die Substrate in die Bindungsspalte von hCatL vorpositioniert.



**Abb. 28** Darstellung des Enzym-Ligand-Komplexes von hCatL mit Abz-Leu-Phe-Lys-Lys-Tyr(3-NO<sub>2</sub>)-Gly (9) (grün). Die berechneten GRID-Felder mit einer Methylgruppenprobe sind blau gefärbt und die einer Ammoniumprobe orangefarben. Die magentafarbenen Punkte sind die Positionen, an denen die günstigsten Wechselwirkungen zwischen den Proben und dem Enzym berechnet wurden.

Unter Verwendung des Programms FLEXIDOCK wurden verschiedene Anordnungsmöglichkeiten dieser Substrate in hCatL berechnet. Die resultierenden Komplexe wurden minimiert. In Tab. 17 sind die berechneten Wechselwirkungsenergien sowie die mit SCORE berechneten K<sub>d</sub>-Werte aufgelistet.

**Tab. 17** Bindungsaffinität der Liganden zu hCatL

Nr.	Peptid	$\Delta E$ [kcal/mol] <sup>a</sup>	$K_d$ [M] <sup>b</sup>
Hexapeptide			
P3 - P2 - P1 - P1'			
(9)	Abz-Leu-Phe-Lys-Lys-Tyr(3-NO <sub>2</sub> )-Gly	-100.79	4.39·10 <sup>-8</sup>
(10)	Abz-Leu-Phe-Arg-Lys-Tyr(3-NO <sub>2</sub> )-Gly	-98.23	5.01·10 <sup>-8</sup>
(11)	Abz-Phe-Phe-Lys-Lys-Tyr(3-NO <sub>2</sub> )-Gly	-97.18	5.91·10 <sup>-8</sup>
(12)	Abz-Phe-Phe-Arg-Lys-Tyr(3-NO <sub>2</sub> )-Gly	-91.04	1.78·10 <sup>-7</sup>
(13)	Abz-Leu-Leu-Arg-Lys-Tyr(3-NO <sub>2</sub> )-Gly	-88.58	2.09·10 <sup>-6</sup>
(14)	Abz-Phe-Leu-Arg-Lys-Tyr(3-NO <sub>2</sub> )-Gly	-84.33	8.41·10 <sup>-6</sup>
(15)	Abz-Leu-Phe-Arg-Ser-Tyr(3-NO <sub>2</sub> )-Gly	-82.71	1.66·10 <sup>-5</sup>
(16)	Abz-Phe-Leu-Arg-Ser-Tyr(3-NO <sub>2</sub> )-Gly	-80.14	8.32·10 <sup>-5</sup>
Pentapeptide			
P2 - P1 - P1'			
(17)	Abz-Phe-Arg-Lys-Tyr(3-NO <sub>2</sub> )-Gly	-92.58	1.05·10 <sup>-7</sup>
(18)	Abz-Phe-Lys-Lys-Tyr(3-NO <sub>2</sub> )-Gly	-89.86	9.71·10 <sup>-7</sup>
(19)	Abz-Leu-Lys-Lys-Tyr(3-NO <sub>2</sub> )-Gly	-85.75	6.26·10 <sup>-6</sup>
(20)	Abz-Phe-Arg-Ser-Tyr(3-NO <sub>2</sub> )-Gly	-82.04	5.08·10 <sup>-6</sup>

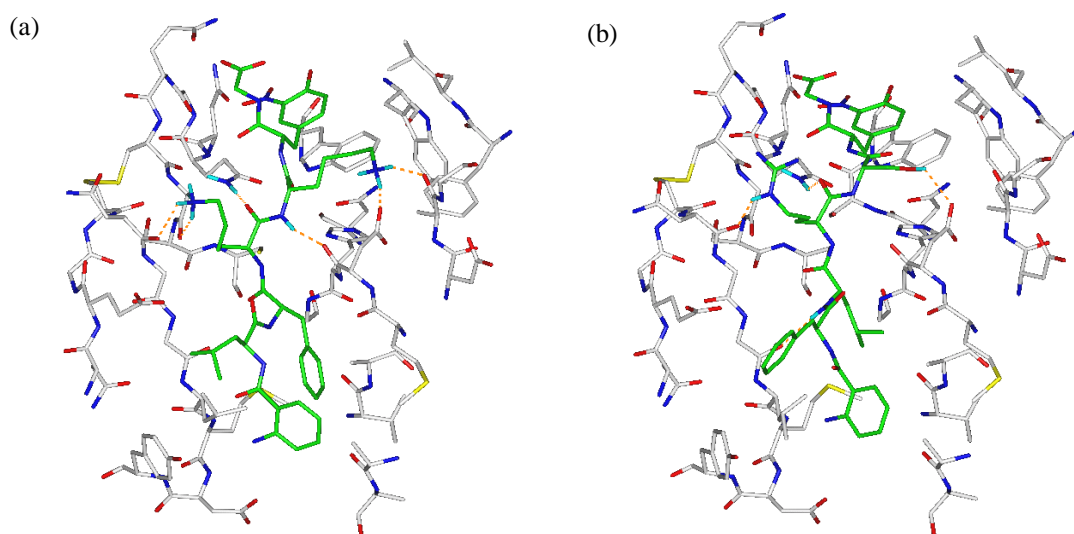
<sup>a</sup> berechnete Wechselwirkungsenergien

<sup>b</sup>  $K_d$ -Werte, berechnet unter Verwendung von SCORE

Die Struktur von hCatL der Enzym-Substrat-Komplexe korrespondiert mit der Proteinstruktur des nativen Cathepsins. Die RMS-Differenzen zwischen hCatL und den Enzym-Substrat-Komplex (9), berechnet über die Hauptkettenatome des Enzyms sowie über die Seitenketten betragen 0.89 Å bzw. 1.98 Å. Die größten Unterschiede treten für die Aminosäureseitenketten Gln19, Leu69 und Asp162 des Enzyms auf, da sich diese Seitenketten für eine optimale Wechselwirkung mit den Ligandenseitenketten ausrichten.

Der Austausch der Aminosäuren in der Peptidsequenz der Liganden bewirkt eine zum Teil starke Veränderung der Affinität des Liganden zum Enzym (Tab. 17). Vergleicht man die berechneten Energiewerte der sich bildenden Enzym-Ligand-Komplexe untereinander, kann man folgende Schlußfolgerungen ziehen. Substrate mit einem Phenylalanin in P2-Position bilden aufgrund der Orientierung der Seitenkette eine stärkere Affinität zum Enzym aus, als solche mit einem Leucin in dieser Position. Lysin ist in P1'-Position gegenüber einem Serin deutlich bevorzugt, was durch die Ausbildung von stabileren Wasserstoffbrückenbindungen zwischen der Lysinseitenkette und dem Enzym begründet werden kann. Im Gegensatz dazu läßt sich diese Abstufung für den P1-Rest (Lysin bzw. Arginin) sowie den P3-Rest (Leucin bzw. Phenylalanin) nicht treffen.

Die unterschiedliche Affinität beruht auf einem variierten Wechselwirkungsverhalten dieser Substrate mit hCatL und soll nachfolgend anhand von zwei Beispielen beschrieben werden. Wird das Peptid Abz-Leu-Phe-Lys-Lys-Tyr(3-NO<sub>2</sub>)-Gly (9) in die Bindungsspalte von hCatL gedockt, richten sich die Seitenketten des Liganden so aus, daß sich zahlreiche Wechselwirkungen zwischen dem Liganden und Enzym ausbilden können (Abb. 29a).



**Abb. 29** Enzym-Ligand-Komplexe von hCatL mit Abz-Leu-Phe-Lys-Lys-Tyr(3-NO<sub>2</sub>)-Gly (9) (a) sowie mit Abz-Phe-Leu-Arg-Ser-Tyr(3-NO<sub>2</sub>)-Gly (16) (b). Die Kohlenstoffatome des Inhibitors sind grün gefärbt. Von hCatL ist nur das katalytisch aktive Bindungszentrum abgebildet.

Mehrere Wasserstoffbrücken wurden zwischen dem Peptid und dem Enzym beobachtet (Abb. 29a). Die Salzbrücke zwischen der Lysinseitenkette in P1'-Position und Asp162 von hCatL hat dabei wesentlichen Einfluß auf die Stabilität des Komplexes. Die Seitenkette des Lysins in P1-Position liegt zwischen den beiden Proteinketten von hCatL, welche die S1-Bindungsstelle bilden. Es bilden sich Wasserstoffbrücken zu Hauptkettenatomen des Proteins aus (Asn66, Gly23). Der Phenylring des P2-Restes (Phenylalanin) liegt optimal in der S2-Bindungstasche von hCatL, wodurch sich hydrophobe Wechselwirkungen zwischen dem P2-Rest und den Seitenketten der S2-Bindungstasche entwickeln können (Ala214, Ala135, Leu69, Met70). Außerdem bilden sich hydrophobe Wechselwirkungen zwischen dem Leucin in P3-Position und Leu69 des Enzyms aus.

Im Vergleich dazu ist die Bindungsaffinität des Liganden Abz-Phe-Leu-Arg-Ser-Tyr(3-NO<sub>2</sub>)-Gly (16) zum Enzym geringer (Abb. 29b). Obwohl nur geringe Differenzen in der Struktur der Peptidrückgrate auftreten, werden die intermolekularen Wechselwirkungen zwischen den Seitenketten des Liganden und hCatL abgeschwächt. Durch den Austausch des Lysins in P1'-Position gegen Serin ist die Stärke der Wasserstoffbrückenbindung reduziert (Abb. 29b). Die Lage des P1-Restes Arginin ist ähnlich mit der des Lysins (P1-Position) des vorher beschriebenen Liganden. Die Seitenkette des P2-Restes Leucin ist in die S2-Bindungstasche des Enzyms ausgerichtet. Allerdings sind aufgrund des geringeren Volumens der Leucinseitenkette die auftretenden hydrophoben Wechselwirkungen zum Enzym abgeschwächt.

Zur Untersuchung der Selektivität dieser Substrate gegenüber Cathepsin L wurden diese Peptide in die Bindungszentren von hCatK, hCatB und hCatS gedockt. Dabei wurde die gleiche Vorgehensweise angewendet, wie sie oben beschrieben wurde. Für hCatB wurden die Koordinaten seiner Röntgenkristallstruktur verwendet (PDB-Eintrag: 1HUC [53]) und für die Cathepsine K und S die der

ermittelten Tertiärstrukturmodelle (siehe Abschnitt 4.2.1 und 4.2.2). Eine Gegenüberstellung der berechneten Wechselwirkungsenergien ist in Tab. 18 zu finden. Anhand dieser Werte ist zu sehen, daß die Substrate im Vergleich zu den anderen Cathepsinen eine deutlich höhere Affinität zu Cathepsin L besitzen. Die größten Unterschiede für die Wechselwirkungsenergie können für das Substrat Abz-Phe-Phe-Lys-Lys-Tyr(3-NO<sub>2</sub>)-Gly (11) beobachtet werden (Tab. 18).

**Tab. 18** Bindungsaffinität der Liganden zu den Cathepsinen L, K, B und S

Peptid	hCatL		hCatK		hCatB		hCatS	
	$\Delta E$ [kcal/mol] <sup>a</sup>	$K_d$ [M] <sup>b</sup>	$\Delta E$ [kcal/mol] <sup>a</sup>	$K_d$ [M] <sup>b</sup>	$\Delta E$ [kcal/mol] <sup>a</sup>	$K_d$ [M] <sup>b</sup>	$\Delta E$ [kcal/mol] <sup>a</sup>	$K_d$ [M] <sup>b</sup>
Hexapeptide								
(9)	-100.79	$4.39 \cdot 10^{-8}$	-92.10	$3.47 \cdot 10^{-7}$	-85.78	$1.45 \cdot 10^{-6}$	-65.91	$2.99 \cdot 10^{-4}$
(10)	-98.23	$5.01 \cdot 10^{-8}$	-89.31	$9.01 \cdot 10^{-7}$	-82.69	$2.98 \cdot 10^{-5}$	-69.14	$1.71 \cdot 10^{-4}$
(11)	-97.18	$5.91 \cdot 10^{-8}$	-84.58	$8.75 \cdot 10^{-6}$	-78.63	$7.63 \cdot 10^{-5}$	-68.49	$1.91 \cdot 10^{-4}$
(12)	-91.04	$1.78 \cdot 10^{-7}$	-86.26	$5.09 \cdot 10^{-6}$	-84.54	$2.24 \cdot 10^{-5}$	-70.44	$8.02 \cdot 10^{-5}$
(13)	-88.58	$2.09 \cdot 10^{-6}$	-82.42	$2.95 \cdot 10^{-5}$	-80.03	$5.18 \cdot 10^{-5}$	-65.89	$3.43 \cdot 10^{-4}$
(14)	-84.33	$8.41 \cdot 10^{-6}$	-83.46	$1.86 \cdot 10^{-6}$	-83.32	$2.03 \cdot 10^{-6}$	-71.39	$8.47 \cdot 10^{-5}$
(15)	-82.71	$1.66 \cdot 10^{-5}$	-80.98	$9.71 \cdot 10^{-5}$	-76.24	$6.62 \cdot 10^{-5}$	-58.43	$4.97 \cdot 10^{-4}$
(16)	-80.14	$8.32 \cdot 10^{-5}$	-78.36	$8.34 \cdot 10^{-5}$	-75.86	$6.43 \cdot 10^{-5}$	-60.34	$4.23 \cdot 10^{-4}$
Pentapeptide								
(17)	-92.58	$1.05 \cdot 10^{-7}$	-82.24	$3.16 \cdot 10^{-5}$	-70.73	$9.74 \cdot 10^{-5}$	-61.09	$5.17 \cdot 10^{-4}$
(18)	-89.86	$9.71 \cdot 10^{-7}$	-82.63	$4.64 \cdot 10^{-5}$	-73.24	$9.13 \cdot 10^{-5}$	-57.84	$6.32 \cdot 10^{-4}$
(19)	-85.75	$6.26 \cdot 10^{-6}$	-86.41	$6.53 \cdot 10^{-6}$	-72.20	$8.18 \cdot 10^{-5}$	-57.63	$6.30 \cdot 10^{-4}$
(20)	-82.04	$5.08 \cdot 10^{-6}$	-78.33	$9.16 \cdot 10^{-5}$	-61.45	$8.35 \cdot 10^{-4}$	-58.48	$6.97 \cdot 10^{-4}$

<sup>a</sup> berechnete Wechselwirkungsenergien

<sup>b</sup> Bindungskonstanten, berechnet unter Verwendung von SCORE

Hervorgerufen werden diese Unterschiede durch die variierten Eigenschaften der einzelnen Bindungspositionen der Cathepsine. Im Gegensatz zu hCatL enthalten die Cathepsine K, B und S in ihrer S1'-Bindungsposition ein Asparagin bzw. Glycin. Dadurch ist die Möglichkeit zur Wasserstoffbrückenbindung zwischen dem P1-Rest Lysin und der Aminosäure in der S1'-Bindungsposition reduziert bzw. unterbunden. Außerdem ist im Vergleich zu hCatL die Größe der S2-Bindungstasche von hCatK eingeschränkt, so daß Wechselwirkungen zu einem sterisch anspruchsvollen Rest (z. B. Phenylalanin) abgeschwächt werden. Die S2-Bindungsposition von Cathepsin B enthält ein Glutamat, wodurch die Affinität von Phenylalanin in P2-Position geringer ist. Basierend auf diesen Ergebnissen wurde das Substrat Abz-Phe-Phe-Lys-Lys-Tyr(3-NO<sub>2</sub>)-Gly (11) im AK Prof. K. Neubert (Universität Halle-Wittenberg) von Herrn Dr. J. Faust synthetisiert. Die enzymkinetischen Studien an diesem Substrat mit den Cathepsinen L, K und B wurden von Herrn Dr. F. Bühling (Universität Magdeburg) durchgeführt. Angaben zur experimentellen Durchführung befinden sich im Anhang (Seite A-15).

### 4.3.2.1 Enzymkinetische Untersuchungen

Die ermittelten kinetischen Daten der proteolytischen Aktivität von Abz-Phe-Phe-Lys-Lys-Tyr(3-NO<sub>2</sub>)-Gly (11) sind in Tab. 19 aufgelistet. Von Cathepsin K wird dieses Substrat nicht hydrolysiert.

**Tab. 19.** Kinetische Daten der Hydrolyse des Substrates Abz-Phe-Phe-Lys-Lys-Tyr(3-NO<sub>2</sub>)-Gly (11) durch die Cathepsine L und B

Enzym	K <sub>m</sub> [M]	k <sub>cat</sub> [1/s]	k <sub>cat</sub> /K <sub>m</sub> [1/M·s]
Cathepsin L	7.6·10 <sup>-8</sup>	0.0322	4.234·10 <sup>5</sup>
Cathepsin B	1.4·10 <sup>-5</sup>	0.6135	4.35·10 <sup>4</sup>

Wie anhand der K<sub>m</sub>-Werte zu sehen ist, besitzt das Substrat im Vergleich zu Cathepsin B eine ca. 184-fach erhöhte Affinität zu Cathepsin L. Die Ergebnisse der theoretischen Vorhersagen bezüglich der bevorzugten Bindungsaffinität von Abz-Phe-Phe-Lys-Lys-Tyr(3-NO<sub>2</sub>)-Gly (11) in hCatL konnte somit bestätigt werden.

### 4.3.2.2 Verifizierungen der Synthesevorschläge

Wie gezeigt werden konnte, ist das Substrat Abz-Phe-Phe-Lys-Lys-Tyr(3-NO<sub>2</sub>)-Gly (11) hochaffin für hCatL. Allerdings zeigen die geringen k<sub>cat</sub>-Werte, daß die proteolytische Aktivität von hCatL gegenüber diesem Peptid nicht sehr hoch ist. Aus diesem Grund wurden weitere Modifizierungen in der Peptidsequenz vorgenommen. Das Lysin in P1-Position wurde gegen ein Asparagin bzw. Glutamin ausgetauscht. Es ist zu erwarten, daß sich ähnlich wie bei den zuerst beschriebenen Liganden zwischen diesen Aminosäuren und dem Enzym stabile Wechselwirkungen ausbilden können. In P2-Position befindet sich ein Phenylalanin. Die hydrophobe Seitenkette dieser Aminosäure richtet sich so in der S2-Bindungstasche aus, daß sie mit den Aminosäuren des Enzyms optimal interagieren kann. Im Unterschied zu den Cathepsinen F und S enthält Cathepsin L in der S3-Bindungsposition neben Leu69 eine saure Aminosäure (Glu63). Zur Ausnutzung dieser Eigenschaft wurden basische Aminosäuren (Lysin, Ornitin) als P3-Rest des Liganden verwendet. Über eine optimale Anordnung der Seitenketten zueinander können sich Wasserstoffbrücken ausbilden, die zur Erhöhung der Bindungsaffinität des Peptides zu hCatL beitragen sollten. Wie schon im vorherigen Abschnitt beschrieben wurde, bilden sich Wasserstoffbrücken zwischen dem Cathepsin und Lysin (P1'-Position) aus. Aufgrund dieser bevorzugten Wechselwirkung wurden auch für die nachfolgend untersuchten Peptide Lysin verwendet.



Die so neu ermittelten Peptidsequenzen wurden in die Bindungsspalte von hCatL gedockt und die Ligandenaffinität der resultierenden hCatL-Substrat-Komplexe untersucht (Tab. 20).

**Tab. 20** Wechselwirkungsverhalten der Liganden in hCatL

Nr.	Peptid	$\Delta E$ [kcal/mol] <sup>a</sup>	$K_d$ [M] <sup>b</sup>
	P3 - P2 - P1 - P1'		
(21)	Abz-Lys-Phe-Gln-Lys-Tyr(3-NO <sub>2</sub> )-GlyNH <sub>2</sub>	-100.56	1.91·10 <sup>-8</sup>
(22)	Abz-Lys-Phe-Asn-Lys-Tyr(3-NO <sub>2</sub> )-GlyNH <sub>2</sub>	-98.79	2.95·10 <sup>-8</sup>
(23)	Abz-Orn-Phe-Gln-Lys-Tyr(3-NO <sub>2</sub> )-GlyNH <sub>2</sub>	-98.76	2.12·10 <sup>-8</sup>
(24)	Abz-Orn-Phe-Asn-Lys-Tyr(3-NO <sub>2</sub> )-GlyNH <sub>2</sub>	-95.11	6.63·10 <sup>-8</sup>

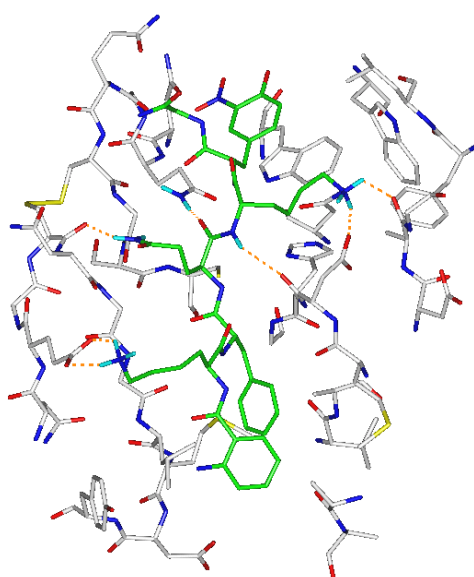
<sup>a</sup> berechnete Wechselwirkungsenergien

<sup>b</sup>  $K_d$ -Werte, berechnet unter Verwendung von SCORE

Aufgrund des Austausches von Lysin gegen Asparagin bzw. Glutamin treten nur geringe Veränderungen in der Orientierung des Peptides im Enzym auf (Abb. 30).

Die Anordnungen der Seitenketten des Lysins, des Glutamins und des Phenylalanins in den Positionen P1', P1 bzw. P2 korrespondieren mit den Orientierungen der Aminosäureseitenketten der anderen Liganden in den entsprechenden Positionen (vgl. Abb. 29 und Abb. 30). Jedoch können sich in diesem Fall stabile intermolekulare Wasserstoffbrücken zwischen dem Lysin in P3-Position und Glu63 des Enzyms ausbilden (Abb. 30). Diese Wechselwirkung trägt zur weiteren Stabilisierung des Enzym-Ligand-Komplexes bei.

Die berechneten Werte der Bindungsaffinität der neu erstellten Peptidsequenzen liegen in der Größenordnung der ersten beschriebenen Substrate (vgl. Tab. 17).



**Abb. 30** Enzym-Ligand-Komplex von hCatL mit Abz-Lys-Phe-Gln-Lys-Tyr(3-NO<sub>2</sub>)-GlyNH<sub>2</sub> (21). Es wurden nur die Aminosäuren des katalytisch aktiven Bindungszentrums sowie die unmittelbar benachbarten dargestellt. Die Kohlenstoffatome des Liganden sind grün gefärbt.

Zum Vergleich wurden diese Substrate wiederum in die Cathepsine K, B und S eingepaßt und deren Affinität gegenüber den Enzymen berechnet (Tab. 21). Anhand der berechneten Werte ist zu sehen, daß die Liganden im Vergleich zu den anderen aufgeführten Cathepsinen ein deutlich günstigeres Wechselwirkungsverhalten mit hCatL entwickeln können. Basierend auf diesen Ergebnissen sollen einiger dieser Substrate in nächster Zeit synthetisiert und deren proteolytische Aktivität getestet werden.

**Tab. 21** Affinität der Liganden zu den Cathepsinen L, K, B und S

Peptid	hCatL		hCatK		hCatB		hCatS	
	$\Delta E$ [kcal/mol] <sup>a</sup>	$K_d$ [M] <sup>b</sup>	$\Delta E$ [kcal/mol] <sup>a</sup>	$K_d$ [M] <sup>b</sup>	$\Delta E$ [kcal/mol] <sup>a</sup>	$K_d$ [M] <sup>b</sup>	$\Delta E$ [kcal/mol] <sup>a</sup>	$K_d$ [M] <sup>b</sup>
(21)	-100.56	$1.91 \cdot 10^{-8}$	-94.07	$2.07 \cdot 10^{-7}$	-83.06	$1.04 \cdot 10^{-5}$	-76.75	$9.54 \cdot 10^{-5}$
(22)	-98.79	$2.95 \cdot 10^{-8}$	-91.81	$8.81 \cdot 10^{-7}$	-82.11	$1.25 \cdot 10^{-6}$	-75.27	$8.12 \cdot 10^{-5}$
(23)	-98.76	$2.12 \cdot 10^{-8}$	-91.40	$7.92 \cdot 10^{-7}$	-80.42	$4.37 \cdot 10^{-6}$	-71.16	$9.75 \cdot 10^{-5}$
(24)	-95.11	$6.63 \cdot 10^{-8}$	-90.53	$9.47 \cdot 10^{-7}$	-78.39	$2.95 \cdot 10^{-5}$	-69.85	$3.64 \cdot 10^{-4}$

<sup>a</sup> berechnete Wechselwirkungsenergien

<sup>b</sup> Bindungsaffinität, berechnet unter Verwendung von SCORE

Aus den Ergebnissen dieser Enzym-Substrat-Komplexe von Cathepsin L lassen sich für die jeweiligen Bindungspositionen im Enzym folgende Schlußfolgerungen ziehen. Die S1-Bindungsstelle von hCatL wird aus Aminosäuren gebildet, die in den Papain-ähnlichen Cysteinproteasen konserviert vorliegen. Deshalb hat diese Bindungsposition keinen wesentlichen Einfluß auf die Spezifität von hCatL. In Analogie zu anderen Cathepsinen dieser Enzymklasse können sich zwischen Arginin, Lysin sowie Glutamin in P1-Position eines Liganden und den Aminosäuren der S1-Bindungsstelle Wasserstoffbrücken ausbilden. Ist der P2-Rest ein Phenylalanin, treten zwischen diesem Rest und den Aminosäuren der S2-Bindungsstelle des Enzyms stabile hydrophobe Wechselwirkungen auf. In P3-Position wird Lysin bevorzugt, so daß die basische Seitenkette dieser Aminosäure mit Glu63 interagieren kann. Da die S1'-Bindungsstelle von hCatL ein Aspartat enthält, können sich stabile Salzbrücken zu der Seitenkette eines Lysins in P1'-Position ausbilden.

### 4.3.3 Erstellung von Peptidsequenzen für hochaffine Cathepsin K-Substrate

In Analogie zu hCatL sollten in einem nächsten Schritt Vorschläge für hochaffine Liganden für Cathepsin K erarbeitet werden. Es wurde dazu das generierte Proteinstrukturmodell von hCatK verwendet. Resultierend auf der Analyse des elektrostatischen Potentials der molekularen Oberfläche von hCatK (Abb. 26c), den Ergebnissen der GRID-Berechnungen (Abb. 27c) und dem Vergleich mit bekannten Enzym-Substrat-Komplexen dieses Cathepsins ist zu erwarten, daß folgende Aminosäurereste mit den Aminosäuren der einzelnen Bindungspositionen stabile Wechselwirkungen

ausbilden können. Da die S1-Bindungsposition konservierte Aminosäuren enthält (Tab. 15), wird in dieser Bindungsstelle in Analogie zu den anderen Cathepsinen Arginin bzw. Lysin bevorzugt, so daß sich elektrostatische Wechselwirkungen zwischen diesem Rest in P1-Position und dem Enzym ausbilden können. Die S2-Bindungstasche von hCatK besitzt hydrophoben Charakter (Abb. 27c). Aus diesem Grund sollten in dieser Bindungsstelle bevorzugt Aminosäuren mit hydrophoben Seitenketten gebunden werden (Prolin, Leucin, Valin) [38,135]. Ist in P2-Position ein Prolin, ist die Affinität des Enzyms zum Liganden am höchsten, wenn in P3-Position ein Glycin auftritt (Glycin > Valin > Phenylalanin) [135]. Aufgrund der Eigenschaften der S3-Bindungsposition ist anzunehmen, daß in P3-Position hydrophobe Reste, aber auch Aminosäuren mit basischen Seitenketten bevorzugt werden (Abb. 26c und Abb. 27c). Die S1'-Bindungsposition von hCatK enthält ein Glutamin (Gln143), dessen hydrophile Seitenkette in der Bindungsspalte des Enzyms liegt (Tab. 15). Vergleicht man diese Bindungsstelle mit der der Cathepsine B, L, H, S und F, ist zu sehen, daß nur hCatF diese Eigenschaften aufweist (Gln140). In den anderen Cathepsinen treten anstelle des Glutamins hydrophobe Reste auf (Tab. 15). Zur Ausnutzung der Eigenschaften der Seitenkette des Gln143 werden als P1'-Reste Glutamin bzw. Glutamat verwendet, so daß sich zwischen Ligand und Enzym Wasserstoffbrückenbindungen ausbilden können. Es resultiert daraus eine allgemeine Peptidsequenz für ein Substrat mit folgenden Modifizierungen:

<b>P3</b>	-	<b>P2</b>	-	<b>P1</b>	-	<b>P1'</b>
Leu		Leu		Arg		Gln
Gly		Pro		Lys		Glu
Val						
His						

Die Liganden für hCatK wurden in der gleichen Art und Weise in die Bindungsspalte des Enzyms eingepaßt, wie es für die Liganden für hCatL beschrieben wurde (Abschnitt 4.3.2). Das Histidin in P3-Position des Liganden liegt in protonierter Form vor. Anhand der Ergebnisse sowie der Berechnungen bevorzugter Wechselwirkungen mit dem GRID-Programm, wurden die Seitenketten der Liganden im Enzym vorpositioniert. Daran anschließend wurden mit dem Docking-Programm FLEXIDOCK verschiedene Anordnungsmöglichkeiten der Liganden im Enzym berechnet und deren Wechselwirkungsverhalten in hCatK untersucht. Die ermittelten Energiewerte sind in Tab. 22 aufgelistet.

**Tab. 22** Bindungsaffinität der Liganden zu hCatK

Nr.	Peptid	$\Delta E$ [kcal/mol] <sup>a</sup>	$K_d$ [M] <sup>b</sup>
P3 - P2 - P1 - P1'			
(25)	Abz-His-Leu-Arg-Glu-Tyr(3-NO <sub>2</sub> )-Gly	-99.42	8.42·10 <sup>-9</sup>
(26)	Abz-His-Leu-Arg-Gln-Tyr(3-NO <sub>2</sub> )-Gly	-95.75	2.39·10 <sup>-8</sup>
(27)	Abz-His-Leu-Lys-Glu-Tyr(3-NO <sub>2</sub> )-Gly	-95.61	3.15·10 <sup>-8</sup>
(28)	Abz-Leu-Leu-Arg-Gln-Tyr(3-NO <sub>2</sub> )-Gly	-95.55	3.99·10 <sup>-8</sup>
(29)	Abz-Leu-Leu-Lys-Glu-Tyr(3-NO <sub>2</sub> )-Gly	-93.69	4.26·10 <sup>-8</sup>
(30)	Abz-Gly-Pro-Arg-Glu-Tyr(3-NO <sub>2</sub> )-Gly	-91.36	2.51·10 <sup>-8</sup>
(31)	Abz-Gly-Pro-Arg-Gln-Tyr(3-NO <sub>2</sub> )-Gly	-90.18	2.99·10 <sup>-7</sup>

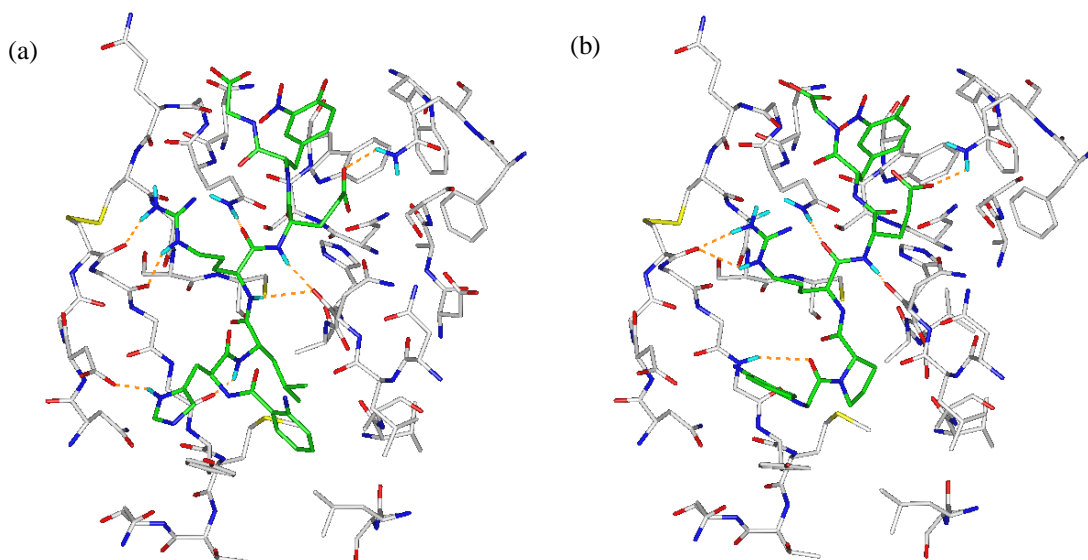
<sup>a</sup> berechnete Wechselwirkungsenergien

<sup>b</sup>  $K_d$ -Werte, berechnet unter Verwendung von SCORE

Die Liganden ordnen sich in hCatK so an, daß diese mit den Aminosäuren des Enzyms interagieren können. Die Proteinstruktur von hCatK im Enzym-Ligand-Komplex ist nahezu identisch mit der Struktur des nativen Cathepsins. Die RMS-Abweichungen zwischen hCatK und dem Komplex (25) wurden über die Hauptkettenatome des Enzyms sowie über die Seitenketten berechnet und betragen 0.68 Å bzw. 2.21 Å. Die größten konformellen Unterschiede treten für die Aminosäureseitenketten Gln143, Asp61 und Leu209 des Enzyms auf.

Vergleicht man die Energie der Enzym-Ligand-Komplexe des hCatK mit den berechneten Werten der hCatL-Komplexe (Tab. 17) sind die Energiedifferenzen innerhalb der hCatK-Komplexe geringer (ca. 10 kcal/mol) (Tab. 22).

Es tritt ein unterschiedliches Dockingverhalten innerhalb der Liganden für hCatK auf, wenn der P2-Rest Leucin gegen ein Prolin ausgetauscht wird. Aus diesem Grund soll das Wechselwirkungsverhalten der Liganden Abz-His-Leu-Arg-Glu-Tyr(3-NO<sub>2</sub>)-Gly (25) und Abz-Gly-Pro-Arg-Glu-Tyr(3-NO<sub>2</sub>)-Gly (30) beschrieben werden (Abb. 31).



**Abb. 31** Enzym-Ligand-Komplexe von hCatK mit Abz-His-Leu-Arg-Glu-Tyr(3-NO<sub>2</sub>)-Gly (25) (a) sowie mit Abz-Gly-Pro-Arg-Glu-Tyr(3-NO<sub>2</sub>)-Gly (30) (b). Die Kohlenstoffatome des Inhibitors sind grün. Von hCatK sind nur die Aminosäuren des katalytisch aktiven Bindungszentrums abgebildet.

Die Anordnungen der Aminosäuren in den Positionen P1 und P1' sind innerhalb der Komplexe sehr ähnlich. Es bilden sich Wasserstoffbrücken zwischen der Carboxylatgruppe der Seitenkette des Glutamats vom Substrat und der Seitenkette von Gln143 aus. Weitere Wasserstoffbrücken können zwischen dem Arginin in P1-Position und den Aminosäuren des Cathepsins (Cys63, Gly64) beobachtet werden (Abb. 31a). Ist in P2-Position ein Leucin, bilden sich hydrophobe Wechselwirkungen zu den Seitenketten der Aminosäuren der S2-Bindungsstelle des Enzyms aus (Met68, Ala134, Leu160, Leu209). Die Seitenkette des Histidins in P3-Position wird so ausgerichtet, daß sich eine Salzbrücke zu der sauren Seitenkette von Asp61 von hCatK ausbilden kann (Abb. 31a). Wird Leucin (P2-Position) gegen Prolin ausgetauscht und ist in P3-Position Glycin vorhanden, treten leichte Veränderungen in der Peptidstruktur des Liganden auf (Abb. 31b). Aufgrund dieser Modifizierungen liegt der Ligand tiefer in der S2-Bindungsstelle des Enzyms. Die van der Waals-Kontakte zwischen dem Pyrrolidinring und den Aminosäuren des Enzyms sind dichter, wodurch die resultierenden hydrophoben Wechselwirkungen stabiler sind. Der Phenylring der Abz-Gruppe richtet sich zur Seitenkette von Tyr67 aus, so daß sich Wechselwirkungen zwischen diesen beiden Aromaten ausbilden können.

Um die Affinität dieser Liganden gegenüber den anderen Cathepsinen vergleichen zu können, wurden diese Peptide in die Bindungsspalte der Cathepsine B, L und S gedockt und die Wechselwirkungsenergien der Peptide im Enzym berechnet (Tab. 23).

**Tab. 23** Bindungsaffinität der Liganden zu den Cathepsinen K, L, B und S

Peptid	hCatK		hCatL		hCatB		hCatS	
	$\Delta E$ [kcal/mol] <sup>a</sup>	$K_d$ [M] <sup>b</sup>	$\Delta E$ [kcal/mol] <sup>a</sup>	$K_d$ [M] <sup>b</sup>	$\Delta E$ [kcal/mol] <sup>a</sup>	$K_d$ [M] <sup>b</sup>	$\Delta E$ [kcal/mol] <sup>a</sup>	$K_d$ [M] <sup>b</sup>
(25)	-99.42	$8.42 \cdot 10^{-9}$	-85.64	$1.41 \cdot 10^{-6}$	-81.87	$7.59 \cdot 10^{-5}$	-78.07	$8.73 \cdot 10^{-5}$
(26)	-95.75	$2.39 \cdot 10^{-8}$	-79.97	$1.29 \cdot 10^{-5}$	-83.97	$4.07 \cdot 10^{-5}$	-76.24	$9.15 \cdot 10^{-5}$
(27)	-95.61	$3.15 \cdot 10^{-8}$	-79.10	$9.12 \cdot 10^{-5}$	-84.42	$7.11 \cdot 10^{-6}$	-75.51	$9.42 \cdot 10^{-5}$
(28)	-95.55	$3.99 \cdot 10^{-8}$	-81.05	$2.24 \cdot 10^{-5}$	-80.94	$7.76 \cdot 10^{-5}$	-75.59	$9.63 \cdot 10^{-5}$
(29)	-93.69	$4.26 \cdot 10^{-8}$	-80.59	$2.57 \cdot 10^{-5}$	-80.12	$7.14 \cdot 10^{-5}$	-73.91	$9.13 \cdot 10^{-5}$
(30)	-91.36	$2.51 \cdot 10^{-8}$	-75.53	$8.50 \cdot 10^{-5}$	-71.50	$1.35 \cdot 10^{-4}$	-65.98	$4.12 \cdot 10^{-4}$
(31)	-90.18	$2.99 \cdot 10^{-7}$	-73.90	$2.45 \cdot 10^{-4}$	-72.39	$6.34 \cdot 10^{-5}$	-65.76	$4.92 \cdot 10^{-4}$

<sup>a</sup> berechnete Wechselwirkungsenergien

<sup>b</sup> Bindungskonstanten, berechnet unter Verwendung von SCORE

Anhand dieser Werte ist zu sehen, daß die Bindungsaffinität der Liganden zu hCatK am höchsten ist. Aus diesem Grund ist anzunehmen, daß diese Peptide im Vergleich zu den Cathepsinen B, L und S spezifisch sind, was über enzymkinetische Untersuchungen an diesen Komplexen nachgewiesen werden sollte. Dazu ausgewählt wurde das Hexapeptid Abz-Leu-Leu-Arg-Gln-Tyr(3-NO<sub>2</sub>)-Gly (28). Es wurde nicht das Substrat Abz-His-Leu-Arg-Glu-Tyr(3-NO<sub>2</sub>)-Gly (25) genutzt. Die Synthese dieses Substrates, mit einem Histidin in der Peptidsequenz gestaltet sich wesentlich aufwendiger, da Histidin zu Nebenreaktionen führen kann.

Das Hexapeptid Abz-Leu-Leu-Arg-Gln-Tyr(3-NO<sub>2</sub>)-Gly (28) wurde im AK Prof. K. Neubert (Universität Halle-Wittenberg) von Herrn Dr. J. Faust synthetisiert (Seite A-15). Allerdings konnte diese Verbindung aufgrund seiner starken Hydrophobizität nicht in Lösung gebracht werden, so daß die Peptidsequenz verändert werden mußte.

#### 4.3.3.1 Überarbeitung der Substratvorschläge für Cathepsin K

Durch einen geeigneten Austausch der Aminosäuren in den Positionen P2 und P3 in der Peptidsequenz wurde versucht, die ausgeprägten hydrophoben Eigenschaften der Liganden zu verringern. Dabei mußte berücksichtigt werden, daß die Bindungsaffinität der Liganden zum Enzym nicht abgeschwächt wird. In Analogie zu den ersten ermittelten Peptidsequenzen wurde in den Positionen P1'- und P1-Positionen Glutamat, Glutamin bzw. Arginin, Lysin verwendet, da anhand des Wechselwirkungsverhalten der Liganden in hCatK gezeigt werden konnte, daß diese Aminosäuren eine hohe Affinität zum Enzym gewährleisten. In P2-Position liegt Leucin vor (hydrophobe Wechselwirkung), das gegen ein Glutamin ausgetauscht wurde. Durch diesen Austausch sollte untersucht werden, ob es möglich ist, daß sich zwischen der Glutaminseitenkette und den Aminosäuren der S2-Bindungsposition stabile Wasserstoffbrücken ausbilden können. Die Ergebnisse

der GRID-Berechnungen zeigen, daß sich in der S2-Bindungstasche neben hydrophoben Wechselwirkungen auch Wechselwirkungen zu einer Carboxylatprobe ausbilden können (Abb. 27c). Leucin in P3-Position wurde durch ein Lysin bzw. Arginin ersetzt, um die Hydrophobizität der Verbindungen zu reduzieren. Aufgrund der Basizität der Seitenketten von Arginin und Lysin ist es möglich, daß sich Wasserstoffbrücken zum Enzym ausbilden können. Die berechneten Wechselwirkungsenergien zwischen Cathepsin K und den so abgeleiteten Peptidsequenzen sind in Tab. 24 aufgeführt.

Da die Synthese des 3-Nitrotyrosinrestes mit einem sehr hohen präparativen Aufwand verbunden ist, wurde bei diesen Peptidsequenzen das Quenchingpaar Abz und 4-Nitrophenylalanin (Phe(4-NO<sub>2</sub>)) verwendet (Rücksprache mit Herrn Dr. J. Faust) [151].

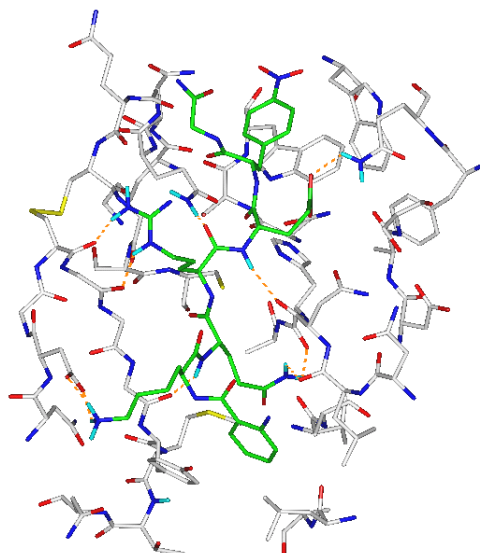
**Tab. 24** Wechselwirkungsverhalten der Liganden in hCatK

Nr.	Peptid	$\Delta E$ [kcal/mol] <sup>a</sup>	$K_d$ [M] <sup>b</sup>
	P3 - P2 - P1 - P1'		
(32)	Abz-Lys-Gln-Arg-Glu-Phe(4-NO <sub>2</sub> )-GlyNH <sub>2</sub>	-105.46	1.38·10 <sup>-8</sup>
(33)	Abz-Lys-Leu-Arg-Glu-Phe(4-NO <sub>2</sub> )-GlyNH <sub>2</sub>	-105.19	1.04·10 <sup>-8</sup>
(34)	Abz-Lys-Leu-Lys-Glu-Phe(4-NO <sub>2</sub> )-GlyNH <sub>2</sub>	-99.14	4.90·10 <sup>-8</sup>
(35)	Abz-Lys-Leu-Lys-Gln-Phe(4-NO <sub>2</sub> )-GlyNH <sub>2</sub>	-98.43	2.04·10 <sup>-8</sup>
(36)	Abz-Arg-Leu-Lys-Glu-Phe(4-NO <sub>2</sub> )-GlyNH <sub>2</sub>	-97.87	2.92·10 <sup>-8</sup>
(37)	Abz-Arg-Leu-Arg-Gln-Phe(4-NO <sub>2</sub> )-GlyNH <sub>2</sub>	-96.99	5.95·10 <sup>-8</sup>

<sup>a</sup> berechnete Wechselwirkungsenergien

<sup>b</sup> K<sub>d</sub>-Werte, berechnet unter Verwendung von SCORE

Die Anordnung der Peptide im Enzym korrespondiert mit der Lage der Liganden in den anderen hCatK-Komplexen, die Leucin in P2-Position enthalten. Zwischen Arginin bzw. Lysin in P3-Position können sich Salzbrücken zu Asp61 von hCatK ausbilden. Interessant ist jedoch der Austausch des P2-Restes gegen ein Glutamin. Aus der berechneten Wechselwirkungsenergie dieses Komplexes ist zu entnehmen, daß die Affinität des Liganden Abz-Lys-Gln-Arg-Glu-Phe(4-NO<sub>2</sub>)-GlyNH<sub>2</sub> (32) zum Cathepsin K stark ausgeprägt ist (Tab. 24). Betrachtet man diesen Komplex genauer, können zahlreiche Interaktionen des Liganden mit hCatK beobachtet werden (Abb. 32).



**Abb. 32** Komplex von hCatK mit Abz-Lys-Gln-Arg-Glu-Phe(4-NO<sub>2</sub>)-GlyNH<sub>2</sub> (32). Es sind nur die Aminosäuren des katalytisch aktiven Bindungszentrums und die unmittelbar benachbart sind, abgebildet. Die Kohlenstoffatome des Liganden sind grün gefärbt.

Wie schon erwähnt wurde, sind die Anordnungen der P1- und P1'-Reste ähnlich im Vergleich zu den anderen hCatK-Komplexen (Abb. 31a und Abb. 32). Es bilden sich Wasserstoffbrücken zwischen dem Enzym und den Seitenketten des Arginins (P1-Position) sowie des Glutamats (P1'-Position) und zu den Peptidhauptkettenatomen aus. Die Seitenkette des Glutamins in P2-Position des Liganden wird so ausgerichtet, daß es zur Ausbildung von Wasserstoffbrücken zu den Carbonyl-Sauerstoffatomen von Leu160 und His162 kommt. Außerdem tritt eine Wasserstoffbrücke zwischen dem Amidproton der Peptidhauptkette des Glutamins und dem Carbonyl-Sauerstoffatom von Gly66 auf. Durch die Lage der Seitenkette des Lysins in P3-Position des Liganden werden stabile Salzbrücken zu der Seitenkette von Asp61 des Enzyms beobachtet (Abb. 32).

Anhand der vergleichenden Untersuchungen dieser Peptide zu den Cathepsinen B, L und S, kann gezeigt werden, daß die Liganden hochaffin für Cathepsin K sind (Tab. 25). Im Vergleich zu Cathepsin K ist die Bindungsaffinität dieser Substrate gegenüber den Cathepsinen B, L und S bedeutend geringer.



**Tab. 25** Affinität der Liganden zu den Cathepsinen K, L, B und S

Peptid	hCatK		hCatL		hCatB		hCatS	
	$\Delta E$ [kcal/mol] <sup>a</sup>	$K_d$ [M] <sup>b</sup>	$\Delta E$ [kcal/mol] <sup>a</sup>	$K_d$ [M] <sup>b</sup>	$\Delta E$ [kcal/mol] <sup>a</sup>	$K_d$ [M] <sup>b</sup>	$\Delta E$ [kcal/mol] <sup>a</sup>	$K_d$ [M] <sup>b</sup>
(32)	-105.46	$1.38 \cdot 10^{-8}$	-84.94	$3.63 \cdot 10^{-6}$	-82.85	$2.48 \cdot 10^{-6}$	-78.77	$4.16 \cdot 10^{-5}$
(33)	-105.19	$1.04 \cdot 10^{-6}$	-85.12	$3.41 \cdot 10^{-6}$	-84.17	$1.98 \cdot 10^{-6}$	-78.15	$2.89 \cdot 10^{-5}$
(34)	-99.14	$4.90 \cdot 10^{-8}$	-79.84	$7.52 \cdot 10^{-6}$	-76.62	$3.61 \cdot 10^{-5}$	-76.90	$3.17 \cdot 10^{-5}$
(35)	-98.43	$2.04 \cdot 10^{-8}$	-81.61	$2.63 \cdot 10^{-6}$	-76.19	$3.59 \cdot 10^{-5}$	-75.43	$6.94 \cdot 10^{-5}$
(36)	-97.87	$2.92 \cdot 10^{-8}$	-78.58	$6.92 \cdot 10^{-6}$	-78.72	$3.01 \cdot 10^{-5}$	-76.51	$5.71 \cdot 10^{-5}$
(37)	-96.99	$5.95 \cdot 10^{-8}$	-79.96	$7.13 \cdot 10^{-6}$	-79.83	$3.21 \cdot 10^{-5}$	-74.89	$6.15 \cdot 10^{-5}$

<sup>a</sup> berechnete Wechselwirkungsenergien

<sup>b</sup>  $K_d$ -Werte, berechnet unter Verwendung von SCORE

Basierend auf den berechneten Wechselwirkungsenergien und den vergleichenden Betrachtungen zu den Cathepsinen L, B und S erweisen sich folgende Aminosäurereste in den einzelnen Pn-Positionen gegenüber dem Enzym begünstigt. Arginin bzw. Lysin in P1-Position können Wasserstoffbrücken zu den Aminosäuren der S1-Bindungsstelle von hCatK ausbilden. Günstige hydrophobe Wechselwirkungen treten zwischen dem P2-Rest und dem Enzym auf, wenn in dieser Position Leucin ist. Wird Leucin gegen Glutamin ausgetauscht, können sich intermolekulare Wasserstoffbrücken ausbilden. Da die S3-Bindungsstelle von hCatK eine saure Aminosäure (Asp61) enthält, sind in P3-Position Histidin oder Lysin bevorzugt. Ist in P1'-Position ein Glutamat bzw. Glutamin, können sich Wasserstoffbrücken zu Gln143 des Enzyms (S1'-Bindungsposition) ausbilden.

Zum Nachweis der hohen Affinität der hier beschriebenen fluorogenen Substrate wurde Abz-Lys-Leu-Arg-Glu-Phe(4-NO<sub>2</sub>)-GlyNH<sub>2</sub> (33) zur Synthese vorgeschlagen (Dr. J. Faust, AK Prof. Dr. K. Neubert) (Seite A-15).

#### 4.3.3.2 Enzymkinetische Untersuchungen

Die enzymkinetischen Studien wurden von Herrn Dr. F. Bühling (Universität Magdeburg) durchgeführt. Das Substrat Abz-Lys-Leu-Arg-Glu-Phe(4-NO<sub>2</sub>)-GlyNH<sub>2</sub> (33) zeigt eine sehr hohe Hintergrundfluoreszenz. Aus diesem Grund ist ein Nachweis der enzymkinetischen Aktivität dieses Substrates, wie im Anhang beschrieben wurde (Seite A-15), nicht möglich. Ursache für diesen negativen Effekt kann der zu geringe Abstand zwischen den beiden fluorogenen Gruppen sein. Zur Untersuchung des Bindungsverhaltens dieses Substrates wurde vorgeschlagen, die Affinität des Liganden zum Enzym über eine andere Vorgehensweise zu ermitteln. Dabei fungiert der Ligand als potentieller Inhibitor des Enzyms.

Erste Untersuchungen mit Abz-Lys-Leu-Arg-Glu-Phe(4-NO<sub>2</sub>)-GlyNH<sub>2</sub> (33) deuten darauf hin, daß das Peptid eine höhere Affinität zu Cathepsin K als zu Cathepsin L besitzt. Da die Untersuchungen mit

diesem Peptid noch nicht abgeschlossen sind, können keine endgültigen Aussagen über die Substratspezifität getroffen werden.

#### 4.3.4 Erstellung neuer Peptidsequenzen für effiziente Cathepsin F-Inhibitoren

Das generierte Strukturmodell von Cathepsin F wurde verwendet, um für dieses Enzym Inhibitoren zu entwickeln. Nach der Charakterisierung des katalytisch aktiven Bindungszentrums von hCatF und anhand der kinetischen Daten, beschrieben durch *Wang et al.* [41], war es möglich, geeignete Aminosäurereste für neue Inhibitoren für hCatF zu ermitteln. Zahlreiche Studien an Enzym-Inhibitor-Komplexen der Papain-ähnlichen Cysteinproteasen zeigten, daß in P1-Position eines Inhibitors ein hydrophober Rest vorliegen sollte [153,154]. Die S2-Bindungstasche von hCatF wird aus überwiegend hydrophoben Aminosäuren gebildet, wodurch der Charakter dieser Bindungsstelle bestimmt wird (Tab. 15, Abb. 27f). Die Charakterisierung unter Verwendung von GRID zeigt, daß durch die Eigenschaften sowie durch die Lage der Aminosäuren dieser Bindungsstelle, die S2-Bindungstasche von hCatF sterisch anspruchsvolle Aminosäureseitenketten aufnehmen kann (3-(1'-Naphthyl)alanin, Tryptophan, Leucin, Isoleucin, Valin). Im Gegensatz zu den Cathepsinen L, B und K enthält die S3-Bindungsposition von hCatF ein Lysin (Lys61) (Tab. 15, Abb. 27f). Zur Stabilisierung möglicher Enzym-Inhibitor-Komplexe dieses Cathepsins wurde für die P3-Position die Succinylgruppe bzw. ein Glutamat verwendet. Deshalb wurden zur Erstellung der Sequenzen von Peptiden folgende Aminosäuren für die jeweiligen Pn-Positionen ausgewählt:

<b>P3</b>	-	<b>P2</b>	-	<b>P1</b>
Suc		1-Nal		Leu
Glu		Trp		
		Leu		
		Ile		
		Val		

Es resultieren daraus zahlreiche Peptide, die in ihrer Bindungsaffinität zu hCatF untersucht wurden (Tab. 26). Die Verbindungen wurden in das katalytisch aktive Bindungszentrum von hCatL gedockt, in dem die Lage der Hauptkettenatome an die Position der Atome von Liganden in bekannten Enzym-Inhibitor-Komplexen verwandter Cysteinproteasen angepaßt wurde (Abschnitt 3.2.2). Die Anordnung der Seitenketten des Liganden wurde so variiert, daß sie optimal in den entsprechenden GRID-Feldern positioniert waren (Abb. 27f). Daran anschließend wurden mit dem Docking-Programm FLEXIDOCK verschiedene Anordnungsmöglichkeiten der Liganden im Enzym berechnet und deren Wechselwirkungsverhalten in hCatK untersucht. Die berechneten Affinitätswerte sind in Tab. 26 aufgelistet.

**Tab. 26** Bindungsaffinität der Liganden zu hCatF

Nr.	Peptid	$\Delta E$ [kcal/mol] <sup>a</sup>	$K_d$ [M] <sup>b</sup>
P3 - P2 - P1			
(38)	Z-Glu-1-Nal-Leu-CONH-(CH <sub>2</sub> ) <sub>2</sub> -Ph	-79.59	1.26·10 <sup>-7</sup>
(39)	Z-Glu-1-Nal-Leu-CONH-(CH <sub>2</sub> ) <sub>2</sub> -1-Imidazol	-77.30	2.80·10 <sup>-7</sup>
(40)	Z-Glu-Trp-Leu-CONH-(CH <sub>2</sub> ) <sub>2</sub> -Ph	-74.85	3.34·10 <sup>-7</sup>
(41)	Suc-1-Nal-Leu-CONH-(CH <sub>2</sub> ) <sub>2</sub> -Ph	-74.51	3.31·10 <sup>-7</sup>
(42)	Suc-1-Nal-Leu-CONH-(CH <sub>2</sub> ) <sub>2</sub> -1-Imidazol	-73.31	5.75·10 <sup>-7</sup>
(43)	Z-Glu-Trp-Leu-CONH-(CH <sub>2</sub> ) <sub>2</sub> -1-Imidazol	-73.27	1.43·10 <sup>-6</sup>
(44)	Suc-Trp-Leu-CONH-(CH <sub>2</sub> ) <sub>2</sub> -Ph	-71.26	2.62·10 <sup>-5</sup>
(45)	Suc-Trp-Leu-CONH-(CH <sub>2</sub> ) <sub>2</sub> -1-Imidazol	-70.70	3.98·10 <sup>-5</sup>
(46)	Z-Glu-Leu-Leu-CONH-(CH <sub>2</sub> ) <sub>2</sub> -Ph	-69.73	1.25·10 <sup>-5</sup>
(47)	Z-Glu-Ile-Leu-CONH-(CH <sub>2</sub> ) <sub>2</sub> -Ph	-68.47	6.37·10 <sup>-5</sup>
(48)	Z-Glu-Val-Leu-CONH-(CH <sub>2</sub> ) <sub>2</sub> -Ph	-67.04	6.14·10 <sup>-5</sup>
P2 - P1			
(49)	Z-1-Nal-Leu-CONH-(CH <sub>2</sub> ) <sub>2</sub> -Ph	-63.61	3.02·10 <sup>-4</sup>
(50)	Z-Trp-Leu-CONH-(CH <sub>2</sub> ) <sub>2</sub> -Ph	-58.78	2.13·10 <sup>-4</sup>

<sup>a</sup> berechnete Wechselwirkungsenergien

<sup>b</sup>  $K_d$ -Werte, berechnet unter Verwendung von SCORE

Die Unterschiede der Proteinstrukturen zwischen dem nativen hCatF und dem Komplex (38) sind sehr gering. Die RMS-Abweichung über die Hauptkettenatome beider Strukturen beträgt 0.71 Å. Wird der RMS-Wert über die Seitenkettenatome berechnet, beträgt dieser Wert 2.49 Å. Es treten nur für die Seitenketten von Lys61 und Met207 variierte Anordnungen auf, die allerdings keinen Einfluß auf die gesamte Proteinstruktur haben.

Anhand der berechneten Wechselwirkungsenergien der Peptide mit hCatF ist zu sehen, daß diese Verbindungen eine hohe Bindungsaffinität zum Enzym besitzen (Tab. 26).

Aufgrund der energetischen Abstufung der Enzym-Inhibitor-Komplexe resultieren folgende Bevorzugen innerhalb der Peptidsequenz der Liganden. Inhibitoren mit sterisch anspruchsvollen Resten in P2-Position (3-(1'-Naphthyl)alanin, Tryptophan) besitzen eine stärkere Affinität zum Enzym. Es bilden sich hydrophobe Wechselwirkungen zu den Aminosäuren der S2-Bindungsposition (Leu67, Pro68, Ala133, Ile159, Ala162, Met207) aus. Wird dieser P2-Rest gegen Leucin, Isoleucin bzw. Valin ausgetauscht, verringert sich die Bindungsaffinität des Liganden in der Reihenfolge Leucin > Isoleucin > Valin (Tab. 26). Wie zu erwarten war, können sich zwischen dem Glutamat bzw. der Succinylgruppe in P3-Position des Liganden und Lys61 von hCatF Salzbrücken ausbilden und so zur Stabilisierung des Komplexes beitragen. Dieser Effekt konnte überprüft werden, in dem Peptidsequenzen untersucht wurden, die diesen P3-Rest nicht enthalten (Verbindungen (49) und (50) in Tab. 26). Die Wechselwirkungsenergien dieser Komplexe sind im Vergleich zu den anderen hCatF-Komplexen deutlich reduziert.

In einem nächsten Schritt wurde versucht, die Peptide in die katalytisch aktiven Bindungszentren der Cathepsine K, L und S zu docken. Da die Größe der S2-Bindungstasche von hCatK eingeschränkt ist,

können die Liganden nur so eingepaßt werden, daß sich energetisch ungünstige Konformationen innerhalb der Peptidstruktur der Liganden ausbilden. Deshalb wurde das Wechselwirkungsverhalten der Inhibitoren gegenüber hCatK nicht weiter untersucht. Anhand der berechneten Werte der Wechselwirkungsenergie der Enzym-Inhibitor-Komplexe von hCatF, hCatL und hCatS ist zu erkennen, daß eine deutlich erhöhte Affinität der Liganden gegenüber hCatF auftritt (Tab. 27).

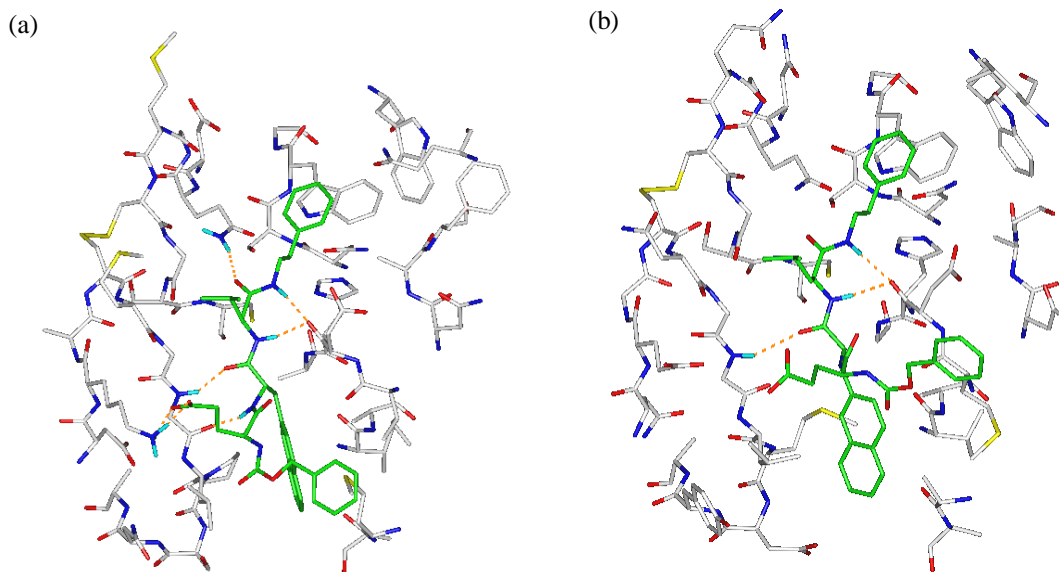
**Tab. 27** Bindungsaffinität der Liganden zu den Cathepsinen F, L und S

Peptid	hCatF		hCatL		hCatS	
	$\Delta E$ [kcal/mol] <sup>a</sup>	$K_d$ [M] <sup>b</sup>	$\Delta E$ [kcal/mol] <sup>a</sup>	$K_d$ [M] <sup>b</sup>	$\Delta E$ [kcal/mol] <sup>a</sup>	$K_d$ [M] <sup>b</sup>
(38)	-79.59	$1.26 \cdot 10^{-7}$	-65.63	$1.15 \cdot 10^{-5}$	-59.38	$2.48 \cdot 10^{-5}$
(39)	-77.30	$2.80 \cdot 10^{-7}$	-58.78	$2.12 \cdot 10^{-5}$	-57.40	$2.45 \cdot 10^{-5}$
(40)	-74.85	$3.34 \cdot 10^{-7}$	-58.03	$6.61 \cdot 10^{-5}$	-58.52	$3.33 \cdot 10^{-5}$
(41)	-74.51	$3.31 \cdot 10^{-7}$	-60.47	$3.34 \cdot 10^{-5}$	-61.74	$1.98 \cdot 10^{-5}$
(42)	-73.31	$5.75 \cdot 10^{-7}$	-55.37	$1.08 \cdot 10^{-4}$	-62.24	$2.14 \cdot 10^{-6}$
(43)	-73.27	$1.43 \cdot 10^{-6}$	-55.80	$4.79 \cdot 10^{-4}$	-58.36	$4.07 \cdot 10^{-6}$
(44)	-71.26	$2.62 \cdot 10^{-5}$	-53.40	$5.14 \cdot 10^{-4}$	-59.56	$2.21 \cdot 10^{-5}$
(45)	-70.70	$3.98 \cdot 10^{-5}$	-50.87	$7.94 \cdot 10^{-4}$	-57.14	$3.16 \cdot 10^{-5}$
(46)	-69.73	$1.25 \cdot 10^{-5}$	-57.00	$4.84 \cdot 10^{-5}$	-56.46	$8.68 \cdot 10^{-5}$
(47)	-68.47	$6.37 \cdot 10^{-5}$	-56.32	$7.07 \cdot 10^{-5}$	-56.31	$7.26 \cdot 10^{-5}$
(48)	-67.04	$6.14 \cdot 10^{-5}$	-54.42	$8.94 \cdot 10^{-5}$	-54.52	$8.71 \cdot 10^{-5}$

<sup>a</sup> berechnete Wechselwirkungsenergien

<sup>b</sup> Bindungsaffinität, berechnet unter Verwendung von SCORE

Das unterschiedliche Wechselwirkungsverhalten der Liganden im Enzym soll anhand des Liganden Z-Glu-1-Nal-Leu-CONH-(CH<sub>2</sub>)<sub>2</sub>-Ph (38), gedockt in hCatF und hCatL näher erläutert werden (Abb. 33). Obwohl die Sequenzhomologie zwischen diesen beiden Cathepsinen sehr hoch ist (57%), wurden unterschiedliche Eigenschaften innerhalb der katalytisch aktiven Bindungszentren dieser Enzyme beobachtet (Abschnitt 4.3.1, 4.3.1.1 und 4.3.1.2). Diese Charakteristika äußern sich auch im Wechselwirkungsverhalten dieses Peptides in beiden Cathepsinen.



**Abb. 33** Enzym-Inhibitor-Komplexe von hCatF (a) und hCatL (b) mit Z-Glu-1-Nal-Leu-CONH-(CH<sub>2</sub>)<sub>2</sub>-Ph (38). Von den Enzymen sind nur die Aminosäuren der katalytisch aktiven Bindungszentren und der Umgebung abgebildet. Die Kohlenstoffatome des Liganden sind grün gefärbt.

Wie in Abb. 33 zu sehen ist, ist die Anordnung des Phenylrings in der S1'-Bindungsposition sowie des P1-Restes Leucin in beiden Cathepsinen sehr ähnlich. Es bilden sich mehrere Wasserstoffbrücken zwischen Ligand und Enzym aus (Abb. 33). Vergleicht man die Anordnung der aromatischen Seitenkette des 3-(1'-Naphthyl)alanins in beiden Cathepsinen, kann man feststellen, daß sich diese Seitenkette in die S2-Bindungstasche von hCatF tiefer einpaßt als im Fall des hCatL-Komplexes. Dadurch nimmt der P2-Rest eine optimale Lage zu den Aminosäuren der S2-Bindungsstelle ein (Leu67, Pro68, Ala133, Met207). Im Fall des hCatL-Komplexes wird die Peptidkonformation des Liganden in den Positionen P2 und P3 verändert. Die 3-(1'-Naphthyl)alaninseitenkette wird um  $\chi_1$  und  $\chi_2$  gedreht, so daß die hydrophoben Wechselwirkungen zum Enzym abgeschwächt werden. Aufgrund der veränderten Peptidstruktur wird die Ausbildung einer Wasserstoffbrücke zwischen dem Amidproton von 3-(1'-Naphthyl)alanin und dem Carbonyl-Sauerstoffatom von Leu67, welche im hCatF-Komplex beobachtet werden konnte (Abb. 33a), unterbunden. Die Bindungsposition S3 von hCatF besteht aus Lys61 und Leu67. Aufgrund der Ausrichtung der Seitenkettenatome von Lys61 können sich stabile Salzbrücken zwischen dieser Aminosäure und dem P3-Rest Glutamat ausbilden (Abb. 33a). Die entsprechende Bindungsposition von hCatL wird aus Glu63 und Leu69 gebildet, wodurch keine intermolekularen Wasserstoffbrücken zu den Aminosäuren der S3-Bindungsstelle von hCatL beobachtet werden konnten (Abb. 33b).

Basierend auf den dargestellten Interaktionen zwischen den Peptiden sowie den berechneten Wechselwirkungsenergien der hCatF-Komplexe kann die unterschiedliche Bindungsaffinität der Liganden begründet werden. Weiterhin kann für die Reste in den Positionen P2 und P3 folgende Bevorzugung festgestellt werden. Es können sich hydrophobe Wechselwirkungen zwischen einem P2-Rest und dem Enzym ausbilden, die um so stärker sind, je größer das sterische Volumen dieses Restes

---

ist (3-(1'-Naphthyl)alanin > Tryptophan > Leucin). Ist in P3-Position ein saurer Rest, bilden sich Wasserstoffbrücken zu Cathepsin F aus, die zur Stabilisierung des sich bildenden Komplexes beitragen (Glutamat > Succinylgruppe).

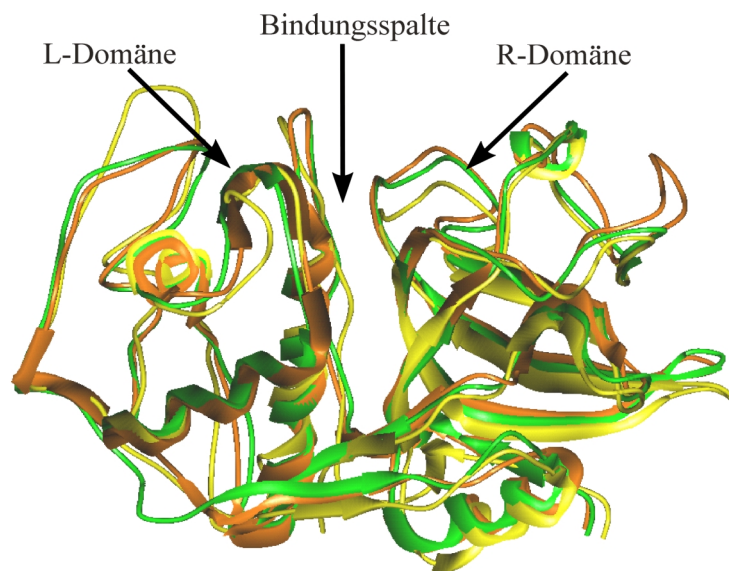
Da im Vergleich zu anderen Cathepsinen gezeigt werden konnte, daß diese Liganden gegenüber hCatF eine hohe Affinität besitzen, wurden zwei dieser Peptide Z-Glu-1-Nal-Leu-CONH-(CH<sub>2</sub>)<sub>2</sub>-Ph (38) und Suc-1-Nal-Leu-CONH-(CH<sub>2</sub>)<sub>2</sub>-Ph (41) zur Synthese vorgeschlagen (Herr Dr. R. Singh, Rigel Inc., South San Francisco, USA). Die enzymkinetischen Untersuchungen werden im AK von Herrn Prof. D. Brömme durchgeführt (Mount Sinai School of Medicine, New York, USA).

## 5 ZUSAMMENFASSUNG UND AUSBLICK

Papain-ähnliche Cysteinproteasen besitzen eine wesentliche Bedeutung beim Umbau zellulärer und extrazellulärer Proteine. Diese Enzyme spielen eine wichtige Rolle in pathologischen Prozessen, die mit einer Gewebedestruktion verbunden sind.

Gegenstand der Arbeit waren Untersuchungen zur strukturellen Charakterisierung von Enzymen der Papain-Superfamilie (Cathepsine) sowie das Design von neuen selektiven Liganden für die Cathepsine L, K und F. Die Ergebnisse der Untersuchungen sind im folgenden zusammengestellt.

1. Proteasen dieser Familie besitzen eine ähnliche Aminosäuresequenz und eine sehr hohe Konformität in ihren dreidimensionalen Proteinstrukturen. Die Proteinstrukturen der Cathepsine K, S, H und F konnten aufgrund der hohen Sequenzhomologie zu verwandten Cysteinproteasen mit bekannter Röntgenkristallstruktur modelliert werden. Die Eigenschaften der Tertiärstrukturmodelle dieser Proteine stehen in sehr hoher Übereinstimmung zu den Charakteristika anderer Mitglieder der Papain-Superfamilie. Zur Veranschaulichung der Gemeinsamkeiten der generierten Strukturmodelle sind in Abb. 34 die Tertiärstrukturen der Cathepsine K, S und F dargestellt. Für eine bessere Übersicht wurde auf die Abbildung der Proteinstruktur des Modells von Cathepsin H verzichtet, da die aktive Form dieses Enzyms eine Minikette enthält.



**Abb. 34** Darstellung der Tertiärstrukturmodelle der Cathepsine K (grün), S (gelb) und F (orange). Die Proteinstrukturen dieser Cathepsine wurden übereinander gelegt.

Die intermolekularen Atomabstände zwischen den Aminosäuren der katalytischen Triade (Cystein, Histidin, Asparagin) korrespondieren mit den Werten, wie sie in

Röntgenkristallstrukturen von verwandten Cysteinproteasen gefunden wurden (Cys25S $\gamma$ ...His162N $\delta$ 1 und His162N $\epsilon$ 2...Asn182O $\delta$ 1 betragen  $3.73 \text{ \AA} \pm 0.32 \text{ \AA}$  bzw.  $3.13 \text{ \AA} \pm 0.24 \text{ \AA}$ ) (Cathepsin K-Numerierung).

Im Gegensatz zu den Cathepsinen K, S und F enthält Cathepsin H eine Minikette, die Bestandteil des Enzyms ist. Die Lage und Anordnung dieser Minikette im Enzym konnte ermittelt werden.

Die Überprüfung der Strukturmodelle erfolgte über verschiedene Methoden (Qualität und Geometrie der Struktur, Berechnung des Energieprofils des Proteinmodells, Untersuchung des Wechselwirkungsverhaltens von ausgesuchten Liganden im Vergleich zu Literaturdaten). Dabei konnte gezeigt werden, daß die generierten Proteinstrukturen dieser Cathepsine eine hohe Genauigkeit aufweisen. Einen weiteren Beweis lieferte der Vergleich des Strukturmodells von Cathepsin K mit seiner später zugänglichen Röntgenstruktur. Die hervorragenden Übereinstimmungen zwischen beiden Tertiärstrukturen dieses Cathepsins unterstreichen, daß die hier aufgezeigte Vorgehensweise zur Ermittlung von Proteinstrukturmodellen einen geeigneten Weg darstellt, um Informationen über den strukturellen Aufbau dieser Proteine zu erhalten.

Basierend auf diesen Ergebnissen bilden die so generierten Tertiärstrukturmodelle der Cathepsine K, S, H und F eine grundlegende Voraussetzung zum strukturbasierten Ligandendesign von catheptischen Cysteinproteasen.

Die Koordinaten der generierten 3D-Strukturmodelle der Cathepsine S, H und F wurden in der Brookhaven Proteindatenbank veröffentlicht (PDB-Einträge: 1BXF (hCatS) [155], 1BZN (hCatH) [156], 1D5U (hCatF) [157]).

2. Die Kenntnis der 3D-Struktur eines Proteins kann wesentlich zum Verständnis der biologischen Funktion des Moleküls beitragen. Ausgehend von den Strukturmodellen der Cathepsine K, S, H und F sowie der Röntgenkristallstrukturen der Cathepsine B und L wurden die Substratbindungsregionen der katalytisch aktiven Zentren dieser Enzyme detailliert untersucht und charakterisiert. Über den Vergleich der Proteinstrukturen dieser Proteasen konnte der Aufbau der einzelnen Bindungspositionen S1-S3 und S1'-S2' ermittelt werden (Tab. 15). Diese Untersuchungen zeigen, daß sich die S2-, S3- bzw. S1'-Bindungsstellen aus zum Teil unterschiedlichen Aminosäuren zusammensetzen. Dadurch werden die Eigenschaften der Bindungspositionen beeinflusst. Diese Effekte konnten für ein effektives Ligandendesign für das jeweilige Enzym ausgenutzt werden.

Zur weiteren Untersuchung der katalytisch aktiven Bindungszentren der Cathepsine B, L, K, H, S und F wurden die elektrostatischen Potentiale der molekularen Oberfläche dieser Enzyme berechnet. Auf der Grundlage dieser Berechnungen wurden die Bindungsspalten der Cathepsine hinsichtlich ihrer Potentialverteilung sowie ihrer Größe eingeteilt. Im Gegensatz zu der hohen Sequenzhomologie dieser Enzyme weisen die berechneten elektrostatischen Oberflächenpotentiale deutliche Unterschiede auf. Der Vergleich der Potentiale der Cathepsine untereinander zeigt, daß



Cathepsin L ein sehr ausgeprägtes elektronegatives Potential besitzt. Diese Eigenschaft nimmt zu Cathepsin B und Cathepsin K ab. Cathepsin F besitzt ein schwach elektronegatives bzw. neutrales Oberflächenpotential. Im Gegensatz dazu besitzt die molekulare Oberfläche von Cathepsin S ein elektropositives Potential. Die unterschiedlichen Eigenschaften der elektrostatischen Potentiale können durch das Auftreten von Aminosäuren mit unterschiedlichen Eigenschaften und der unterschiedlichen Anordnung im Enzym begründet werden. Anhand der berechneten Oberflächenpotentiale können Rückschlüsse auf mögliche elektrostatische Interaktionen zwischen Ligand und Enzym gezogen werden.

Zur Bestimmung von bevorzugten Wechselwirkungen, die innerhalb der Substratbindungsregionen von Enzymen auftreten können, wurden unter Verwendung von mehreren funktionellen Gruppen als Proben die Bindungsspalten der Cathepsine L, B, K, H, S und F bezüglich ihrer Eigenschaften abgetastet. Es resultierten daraus die Bereiche, in denen günstige Wechselwirkungen zwischen diesen Gruppen und dem Enzym zu erwarten sind. Auf diese Weise wurden hydrophobe und elektrostatische Eigenschaften der Cathepsine analysiert und zur Positionierung der neuen Liganden im Enzym genutzt.

3. Zum rationalen Design von neuen Liganden für die Cathepsine L, K und F wurden die Röntgenstruktur von Cathepsin L sowie die berechneten Tertiärstrukturmodelle der Cathepsine K und F verwendet. Basierend auf den Studien der katalytisch aktiven Bindungszentren dieser Cathepsine sowie dem Vergleich mit verwandten Enzym-Ligand-Komplexen wurden neue Peptidsequenzen für Liganden erstellt. Diese Sequenzen sollten sich durch eine hohe Selektivität für eines dieser Cathepsine auszeichnen. Die Liganden wurden hinsichtlich ihres Docking- und Wechselwirkungsverhaltens zum Enzym untersucht. Dazu wurden die Liganden in die Cathepsine gedockt und die Bindungsaffinität dieser Peptide zum Protein analysiert. Um Aussagen über die Selektivität dieser Liganden treffen zu können, wurden vergleichende Berechnungen an anderen Papain-ähnlichen Cysteinproteasen durchgeführt.

Um die Ausbildung von stabileren Interaktionen zwischen Ligand und Enzym zu erreichen, wurden zahlreiche Modifizierungen in der Peptidsequenz der Liganden durchgeführt. Auf der Grundlage dieser Berechnungen resultieren folgende Reste in den einzelnen Pn-Positionen, die bevorzugt mit den Aminosäuren der Sn-Bindungspositionen des Enzyms interagieren sollten:

	<b>Cathepsin L</b>	<b>Cathepsin K</b>	<b>Cathepsin F</b>
<b>P1</b>	Arg, Lys, Gln	Arg, Lys	Leu
<b>P2</b>	Phe	Leu, Gln	1-Nal, Trp
<b>P3</b>	Leu, Phe, Lys	His, Lys	Glu, Suc
<b>P1'</b>	Lys	Glu, Gln	

Einige Liganden, die eine deutlich bevorzugte Bindungsaffinität zu einem Cathepsin aufweisen, wurden zur Synthese vorgeschlagen.

Aufgrund der theoretischen Untersuchungen sollte das Substrat Abz-Phe-Phe-Lys-Lys-Tyr(3-NO<sub>2</sub>)-Gly (11) die höchste Affinität zu Cathepsin L im Vergleich zu den Cathepsinen B und K aufweisen. Diese Vorhersage konnte durch die Ergebnisse der enzymkinetischen Untersuchungen bestätigt werden. Danach wird das Substrat deutlich besser in Cathepsin L gebunden als in Cathepsin B. Für Cathepsin K konnte keine proteolytische Aktivität nachgewiesen werden.

Im Fall des Cathepsin K wurden die Substrate Abz-Leu-Leu-Arg-Gln-Tyr(3-NO<sub>2</sub>)-Gly (28) sowie Abz-Lys-Leu-Arg-Glu-Phe(4-NO<sub>2</sub>)-GlyNH<sub>2</sub> (33) in ihrer proteolytischen Aktivität untersucht. Erste Untersuchungen deuten darauf hin, daß Abz-Lys-Leu-Arg-Glu-Phe(4-NO<sub>2</sub>)-GlyNH<sub>2</sub> (33) eine höhere Bindungsaffinität zu Cathepsin K als zu Cathepsin L besitzt.

Weiterhin konnten Inhibitoren generiert werden, die im Vergleich zu den Cathepsinen L, K und S eine ausgeprägte Affinität zu Cathepsin F besitzen. Zwei dieser Peptide (Z-Glu-1-Nal-Leu-CONH-(CH<sub>2</sub>)<sub>2</sub>-Ph (38), Suc-1-Nal-Leu-CONH-(CH<sub>2</sub>)<sub>2</sub>-Ph (41)) sollen in der nächsten Zeit synthetisiert und getestet werden.

Anhand der Untersuchungen konnte nachgewiesen werden, daß die aus dem Design neuer Liganden getroffenen Vorhersagen bezüglich ihrer Selektivität gut mit den experimentell ermittelten Werten übereinstimmen. Daher stellt die in dieser Arbeit aufgeführte Vorgehensweise eine geeignete Methode zur Entwicklung neuer hochaffiner Liganden für ein Enzym dar. Auf diese Befunde stützend, soll in weiteren Arbeiten versucht werden, selektive kompetitive Inhibitoren zu generieren, die auf den hier diskutierten Peptidsequenzen beruhen.

## 6 LITERATURVERZEICHNIS

- 1 Polgar, L. General Aspects of Proteases. In Mechanismus of Protease Action. (Polgar, L., ed.) Boca Ratan, FL: CRC Press **1989**, pp. 43-76.
- 2 Polgar L. Metalloproteases. In Mechanismus of Protease Action. (Polgar, L., ed.) Boca Ratan, FL: CRC Press **1989**, pp. 208-210.
- 3 Rawlings, N. D.; Barrett, A. J. *Methods Enzymol.* **1994**, *244*, 461-486.
- 4 Berti, P. J.; Storer, A. C. *J. Mol. Biol.* **1995**, *246*, 273-283.
- 5 Kirschke, H.; Barrett, A. J. Chemistry of Lysosomal Proteases. In Lysosomes their Role in Protein Breakdown. (Glaumann, H.; Ballard, F. J., eds.) Acad. Press **1987**, pp. 193-238.
- 6 Willstätter, R.; Barmann, E. *Hoppe-Seyler's Z. Physiol. Chem.* **1929**, *180*, 127-143.
- 7 Riese, R. J.; Wolff, P. R.; Brömme, D.; Natkin, L. R.; Villadangos, J. A.; Ploegh, H. L.; Chapman, H. A. *Immunity* **1996**, *4*, 357-366.
- 8 Tezuka, K.; Tezuka, Y.; Maejima, A.; Sato, T.; Nemoto, K.; Kamioka, H.; Hakeda, Y.; Kumegawa, M. *J. Biol. Chem.* **1994**, *269*, 1106-1109.
- 9 Krieger, T. J.; Hook, V. Y. H. *J. Biol. Chem.* **1991**, *266*, 8376-8383.
- 10 Mason, R. W.; Johnson, D. A.; Barrett, A. J.; Chapman, H. A. *Biochem. J.* **1986**, *233*, 925-927.
- 11 Delaisse, J. M.; Pascale, L.; Vaes, G. *Biochem. J.* **1991**, *279*, 167-174.
- 12 Golde, T. E.; Estus, S.; Younkin, L. H.; Selkoe, D. J.; Younkin, S. G. *Science*, **1992**, *255*, 728-730.
- 13 Mort, J. S.; Recklies, A. D.; Poole, A. R. *Arthritis Rheum.* **1984**, *27*, 509-515.
- 14 Sloane, B. F. *Semin. Cancer Biol.* **1990**, *1*, 137-152.
- 15 Varughese, K. L.; Ahmed, F. F.; Careys, P. R.; Hasnain, S.; Huber, C. P.; Storer, A. C. *Biochemistry* **1989**, *28*, 1330-1332.
- 16 Mason, R. W.; Wilcox, D. *Adv. Cell Mol. Biol. Membr.* **1993**, *1*, 81-116.
- 17 Storer, A. C.; Menard, R. *Methods Enzymol.* **1994**, *244*, 486-500.
- 18 Turk, D.; Guncar, G.; Podobnik, M.; Turk, B. *Biol. Chem. Hoppe-Seyler* **1998**, *379*, 137-147.
- 19 Kamphuis, I. G.; Kalk, K. H.; Swarte, M. B. A.; Drenth, J. *J. Mol. Biol.* **1984**, *179*, 233-256.
- 20 Turk, B.; Turk, V.; Turk, D. *Biol. Chem. Hoppe-Seyler* **1997**, *378*, 141-150.
- 21 Storer, A. C.; Menard, R. *Methods Enzymol.* **1994**, *244*, 486-500.
- 22 Fox, T.; de Miguel, E.; Mort, J. S.; Storer, A. C. *Biochemistry* **1992**, *31*, 12571-12576.
- 23 Tao, K.; Stearns, N. A.; Dong, J.; Wu, Q.; Sahagian, G. G. *Arch. Biochem. Biophys.* **1994**, *311*, 19-27.
- 24 Mach, L.; Mort, J. S.; Glössl, J. *J. Biol. Chem.* **1994**, *269*, 13036-13040.
- 25 Mach, L.; Mort, J. S.; Glössl, J. *J. Biol. Chem.* **1994**, *269*, 13030-13035.
- 26 Kirschke, H.; Barrett, A. J.; Rawlings, N. D. Proteinases I: Lysosomal Cysteine Proteases. In Protein Profile (Sheterline, P., ed.) Acad. Press **1995**, Vol. 2, pp. 1587-1646.
- 27 Kirschke, H.; Langner, J.; Wiederanders, B.; Ansorge, S.; Bohley, P.; Hanson, H. *Acta Biol. Med. Ger.* **1977**, *36*, 185-199.
- 28 Chan, S. J.; San Segrundo, B.; McCormick, M. B.; Steiner, D. F. *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 7721-7725.
- 29 Gal, S.; Gottesmann, M. M. *Biochem. J.* **1988**, *255*, 303-306.

- 30 Joseph, L. J.; Chang, L. C.; Stamenkovich, D.; Sukhatme, V. P. *J. Clin. Invest.* **1988**, *81*, 1621-1629.
- 31 Fuchs, R.; Machleidt, W.; Gassen, H. G. *Biol. Chem. Hoppe-Seyler* **1988**, *369*, 469-475.
- 32 Ritonja, A.; Popovic, T.; Kotnik, M.; Machleidt, W.; Turk, V. *FEBS Lett.* **1988**, *228*, 341-345.
- 33 Wiederanders, B.; Brömme, D.; Kirschke, H.; von Figura, K.; Schmidt, B.; Peters, C. *J. Biol. Chem.* **1992**, *267*, 13708-13713.
- 34 Shi, G. P.; Munger, J. S.; Meara, J. P.; Rich, D. H.; Chapman, H. A. *J. Biol. Chem.* **1992**, *267*, 7258-7262.
- 35 Paris, A.; Strukelj, B.; Pungercar, J.; Renko, M.; Dolenc, I.; Turk, V. *FEBS Lett.* **1995**, *369*, 326-330.
- 36 Velasco, G.; Ferrando, A. A.; Puente, X. S.; Sanchez, L. M.; Lopez-Otin, C. *J. Biol. Chem.* **1994**, *269*, 27136-27142.
- 37 Inaoka, T.; Bilbe, G.; Ishibashi, O.; Tezuka, K.; Kumegawa, M.; Kokubo, T. *Biochem. Biophys. Res. Commun.* **1995**, *206*, 89-96.
- 38 Brömme, D.; Okamoto, K.; Wang, B. B.; Biroc, S. *J. Biol. Chem.* **1996**, *271*, 2126-2132.
- 39 Linnervers, C.; Smeekens, S. P.; Brömme, D. *FEBS Lett.* **1997**, *405*, 253-259.
- 40 Santamaria, I.; Velasco, G.; Pendas, A. M.; Paz, A.; Lopez-Otin, C. *J. Biol. Chem.*, **1999**, *274*, 13800-13809.
- 41 Wang, B.; Shi, G.-P.; Yao, P. M.; Li, Z.; Chapman, H. A.; Brömme, D. *J. Biol. Chem.* **1998**, *273*, 32000-32008.
- 42 Santamaria, I.; Velasco, G.; Cazorla, M.; Fueyo, A.; Campo, E.; Lopez-Otin, C. *Cancer Res.* **1998**, *58*, 1624-1630.
- 43 Santamaria, I.; Velasco, G.; Pendas, A. M.; Fueyo, A.; Lopez-Otin, C. *J. Biol. Chem.* **1998**, *273*, 16816-16823.
- 44 Kirschke, H.; Wiederanders, B.; Brömme, D.; Rinne, A. *Biochem. J.* **1989**, *264*, 467-473.
- 45 Karrer, K. M.; Pfeiffer, S. L.; di Thomas, M. E. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 3063-3067.
- 46 Cygler, M.; Mort, J. S. *Biochimie* **1997**, *79*, 645-652.
- 47 Schechter, I.; Berger, A. *Biochem. Biophys. Res. Commun.* **1967**, *27*, 157-162.
- 48 Drenth, J.; Kalk, K. H.; Swen, H. M. *Biochemistry* **1976**, *15*, 3731-3738.
- 49 Turk, D.; Podobnik, M.; Popovic, T.; Katunuma, N.; Bode, W.; Turk, V. *Biochemistry* **1995**, *34*, 4791-4797.
- 50 Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J. Mol. Biol.* **1977**, *112*, 535-542.
- 51 <http://rutgers.rcsb.org/pdb>
- 52 Otto, K. *Hoppe-Seylers Z. Physiol. Chem.* **1967**, *348*, 1449-1460.
- 53 Musil, D.; Zucic, D.; Turk, D.; Eng, R. A.; Mayr, I.; Huber, R.; Popovic, T.; Turk, V.; Towatari, T.; Katunuma, N.; Bode, W. *EMBO J.* **1991**, *10*, 2321-2330.
- 54 Illy, C.; Quraishi, O.; Wang, J.; Purisima, E.; Vernet, T.; Mort, J. S. *J. Biol. Chem.* **1997**, *272*, 1197-1202.
- 55 Nägler, D. K.; Storer, A. C.; Portaro, F. C. V.; Carmona, E.; Juliano, L.; Menard, R. *Biochemistry*, **1997**, *36*, 12608-12615.
- 56 Turk, D.; Podobnik, M.; Kuhelj, R.; Dolinar, M.; Turk, V. *FEBS Lett.* **1996**, *384*, 211-214.
- 57 Cygler, M.; Sivaraman, J.; Grochulski, P.; Coulombe, R.; Storer, A. C.; Mort, J. S. *Structure* **1996**, *4*, 405-416.
- 58 Schweiger, A.; Stabuc, B.; Popovic, T.; Turk, V.; Kos, J. *J. Immunol. Methods* **1997**, *201*, 165-232.

- 59 Fruton, J. S.; Irving, G. W.; Bergmann, M. *J. Biol. Chem.* **1941**, *138*, 9374-9381.
- 60 Baudys, M.; Meloun, B.; Gan-Erdene, T.; Fusek, M.; Mares, M.; Kosta, V.; Pohl, J.; Blake, C. C. *Biomed. Biochim. Acta* **1991**, *50*, 569-637.
- 61 Guncar, G.; Podobnik, M.; Pungercar, J.; Strukelj, B.; Turk, V.; Turk, D. *Structure* **1998**, *6*, 51-61.
- 62 Barrett, A. J.; Kirschke, H. *Methods Enzymol.* **1981**, *80*, 535-561.
- 63 Kirschke, H.; Langner, J.; Wiederanders, B.; Ansorge, S.; Bohley, P. *Eur. J. Biochem.* **1997**, *74*, 293-301.
- 64 Kirschke, H.; Kumbhavi, A. A.; Bohley, P.; Barrett, A. J. *Biochem. J.* **1982**, *201*, 367-372.
- 65 Bohley, P.; Kirschke, H.; Langner, J.; Miehe, M.; Riemann, S.; Slama, Z.; Schon, E.; Wiederanders, B.; Ansorge, S. Intracellular Proteolysis. In *Biological Functions of Proteins* (Holzer, H.; Tscheche, H., eds.) Springer-Verlag, Berlin **1979**, p. 17-18.
- 66 Millett, A. J.; Breen, S. A.; Loveday, B. E.; Clarkson, P. E.; Simpson, C. A.; Waterton, J. C.; Johnstone, J. *Bone*, **1997**, *5*, 465-471.
- 67 Stebbins, J. W.; Jaffe, H.; Fales, H. M.; Möller, J. R. *Biochemistry* **1997**, *36*, 2221-2226.
- 68 Yagel, S.; Warner, A. H.; Nellans, H. N.; Lala, P. K.; Waghorne, C.; Denhard, D. T. *Cancer Res.* **1989**, *49*, 3553-3557.
- 69 Kominami, E.; Tsukahara, T.; Hara, K.; Katunuma, N. *FEBS Lett.* **1988**, *231*, 225-228.
- 70 Tao, K.; Stearns, N. A.; Dong, J.; Wu, Q. L.; Sahagian, G. G. *Arch. Biochem. Biophys.* **1994**, *311*, 19-27.
- 71 Brömme, D.; Steinert, A.; Friebe, S.; Fittkau, S.; Wiederanders, B.; Kirschke, H. *Biochem. J.* **1989**, *264*, 475-481.
- 72 Turnsek, T.; Kregar, I.; Lebez, D. *Biochem. Biophys. Acta* **1975**, *403*, 514-520.
- 73 Petanceska, S.; Devi, L. *J. Biol. Chem.* **1992**, *267*, 26038-26043.
- 74 Quian, F.; Chan, S. J.; Gong, Q.; Bajkowski, A. S.; Steiner, D. F.; Frankfater, A. *Biomed. Biochim. Acta* **1991**, *50*, 4-6.
- 75 Brömme, D.; McGrath, M. E. *Protein Sci.* **1996**, *5*, 789-791.
- 76 Kirschke, H.; Wiederanders, B. *Methods Enzymol.* **1994**, *244*, 500-512.
- 77 McGrath, M. E.; Palmer, J. T.; Brömme, D.; Somoza, J. R. *Protein Sci.* **1998**, *7*, 1294-1302.
- 78 Shi, G. P.; Chapman, H. A.; Bhairi, S.; DeLeeuw, C.; Reddy, V. Y.; Weiss, S. J. *FEBS Lett.* **1995**, *357*, 129-134.
- 79 Li, Y. P.; Alexander, M. B.; Wucherpfennig, A. L.; Yelick, P.; Chen, W.; Stashenko, P. *J. Bone Mineral Res.* **1995**, *10*, 1197-1202.
- 80 Goto, T.; Tsukuba, T.; Kiyoshima, T.; Nishimura, Y.; Kato, K.; Yamamoto, K.; Tanaka, T. *Histochemistry* **1993**, *99*, 411-414.
- 81 Brömme, D.; Okamoto, K. *Biol. Chem. Hoppe-Seyler* **1995**, *376*, 379-384.
- 82 Drake, F. H.; Dodds, R. A.; James, I. E.; Connor, J. R.; Debouck, C.; Richardson, S.; Lee-Rykaczewski, E.; Coleman, L.; Rieman, D.; Barthlow, R.; Hastings, G.; Gowen, M. *J. Biol. Chem.* **1996**, *271*, 12511-12516.
- 83 Kafienah, W.; Brömme, D.; Buttle, D. J.; Croucher, L. J.; Hollander, A. P. *Biochem. J.* **1998**, *331*, 727-732.
- 84 Hou, W.-S.; Brömme, D.; Zhao, Y.; Mehler, E.; Dushey, C.; Weinstein, H.; Miranda, C. S.; Fraga, C.; Greig, F.; Carey, J.; Rimoin, D. L.; Desnick, R. J.; Gelb, B. D. *J. Clin. Invest.* **1999**, *103*, 731-738.
- 85 Bühling, F.; Gerber, A.; Häckel, C.; Krüger, S.; Köhnlein, T.; Brömme, D.; Reinhold, D.; Ansorge, S.; Welte, T. *Am. J. Respir. Cell Mol. Biol.* **1999**, *20*, 612-619.

- 86 Haeckel, C.; Krueger, S.; Buehling, F.; Broemme, D.; Franke, K.; Schuetze, A.; Roese, I.; Roessner, A. *Dev. Dyn.* **1999**, *216*, 89-95.
- 87 McQueeney, M. S.; Amegadzie, B. Y.; D'Alessio, K.; Hanning, C. R.; McLaughlin, M. M.; McNulty, D.; Carr, S. A.; Ijames, C.; Kurdyla, J.; Jones, C. S. *J. Biol. Chem.* **1997**, *272*, 13955-13960.
- 88 Baron, R. *Anat. Rec.* **1989**, *224*, 317-324.
- 89 Nägler, D.; Sulea, T.; Menard, R. *Biochem. Biophys. Res. Commun.* **1999**, *257*, 313-318.
- 90 Wex, T.; Levy, B.; Wex, H.; Brömme, D. *Biochem. Biophys. Res. Commun.* **1999**, *259*, 401-407.
- 91 Bairoch, A.; Apweiler, R. *J. Mol. Med.* **1997**, *75*, 312-316.
- 92 <http://www.ebi.ac.uk/swissprot.html>
- 93 Johnson, M. S.; Srinivasan, N.; Sowdhamini, R.; Blundell, T. L. *Crit. Rev. Biochem. Mol. Biol.* **1994**, *29*, 1-68.
- 94 <http://pipe.rockefeller.edu/modbase/>
- 95 Sternberg, M. J. E.; Islam, S. A. *Protein Eng.* **1990**, *4*, 125-131.
- 96 Sander, C.; Schneider, R. *Proteins* **1991**, *9*, 56-68.
- 97 Thompson, J. D.; Higgins, D. G.; Gibson, T. J. *Nucl. Acids Res.* **1994**, *22*, 4673-4680.
- 98 Saqi, M. A. S. Identification of Protein Sequence Motifs. In Protein Structure Prediction (Sternberg, M. J. E., ed.) Oxford University Press, **1996**, pp 65-78.
- 99 King, R. D. Prediction of secondary structure. In Protein Structure Prediction (Sternberg, M. J. E., ed.) Oxford University Press, **1996**, pp 79-99.
- 100 Rost, B. *Methods Enzymol.* **1996**, *266*, 525-539.
- 101 <http://dodo.cpms.columbia.edu/predictprotein/>
- 102 Rost, B.; Sander, C. *J. Mol. Biol.* **1993**, *232*, 584-599.
- 103 Blundell, T.; Carney, D.; Gardner, S.; Hayes, F.; Howlin, B.; Hubbard, T.; Overington, J.; Singh, D. A.; Sibanda, B. L.; Sutcliffe, M. *Eur. J. Biochem.* **1988**, *172*, 513-520.
- 104 Sali, A.; Blundell, T. L. *J. Mol. Biol.* **1993**, *234*, 779-815.
- 105 Srinivasan, N.; Guruprasad, K.; Blundell, T. L. Comparative Modelling of Proteins. In Protein Structure Prediction (Sternberg, M. J. E., ed.) Oxford University Press, **1996**, pp 112-140.
- 106 TRIPOS Associates Inc., 1699 S. Hanley Road, Suite 303, St. Louis, MO 63144, USA.
- 107 <http://www.tripos.com>
- 108 Altschul, S. F.; Madden, T. L.; Schäffer, A. A.; Jinghui, Z.; Zhang, Z.; Zhang, Z.; Miller, W.; Lipman, D. J. *Nucleic Acids Res.* **1997**, *25*, 2289-3402.
- 109 <http://www.ncbi.nlm.nih.gov/BLAST>
- 110 Needleman, S. B.; Wunsch, C. D. *J. Mol. Biol.* **1970**, *48*, 443-453.
- 111 Sutcliffe, M. J.; Haneef, I.; Carney, D.; Blundell, T. L. *Protein Eng.* **1987**, *1*, 377-384.
- 112 Brooks, B. R.; Brucoleri, R. E.; Olafson, B. D.; States, D.; Swaminathan, S.; Karplus, M. *J. Comp. Chem.* **1983**, *3*, 187-217.
- 113 Gasteiger, J.; Marsili, M. *Tetrahedron* **1980**, *36*, 3219-3238.
- 114 Clark, M.; Cramer III, R. D.; VanOpdenbosch, N. *Comput. Chem.* **1989**, *10*, 982-1012.
- 115 Sippl, M. *Proteins* **1993**, *17*, 355-362.
- 116 Sippl, M. *J. Comp. Aided-Mol. Des.* **1993**, *7*, 473-501.
- 117 Varughese, K. I.; Su, Y.; Cromwell, D.; Hasnain, S.; S.; Xuong, N.-H. *Biochemistry* **1992**, *31*, 5172-5176.

- 118 Coulombe, R.; Grochulski, P.; Sivaraman, J.; Menard, R.; Mort, J. S.; Cygler, M. *EMBO J.* **1996**, *15*, 5492-5503.
- 119 Laskowski, R. A.; McArthur, M. W.; Moss, D. S.; Thornton, J. M. *J. Appl. Cryst.* **1993**, *26*, 283-289.
- 120 Stryer, L. Struktur und Funktion von Proteinen. In Biochemie. (Wigger, F.; Behncke, M.; Hrsg.) Spektrum Akademischer Verlag GmbH Heidelberg, Berlin, Oxford **1994**, pp. 15-44.
- 121 Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. *J. Mol. Biol.* **1982**, *161*, 268-288.
- 122 Goodsell, D. S.; Olson, A. J. *Proteins* **1990**, *8*, 195-202.
- 123 Judson, R. Genetic Algorithms and their Use in Chemistry. In Reviews in Computer Chemistry (Lipkowitz, K. B.; Boyd, D. B., eds) VCH Publishers **1997**, Vol. 10, pp 1-73.
- 124 Wang, R.; Liu, L.; Lai, L.; Tang, Y. *J. Mol. Model.* **1998**, *4*, 379-394.
- 125 Fischer, E. *Ber. Dtsch. Chem. Ges.* **1894**, *27*, 2984-2993.
- 126 Sowdhamini, R.; Srinivasan, N.; Guruprasad, K.; Rufino, S.; Dhanaraj, V.; Wood, S. P.; Emsley, J.; White, H. E.; Blundell, T. *Pharm. Acta Helv.* **1995**, *69*, 185-192.
- 127 Goodford, P. J. *J. Med. Chem.* **1985**, *28*, 849-857.
- 128 Heiden, W.; Goetze, T.; Brickmann, J. *J. Comput. Chem.* **1993**, *14*, 246-250.
- 129 Connolly, M. L. *J. Appl. Crystallogr.* **1983**, *16*, 548-558.
- 130 Yamashita, D. S.; Smith, W. W.; Zhao, B.; Janson, C. A.; Tomaszek, T. A.; Bossard, M. J.; Levy, M. A.; Oh, H.-J.; McQueney, M.; D'Alessio, K. J.; Amegadzie, B. Y.; Hanning, C. R.; Abdel-Meguid, S. S.; Desjarlais, R. L.; Gleason, J. G.; Veber, D. F. *J. Am. Chem. Soc.* **1997**, *119*, 11351-11352.
- 131 LaLonde, J. M.; Zhao, B.; Smith, W. W.; Janson, C. A.; Desjarlais, R. L.; Tomaszek, T. A.; Carr, S. A.; Thompson, S. K.; Yamashita, D. S.; Veber, D. F.; Abdel-Meguid, S. S. *J. Med. Chem.* **1998**, *41*, 4567-4576.
- 132 McGrath, M. E.; Klaus, J. L.; Barnes, M. G.; Brömme, D. *Nature Struct. Biol.* **1997**, *4*, 105-109.
- 133 Taylor, M. A. J.; Pratt, K. A.; Revell, D. F.; Baker, K. C.; Sumner, I. G.; Goodenough, P. W. *Protein Eng.* **1992**, *5*, 455-459.
- 134 Kamphuis, I. G.; Drenth, J.; Baker, E. N. *J. Mol. Biol.* **1985**, *182*, 317-329.
- 135 Aibe, K.; Yazawa, H.; Abe, K.; Teramura, K.; Kumegawa, M.; Kawashima, H.; Honda, K.; *Biol. Pharm. Bull.* **1996**, *19*, 1026-1031.
- 136 Brömme, D.; Klaus, J. L.; Okamoto, K.; Rasnick, D.; Palmer, J. T. *Biochem. J.* **1996**, *315*, 85-89.
- 137 Takio, K.; Towatari, T.; Katunuma, N.; Teller, D. C.; Titani, K. *Proc. Natl. Acad. Sci USA* **1983**, *80*, 3666-3670.
- 138 Baudys, M.; Meloun, B.; Gan-Erdene, T.; Fusek, M.; Mares, M.; Kostka, V.; Pohl, J.; Blake, C. C. *Biomed. Biochim. Acta* **1991**, *50*, 569-577.
- 139 Sivaraman, J.; Lalumiere, M.; Menard, R.; Cygler, M. *Protein Sci.* **1999**, *8*, 283-290.
- 140 persönliche Mitteilung
- 141 McGrath, M. E.; Eakin, A. E.; Engel, J. C.; McKerrow, J. H.; Craik, C. S.; Fletterick, R. *J. Mol. Biol.* **1995**, *247*, 251-259.
- 142 Gillmor, S. A.; Craik, C.; Fletterick, R. *J. Protein Sci.* **1997**, *6*, 1603-1611.
- 143 O'Hara, B. P.; Hemmings, A. M.; Buttle, D. J.; Pearl, L. *Biochemistry* **1995**, *34*, 13190-13195.
- 144 Schroeder, E.; Phillips, C.; Garman, E.; Harlos, K.; Crawford, C. *FEBS Lett.* **1993**, *315*, 38-42.

- 145 Jia, Z.; Hasnain, S.; Hiram, T.; Lee, X.; Mort, J. S.; To, R.; Huber, R. C. *J. Biol. Chem.* **1995**, *270*, 5527-5533.
- 146 Zhao, B.; Janson, C. A.; Amegadzie, B. Y.; D'Alessio, K.; Griffin, C.; Hanning, C. R.; Jones, C.; Kurdyla, J.; McQueney, M.; Qiu, X.; Smith, W. W.; Abdel-Meguis, S. S. *Nat. Struct. Biol.* **1997**, *4*, 109-111.
- 147 Thompson, S. K.; Halbert, S. M.; Bossard, M. J.; Tomaszek, T. A.; Levy, M. A.; Zhao, B.; Smith, W. W.; Abdel-Meguis, S.; Janson, C. A.; D'Alessio, K. J.; McQueney, M.; Amegadzie, B. Y.; Hanning, C. R.; DesJarlais, R. L.; Briand, J.; Sarkar, S. K.; Huddleston, M. J.; Ijames, C. F.; Carr, S. A.; Garnes, K. T.; Shu, A. *Proc. Nat. Acad. Sci. USA* **1994**, *94*, 14249-14254.
- 148 DesJarlais, R. L.; Yamashita, D. S.; Oh, H.-J.; Uzinskas, I. N.; Erhard, K. F.; Allen, A. C.; Haltiwanger, R. C.; Zhao, B. G.; Smith, W. W.; Abdel-Meguis, S. S.; D'Alessio, K. J.; McQueney, M.; Tomaszek, T. A.; Levy, M. A.; Veber, D. F. *J. Am. Chem. Soc.* **1998**, *120*, 9114-9115.
- 149 Yamamoto, A.; Hara, T.; Tomoo, K.; Ishida, T.; Fujii, T.; Hata, Y.; Murata, M.; Kitamura, K. *J. Biochem.* **1997**, *121*, 974-977.
- 150 Brömme, D.; Bonneau, P. R.; Lachance, P.; Storer, A. *J. Biol. Chem.* **1994**, *269*, 30238-30242.
- 151 Gershkovich, A. A.; Kholodovych, V. V. *J. Biochem. Biophys. Methods* **1996**, *33*, 135-162.
- 152 Knight, C. G. *Methods Enzymol.* **1995**, *248*, 18-34.
- 153 Bossard, M. J.; Tomaszek, T. A.; Thompson, S. K.; Amegadzie, B. Y.; Hannings, C. R.; Jones, C.; Kurdyla, J. T.; McNulty, D. E.; Drake, F. H.; Gowen, M.; Levy, M. A. *J. Biol. Chem.* **1996**, *271*, 12517-12524.
- 154 Votta, B. J.; Levy, M. A.; Badger, A.; Bradbeer, J.; Dodds, R. A.; James, I. E.; Thompson, S.; Bossard, M. J.; Carr, T.; Connor, J. R.; Tomaszek, T. A.; Szweczuk, L.; Drake, F. H.; Veber, D. F.; Gowen, M. *J. Bone Min. Res.* **1997**, *12*, 1396-1406.
- 155 Fengler, A.; Brandt, W. *Prot. Eng.* **1998**, *11*, 1007-1013.
- 156 Fengler, A.; Brandt, W. *J. Mol. Model.* **1999**, *5*, 177-188.
- 157 Fengler, A.; Brandt, W. Development and Validation of Homology Models of Human Cathepsins K, S, H, and F. In *Advances in Experimental Medicine and Biology: Cellular Peptidases in Immune Functions and Diseases (2)* (Ansorge, S.; Langner, J., eds.) Plenum Press, im Druck.
- 158 LaLonde, J. M.; Zhao, B.; Janson, C. A.; D'Alessio, K. J.; McQueney, M. S.; Orsini, M. J.; Debouck, C. M.; Smith, W. W. *Biochemistry* **1999**, *38*, 862-869.
- 159 Guncar, G.; Pungercic, G.; Klemencic, I.; Turk, V.; Turk, D. *EMBO J.* **1999**, *18*, 793-803.
- 160 Selzer, P. M.; Chen, X.; Chan, V. J. Cheng, M.; Kenyon, G. L.; Kuntz, I. D.; Saknari, J. A.; Cohen, F. E.; McKerrow, J. H. *Exp. Parasitol.* **1997**, *87*, 212-221.
- 161 Groves, M. R.; Coulombe, R.; Jenkins, J.; Cygler, M. *Proteins: Struct. Funct. Genet* **1998**, *32*, 504-514.
- 162 Yamamoto, A.; Matsumoto, K.; Ohishi, H.; Ishida, T.; Inoue, M.; Kitamura, K.; Mizuno, H. *J. Biol. Chem.* **1991**, *266*, 14771-14777.
- 163 Yamamoto, A.; Tomoo, K.; Doi, M.; Ohishi, H.; Inoue, M.; Ishida, T.; Yamamoto, D.; Okamoto, H.; Okada, Y. *Biochemistry* **1992**, *31*, 11305-11309.
- 164 Pickersgill, R. W.; Harris, G. W.; Garman, E. *Acta Crystallogr., Sect. B* **1992**, *48*, 59-67.
- 165 Kim, M.-J.; Yamamoto, D.; Matsumoto, K.; Inoue, M.; Ishida, T.; Mizuno, H.; Sumiya, S.; Kitamura, K. *Biochem. J.* **1992**, *287*, 797-803.



- 
- 166 Stubbs, M. T.; Laber, B.; Bode, W.; Huber, R.; Jerala, R.; Lenarciv, B.; Turk, V. *EMBO J.* **1990**, *9*, 1939-1947.
- 167 Priestle, J. P.; Ford, G. C.; Glor, M.; Mehler, E. L.; Smit, J. D. G.; Thaller, C.; Jansonsius, J. N. *Acta Crystallogr., Sect. A (Suppl.)* **1984**, *40*, 17-24.
- 168 Maes, D.; Bouckaert, J.; Poortmans, F.; Wyns, L.; Looze, Y. *Biochemistry* **1996**, *35*, 16292-16298.
- 169 Baker, E. N.; Dodson, E. J. *Acta Crystallogr., Sect. A* **1980**, *36*, 559-572.

**Tab. A-1** Proteinstrukturen von Cysteinproteasen der Papain-Superfamilie C1, eingetragen in der Brookhaven Proteindatenbank [50,51] (Stand November 1999)

Enzym	PDB-Eintrag	Organismus	Auflösung [Å]	Literatur	Datum der Verfügbarkeit	Bemerkungen
Cathepsin B	1CSB	Homo sapiens	2.10	[49]	03.04.1996	Komplex mit CA030
	1HUC	Homo sapiens	2.10	[53]	26.01.1995	-
	1CPJ	Rattus norvegicus	2.20	[145]	07.12.1995	Mutation: S115A
	1CTE	Rattus norvegicus	2.10	[145]	31.07.1995	Mutation: S115A, Thiolgruppe von Cys29 ist blockiert
	1THE	Rattus norvegicus	1.90	[145]	10.03.1996	Komplex mit einem Chlormethylketon-Inhibitor, Mutation: S115A
Procathepsin B	1QDQ	Bos taurus	2.18	[149]	ca. 07. 2000	Komplex mit CA074
	1MIR	Rattus norvegicus	2.80	[57]	01.11.1997	Mutationen: C29S, S115A
	1PBH	Homo sapiens	3.20	[56]	25.02.1998	-
	2PBH	Homo sapiens	3.30	[56]	25.02.1998	-
	3PBH	Homo sapiens	2.50	[56]	25.02.1998	-
Cathepsin F	1D5U	Homo sapiens	-	[157]	ca. 2000	Strukturmodell
Cathepsin H	1BZN	Homo sapiens	-	[156]	18.10.1999	Strukturmodell mit Minikette
	8PCH	Sus scrofa	2.10	[61]	09.12.1998	Röntgenkristallstruktur mit Minikette
Cathepsin K	1ATK	Homo sapiens	2.20	[146]	04.04.1998	Komplex mit E-64
	1AU0	Homo sapiens	2.60	[130]	14.10.1998	Komplex mit einem Diacylaminomethylketon-Inhibitor
	1AU2	Homo sapiens	2.60	[130]	14.10.1998	Komplex mit einem Propanon-Inhibitor
	1AU3	Homo sapiens	2.50	[130]	14.10.1998	Komplex mit einem Pyrrolidinon-Inhibitor
	1AU4	Homo sapiens	2.30	[130]	14.10.1998	Komplex mit einem Pyrrolidinon-Inhibitor
	1AYU	Homo sapiens	2.20	[147]	25.11.1998	Komplex mit einem Biscarbohydrazid-Inhibitor
	1AYV	Homo sapiens	2.30	[147]	25.11.1998	Komplex mit einem Thiazolhydrazid-Inhibitor
	1AYW	Homo sapiens	2.40	[147]	25.11.1998	Komplex mit einem Benzyloxycarbohydrazid-Inhibitor
		1BGO	Homo sapiens	2.30	[148]	08.06.1999
Procathepsin K	1MEM	Homo sapiens	1.80	[132]	14.08.1998	Komplex mit einem Vinylsulfoninhibitor
	1BY8	Homo sapiens	2.80	[158]	24.10.1999	-
	7PCK	Homo sapiens	3.20	[139]	25.10.1999	-

Fortsetzung der Tab. A-1

Enzym	PDB-Eintrag	Organismus	Auflösung [Å]	Literatur	Datum der Verfügbarkeit	Bemerkungen
Cathepsin L	1ICF	Homo sapiens	2.20	[159]	ca. 01. 2000	-
Cathepsin L-ähnlich	1BMJ	Leishmania major	-	[160]	29.07.1998	Strukturmodell
Procathepsin L	1CJL	Homo sapiens	2.20	[118]	12.08.1997	Mutationen: F(78P)L, C25S, T110A, E176G, D178G
	1CS8	Homo sapiens	1.80	[161]	23.08.1999	Mutation: T110A
Cathepsin S	1BXF	Homo sapiens	-	[155]	07.10.1998	Strukturmodell
Papain	1BP4	Carica papaya	2.20	[131]	12.08.1999	Komplex mit einem Aldehyd-Inhibitor
	1BQI	Carica papaya	2.50	[131]	16.08.1999	Komplex mit einem Methoxymethylketon-Inhibitor
	1PAD	Carica papaya	2.80	[48]	12.04.1977	Komplex mit Chlormethylketon-Inhibitor
	1PE6	Carica papaya	2.10	[162]	15.04.1993	Komplex mit E-64c
	1PIP	Carica papaya	1.70	[163]	31.10.1993	Komplex mit Suc-Gln-Val-Val-Ala-Ala-pNA
	1POP	Carica papaya	2.10	[144]	31.10.1993	Komplex mit Leupeptin
	1PPN	Carica papaya	1.60	[164]	31.01.1991	Thiolgruppe von Cys25 ist oxidiert
	1PPP	Carica papaya	1.90	[165]	31.01.1991	Komplex mit E-64c
	1STF	Carica papaya	2.37	[166]	31.01.1991	Komplex mit Stefin B
	2PAD	Carica papaya	2.80	[48]	04.12.1977	Komplex mit Chlormethylketon-Inhibitor
	4PAD	Carica papaya	2.80	[48]	04.12.1977	Komplex mit Chlormethylketon-Inhibitor
	5PAD	Carica papaya	2.80	[48]	04.12.1977	Komplex mit Chlormethylketon-Inhibitor
	6PAD	Carica papaya	2.80	[48]	04.12.1977	Komplex mit Chlormethylketon-Inhibitor
	9PAP	Carica papaya	1.65	[19]	24.10.1986	Thiolgruppe von Cys25 ist oxidiert
2-Hydroxyethyl-papain	1PPD	Carica papaya	2.00	[167]	02.01.1985	-
Chympopapain	1YAL	Carica papaya	1.70	[168]	23.12.1996	Thiolgruppen von Cys25 und Cys117 sind mit einer Methylgruppe geschützt
Actinidin	1AEC	Acinidia chinensis	1.86	[117]	31.10.1993	Komplex mit E-64
	2ACT	Acinidia chinensis	1.70	[169]	07.03.1980	-
Gycylendopeptidase	1GEC	Carica papaya	2.10	[143]	07.12.1995	Komplex mit Z-Leu-Val-Gly-Methylen-Inhibitor
Cruzain	1AIM	Trypanosoma cruzi	2.00	[141]	22.10.1997	Komplex mit einem Fluormethylketon-Inhibitor
	2AIM	Trypanosoma cruzi	2.20	[142]	22.10.1997	Komplex mit einem Fluormethylketon-Inhibitor

**Tab. A-2** Allgemeine Informationen über die Cathepsine B, H, L, S, K und F, entnommen aus der Swiss-Prot-Datenbank [91,92] und der TrEMBL-Datenbank [92]

	Cathepsin B	Cathepsin H	Cathepsin L	Cathepsin S	Cathepsin K	Cathepsin F
Zugriffsnummer	P07858	P09668	P07711	P25774	P43235	O95240 <sup>b</sup>
E.C.-Nummer	3.4.22.1	3.4.22.16	3.4.22.15	3.4.22.27	3.4.22.38	
Synonyme	Cathepsin B1 APP Secretase		Major Excreted Protein (MEP)		Cathepsin O2	
Organismus	Mensch	Mensch	Mensch	Mensch	Mensch	Mensch
Sequenzlänge <sup>a</sup>	339 aa	335 aa	333 aa	331 aa	329 aa	489 aa <sup>c</sup>
Molmasse <sup>a</sup> [Da]	37,807	37,403	37,564	37,479	36,966	53,365 <sup>c</sup>
Signalsequenz	Met1-Ala17	Met1-Ala22	Met1-Ala17	Met1-Ala16	Met1-Ala15	Met1-Ala19 <sup>c</sup>
Propeptidsequenz	Arg18-Lys79	Ala23-Ser97	Thr18-Glu113	Gln17-Ile114	Leu16-Arg114	Ala20-Leu270 <sup>c</sup>
Minikette		Glu98-Thr105				
Gesamtkette	Leu80-Asp333	Tyr116-Val335	Ala114-Val333	Leu115-Ile331	Ala115-Met329	Ala271-Asp484 <sup>c</sup>
leichte Kette	Leu80-Asn126	Tyr116-Asn292	Glu289-Asp291			
schwere Kette	Val129-Asp333	Gly293-Val335	Asn292-Val333			
Propeptid	Gln334-Ile339					
katalytische Triade	Cys108 His278 Asn298	Cys141 His281 Asn301	Cys138 His276 Asn300	Cys139 His278 Asn298	Cys139 His276 Asn296	Cys295 <sup>c</sup> His411 <sup>c</sup> Asn431 <sup>c</sup>
Disulfidbrücken	Cys93-Cys122 Cys105-Cys150 Cys141-Cys207 Cys142-Cys146 Cys179-Cys211 Cys187-Cys198	Cys102-Cys327 Cys138-Cys181 Cys172-Cys214 Cys272-Cys322	Cys135-Cys178 Cys169-Cys211 Cys269-Cys322	Cys126-Cys224 Cys136-Cys180 Cys170-Cys213 Cys272-Cys320	Cys136-Cys177 Cys170-Cys210 Cys269-Cys318	Cys292-Cys333 <sup>c</sup> Cys326-Cys366 <sup>c</sup> Cys403-Cys452 <sup>c</sup>
weitere bekannte Sequenzen (Organismen)	Rind, Huhn, Maus, Ratte, japanische Wachtel	Ratte, Maus	Rind, Ratte, Huhn, Maus, Blutegel, Katze, Schwein, Schaf	Rind, Ratte	Maus, Hase	

<sup>a</sup> diese Angaben sind vom Präproenzym; <sup>b</sup> Zugriffsnummer der TrEMBL-Datenbank; <sup>c</sup> entnommen aus [40]

**Tab. A-3** Proteinsequenzhomologie und -identität der Cathepsine K, L, F, S, H, B sowie Papain und Actinidin (ermittelt unter Verwendung von BLAST [108] und CLUSTAL W [97])

Identität \ Homologie	hCatK	hCatL	hCatF	hCatS	hCatH	hCatB	Papain	Actinidin
hCatK	100	58	39	57	46	21	41	38
hCatL	74	100	41	56	40	28	38	40
hCatF	56	57	100	33	38	25	35	40
hCatS	71	73	54	100	44	28	40	46
hCatH	61	62	54	60	100	28	41	39
hCatB	43	43	42	47	47	100	26	21
Papain	57	54	51	57	55	41	100	49
Actinidin	59	58	52	62	51	44	58	100

**Tab. A-4** Atomabstände zwischen den Aminosäuren der katalytischen Triade, gemessen in den Proteinstrukturen von lysosomalen Cysteinproteasen

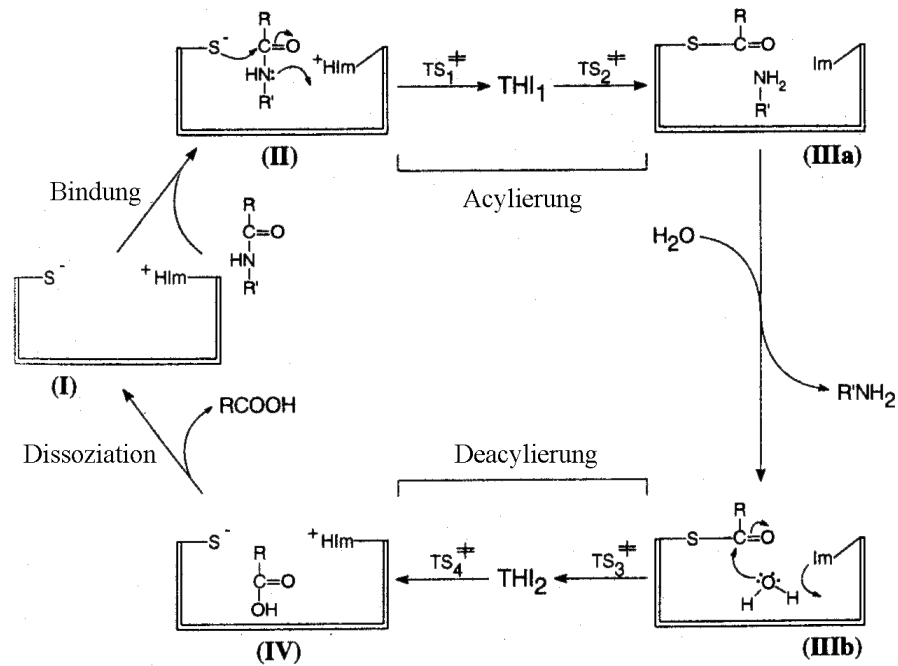
Enzym	PDB-Eintrag	CysSγ...HisNδ1-Abstand [Å]	HisNε2...AsnOδ1-Abstand [Å]
Röntgenkristallstrukturen			
Cathepsin B	1HUC	3.48	2.53
Cathepsin H	8PCH	5.68 <sup>a</sup>	7.42 <sup>a</sup>
Cathepsin K	1MEM	3.65	2.77
Cathepsin L	1CJL	3.32 <sup>b</sup>	2.81
		3.34 <sup>c</sup>	
Papain	9PAP	3.67	2.77
Actinidin	1AEC	3.88	2.88
Strukturmodelle			
Cathepsin K	-	3.75	3.26
Cathepsin S	1BXF	3.99	3.01
Cathepsin H	1BZN	3.78	3.37
Cathepsin F	1D5U	3.41	2.91

<sup>a</sup> Seitenkette von His166 ist um ca. 90° gedreht

<sup>b</sup> Röntgenstruktur von Cathepsin L mit Ser25

<sup>c</sup> Ser25 gegen ein Cystein ausgetauscht

## 7.2 Schemata



**Schema A-1** Darstellung der verschiedenen Schritte, die bei der Hydrolyse von Substraten durch Cysteinteasen auftreten (in Anlehnung an [17]).

Der erste Schritt der Reaktion korrespondiert mit der Assoziation (bzw. nichtkovalente Bindung) des freien Enzyms (I) und des Substrates zur Bildung des Michaelis-Komplexes (II). Daran schließt sich die Acylierung des Enzyms (IIIa), mit der Ausbildung des Acylenzyms und der Freisetzung des Amins  $R'NH_2$  an. Das Acylenzym (IIIb) reagiert mit einem Wassermolekül, und es entsteht die Carbonsäure (Deacylierungsprozeß IV). Die Abdiffusion dieses Produktes bewirkt die Regenerierung des freien Enzyms. Es wird angenommen, daß die Acylierung und Deacylierung die Bildung tetrahedraler Intermediate ( $THI_1$  und  $THI_2$ ) beinhaltet. Zusätzlich treten im Katalysemechanismus vier Übergangszustände ( $TS_1^\ddagger$ ,  $TS_2^\ddagger$ ,  $TS_3^\ddagger$ ,  $TS_4^\ddagger$ ) auf.

Für die berechneten Wechselwirkungsenergien ( $\Delta E$ ) wird folgende Näherung angenommen:

$$\Delta G_0 = -RT \ln K = \Delta H_0 - T\Delta S_0 \approx \Delta E$$

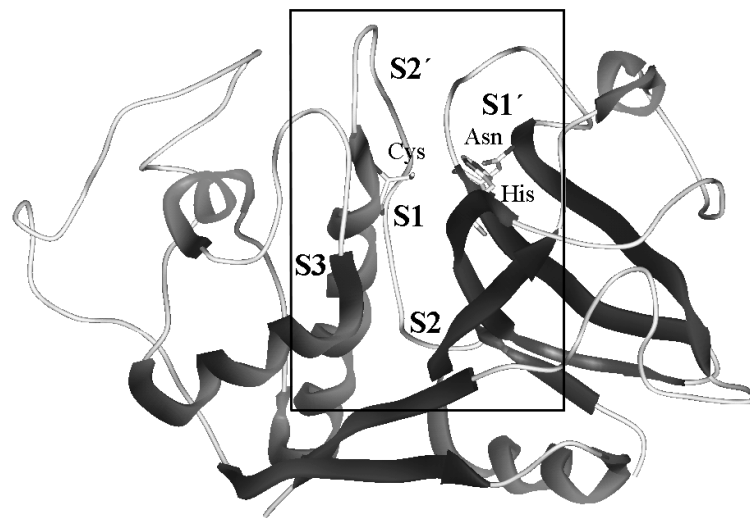
Hierbei ist  $\Delta G_0$  die Differenz in der Freien Enthalpie (Gibbs'sche Freie Enthalpie) zwischen den Spezies des Gleichgewichtes, definiert durch die Gleichgewichtskonstante  $K$ .  $\Delta H_0$  ist die Enthalpiedifferenz und  $\Delta S_0$  die entsprechende Entropiedifferenz.  $R$  stellt die Gaskonstante und  $T$  die Temperatur des Systems dar.

In der vorliegenden Arbeit werden zahlreiche Untersuchungen über das Wechselwirkungsverhalten von Liganden im Enzym diskutiert. Die wesentliche Rolle spielt dabei die Bindungsaffinität des Liganden im Enzym, d. h. es kann nur die Stärke der Bindung der Liganden untersucht werden. Im Katalysezyklus wird die Geschwindigkeitskonstante auch durch die Geschwindigkeitskonstante nachfolgender Prozesse bestimmt. Anhand der theoretischen Untersuchungen können keine Aussagen über  $k_{\text{cat}}$  getroffen werden. Es lassen sich auch keine vollständigen Schlußfolgerungen über  $K_m$  treffen, da die Michaelis-Konstante von zahlreichen Effekten während der Hydrolyse eines Substrates abhängig ist, die in die theoretischen Studien nicht mit einbezogen werden können.

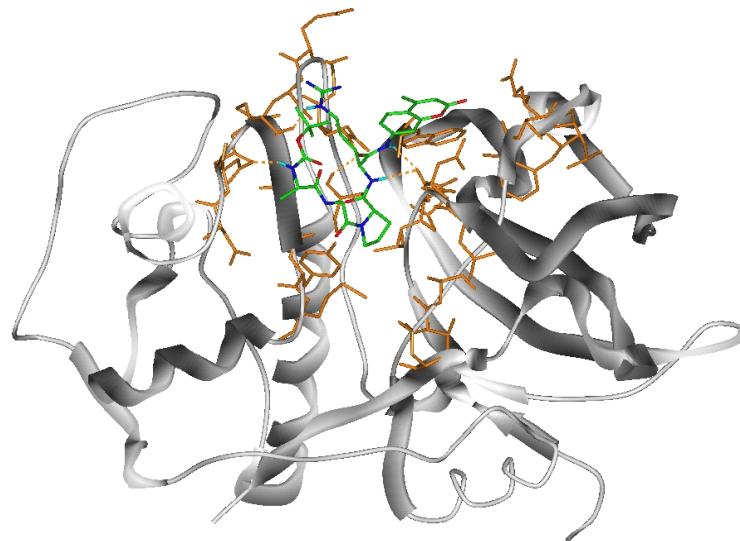
Nur unter bestimmten Bedingungen ist  $K_i$  proportional  $K_m$  zu setzen. Für die untersuchten Substrate sind die geschwindigkeitsbestimmenden Schritte nicht bekannt, so daß nicht gesagt werden kann, inwieweit die Reaktionsgeschwindigkeitskonstante beeinflußt wird. Es wurde anhand des Vergleiches der berechneten Werte mit experimentell ermittelten Daten angenommen, daß für alle untersuchten Verbindungen gilt:

$$K_m \sim K_i \text{ bzw. } K_m \sim K_i \approx K_d$$

### 7.3 Abbildungen



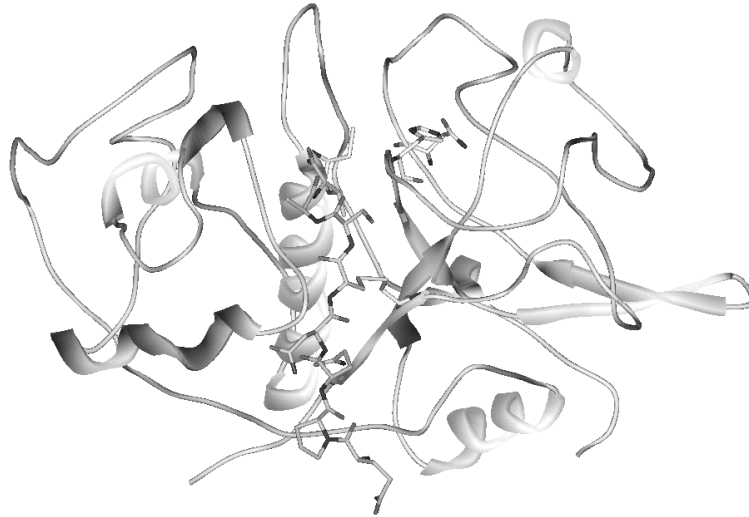
**Abb. A-1** Proteinstruktur von Papain (PDB-Eintrag: 9PAP [19]) mit Kennzeichnung der Bindungspositionen S1-S3 und S1'-S2' für Substrate catheptischer Cysteinproteasen. Die Aminosäuren der katalytischen Triade (Cystein, Histidin, Asparagin) sind eingeblendet.



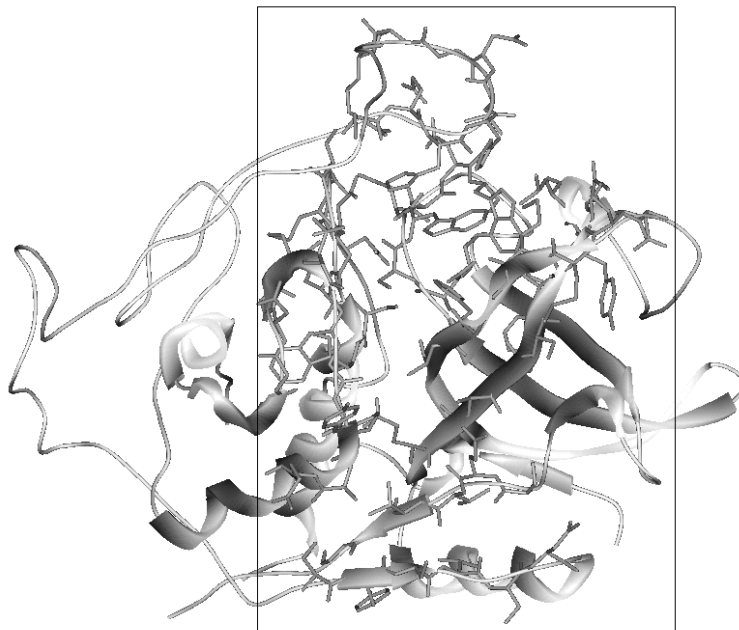
**Abb. A-2** Enzym-Substrat-Komplex von hCatK mit Boc-Ala-Gly-Pro-Arg-MCA. Es sind nur die Aminosäurereste von hCatK sichtbar, die im katalytisch aktiven Bindungszentrum des Cathepsins liegen bzw. die unmittelbar benachbart sind. Die Atome des Enzyms sind orange gefärbt und die des Substrates grün.

Diese Abbildung dient zur Orientierung über die Lage eines Liganden in der Bindungsspalte der Cathepsine. Als Beispiel wurde die Tertiärstruktur von Cathepsin K dargestellt.





**Abb. A-3** Darstellung der Proteinstruktur von hCatH mit seiner Minikette. Die Atome der Minikette sind dunkel hervorgehoben. Die Aminosäuren der katalytischen Triade (Cys26, His166, Asn186) sowie Cys212 wurden mit dargestellt.



**Abb. A-4** Proteinstruktur von Cathepsin B (PDB-Eintrag: 1HUC [53]). Von den Aminosäuren, die sich innerhalb der Umrahmung befinden, wurde das EP der molekularen Oberfläche des Enzyms berechnet. Für Cathepsin B sind diese Aminosäuren eingeblendet. In Analogie dazu wurden für die gleichen Bereiche die EP der Cathepsine L, K, H, S und F berechnet.

	10	20	30	40	50	60
			*			
hCatK	APDSVDYRKK	GYVTPVKNQG	QCGSCWAFSS	VGALEGQLKK	KTGKLLNLSL	QNLVDCVSEN
PHD	----HHHH--	----E-----	--HHHHHHHH	HHHHH-EEE-	----EEEE--	HHHH-----
1MEM	-----	-----	----HHHHHH	HHHHHHHHHH	HH-----H	HHHHHH----
<b>Modell</b>	-----	-----	---- <b>HHHHHH</b>	<b>HHHHHHHHHH</b>	<b>HH-----H</b>	<b>HHHHHH----</b>
	70	80	90	100	110	120
hCatK	DGCGGGYMTN	AFQYVQKNRG	IDSEDAYPYV	GQEESCMYNP	TGKAAKCRGY	REIPEGNEKA
PHD	----HHHHH	HHHHHHHH--	-----	-----	----EEE--	EE----HHHH
1MEM	-HHH---HHH	HHHHHHHH--	-----	-----H	HH-----E	EE----HHH
<b>Modell</b>	<b>-HHH---HHH</b>	<b>HHHHHHHH--</b>	-----	-----	----- <b>EE</b>	<b>EE----HHH</b>
	130	140	150	160	170	180
				*		
hCatK	LKRAVARVGP	VSVAIDASLT	SFQFYSGKVY	YDESCNSDNL	NHAVLAVGYG	IQKGNKHWI I
PHD	HHHHHHH--	EEEEEE--H	HHEEE--EE	E-----	--EEEEEE--	-----EEEE
1MEM	HHHHHHH--	EEEE--H	HH-----	-----	-EEEEEEEE	EE--EEEE
<b>Modell</b>	<b>HHHHHHH--</b>	<b>EEEE--H</b>	<b>HH-----</b>	-----	<b>-EEEEEEEE</b>	<b>EE--EEEE</b>
	190	200	210	220		
	*					
hCatK	KNSWGENWGN	KGYILMARNK	NNACGIANLA	SFPKM		
PHD	E-----	--EEEE--	---E--EE--	-----		
1MEM	E-----	--EEEE--	---HH-----	EEE--		
<b>Modell</b>	<b>E-----</b>	<b>--EEEE--</b>	<b>---HH-----</b>	<b>EEEE-</b>		

**Abb. A-5** Aminosäuresequenz von hCatK. Die Sekundärstrukturelemente, die aus der Berechnung mit PHD resultieren sind grau, die der Röntgenstruktur dieses Enzyms (PDB-Eintrag: 1MEM [132]) sind schwarz und die Strukturelemente des Strukturmodells von hCatK fett hervorgehoben (H = Helix, E =  $\beta$ -Faltblattstrang). Die Aminosäurereste der katalytischen Triade sind mit einem Stern gekennzeichnet.

	10	20	30	40	50	60
			*			
hCatS	LPDSVDWREK	GCVTEVKYQG	SCGACWAFSA	VGALEAQLKL	KTGKLVTLA	QNLVDCSTEK
PHD	-----	-----	--HHHHHHH	HHHHHHEEEE	----EEEE--	HHHH-----
<b>Modell</b>	----- <b>HHH-</b>	-----	---- <b>HHHHHH</b>	<b>HHHHHHH</b> ----	----- <b>H</b>	<b>HHHHHH----</b>
	70	80	90	100	110	120
hCatS	YGNGKCGNGF	MTTAFQYIID	NKGIDSDASY	PYKAMDQKCQ	YDSKYRAATC	SKYTELPYGR
PHD	-----HH	HHHHHHHHHH	H-----	-----	-----EEEE	E-EE-----
<b>Modell</b>	-----	<b>HHHHHHHHHH</b>	<b>H-----</b>	-----	-----	---- <b>EE-----</b>
	130	140	150	160	170	180
					*	
hCatS	EDVLKEAVAN	KGPVSVGVDA	RHPSFFLYRS	GVYYEPSCTQ	NVNHGVLVVG	YGDLNGKEYW
PHD	HHHHHHHHHH	---EEEE--	--HHHEEE--	-EE-----	---EEEEEE	E-----EE
<b>Modell</b>	<b>HHHHHHHHHH</b>	---- <b>EEEE--</b>	--- <b>HHHH--</b>	-----	---- <b>EEEEEE</b>	<b>EE-----EE</b>
	190	200	210	220		
	*					
hCatS	LVKNSWGHNF	GEEGYIRMAR	NKGNHCGIAS	FPSYPEI		
PHD	EEE-----	---EEEE--	-----EE--	-----		
<b>Modell</b>	<b>EEE-----</b>	---- <b>EEEE--</b>	---- <b>HHH--</b>	<b>--EE---</b>		

**Abb. A-6** Aminosäuresequenz von hCatS. Die darauffolgenden Zeilen enthalten die Sekundärstrukturelemente von hCatS, die mit PHD (grau) und anhand des Proteinmodells (fett) mit DSSP berechnet wurden (H = Helix, E =  $\beta$ -Faltblattstrang). Die Aminosäuren der katalytischen Triade sind mit einem Stern gekennzeichnet.

	10	20	30	40	50	60
hCatH	YPPSVDWRKK	GNFVSPVKNQ	GACGSCWTF <sup>*</sup> S	TTGALES <sup>*</sup> AI <sup>*</sup> A	IATGKMLSLA	EQQLVDCAQD
PHD	---HHH---	-----	---HHHHHHH	HHHHHHHHHH	HH---EEEE-	-HHHHH----
<b>Modell</b>	<b>-----HHH</b>	<b>H-----</b>	<b>-----HHHHH</b>	<b>HHHHHHHHHH</b>	<b>HHH-----</b>	<b>-HHHHH----</b>
	70	80	90	100	110	120
hCatH	FNNYGCQGG <sup>*</sup> L	PSQAF <sup>*</sup> EYILY	NKGIMGEDTY	PYQ <sup>*</sup> GKDG <sup>*</sup> YCK	FQPGKAIGFV	KDVANIT <sup>*</sup> IYD
PHD	-----HH	HHHHHHHHHH	H-----	-----	-----EEE	E--EE---H
<b>Modell</b>	<b>----HHH---</b>	<b>--HHHHHHHH</b>	<b>H-----</b>	<b>-----</b>	<b>-----EEE</b>	<b>E--EE---H</b>
	130	140	150	160	170	180
hCatH	EEAMVEAVAL	YNPVSFAFEV	TQDFMMYRTG	IYSSTSCHKT	PDKVNHAVLA	VG <sup>*</sup> YGEKNGIP
PHD	HHHHHHHHHH	---EEEE--	-HHHHEEE--	EEEE-----	-----EEEE	EEE-----
<b>Modell</b>	<b>-HHHHHHHHH</b>	<b>---EEEE--</b>	<b>--HHHH----</b>	<b>-----</b>	<b>-----EEE</b>	<b>--EEEE--EE</b>
	190	200	210	220		
hCatH	YWIVKNSWGP	QWGMNGYFLI	ERGKNMCGLA	ACASYPIPLV		
PHD	EEEE-----	-----EEEE	E-----EEEE	E-----		
<b>Modell</b>	<b>EE-----</b>	<b>-----</b>	<b>-----</b>	<b>-EEEE-----</b>		

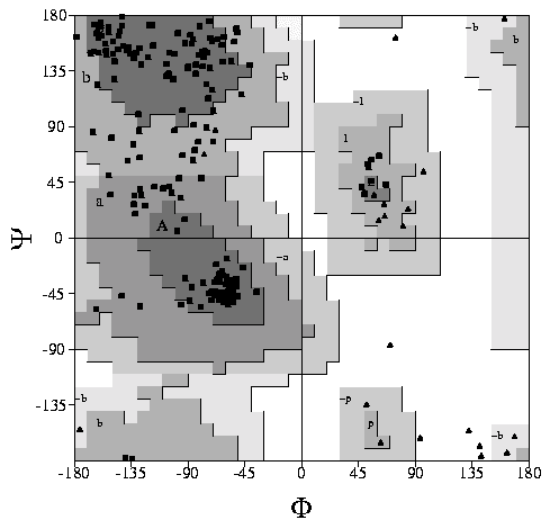
**Abb. A-7** Aminosäuresequenz von hCatH mit den Sekundärstrukturelementen, bestimmt über PHD (grau) sowie vom 3D-Strukturmodell von hCatH, ermittelt über DSSP (fett) (H = Helix, E =  $\beta$ -Faltblattstrang). Die Aminosäurereste der katalytischen Triade sind mit einem Stern gekennzeichnet.

	10	20	30	40	50	60
hCatH	YPPSVDWRKK	GNFVSPVKNQ	GACGSCWTF <sup>*</sup> S	TTGALES <sup>*</sup> AI <sup>*</sup> A	IATGKMLSLA	EQQLVDCAQD
pCatH	YPPSMDWRKK	GNFVSPVKNQ	GSCGSCWTF <sup>*</sup> S	TTGALES <sup>*</sup> AVA	IATGKMLSLA	EQQLVDCAQN
	70	80	90	100	110	120
hCatH	FNNYGCQGG <sup>*</sup> L	PSQAF <sup>*</sup> EYILY	NKGIMGEDTY	PYQ <sup>*</sup> GKDG <sup>*</sup> YCK	FQPGKAIGFV	KDVANIT <sup>*</sup> IYD
pCatH	FNNHGCQGG <sup>*</sup> L	PSQAF <sup>*</sup> EYIRY	NKGIMGEDTY	PYKGQDDHCK	FQPDKAIAFV	KDVANITMND
	130	140	150	160	170	180
hCatH	EEAMVEAVAL	YNPVSFAFEV	TQDFMMYRTG	IYSSTSCHKT	PDKVNHAVLA	VG <sup>*</sup> YGEKNGIP
pCatH	EEAMVEAVAL	YNPVSFAFEV	TNDFLMYRKG	IYSSTSCHKT	PDKVNHAVLA	VG <sup>*</sup> YGEENGIP
	190	200	210	220		
hCatH	YWIVKNSWGP	QWGMNGYFLI	ERGKNMCGLA	ACASYPIPLV		
pCatH	YWIVKNSWGP	QWGMNGYFLI	ERGKNMCGLA	ACASYPIPLV		

**Abb. A-8** Sequenzalignment der Aminosäuresequenzen von hCatH und pCatH (CLUSTAL W [97]). Die Aminosäuren der katalytischen Triade sind mit einem Stern gekennzeichnet. Identische Aminosäuren wurden fett hervorgehoben.

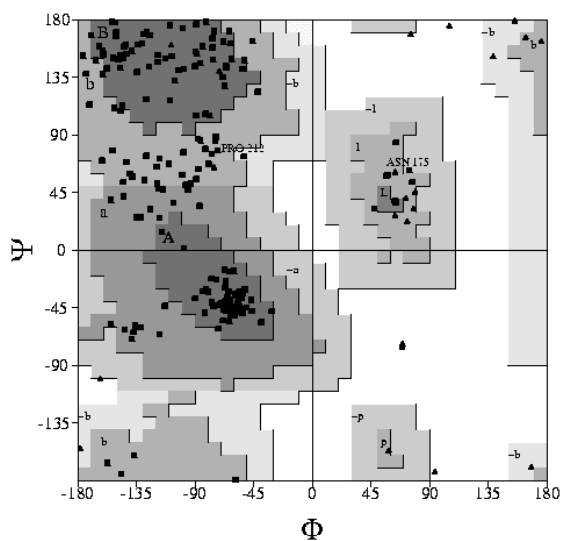
	10	20	30	40	50	60
hCatF	APPEWDWRSK	GAVTKVKDQG	MCGSCWAFSV	TGNVEGQWFL	NQGTLLSLSE	QELLDCKMD
PHD	-----	-----	--HHHHHHHH	HHHHHHHHHH	----EEEE-H	HHHH-----
<b>Modell</b>	<b>----EEHHHH</b>	<b>-----</b>	<b>----HHHHHH</b>	<b>HHHHHHHHHH</b>	<b>HH-----H</b>	<b>HHHHHH-----</b>
	70	80	90	100	110	120
hCatF	KACMGGLPSN	AYSAIKNLGG	LETEDDYSYQ	GHMQSCNFSA	EKAKVYINDS	VELSQNEQKL
PHD	----HHHHH	HHHHHHHH--	-----	-----	---EEEE-E	EE---HHHHH
<b>Modell</b>	<b>-HHH--HHH</b>	<b>HHHHHHHH--</b>	<b>-----</b>	<b>-----H</b>	<b>HH-----EEEE</b>	<b>EE-----HHHH</b>
	130	140	150	160	170	180
hCatF	AAWLAKRGPI	SVAINAFGMQ	FYRHGISRPL	RPLCSPWLID	HAVLLVGYGN	RSDVPFVAIK
PHD	HHHHHH---E	EEEE---EE	EEEEEEEE--	-----	EEEEEEEE--	----EEEEEE
<b>Modell</b>	<b>HHHHHH---E</b>	<b>EEEE-----</b>	<b>-----</b>	<b>-----</b>	<b>EEEEEEEEEE</b>	<b>----EEEEEE</b>
	190	200	210	220		
	*					
hCatF	NSWGTDWGEK	GYYYLHRGSG	ACGVNTMASS	AVVD		
PHD	-----	-EEEEEE--	-EEEEEE--	----		
<b>Modell</b>	<b>-----</b>	<b>-EEEEEE--</b>	<b>-----EEE</b>	<b>EEE-</b>		

**Abb. A-9** Proteinsequenz von humanem Cathepsin F mit seinen Sekundärstrukturelementen, ermittelt aus der Primärstruktur unter Verwendung von PHD und berechnet anhand des resultierenden Strukturmodells von hCatF unter Verwendung von DSSP (H = Helix, E =  $\beta$ -Faltblattstrang).



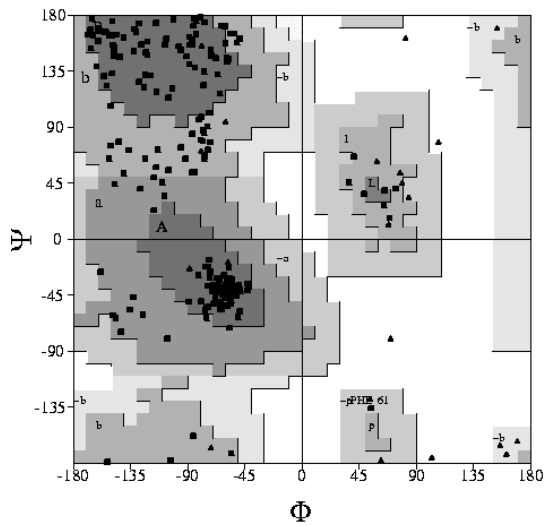
- 80.6% der Aminosäuren liegen in den Bereichen, die am stärksten für eine optimale Proteinstruktur begünstigt sind (A, B, L),
- 18.9% befinden sich in zusätzlich erlaubten Regionen (a, b, l, p),
- 0.5% in tolerierten Regionen (~a, ~b, ~l, ~p).

**Abb. A-10** Ramachandran-Plot der modellierten Proteinstruktur von Cathepsin K.



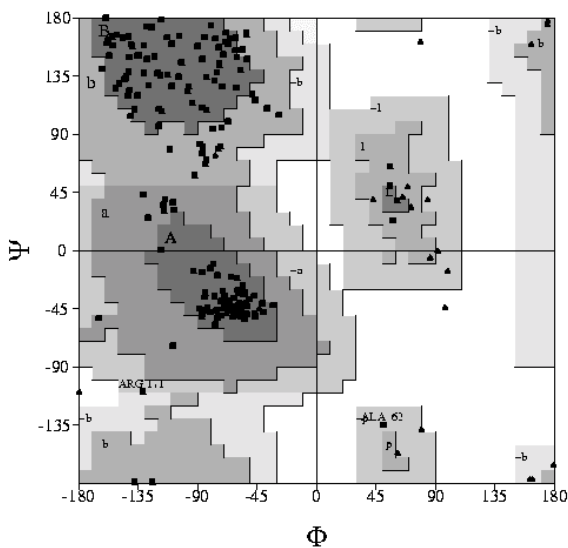
- 64.5% der Aminosäuren liegen in den Bereichen, die am stärksten für eine optimale Proteinstruktur begünstigt sind (A, B, L),
- 35.0% befinden sich in zusätzlich erlaubten Regionen (a, b, l, p),
- 0.5% in tolerierten Regionen (~a, ~b, ~l, ~p).

**Abb. A-11** Ramachandran-Plot des 3D-Strukturmodells von Cathepsin S.



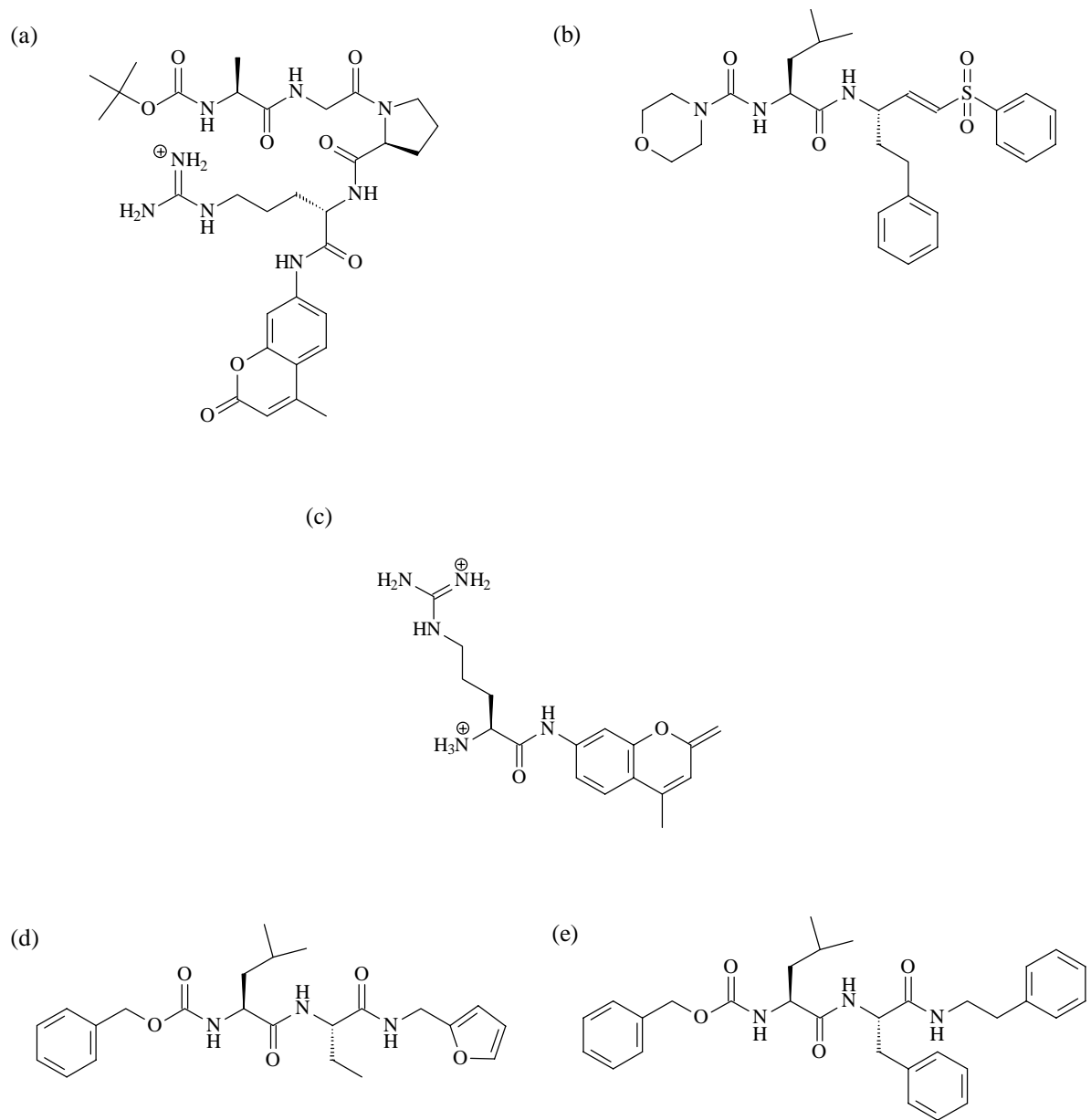
- 68.9% der Aminosäuren liegen in den Bereichen, die am stärksten für eine optimale Proteinstruktur begünstigt werden (A, B, L),
- 29.0% befinden sich in zusätzlich erlaubten Regionen (a, b, l, p),
- 2.1% in tolerierten Regionen (~a, ~b, ~l, ~p).

Abb. A-12 Ramachandran-Plot des Proteinstrukturmodells von Cathepsin H.



- 81.3% der Aminosäuren liegen in den Bereichen, die am stärksten für eine optimale Proteinstruktur begünstigt werden (A, B, L),
- 17.6% befinden sich in zusätzlich erlaubten Regionen (a, b, l, p),
- 1.1% in tolerierten Regionen (~a, ~b, ~l, ~p).

Abb. A-13 Ramachandran-Plot der modellierten Proteinstruktur von Cathepsin F.



**Abb. A-14** Strukturformeln der Liganden für die Verifizierung der Strukturmodelle der Cathepsine K, S, H und F: Boc-Ala-Gly-Pro-Arg-MCA (a), Mu-Leu-Hph- $\Psi$ (CH=CH-SO<sub>2</sub>-Ph) (b), Arg-NMec (c), Z-Leu-Abu-CONH-CH<sub>2</sub>-2-furan (d), Z-Leu-Phe-CONH-(CH<sub>2</sub>)<sub>2</sub>-Ph (e).

## 7.4. Synthese der Quenchingsubstrate

Die Verbindungen (Abz-Phe-Phe-Lys-Lys-Tyr(3-NO<sub>2</sub>)-Gly (11), Abz-Leu-Leu-Arg-Gln-Tyr(3-NO<sub>2</sub>)-Gly (28) und Abz-Lys-Leu-Arg-Glu-Phe(4-NO<sub>2</sub>)-GlyNH<sub>2</sub>) (33) wurden im AK Prof. K. Neubert (Universität Halle-Wittenberg) von Herrn Dr. J. Faust synthetisiert. Dazu wurde ein Peptidsynthesizer 431A der Firma Applied Biosystems verwendet. Es wurde eine Festphasenpeptidsynthese an Wang-Harz unter Verwendung der Fmoc-Technik mit HBTU als Kopplungsreagenz durchgeführt. Das Fluorophor Abz wurde als Boc-Abz-OH am Peptidsynthesizer an die Verbindung gekoppelt.

## 7.5 Enzymkinetische Untersuchungen

### *Beschreibung zu 4.3.2.1*

Zur Bestimmung der proteolytischen Aktivität des fluorogenen Substrates wurden Lösungen der Cathepsine L, B und K verwendet. Die Pufferlösung (Acetatpuffer 50 mM, pH-Wert = 5.5) enthält EDTA und DTT. Eine Lösung des Quenchsubstrates Abz-Phe-Phe-Lys-Lys-Tyr(3-NO<sub>2</sub>)-Gly (11) wurde mit den Cathepsinen in einer Konzentration von 1 µM inkubiert. Der Fluoreszenzanstieg dieser Lösungen (Reaktionsgeschwindigkeit über 1 min) wurde am Fluorimeter des Typs LS50, Perkin Elmer gemessen. Der Konzentrationsbereich des Substrates für hCatL lag in einem Bereich von 0.025 µM-0.5 µM. Die effektive Konzentration des Enzyms wurde durch Titration mit E-64 ermittelt. Zur Bestimmung der Fluoreszenz bei vollständigem Umsatz wurde das Substrat (1 µM) über Nacht mit den Cathepsinen inkubiert.



## **EIDESSTATTLICHE ERKLÄRUNG**

Hiermit erkläre ich, Annett Fengler, daß ich die vorliegende Dissertation selbständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe. Die den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen sind als solche gekennzeichnet. Diese Arbeit wurde an keiner anderen Einrichtung zur Begutachtung vorgelegt.

Annett Fengler

# LEBENS LAUF

## PERSÖNLICHE DATEN:

Name: Annett Fengler  
Geburtsdatum: 05.09.1970  
Geburtsort: Halle (Saale)  
Staatsangehörigkeit: deutsch  
Familienstand: ledig

## SCHULAU SBILDUNG:

1977-1987 Polytechnische Oberschule in Halle (Saale)  
1987-1990 Berufsschule, Buna AG, Schkopau  
Juli 1990 Abitur

## BERUFSAUSBILDUNG:

1987-1990 Berufsschule, Buna AG, Schkopau  
Juli 1990 Facharbeiter für chemische Produktion

## STUDIUM:

Sept. 1990 - Juli 1995 Studium an der Martin-Luther-Universität Halle-Wittenberg,  
Fachrichtung Chemie  
Sept. 1992 Diplom-Vorprüfung  
Sept. 1994 Diplom-Hauptprüfung  
Okt. 1994 - Juli 1995 Diplomarbeit unter Anleitung von PD Dr. habil. R. Friedemann  
Aug. 1995 - Sept. 1996 Anstellung als wissenschaftlicher Mitarbeiter am Institut für  
Organische Chemie der Martin-Luther-Universität Halle-  
Wittenberg  
Okt. 1996 - März 1999 Promotionsstipendiatin (Graduiertenstipendium des Landes  
Sachsen-Anhalt) am Institut für Biochemie der Martin-Luther-  
Universität Halle-Wittenberg im Arbeitskreis von  
PD Dr. habil. W. Brandt  
Nov. 1998 - Febr. 1999 Forschungsaufenthalt am Department of Physiology &  
Biophysics und am Department of Human Genetics der Mount Sinai  
School of Medicine, New York (USA)  
in den Arbeitskreisen von Prof. Dr. E. Mehler und  
Prof. Dr. D. Brömme  
seit April 1999 Wissenschaftlicher Mitarbeiter am Institut für Biochemie der  
Martin-Luther-Universität Halle-Wittenberg