

**From the test tube to the World Wide Web**

—

**The cleavage specificity of the proteasome**

**Vom Reagenzglas bis ins Internet**

—

**Die Schnittspezifität des Proteasoms**

**DISSERTATION**

**der Fakultät für Chemie und Pharmazie  
der Eberhard-Karls-Universität Tübingen**

**zur Erlangung des Grades eines Doktors  
der Naturwissenschaften**

**2001**

**vorgelegt von  
Alexander Konrad Nußbaum**

II

Tag der mündlichen Prüfung: 16.03.2001

Dekan: Prof. Dr. H. Probst

1. Berichterstatter: Prof. Dr. H.-G. Rammensee

2. Berichterstatter: Prof. Dr. G. Jung

3. Berichterstatter: Prof. Dr. S. Jentsch

To Paul, Jutta & Jörg

*It's never too late to have a happy childhood.*

## PREFACE

This dissertation deals with **proteasomes** (from 'protease' and Greek 'soma' = protein-chopping body) and their role in the regulation of immune responses. Proteasomes are barrel-shaped molecular machines (**enzymes**) that are found in every cell of the body. Their job is to chop up proteins, much like a garden shredder that cuts twigs and branches into small pieces. The small protein pieces can be transported to the cell surface to be presented to T-cells, immune cells that constitute a part of the white blood cells. If a body cell is 'sick' (i.e. it has turned into a tumor cell or is infected by pathogens such as viruses or bacteria), the protein fragments on the cell surface look different. They therefore can activate T-cells to kill the diseased cell for the good of the whole organism. During the research for my Diploma thesis (April-Dec. 1997) and my Ph.D. thesis (Jan. 1998-Dec. 2000) I tried to find out more about how exactly proteins are cleaved by proteasomes. I was lucky and could determine some of the rules that proteasomes follow to chop up proteins. These rules were used as a basis for the prediction of proteasome cleavages. My results have important implications for vaccine development and the prediction of immune responses.

The dissertation is divided into 4 main parts: **Introduction**, **Results**, **Summary** and **References**.

The **Introduction** provides an extensive, state-of-the-art synopsis on proteasomes and their function within the immune system. My own published and unpublished results, described in detail in the Results part, are briefly mentioned in the Introduction. Specific background information on the single projects of the Results part are given in the introductions to the five Results sections.

The **Results** part is also divided into sections (2.1-2.5), which represent the main accomplishments of my Ph.D. thesis project and are based on publications (2.1-2.3) and submitted manuscripts (2.4 and 2.5) that resulted from my research. They are presented here in chronological order and, as they are thematically linked and interdependent, were chosen to make up the main body of my thesis. Each section of the Results part is subdivided into parts specific for each sub-section: Summary, Introduction, Materials & Methods, Results, Discussion, References, Abbreviations and Participating researchers. Therefore, no separate sections for Materials & Methods and Discussion exist. Please note that the five sections of the Results part reflect the state-of-the-art of the respective

## VI

publishing date, i.e. late 1998 for sections 2.1 and 2.2, and early 2000 for 2.3. Sections 2.4 and 2.5 have only been submitted for publication recently and therefore reflect current state of the art. - I chose this format for the Results section because it adequately mirrors the chronological course and rationale of the entire thesis project. Besides, it corresponds to a common, international format for the presentation of scientific results and thus guarantees this work to be comprehensible for international readers.

The **Summary** at the end of this dissertation combines the results of the five parts of the Result section and puts them into perspective of future developments. Moreover, projects that are still ongoing or were not mentioned for space limitations are briefly outlined here.

**References** are given at the end of this thesis for the Introduction and the Summary. Additional references are provided at the end of each section of the Results part for the specific projects.

Enjoy reading!

Alexander K. Nussbaum

## TABLE OF CONTENTS

<b>Preface</b>	<b>V</b>
<b>Table of contents</b>	<b>1</b>
<b>1 Introduction</b>	<b>6</b>
<b>1.1 The Immune System</b>	<b>6</b>
1.1.1 Function	6
1.1.2 Overview	6
1.1.3 Cellular components and the role of peptides	6
1.1.3.1 MHC-molecules	7
1.1.3.2 Antigen recognition by T-lymphocytes	8
<b>1.2 Antigen Processing</b>	<b>9</b>
1.2.1 MHC class II pathway	9
1.2.2 MHC class I pathway	10
<b>1.3 The Proteasome</b>	<b>11</b>
1.3.1 Cellular roles of proteasomes	11
1.3.2 Proteasome species: Structure and function	12
1.3.3 Three-dimensional structure and catalytic mechanism of the 20S proteasome	14
1.3.4 Topology of the active sites	16
1.3.5 Cleavage specificity	18
1.3.6 Fragment length distribution	21
1.3.7 Modulation by interferon $\gamma$ (IFN- $\gamma$ )	23
1.3.7.1 PA28, the activator/regulator of mammalian proteasomes	23
1.3.7.2 Exchange of constitutive for IFN- $\gamma$ -inducible active site $\beta$ -subunits	26
1.3.7.3 Differential processing of CTL epitopes by constitutive and immuno-proteasomes	28
1.3.8 Protein degradation by proteasomes	29
1.3.8.1 The narrow gate: It is easier for a camel...	29
1.3.8.2 Processive degradation or fierce chopping?	30
<b>1.4 From Proteasomal Degradation Products to CTL Epitopes</b>	<b>33</b>
1.4.1 What makes proteasomal products good CTL epitopes?	33
1.4.2 Additional selection criteria	33
1.4.2.1 TAP specificity	34
1.4.2.2 To trim or not to trim, a nearly philosophical question	35
<b>1.5 The Prediction of CTL Epitopes</b>	<b>35</b>

## 1.6 Goal and Experimental Outline of the Thesis Project \_\_\_\_\_ 37

## 2 Results \_\_\_\_\_ 39

<b>2.1 Contribution of proteasomal <math>\beta</math>-subunits to the cleavage of peptide substrates analyzed with yeast mutants _____</b>	<b>39</b>
2.1.1 Summary _____	39
2.1.2 Introduction _____	39
2.1.3 Materials & Methods _____	42
2.1.3.1 Generation of yeast 20S proteasome mutants _____	42
2.1.3.2 Purification of proteasomes _____	42
2.1.3.3 Measurement of proteasomal activities against substrates with fluorogenic leaving group _____	42
2.1.3.4 Inhibitors _____	42
2.1.3.5 Measurement of non-fluorogenic leaving group by azo coupling _____	42
2.1.3.6 Determination of proteasomal cleavages in unmodified 19-25mer peptides _____	43
2.1.3.7 Separation of cleavage products and online analysis by mass spectrometry _____	43
2.1.4 Results _____	44
2.1.4.1 Cleavage of small fluorogenic substrates _____	44
2.1.4.2 Proteasome subunits are differentially inhibited by lactacystin, LLnL and DCI _____	45
2.1.4.3 "BrAAP" activity as measured with the Z-GAPLA-pAB substrate is catalyzed by the $\beta$ 1/Pre3 active site _____	47
2.1.4.4 Analysis of the degradation of peptide substrates _____	48
2.1.4.5 Degradation of the JAK1-21mer peptide _____	49
2.1.4.6 Degradation of the pp89-25mer peptide _____	50
2.1.4.7 Degradation of the Influenza NP 24mer peptide _____	53
2.1.5 Discussion _____	55
2.1.5.1 Cleavage specificity of the proteasomal subunits _____	55
2.1.5.2 Influence of inhibitors on distinct active sites _____	57
2.1.5.3 No additional catalytic sites in the yeast proteasome _____	58
2.1.5.4 Generation of dual cleavage products _____	59
2.1.6 References _____	61
2.1.7 Abbreviations _____	64
2.1.8 Participating researchers _____	65
<b>2.2 The cleavage motifs of yeast 20S proteasomes deduced from digests of enolase-1 _____</b>	<b>66</b>
2.2.1 Summary _____	66
2.2.2 Introduction _____	66
2.2.3 Materials & Methods _____	68



2.2.3.1	Generation of mutant proteasomes and purification of yeast 20S proteasomes	68
2.2.3.2	HPLC-Separation of enolase digests	68
2.2.3.3	Fragmentation of degradation products	68
2.2.3.4	Digestions	69
2.2.3.5	MALDI MS analysis	69
2.2.3.6	N-terminal sequencing (Edman degradation)	69
2.2.3.7	Statistical analysis	69
2.2.4	Results	70
2.2.4.1	Degradation of enolase-1	70
2.2.4.2	Digestion map of enolase-1 by wt yeast 20S proteasomes	71
2.2.4.3	Digestion of enolase-1 by mutant yeast 20S proteasomes	72
2.2.4.4	P1 amino acids and statistical analysis of flanking residues	74
2.2.4.5	Specificity of the wildtype proteasome	76
2.2.4.6	Specificity of the $\beta$ 5/Pre2 active site	76
2.2.4.7	Certain cleavages are not restricted to a single active site	77
2.2.4.8	Specificity of the $\beta$ 2/Pup1 active site	77
2.2.4.9	Specificity of the $\beta$ 1/Pre3 active site	78
2.2.5	Discussion	78
2.2.5.1	Protein degradation and fragment length	79
2.2.5.2	Specificity and selectivity of the active $\beta$ -subunits	80
2.2.6	References	82
2.2.7	Abbreviations	85
2.2.8	Participating researchers	85
2.2.9	Additional data: Cleavage motif of human erythrocyte 20S proteasomes	86
2.2.9.1	Fractionation of the degradation products by HPLC	86
2.2.9.2	Cleavage map	87
2.2.9.3	Fragment length distribution	88
<b>2.3</b>	<b>An algorithm for the prediction of proteasomal cleavages</b>	<b>90</b>
2.3.1	Summary	90
2.3.2	Introduction	90
2.3.3	Materials & Methods/Procedures	93
2.3.3.1	The experimental training data	93
2.3.3.2	The model	95
2.3.3.3	The learning process	97
2.3.3.4	Testing the model	98
2.3.4	Results	100
2.3.4.1	Reproduction performance	100

2.3.4.2	Computed affinities	103
2.3.4.3	Number of cleavages predicted in proteins	104
2.3.4.4	Predicting the generation of MHC I ligands	106
2.3.4.5	Refined model	108
2.3.5	Discussion	109
2.3.6	References	113
2.3.7	Abbreviations	117
2.3.8	Participating researchers	117

## 2.4 PAProC: A Prediction Algorithm for Proteasomal Cleavages available

	<b>on the WWW</b>	<b>118</b>
2.4.1	Summary	118
2.4.2	Introduction	118
2.4.3	Materials & Methods	120
2.4.3.1	Experimental training data and algorithm	120
2.4.3.2	Server requirements and programming language	121
2.4.3.3	Comparison of MHC I binding prediction and proteasomal cleavage prediction of CTL epitopes	121
2.4.4	Results	121
2.4.4.1	Sequence submission	121
2.4.4.2	Results output	123
2.4.4.3	Prediction results – mutated peptides containing CTL epitopes	124
2.4.4.4	Prediction results – proteins linked to disease	125
2.4.4.5	Prediction results – combination of SYFPEITHI and PAProC	126
2.4.5	Discussion	127
2.4.5.1	Benefits of PAProC	128
2.4.5.2	Prediction success rate of PAProC	128
2.4.5.3	Future developments	129
2.4.5.4	Distant goal	130
2.4.6	References	130
2.4.7	Abbreviations	133
2.4.8	Participating researchers	134

## 2.5 Discrete cleavage motifs of constitutive and immuno-proteasomes

	<b>revealed by quantitative analysis of cleavage products</b>	<b>135</b>
2.5.1	Summary	135
2.5.2	Introduction	135
2.5.3	Materials & Methods	137
2.5.3.1	Purification of 20S proteasomes	137

2.5.3.2	Immunoblotting _____	137
2.5.3.3	Measurement of proteasomal activities against substrates with fluorogenic leaving group _____	138
2.5.3.4	In vitro degradation of enolase 1 _____	138
2.5.3.5	Separation and analysis of cleavage products _____	138
2.5.3.6	Statistical analysis - frequencies of amino acids _____	139
2.5.3.7	Statistical analysis - comparison of amino acid characteristics _____	139
2.5.4	Results _____	139
2.5.4.1	Isolation of proteasomes _____	139
2.5.4.2	Digestion of Enolase _____	141
2.5.4.3	Quantification of digestion profiles _____	143
2.5.4.4	Analysis of cleavage site-usage _____	147
2.5.4.5	Effect of immuno-subunit incorporation on cleavage site selection. _____	149
2.5.5	Discussion _____	150
2.5.6	References _____	155
2.5.7	Abbreviations _____	158
2.5.8	Participating researchers _____	158
<b>3</b>	<b>Summary _____</b>	<b>160</b>
<b>3.1</b>	<b>Summary &amp; Perspectives _____</b>	<b>160</b>
3.1.1	Summary of Results section 2.1 _____	160
3.1.2	Summary of Results section 2.2 _____	161
3.1.3	Summary of Results sections 2.3 and 2.4 _____	163
3.1.4	Summary of Results section 2.5 _____	165
<b>3.2</b>	<b>Unmentioned and Ongoing Projects _____</b>	<b>166</b>
3.2.1	Insensitivity of $\beta$ -peptides to proteasomal cleavage _____	166
3.2.2	Directionality – a story with an end? _____	166
3.2.3	The Gly-Ala story _____	167
3.2.4	When does the PA28-effect kick in? _____	168
<b>4</b>	<b>References _____</b>	<b>170</b>
<b>5</b>	<b>Thanks to... _____</b>	<b>188</b>
<b>6</b>	<b>Abbreviations _____</b>	<b>190</b>
<b>7</b>	<b>My Academic Teachers _____</b>	<b>192</b>
<b>8</b>	<b>CV + Publications _____</b>	<b>193</b>

# 1 Introduction

## 1.1 The Immune System

### 1.1.1 Function

Every organism has to protect itself against continuous attack by foreign organisms, such as parasites, fungi, bacteria or viruses. In the case of prions, even foreign proteins seem to be able to act as pathogens. The immune system was evolved for the defense against these pathogenic organisms. Its two essential characteristics are the recognition of the pathogen and its removal.

### 1.1.2 Overview

The immune system can be divided into the **innate immune system** and the **adaptive (or specific) immune system**. The former contains anatomical barriers such as the skin and mucosae, physiological barriers such as lysozyme, interferons and complement, endocytic and phagocytic cells and natural killer cells. Besides, inflammatory responses - vasodilatation, extravasation of inflammatory cells, acute phase proteins and chemical messengers such as histamine and bradykinin – belong to the innate immune response. The adaptive immune response is characterized by four main features: specificity, diversity, memory and the recognition of self, altered self and non-self. The cells of the adaptive immune system which guarantee these attributes are, on the one hand, **professional antigen-presenting cells (APC)** and, on the other hand, **lymphocytes**. The **B-lymphocytes**, together with the antibodies they secrete, constitute the main effectors of the **humoral immune response**, whereas **T-lymphocytes** represent the main effectors of **cellular defense** mechanisms. **Cytotoxic T-lymphocytes (CTL)** kill foreign cells, mutated cells or virus-infected cells, whereas **T-helper (T<sub>h</sub>)** cells mainly activate and regulate B-cells, CTL and APC through chemical messengers (**cytokines**).

### 1.1.3 Cellular components and the role of peptides

Cells of the adaptive immune system, especially T-cells, screen host cells for two properties: First, is it a host (**self**) cell or a foreign (**non-self**) cell? Second, in case the T cell is looking at self: Is the host cell healthy or is it threatened by virus-infection or transformation? Malignant transformation of a cell can be the result of mutations that will lead to changes in the characteristics of the host cell (**altered self**).

All the decisions about self, foreign and disturbed self are made with the help of the **major histocompatibility system (MHC)**. The peptide-binding and –presenting MHC-cell surface molecules can be viewed as passports for host cells. **MHC class I (MHC I)** molecules can be found on the surface of most nucleated cells, whereas **MHC class II (MHC II)** molecules are mainly found on the surface of APC (dendritic cells, B-cells, macrophages), but also on inflamed and thymus epithelial cells. Self and non-self are distinguished by characteristics of both the MHC molecules and the peptides presented by them; the distinction between healthy self and altered self depends on the nature of the peptides presented in the MHC binding groove and on the presence of signals activating APC.

### 1.1.3.1 MHC-molecules

MHC II molecules consist of two chains ( $\alpha$ ,  $\beta$ ) whose C-termini are fixed in the cell membrane and whose N-terminal domains form the **peptide binding groove**. MHC I molecules are comprised of a heavy  $\alpha$ -chain which is non-covalently linked to  $\beta$ 2-microglobulin ( $\beta$ 2m). Here, the peptide binding groove is only made up of the  $\alpha$ -chain. One of the main differences between MHC I and MHC II molecules, besides their different cellular distribution (see 1.1.3) and peptide-loading (see 1.2), is the structure of the peptide binding groove. In both cases, the groove is made of a “floor” of  $\beta$ -sheet structure and two lining “walls” of  $\alpha$ -helices. However, whereas the groove is open to both sides for MHC II molecules, it is closed for MHC I molecules (Bjorkman et al., 1987; Brown et al., 1993). As a result, MHC I can only accommodate peptides of 8-10 aa in length (Falk et al., 1991; Madden et al., 1991), whereas MHC II can hold peptides of up to 15-20 aa. In both cases, however, the ligand interacts with the surface of the binding groove only over a stretch of about 9 aa. Non-covalent (hydrophobic) interactions and electrostatic forces (hydrogen bonds) fix the peptide to the MHC molecule. This is achieved by means of particular aa of the peptide, the so-called **anchor-residues**, that fit snugly into complementary **pockets** of the MHC binding groove (Rammensee et al., 1993). A peculiarity of MHC molecules is their enormous **polymorphism** (Parham et al., 1995). This polymorphism entails that different MHC alleles vary in their pockets and therefore require different **binding motifs**, i.e. combinations of anchor residues, for binding of their peptide ligands. Most MHC I binding motifs require hydrophobic (aromatic and branched-chain) aa in P9, i.e. at the C-terminus of the peptide. Ligands of the human MHC I molecules HLA-A\*0201, for example, need the aa L or V in P2 (position 2) and V or L in P9. HLA-A3 and HLA-A11 require L, V or M in P2 and K in P9 of the presented peptide

(Rammensee et al., 1999). Thus, the binding motifs clearly represent a strict selection criterion for peptide sequences to be presented by MHC molecules (see also 1.2.2 and Figure 1.2-1).

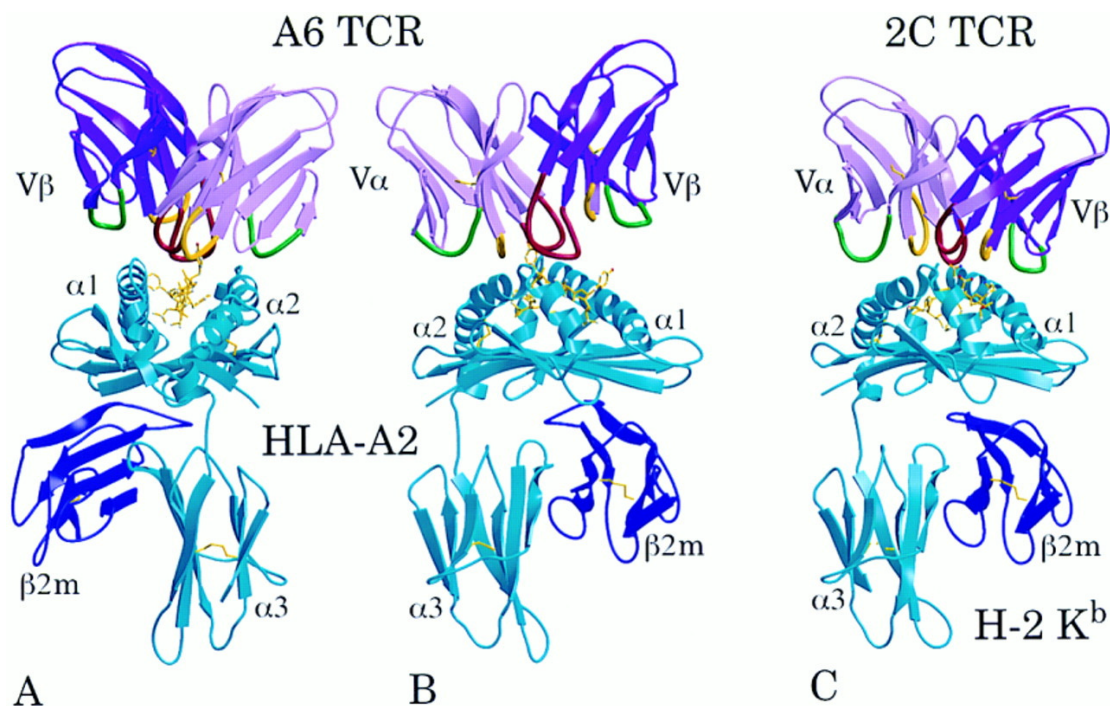
### 1.1.3.2 Antigen recognition by T-lymphocytes

T-lymphocytes are selected according to two criteria during their maturation in the thymus: They have to be able to recognize self-MHC (**positive selection**) and they must not be activated by MHC-bound self-peptides (**negative selection**). The former requirement guarantees that T cells "see" presented peptides only in the context of self-MHC, named **restriction for self-MHC** or **self-restriction** (Zinkernagel and Doherty, 1974). Autoreactive T cells specific for self-MHC and self-peptide are thus eliminated, leading to **self-tolerance**. The selection of a self-restricted, self-tolerant **T cell repertoire** (i.e. pool of available T cells) guarantees that the cellular immune response only hits foreign organisms, virus-infected or tumor cells.

MHC/peptide-complexes are recognized by the **T cell receptor (TCR)**, the antigen receptor of T cells. It consists of two chains,  $\alpha$  and  $\beta$ , which are linked via a disulfide bridge. Both chains are made up of two **immunoglobulin (Ig) domains**. The two outer Ig domains of each TCR chain ( $V\alpha + V\beta$ ) build the antigen recognition site which interacts with and recognizes the MHC/peptide-complex (see Figure 1.1-1). The actual MHC ligand (peptide) that is recognized by T cells and activates them is called **T cell epitope**.

The variability of the TCR ( $10^{16}$ ) is so large that there should exist T cells recognizing all non-self or altered self MHC ligands. However, stochastic processes in T cell maturation and clonal T cell expansion lead to diverse T cell repertoires in individual members of a population.

The role of the MHC molecule as a "passport" was already hinted at earlier. However, eventually it is the nature of the peptide that decides whether T cells will be activated or inactivated (**anergized**), i.e. whether an MHC-bound peptide functions as a T cell epitope or just as an MHC ligand, respectively. Thus the peptide plays a crucial role in the recognition of self and non-self and the regulation of immune responses.



**Figure 1.1-1: Crystal structures of TCR recognizing a MHC/peptide-complex**

Front (A) and side (B) view of the human MHC molecule HLA-A2; side view (C) of the mouse MHC molecule H-2K<sup>b</sup>. MHC and TCR are drawn as ribbon models; the MHC-bound peptide (yellow) is drawn as stick-model. The peptide-binding groove of MHC molecules is made up of the  $\alpha 1$  and  $\alpha 2$  domains of the MHC I heavy  $\alpha$ -chain. [The pictures are taken from Garboczi et al. (1996) for (A), (B) and from Garcia et al. (1996) for (C).]

## 1.2 Antigen Processing

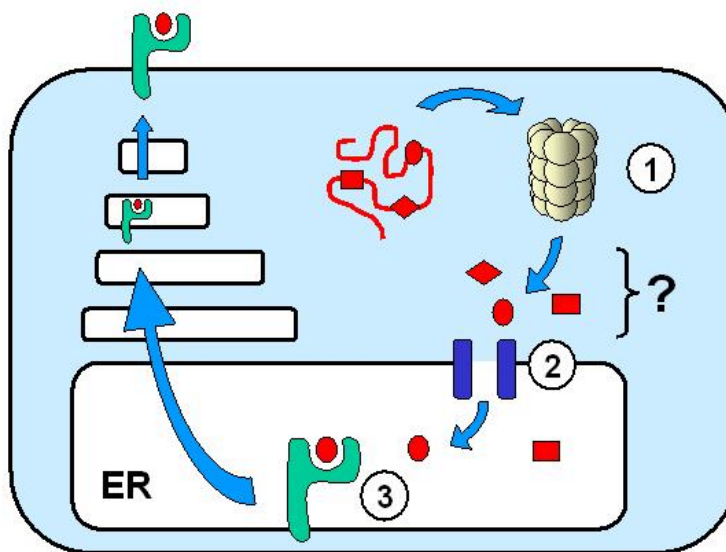
### 1.2.1 MHC class II pathway

The antigen processing pathway leading to the loading of MHC II molecules is also called **endosomal pathway**. Here, membrane proteins are endocytosed or extracellular antigens are taken up either in an unspecific (phagocytosis, pinocytosis; e.g. by macrophages) or specific fashion (receptor-mediated endocytosis; e.g. via membrane-bound antibodies on B cells). The internalized phagosomes or clathrin-coated vesicles then fuse with lysosomes, whose proteolytic enzymes (cathepsins and probably some aminopeptidases) and low pH will start the degradation of the antigens. So-called MHC II-loading compartments (**MIIIC**), formed by fusion of lysosomal vesicles with MHC II-containing trans-Golgi vesicles, represent the sites where the actual loading of MHC II molecules with peptides takes place. The final, trimolecular complex of MHC II  $\alpha$ - and  $\beta$ -

chains and the peptide is then transported to the cell surface (Cresswell, 1994; Watts, 1997).

### 1.2.2 MHC class I pathway

The **class I pathway of antigen processing**, also called the **cytosolic pathway**, leads to the loading of MHC I molecules by peptides mainly derived from cytosolic proteins (Pamer and Cresswell, 1998; Früh and Yang, 1999; van Endert, 1999). In healthy cells, these peptides stem solely from intracellular host (self) proteins. On the contrary, in transformed (mutated) or virus-infected cells, mutated (=altered self) or viral (=non-self) peptides, respectively, can make it onto MHC I and to the cell surface. The first step in this pathway is the degradation of proteins by the barrel-shaped, multicatalytic proteasome complex. The proteasome chops up proteins into small protein pieces, some of which will make their way - probably assisted by chaperones - to the **transporter associated with antigen processing (TAP)** in the ER membrane.



**Figure 1.2-1: Schematic overview over antigen processing in the class I pathway.**

A cytosolic protein is targeted for degradation by the proteasome (1). The nature of the resulting peptide fragments is not completely understood (?). Some proteasomal cleavage products are translocated by TAP (2) into the endoplasmic reticulum (ER) and loaded onto MHC class I molecules (3). MHC/peptide-complexes on the cell surface are screened by CTL for foreign or mutated peptides (not shown).

TAP preferentially translocates peptides of 8-16 aa with hydrophobic or basic C-termini into the ER lumen (Schumacher et al., 1994; Momburg et al., 1994; van Endert, 1999; Uebel and Tampe, 1999). It was only recently determined that the major substrates for



TAP *in vivo* are derived from newly synthesized proteins (Reits et al., 2000). An intricate army of ER-resident chaperones (calreticulin, calnexin, tapasin and perhaps PDI and gp96) will then help to load fitting peptides into the binding groove of MHC I molecules (Lammert et al., 1997a, Lammert et al., 1997b; Sadasivan et al., 1996). The fate of peptides that do not bind to MHC is not completely understood (Falk et al., 1990), but they are probably removed from the ER by retrograde transport into the cytosol, which is most likely mediated by the Sec61p channel (Koopmann et al., 2000). An overview over MHC I processing pathway is shown in Figure 1.2-1.

## 1.3 The Proteasome

### 1.3.1 Cellular roles of proteasomes

Proteasomes are the key proteases for non-lysosomal proteolysis in eukaryotic cells. Most (short-lived<sup>1</sup>) proteins, in order to be recognized and degraded by proteasomes, have to be tagged by ubiquitin (Hershko and Ciechanover, 1998; Hershko et al., 2000). Signals for ubiquitination are manifold (reviewed in Ciechanover, 1998): They comprise primary (substrate-inherent, structural) motifs or secondary (post-translational) modifications. Examples for primary motifs include the **N-end rule pathway** (i.e. protein stability is influenced by its N-terminal aa residue; Bachmair et al., 1986; reviewed in Varshavsky, 1996; Varshavsky et al., 2000) and the **destruction box**, a nine-amino-acid motif found in mitotic cyclins and certain other cell-cycle regulators. A secondary modification that can trigger ubiquitin-dependent degradation is phosphorylation (e.g. in I $\kappa$ B $\alpha$ ; Yaron et al., 1997), for example at Pro(P)-, Glu(E)-, Ser(S) and Thr(T)-rich (**PEST**-) sequences. Proteins can also be targeted for degradation by association with an ancillary protein (e.g. p53 is degraded after binding to the human papilloma virus oncoprotein E6; Scheffner et al., 1993). It was shown recently that a large fraction of the substrates for ubiquitin-dependent proteasomal degradation are newly synthesized proteins, so-called **DRiPs** (for Defective Ribosomal initiation Products) (Schubert et al., 2000). Proteasomes degrade some proteins devoid of ubiquitination-tags, for example ODC (Murakami et al., 1992), damaged calmodulin (Tarcza et al., 2000) and p21<sup>Cip1</sup> (Sheaff et al., 2000). This is speculated to reflect a more ancient mode of substrate recognition by proteasomes,

---

<sup>1</sup> It must be noted that to this date, it is still disputed whether ubiquitin conjugation is essential in the degradation of long-lived proteins (Rock and Goldberg, 1999).

mediated by sequence signals ("class I degrons") inherent in the substrate sequence itself (Verma and Deshaies, 2000). Although some examples for this hypothesis have lately been identified (see above), it still must be assumed that by far the main part of protein turnover in eucaryotic cells is ubiquitin-dependent.

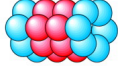


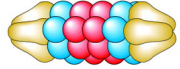
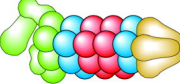
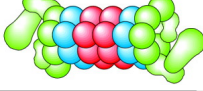
Proteasomes are not limited to purely metabolic functions. Proteasomal protein degradation also takes over regulatory roles such as periodic cyclin degradation (Glotzer et al., 1991; Seufert et al., 1995), the removal of misfolded proteins (Seufert and Jentsch, 1990) or the activation of the transcription factor NF- $\kappa$ B (Palombella et al., 1994). Proteasomes are mainly localized within the cytosol and the nucleus of cells, but they are also found in association with the endoplasmic reticulum (ER) (Rivett et al., 1992; Newman et al., 1996; Enenkel et al., 1998; Brooks et al., 2000) and the cytoskeleton (Scherrer and Bey, 1994).

The importance of ubiquitin-dependent protein degradation in MHC I antigen presentation was suggested by the fact that enhanced protein degradation rescued the defective presentation of viral epitopes in vaccinia-infected cells (Townsend et al., 1988) and was elegantly demonstrated by the use of mutant cell lines with defects in the ubiquitin-dependent degradation pathway (Michalek et al., 1993; Grant et al., 1995). The fundamental role of proteasomes themselves for the supply of MHC I ligands was confirmed by the use of proteasome inhibitors: In living cells, peptide aldehyde inhibitors and the very proteasome-specific *Streptomyces* compound lactacystin- $\beta$ -lactone completely blocked the presentation of peptides from several different proteins, including ovalbumin,  $\beta$ -galactosidase, and influenza viral antigens (Rock et al., 1994; Harding et al., 1995; Grant et al., 1995; Cerundolo et al., 1997).

### 1.3.2 Proteasome species: Structure and function

Proteasomes exist more or less in two ways: As the free catalytic (i.e. proteolytically active) core of all proteasome species, the barrel-shaped **20S proteasome** (named after its sedimentation coefficient), and as associations of 20S proteasomes with regulatory protein complexes. 20S proteasomes exist in a **constitutive** form and as so-called **immuno-proteasomes**, in which the active site subunits have been exchanged for IFN- $\gamma$ -inducible **immuno-subunits**. 20S proteasomes can be capped on both ends of their cylindrical structure by regulatory protein complexes, the **19S regulatory particle (19S cap, PA700)** or the IFN- $\gamma$ -inducible activator **PA28 (11S regulator)**. 20S proteasomes that are singly or doubly capped by the 19S regulatory particle are called **26S proteasomes**

(Figure 1.3-1). The 19S particle consists of a 20S-proximal 'base', which contains all six proteasomal ATPases and can activate the degradation of peptides and non-ubiquitinated proteins, and of a 'lid', which is required for ubiquitin-dependent proteolysis (Glickman et al., 1998). As the 19S cap confers recognition of ubiquitin (Deveraux et al., 1994; Ferrell et al., 1996), 26S proteasomes are thought to be responsible for the main part of protein degradation *in vivo* (DeMartino and Slaughter, 1999; Ferrell et al., 2000). However, 20S proteasomes might execute parts of the ubiquitin-independent protein turnover (1.3.1). Additionally, there exist **hybrid proteasomes** that are capped, on the one end, by 19S and, on the other end, by PA28 (Hendil et al., 1998; Tanahashi et al., 2000). These hybrid proteasomes might play an important role *in vivo* as they possibly link protein degradation to immune responses (see 1.3.8).

Type of Complexes	Model of Proteasome, PA700 PA28 and Their Complexes		Relative Content (%)
20S Proteasome	A		31 ± 4
PA700	B		10 ± 1
PA28	C		15 ± 10
PA28-20S -PA28	D		15 ± 7
PA700-20S -PA28	E		18 ± 7
PA700-20S -PA700	F		11 ± 2
<b>Total</b>			<b>100</b>

**Figure 1.3-1: The different proteasome species within cells.**

Relative distributions of proteasomes and their regulators, PA28 and PA700 (= 19S), in HeLa cells. Immuno-proteasomes are missing from the figure; they can be a part of any proteasome complex containing 20S proteasomes. [The graph was taken from Tanahashi et al. (2000).]

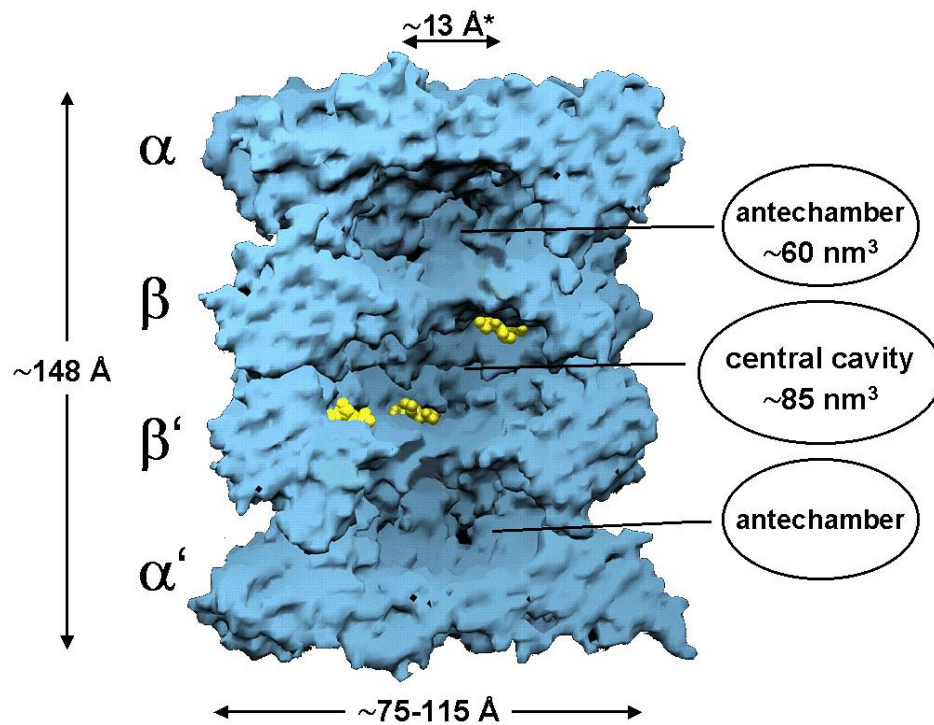
The two most abundant proteasome species in cells are 20S and 26S proteasomes. Two recent reports show that free 20S is more abundant than 26S (Tanahashi et al., 2000;

Brooks et al., 2000). Experimental artifacts from cell fractionation methods (centrifugation, gel filtration, Western blotting) might lead to overestimation of the amount of free 20S, by counting molecules dissociated from larger proteasome complexes due to the purification conditions. However, the same results were obtained by immunofluorescent labeling of cells, confirming that free 20S really exist. Nevertheless, it is still unclear whether 20S proteasomes are functional *in vivo* or whether they merely represent assembly/disassembly intermediates on the way to/from 26S proteasomes. Brooks and colleagues (2000) propose that the existence of free 20S proteasomes permits the formation of PA28-20S complexes after IFN- $\gamma$ -induced synthesis of PA28. Only recently, an inhibitor of proteasomes, PI31, has been identified and described to compete for binding to 20S proteasomes with PA28 (Zaiss et al., 1999).

### 1.3.3 Three-dimensional structure and catalytic mechanism of the 20S proteasome

20S proteasomes have been found in all eucaryotes from yeast to man, in archaebacteria and in some eubacteria (Tamura et al., 1995). The quarternary structure of 28 subunits, which are subdivided into the homologous  $\alpha$ - and  $\beta$ -subunits, is highly conserved in all cases. However, the differentiation of the 20S subunits has increased during evolution: Archaebacteria (*Thermoplasma*) only possess one kind of  $\alpha$ - and  $\beta$ -subunits, some eubacteria have two kinds whereas yeast already has 7 kinds each. Mammalian genomes encode 3 additional, IFN- $\gamma$ -inducible  $\beta$ -subunits, giving rise to altogether 7  $\alpha$ - and 10  $\beta$ -subunits for mammalian proteasomes. The number of proteolytically active  $\beta$ -subunits, however, has decreased in evolution: All  $\beta$ -subunits of *Thermoplasma* proteasomes are active (i.e.  $2 \times 7 = 14$  active sites per 20S particle), whereas only 3 different active  $\beta$ -subunits exist in eucaryotic proteasomes (i.e.  $2 \times 3 = 6$  active sites per 20S particle).

The proteolytic centers (active sites) are located on the inside of the 20S cylinder and are embodied by three different  $\beta$ -subunits in eucaryotic proteasomes. The two identical halves ( $\alpha_7\beta_7$ ) of 20S proteasomes are joined head-to-head with rotational symmetry (C2-axis along the  $\beta_7\beta'_7$ -contact area) (Figure 1.3-2).

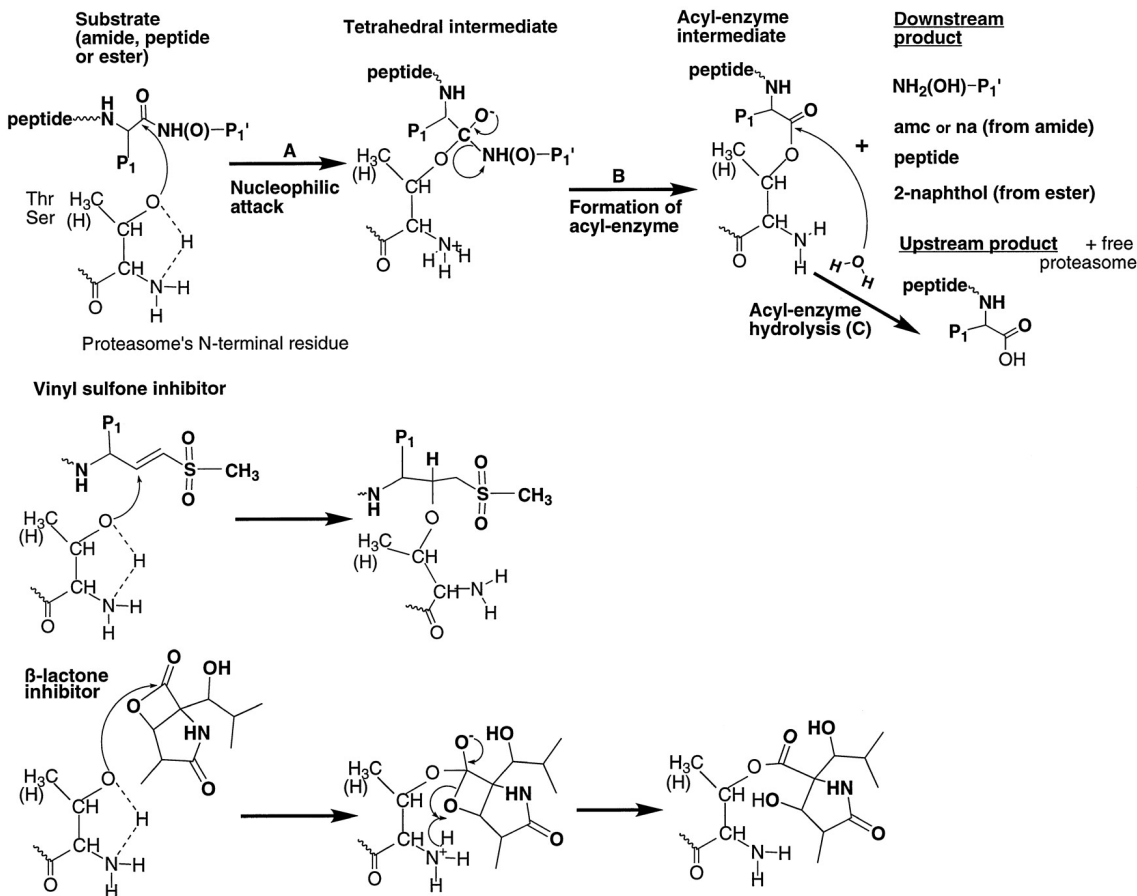


**Figure 1.3-2: X-ray structure of yeast 20S proteasome**

The picture shows the back half of the 20S cylinder. The given dimensions of the complex were for *Thermoplasma* proteasomes, whose quaternary structure is very similar to that of yeast proteasomes (taken from Baumeister et al., 1998). The locations of the active sites are marked by the binding of inhibitor molecules (yellow). \*The opening in the  $\alpha$ -rings of yeast proteasomes is only formed upon activation by SDS or binding of the regulators 19S or PA28 (Whitby, 2000; Groll, 2000). See Figure 1.3-4 for a more schematic view. [The x-ray picture is taken from Groll et al. (1997).]

Proteasomes belong to the small group of **threonine-proteases** within the family of **N-terminal nucleophile (Ntn) hydrolases** (Seemüller et al., 1995). The nucleophile for peptide-bond hydrolysis is the  $\gamma$ -O-atom of the N-terminal threonine in the active  $\beta$ -subunits. This threonine residue is only liberated after autocatalytic processing of propeptides (Seemüller et al., 1996), a mechanism that guarantees proteasome halves not to become active before their assembly into complete proteasome cylinders. The N-terminal amino group acts as proton-acceptor-donor when the substrate peptide bond is attacked. As for other Ntn-hydrolases, a long-lived acyl-enzyme intermediate is postulated (Löwe et al., 1995). The proposed catalytic mechanism is depicted in Figure 1.3-3. The cleavage product with the newly made N-terminus should then diffuse away from the active site, thus rendering it less likely to be cleaved again by the same active site. This assumption is in line with the results of kinetic modeling which suggest that the C-termini of internal protein fragments are generated before the N-termini (Holzhütter and Kloetzel,

2000).

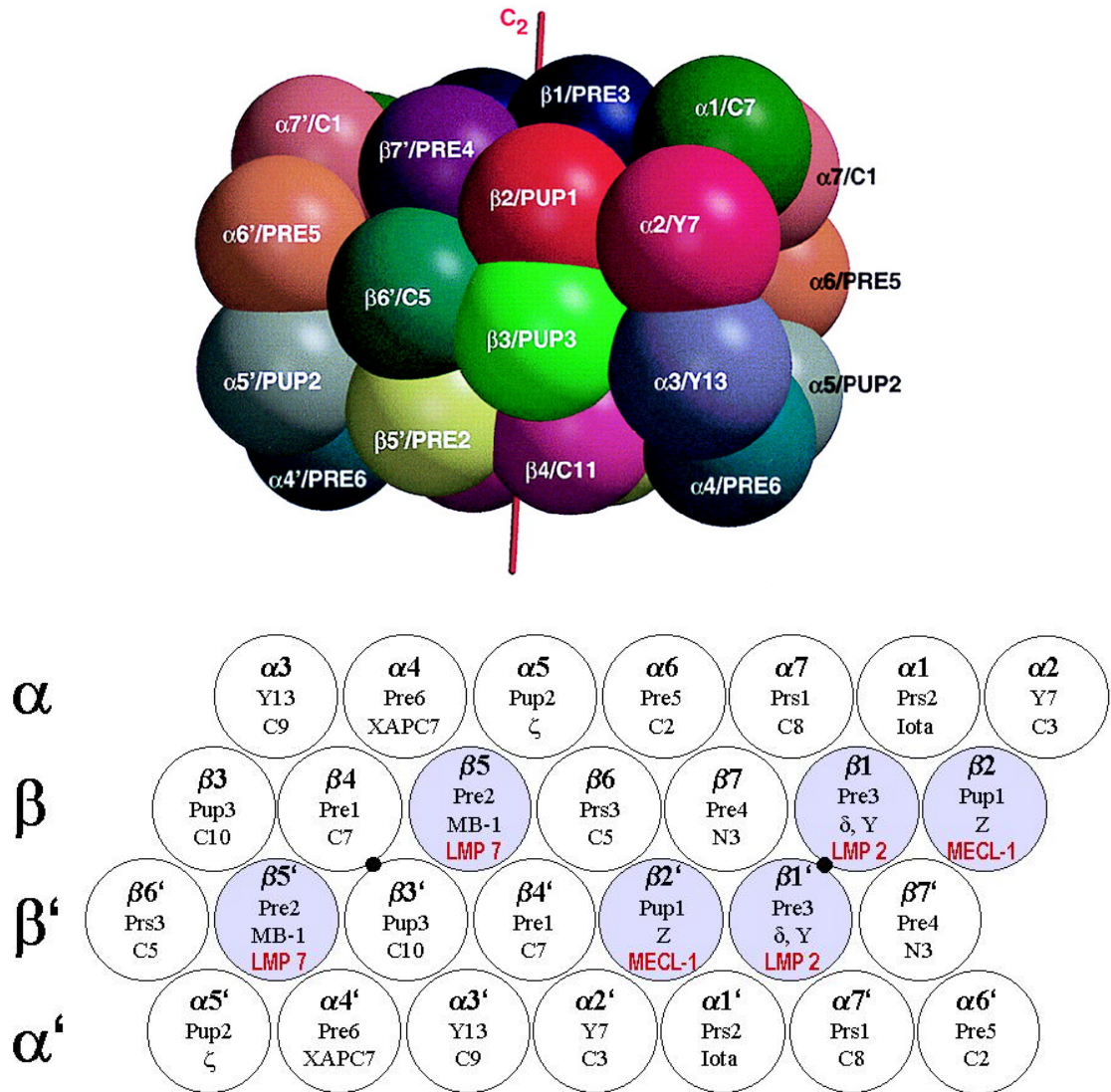


**Figure 1.3-3: Catalytic mechanism of proteasomes**

The chemical cleavage reactions are shown for a substrate peptide bond (top), a vinyl sulfone inhibitor (center) and lactacystin, a  $\beta$ -lactone inhibitor (bottom). [The figure is taken from Kisselev et al. (2000).]

### 1.3.4 Topology of the active sites

The three-dimensional structure of a eucaryotic 20S proteasome (yeast) was only solved recently by X-ray crystallography (Groll et al., 1997; Figure 1.3-2). The X-ray image made it possible to determine the precise positioning of the single subunits in relation to each other (schematic pictures in Figure 1.3-4). The subunit topology of human 20S proteasomes was first reported to slightly differ (Kopp et al., 1997), but was later confirmed by the same group to be identical to that of yeast (Dahlmann et al., 2000). Therefore, the subunit arrangement found for yeast proteasomes seems to be valid for all eucaryots.



**Figure 1.3-4: Topology of  $\alpha$ - and  $\beta$ -subunits in yeast 20S proteasomes**

(Top) 3-D picture of subunit topology; nomenclature: systematic/yeast. (Bottom) Rolled-open 20S cylinder revealing the relative positions of the subunits; nomenclature: systematic, yeast, human, IFN- $\gamma$ -inducible (from top to bottom). C2: symmetry axis through the cylinder. [The picture on the top was taken from Groll et al. (1999).]

For reference and to avoid confusion early on, an overview on the nomenclature and the cleavage specificities of the  $\beta$ -subunits that harbor active sites is shown in Table 1.3-1.

**Table 1.3-1: Nomenclature and cleavage specificity of the active site  $\beta$ -subunits in yeast and human 20S proteasomes**

Nomenclature				Specificity*	
systematic		human		yeast	
constitutive	IFN- $\gamma$ - inducible	constitutive	IFN- $\gamma$ - inducible	(only const.)	(for const. subunits)
$\beta$ 1	$\beta$ 1i	$\delta/\gamma$	LMP2	Pre3	post-acidic
$\beta$ 2	$\beta$ 2i	Z	MECL-1	Pup1	T-like
$\beta$ 5	$\beta$ 5i	X	LMP7	Pre2	ChT-like

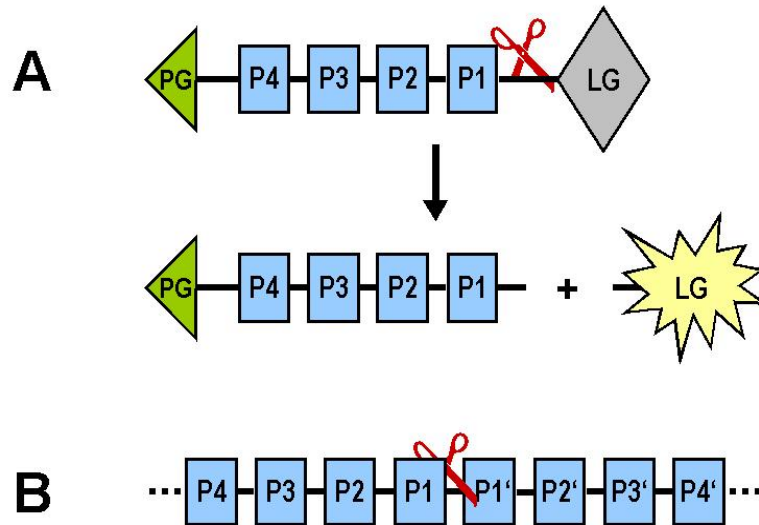
\* The cleavage specificities relate to the constitutive subunits. Check section 1.3.7.2 for more detail.

### 1.3.5 Cleavage specificity

As proteasomal degradation products are the main source for MHC I ligands, the most interesting aspect for immunologists in the investigation of eucaryotic proteasomes is their cleavage specificity. Characterization of proteasomal cleavage specificity has been and is still looked at using artificial tri- or tetrapeptides including a fluorogenic leaving group (see Figure 1.3-5).

Several distinct proteolytic specificities have been described: the chymotrypsin (ChT) -like activity (cleaving after hydrophobic - both branched chain and aromatic - aa residues), the trypsin (T) -like activity (cleaving after the basic aa Arg, Lys and His) and the peptidylglutamyl-peptide hydrolyzing (PGPH) or post-acidic or caspase-like activity (cleaving after the acidic aa Asp and Glu). In addition, two more activities have been postulated: SNAAP (small neutral amino acid peptidase) and BrAAP (branched-chain amino acid peptidase) (Orlowski et al., 1993). The differential influence of inhibitors on some of these activities led to the assumption that they must origin from distinct active sites (Orlowski, 1990; Rivett, 1989; Orlowski et al., 1993). However, the crystallographic structure showed and it is today accepted that only three different active sites exist.





**Figure 1.3-5: Fluorogenic model substrates used for characterization of proteasomal cleavage specificities.**

(A) The peptide is cleaved between the most C-terminal aa (P1) and the fluorogenic leaving group (LG). The cleavage specificity can only be determined by P-residues. (B) A normal peptide contains aa C-terminal of the cleavage site (P'-aa) which influence cleavage site selection. PG: protective group; P4-P4': aa-string (nomenclature according to Schechter and Berger, 1967<sup>2</sup>); LG: leaving group that fluoresces after being cleaved off; pair of scissors symbolize a proteasomal cleavage.

Analyses of mutant yeast proteasomes allowed the assignment of proteolytic activities to certain proteasomal  $\beta$ -subunits (Heinemeyer et al., 1993; Hilt et al., 1993; Enenkel et al., 1994; Arendt and Hochstrasser, 1997):  **$\beta$ 1/Pre3**  $\leftrightarrow$  **post-acidic**,  **$\beta$ 2/Pup1**  $\leftrightarrow$  **T-like**,  **$\beta$ 5/Pre2**  $\leftrightarrow$  **ChT-like**. Today it is clear that results obtained with short fluorogenic substrates only partially reflect physiological reality, because proteasomes select cleavage sites not only according to chemical properties of the P1-aa. Positions beyond P1 influence cleavage site selection (Ustrell et al., 1995ba; Ustrell et al., 1995ab; Cardozo et al., 1994; Ehring et al., 1996). Our own experiments with natural peptides and wild-type and mutant yeast 20S proteasomes showed that the cleavage specificities of the three active sites partially overlap for cleavages after some hydrophobic aa (L, Y, A, M). We could also demonstrate that the postulated BrAAP-activity was mainly inherent to  $\beta$ 1,

<sup>2</sup> The positions relative to a cleavage site ( $\downarrow$ ) are numbered as described by Schechter and Berger, 1967:

P<sub>n</sub>-P<sub>n-1</sub>-...-P<sub>2</sub>-P<sub>1</sub>  $\downarrow$  P<sub>1</sub>'-P<sub>2</sub>'-...-P<sub>n-1</sub>'-P<sub>n</sub>'.

which was subsequently confirmed in inhibition studies by McCormack et al. (1998) and Cardozo et al. (1999) and by crystallographic studies by Groll et al. (1999). We also found that some cleavages are exclusively performed by one of the active sites (Dick et al., 1998; for more details, please refer to 2.1). Later, we could for the first time verify the reported P1-preferences of yeast 20S proteasome active sites in *in vitro* digests of a large, unmodified protein, enolase-1 from yeast (436 aa). More importantly, we were able to expand the cleavage motifs for the three active sites to residues in a window from P6-P6' and, using wild-type and mutant 20S proteasomes, could determine the contribution of each active subunit to enolase degradation (Nussbaum et al., 1998; for more details, please refer to 2.2). In short, ChT-like cleavages after hydrophobic aa contributed most to enolase degradation. This finding is paralleled by the observation that a mutation of the critical Thr residue to Ala in Pre2, the yeast  $\beta$ -subunit harboring the ChT-like activity, is lethal for yeast cells (Heinemeyer et al., 1997; Groll et al., 1999). However, this effect is most likely due to ineffective autocatalytic propeptide cleavage in Pre2, resulting in impaired proteasome assembly.

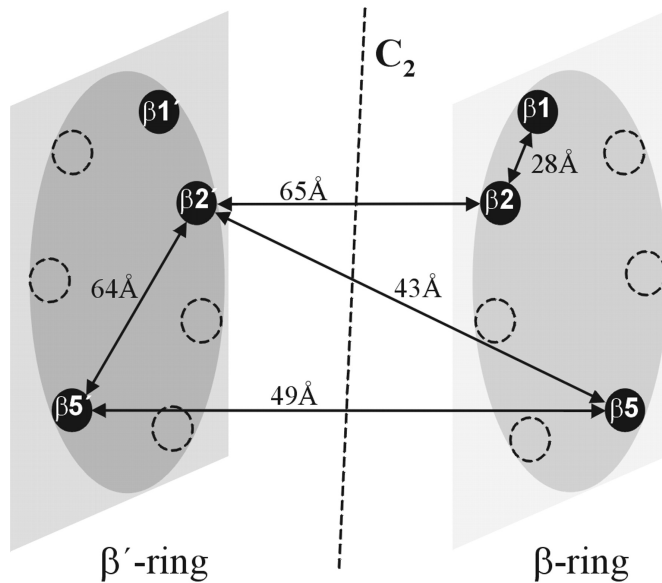
We also found that cleavage site selection is affected by aa in P4 (Pro!) and saw an important contribution of P1'. Several reports confirmed the importance of especially Pro at P4 for proteasomal cleavage site selection: Glas et al. (1998) found that P4 in peptide-based proteasome inhibitors had a major influence on their binding to the active sites. Shimbara et al. (1998) observed that Pro residues at P4 of MHC I ligands conferred protection against internal cleavages by rat liver proteasomes, supporting our findings that Pro is an unfavorable P1-aa and directs cleavages to peptide bonds 3 aa C-terminal of Pro. Finally, the Pro-effect was recently verified using systematically mutated peptides as substrates for *in vitro* digests using human blood proteasomes (Miconnet et al., 2000). Mammalian 20S proteasomes preferentially cleave natural polypeptides *in vitro* after aromatic (F, Y), aliphatic (L, I, V), basic (K, R) or acidic (D, E) residues. However, not each single one of these residues is always used. We were able to confirm and extend these results by enolase-degradation experiments using human erythrocyte 20S proteasomes (see 2.2.9) and constitutive and immuno-proteasomes from B-cell lymphoma cell lines (see 2.5). Our own lab also compared the cleavage specificities of human erythrocyte 20S and 26S proteasomes in *in vitro* degradation of bovine  $\beta$ -casein and found an overlapping, but distinct set of cleavage sites (Emmerich et al., 2000). However, the same rules (cleavage motifs) were employed by both 20S and 26S proteasomes to generate these cleavages. Due to some experimental limitations (e.g. the normalization of the two different proteasome batches, the unusual substrate casein and experimental

variance), it is difficult to draw final conclusions from these results. An additional argument for similar cleavage specificities is the fact that 26S proteasomes, like 20S proteasomes, were able to generate the exact MHC I ligand SIINFEKL from a whole protein *in vivo*, even without the natural flanking sequences (Ben-Shahar et al., 1999). To gain more insight into this question, the cleavage specificities of 20S and 26S proteasomes should be compared systematically using a panel of peptide substrates.

### 1.3.6 Fragment length distribution

Proteasomes from different species do not degrade peptides and proteins into single aa, but into fragments of about 3-30 aa in length. *Thermoplasma* proteasomes generate fragments in the range of 3-30 aa from bovine  $\beta$ -casein (Kisselev et al., 1998). As the degradation products of *Thermoplasma* proteasomes were previously found to average 7-8 aa, a **molecular ruler hypothesis** was coined (Wenzel et al., 1994): Either the distance of active sites or binding constraints for the substrate to the inner surface of the proteasome were thought to determine the fragment length. In the small community of proteasome researchers, the term 'molecular ruler' was almost exclusively linked to the former model (fragment length determined by the distance of active sites), especially after the distance between two active site threonines in the *Thermoplasma* proteasome had been determined to be 28 Å or 2.8nm (Löwe et al., 1995), a stretch just fitting a 7-8mer peptide in extended conformation. For eucaryotic proteasomes, distinct fragment lengths of 8-9 and 14-15 aa were at first reported (Niedermann et al., 1996). Later, a wide fragment length distribution similar to that of *Thermoplasma* proteasomes was detected for yeast (Nussbaum et al., 1998) and mammalian 20S proteasomes (Kisselev et al., 1999b; our own unpublished results, see 2.2.9 and 2.5). The predominating and average fragment lengths still range between 7-9 aa, but no particular fragment lengths really stick out. In order to describe the shape of the fragment length curve (starting at 3 aa, with a peak at 7-9 aa and tailing out around 30 aa) the term **log-normal distribution** was used (Kisselev et al., 1998). Also, the degradation of denatured lysozyme by bovine constitutive and immuno-proteasomes yielded final fragments of 6-20 aa, without the prevalence of a particular size (Wang et al., 1999). However, the strongest piece of evidence against a molecular ruler based on the distance of the active site threonines came from our own experiments using mutant and wt yeast 20S proteasomes: Both wt proteasomes (altogether 6 active sites) and mutants with reduced numbers of active sites due to T1A-point mutations (4 or 2 active sites left) degraded enolase into fragments of the same range, 3-25 aa (Nussbaum et al., 1998; please refer to 2.2 for details). Therefore, the

distance of active sites does not influence the size of the fragments (see Figure 1.3-6 for distances of active sites in eucaryotic proteasomes).



**Figure 1.3-6: Distances of active sites in yeast proteasomes**

The distances between the active sites in yeast proteasomes range from 28-65 Å. Even proteasomes harboring only  $\beta 5$  as active subunit (distance: 49 Å, fitting about 12 aa) generate fragments from 3-25 aa, arguing against the classical interpretation of the molecular ruler hypothesis (Nussbaum et al., 1998; refer to 2.2 for details). [The illustration was taken from Loidl et al. (1999).]

It was proposed before that binding of the regulator PA28 caused proteasomes to directly generate MHC I ligands by coordinated dual cleavages (Dick et al., 1996). Only later, the contribution of the different active sites to the generation of some of the dual cleavage products could be determined using mutant yeast proteasomes. These data, together with the crystallographic structure of yeast proteasomes, revealed that not all previously identified dual cleavage products can be generated by two simultaneous cleavages (Dick et al., 1998). The sheer distances between the active site threonines should theoretically permit the generation of dual cleavage products from 7 to 16 aa in length (see Figure 1.3-6). Yet, it must be emphasized that these distances do not reflect the shortest possible stretch of aa that could be cleaved by two active sites simultaneously. Steric requirements must also be taken into account, for example the correct orientation of the substrate aa chain towards and its interaction with the inner surface of the substrate binding pocket, probably over a stretch of 6-10 aa up- and downstream of the cleavage site (Groll et al., 1997; Nussbaum et al., 1998). This could require some bending of the aa chain that would substantially increase the length of dual cleavage products. However, as the bulk of

proteasomal cleavage products does fall into the size range of 7-9 aa, it is likely that many of them are produced by *consecutive* cleavages at nearby or distant active sites.

Both 20S and 26S proteasomes from rabbit degraded several denatured or unfolded proteins (IGF, ovalbumin, casein) into fragments of 3-22 aa, though at the same time producing a different peptide footprint in HPLC-chromatograms (Kisselev et al., 1999b). A comparison of the protein degradation by human 20S and 26S proteasomes performed in our lab showed that 26S proteasomes degrade bovine  $\beta$ -casein into slightly longer fragments (range: 4-36 aa; average length: 15-18 aa) than 20S proteasomes (3-31 aa, 7-10 aa). (Emmerich et al., 2000). The pool of generated peptides was only partially identical between the two proteasome species. However, as stated above, these discrepancies could be explained by the experimental difficulties to normalize different proteasome batches. The fact the cleavage site selection followed similar rules for both 20S and 26S proteasomes implies that the cleavage preferences of 20S proteasomes are not changed after binding of the 19S regulatory particle. Because of the aforementioned and due to the lack of *in vitro* substrates for 26S proteasomes, it seems fair to study 20S proteasomes *in vitro* to gain insight in the *in vivo* functions of 26S proteasomes.

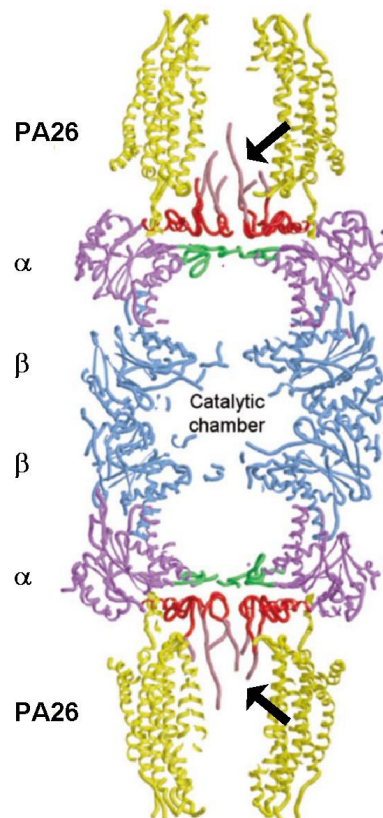
### 1.3.7 Modulation by interferon $\gamma$ (IFN- $\gamma$ )

The chemical messenger IFN- $\gamma$ , which is mainly produced by CTL, NK and inflammatory T ( $T_H1$ ) cells, regulates several immunologically important processes such as the activation of macrophages and NK cells, B cell differentiation, antibody isotype switching and the suppression of helper T ( $T_H2$ ) cell responses. In addition, IFN- $\gamma$  can directly clear virus-infection in a noncytopathic way. Antigen presentation is boosted by IFN- $\gamma$  in two ways: On the one hand, the transcription of MHC I and II genes is enhanced, resulting in increased surface expression of MHC I and II. On the other hand, numerous components of the MHC I antigen processing machinery (TAP, PA28 and immuno-subunits of the immuno-proteasome) are induced. It is assumed that these components have evolved to "fine-tune" the protein degradation machinery for its role in immune responses against viruses. The influence of PA28 and the incorporation of immuno-subunits on protein degradation by proteasomes is discussed in detail below.

#### 1.3.7.1 PA28, the activator/regulator of mammalian proteasomes

The proteasome modulator PA28 (also called 11S regulator) was discovered through its ability to stimulate proteasomal hydrolysis of short fluorogenic peptide substrates (Dubiel et al., 1992). PA28 consists of two homologous subunits, PA28 $\alpha$  and PA28 $\beta$  (Mott et al.,

1994; Ahn et al., 1995; Song et al., 1996), which form a hetero-hexameric ring of  $\alpha_3\beta_3$  stoichiometry. Recombinant PA28 $\alpha$  or PA28 $\beta$  form homo-heptameric structures which are also able to stimulate peptide hydrolysis, though less efficiently than PA28 $\alpha/\beta$  hetero-hexamers (Stohwasser et al., 2000). In the electron microscopic picture, PA28 can be seen attached to the two ends of the 20S cylinder (Gray et al., 1994). Very recently, the three-dimensional structure of a PA28-20S complex was solved (Figure 1.3-7): The homo-heptameric ring of *Trypanosoma brucei* PA26 (equivalent of mammalian PA28; Yao et al., 1999) binds to both ends of the 20S cylinder via the C-terminal interaction loops in each PA28 monomer. These interactions lead to rearrangements in the N-termini of the 20S proteasome  $\alpha$ -rings which cause them to open (see Figure 1.3-8).



**Figure 1.3-7: Complex of *Trypanosoma brucei* PA26 and yeast 20S proteasomes**

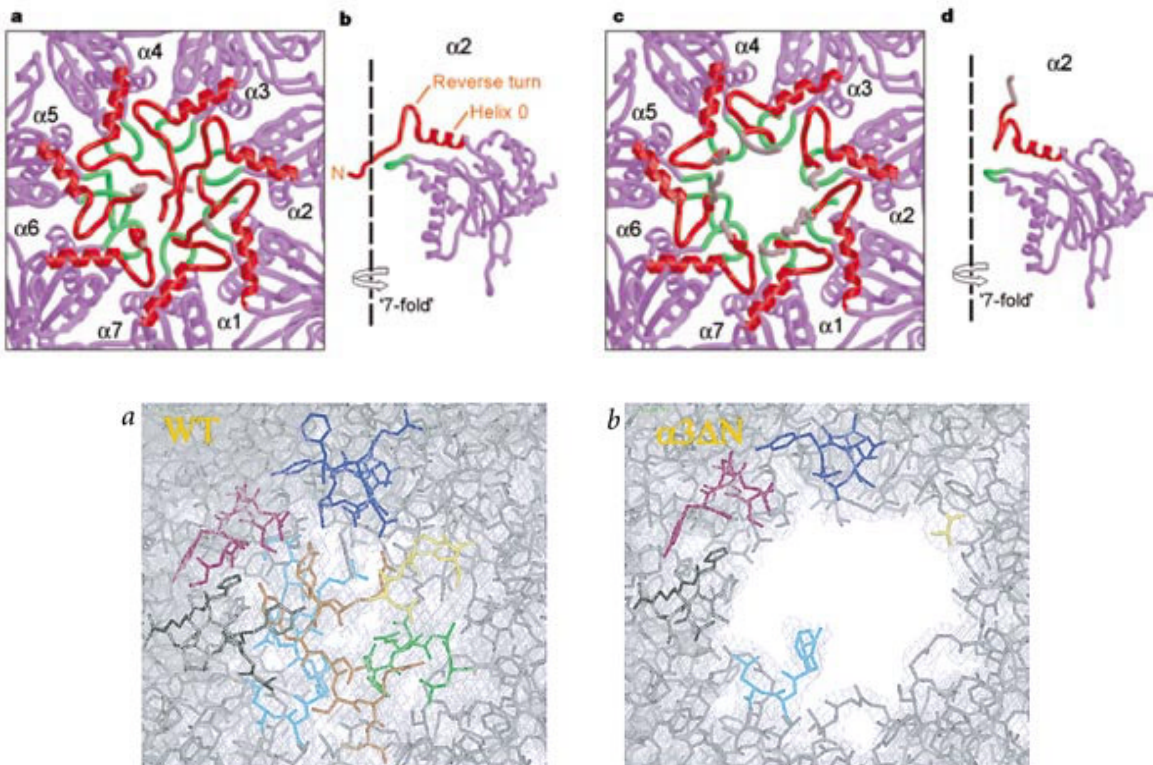
The homo-heptameric PA26 (yellow) from *Trypanosoma brucei* can be seen attached to both ends of the 20S cylinder. The N-termini of the 20S  $\alpha$ -subunits (red and pink) become rearranged (arrows), thereby opening the  $\alpha$ -ring. The 20S cylinder is clipped to visualize the central channel. [Image taken from Whitby et al. (2000).]

Both PA28 $\alpha$  and PA28 $\beta$  are strongly induced by IFN- $\gamma$  and IFN- $\alpha$  (Ahn et al., 1995), a fact that early on pointed towards an immuno-modulatory role of PA28. Experiments using

short fluorogenic substrates initially showed an influence of PA28 on the cooperativity of the active sites of 20S proteasomes (Ma et al., 1992; Yukawa et al., 1993). PA28 was therefore believed to function as a positive, allosteric regulator. Moreover, a change of the proteasomal cleavage pattern was observed in natural peptides in the presence of PA28 (Groettrup et al., 1995). Later results, however, demonstrated a different aspect: PA28 induced stronger immunologically relevant cuts, whereas the cleavage pattern was not changed (Dick et al., 1996; Niedermann et al., 1997). Hence, PA28 caused better lysis of virus-infected cells by CTL (Groettrup et al., 1996b). PA28 $\beta$ -deficient mice lack PA28 $\alpha/\beta$ -complexes and display impaired CTL responses and immuno-proteasome assembly (Preckel et al., 1999), although a role for PA28 in immuno-proteasome assembly was not observed by a different group (Groettrup, personal communications). A recent report showed comparable activation of all three proteasomal activities by PA28 $\alpha/\beta$  hetero-hexamers without a change in the maximal activities. Kinetic modeling led to the assumption that PA28 exerts its effect by enhancing the uptake and release of short peptides, possibly through opening of the  $\alpha$ -rings of the 20S proteasome (Stohwasser et al., 2000). This hypothesis was now confirmed by solving the crystallographic structure of a PA28-20S complex: The binding of PA28 induces an opening in the  $\alpha$ -rings of 20S proteasomes which will probably improve access of substrates to active sites or egress of degradation products from the proteolytic chamber (Whitby et al., 2000; Figure 1.3-8, top). Interestingly, both 'open' and 'closed' conformations of the  $\alpha$ -rings of yeast 20S proteasomes were detected by atomic force microscopy in the absence of PA28 (Osmulski and Gaczynska, 2000). The open conformation could be stabilized by addition of a short fluorogenic peptide or denatured lysozyme, but not by addition of active site inhibitors. The latter stabilized the closed conformation, suggesting that substrate binding to the active sites could regulate access to the proteolytic chamber. Latest findings support the concept that the N-termini of the  $\alpha$ -rings function as regulators of substrate entry: When the  $\alpha$ 3-subunit N-terminus was shortened by 9 aa, yeast 20S proteasomes were as active as SDS- or 19S-activated 20S proteasomes, suggesting that the 19S cap induces the  $\alpha$ -rings to open *in vivo* (Groll et al., 2000; see 1.3.8.1 and Figure 1.3-8, bottom).

There also exists a highly conserved constitutive PA28 subunit, called PA28 $\gamma$  or Ki antigen, which is almost exclusively found in the nucleus of cells and which is replaced by PA28 $\alpha$  and PA28 $\beta$  after IFN- $\gamma$  induction (Jiang and Monaco, 1997; Tanahashi et al., 1997). Mice lacking PA28 $\gamma$  show retarded growth (Murata et al., 1999), but its role for the

modulation of 20S proteasomes is not yet understood.



**Figure 1.3-8: Open conformations of yeast 20S proteasome  $\alpha$ -rings**

(Top) Open conformation induced by binding of *Trypanosoma brucei* PA26 to yeast 20S proteasomes [taken from Whitby et al. (2000)]: (A, B) 20S proteasome without PA26; (C, D) after binding of PA26. After binding of PA26, the N-termini of the  $\alpha$ -ring subunits become flexible and open up a pore in the  $\alpha$ -ring. (Bottom) Electron density maps of the  $\alpha$ -rings in wt and mutated yeast 20S proteasomes [taken from Groll et al. (2000)]: (A) wt yeast proteasome  $\alpha$ -ring; (B) open conformation in yeast proteasomes with 9 aa deleted at the  $\alpha 3$ -N-terminus.

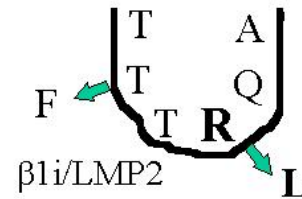
### 1.3.7.2 Exchange of constitutive for IFN- $\gamma$ -inducible active site $\beta$ -subunits

In infection, interferons  $\alpha$  and  $\gamma$  induce the synthesis of new proteolytically active proteasomal  $\beta$ -subunits (so-called **immuno-subunits**) in mammals:  **$\beta 5i$  / low-molecular-weight protein 7 (LMP) 7** (replaces  $\beta 5/X$ ),  **$\beta 1i$  / LMP2** (replaces  $\beta 1/\delta$ ) and  **$\beta 2i$  / multicatalytic endopeptidase complex-like 1 (MECL-1)** (replaces  $\beta 2/Z$ ). These subunits are then incorporated into newly assembled 20S proteasomes, yielding so-called **immuno-proteasomes**. The assembly of 20S proteasomes is tightly regulated, thus securing that homogeneous subsets of constitutive or immuno-proteasomes are formed:  $\beta 1i/LMP2$  and  $\beta 2i/MECL-1$  are mutually required for incorporation into 20S proteasomes



(Stohwasser et al., 1997), probably due to their direct neighborhood in the  $\beta$ -rings (Schmidtke et al., 1999), and the three immuno-subunits are incorporated cooperatively (Griffin et al., 1998). However, this regulation seems to be leaky in one direction:  $\beta 5i/LMP7$  can also be incorporated into  $\beta 1/\delta$ - $\beta 2/Z$ -containing proteasomes (Kingsbury et al., 2000). Latest results describe that up to six subtypes of 20S proteasomes can coexist in one kind of rat tissue (Dahlmann et al., 2000). These subtypes resemble constitutive, immuno- and intermediate-type proteasomes differing with regard to their enzymatic characteristics.

As the secretion of IFN- $\gamma$  coincides with the activation of CTL and APC, it was assumed that the induction of immuno-subunits improves antigen processing by quantitative or qualitative changes in the peptide pool generated by proteasomes. The exchange  $\beta 1/\delta \rightarrow \beta 1i/LMP2$  reduces post-acidic cleavages in favor of ChT-like ones, an immunologically favorable modulation because mouse and human MHC I molecules do not bind peptides with acidic C-termini (Figure 1.3-9). Besides, TAP has a low affinity for peptides with acidic C-termini. First studies using fluorogenic peptides, inhibitory peptide analogs or naturally occurring peptides as substrates showed a prominent influence of the immuno-subunits onto cleavage site selection (Gaczynska et al., 1993; Gaczynska et al., 1994; Boes et al., 1994; Eleuteri et al., 1997; Orłowski et al., 1997). To sum up, post-acidic cleavages were reduced and ChT-like cleavages after branched chain and aromatic aa were enhanced for immuno-proteasomes. At first, a correlation between selected cleavage sites in a naturally occurring peptide and the P1-preferences observed with fluorogenic substrates could not be found (Kuckelkorn et al., 1995). However, later results confirmed the P1-preferences in polypeptide digests: Cleavages after hydrophobic residues in denatured lysozyme (129 aa) were augmented with bovine immuno-proteasomes from spleen, as compared to constitutive proteasomes from pituitary, whereas post-acidic ones were decreased (Cardozo and Kohanski, 1998; Wang et al., 1999). However, in these studies only few individual fragments were identified in a purely *qualitative* way by mass spectrometry and there was no statistical analysis of cleavage site selection. We have just completed the analysis of human constitutive and immuno-proteasome cleavages in digests of enolase. In order to arrive at accurate cleavage motifs spanning several P- and P'-residues, we for the first time identified more than 100 fragments each in a *quantitative* way. We did not find differences in fragment lengths, but confirmed the augmented preference of immuno-proteasomes for hydrophobic aa in P1 (Toes, submitted for publication; see 2.5 for details). Our data will be used for the prediction of constitutive and immuno-proteasome cleavages.

$\beta 5 / \text{Pre}2 / \text{MB-1} || \beta 5 \text{i} / \text{LMP}7$  $\beta 2 / \text{Pup}1 / \text{Z} || \beta 2 \text{i} / \text{MECL-1}$  $\beta 1 / \text{Pre}3 / \delta$ 

**Figure 1.3-9: Active site surfaces and their influence by exchange for immuno-subunits**

Changes in the cleavage preferences of immuno-proteasomes can be explained by the exchanges of few aa residues in  $\beta 1 \text{i} / \text{LMP}2$  as compared to  $\beta 1 / \delta$  (Groll et al., 1997). The exchange  $\text{R} \rightarrow \text{L}$  abrogates the preference of this pocket to accommodate acidic aa; the exchange  $\text{T} \rightarrow \text{F}$  reduces space, which could be the reason for the preference of immuno-proteasomes for the aa Gly in  $\text{P}1'$  (see 2.5 for detail). The aa lining the pockets of the other two active site subunits are not changed by the exchange for  $\text{IFN-}\gamma$  inducible subunits. [Graph reproduced with the permission of T.P. Dick.]

### 1.3.7.3 Differential processing of CTL epitopes by constitutive and immuno-proteasomes

Soon after the  $\text{IFN-}\gamma$  inducible proteasome subunits had been discovered, it was shown that in cell lines they were not generally required for stable MHC I surface expression and for processing and presentation of antigenic peptides from influenza virus and an intracellular protein (Momburg et al., 1992; Arnold et al., 1992). However, mice lacking LMPs showed deficiencies in the immune system: LMP7-deficient mice exhibit reduced MHC I surface expression and presentation of the endogenous antigen HY (Fehling et al., 1994). LMP2-deficient mice display lower numbers of  $\text{CD}8^+$  T lymphocytes, diminished presentation of an Influenza epitope and a reduced CTL-response against Influenza virus infection (Van Kaer et al., 1994).

The influence of immuno-subunit incorporation onto antigen processing has been studied extensively. However, no general rule has been found for the presence of immuno-proteasomes and CTL-activation. Whereas the presentation of some CTL-epitopes seems to profit from immuno-proteasomes (Van Kaer et al., 1994; Boes et al., 1994; Fehling et al., 1994; Schwarz et al., 2000; Sijts et al., 2000a; Sijts et al., 2000b), others seem to be preferentially destroyed by immuno-proteasomes (Morel et al., 2000). Interestingly, many epitopes profiting from immuno-proteasome processing are derived from viral proteins, whereas detrimental effects of immuno-proteasomes have been reported for a ubiquitously expressed self-protein (see Stoltze et al., 2000a for review). This could reflect the evolutionary adaptation of the antigen processing machinery (specifically: the

proteasome) to fight viruses. However, a recent publication shows the generation of one viral epitope and the destruction of another by immuno-proteasomes (van Hall et al., 2000). This differential behavior and the opposing results summarized above imply that it is only the aa sequence of the substrate that determines whether an epitope will be preferentially generated or destroyed by the cleavage preferences of immuno-proteasomes.

Recently it was reported that immuno-proteasomes (and PA28) are up-regulated in dendritic cells, especially after their transition from immature, antigen capturing cells to mature, antigen presenting and T cell-priming cells (Macagno et al., 1999; Morel et al., 2000). This could mean that CTL are predominantly activated against epitopes generated by immuno-proteasomes. The consequences of this are discussed in 2.5.5.

### 1.3.8 Protein degradation by proteasomes

#### 1.3.8.1 *The narrow gate: It is easier for a camel...*

The diameter of the *Thermoplasma* proteasome  $\alpha$ -rings, the proposed substrate entry gates, is only 13 Å, a hole barely allowing the simultaneous passing of 2-3 extended polypeptide chains. In analogy to a biblical metaphor, it is hence easier for a camel to be threaded through the eye of a needle than for a folded protein to enter proteasomes. In other words, substrates probably have to be completely unfolded before they get access to the proteolytic chamber of 20S proteasomes. This unfolding is probably exerted by regulatory protein complexes binding to both sides of the 20S cylinder. In the archaeobacterium *Methanococcus jannaschii*, a proteasome regulator (PAN = proteasome-activating nucleotidase) with chaperone-like and unfoldase activity was just recently identified (Zwickl et al., 1999; Benaroudj and Goldberg, 2000). In yeast proteasomes, the N-terminal residues of the  $\alpha$ -rings were found to seal the 20S-cylinder completely (Groll et al., 1997), raising the question how substrate entry is possible at all. Two very recent publications have shed light on this problem: A homo-heptameric bacterial (*Trypanosoma brucei*) PA28 induced the rearrangement of several  $\alpha$ -subunit N-termini, creating a 13 Å opening in the  $\alpha$ -rings (Whitby et al., 2000; see Figure 1.3-8, top). The effect of a 9 aa deletion in the  $\alpha$ 3-subunit N-terminus was similar and paralleled activation of 20S proteasomes by the 19S cap or the detergent SDS (Groll et al., 2000; see Figure 1.3-8, bottom). For 26S proteasomes, the 19S cap, in addition to its possible role in the  $\alpha$ -ring opening, is thought to exert unfoldase activity through its numerous ATPases (DeMartino and Slaughter, 1999). However, until now only chaperone-like activity (i.e. prevention of

protein aggregation and refolding) has been demonstrated for the base of the 19S cap (Braun et al., 1999; Strickland et al., 2000). It is conceivable that especially hybrid proteasomes (see 1.3.2) could link protein degradation (through ubiquitin-recognition by the 19S lid) and efficient generation of antigenic peptides (through activation of proteolysis by PA28). Whitby et al., 2000 speculate that in hybrid proteasomes the 19S cap could regulate substrate entry, whereas PA28 would regulate fragment exit resulting in longer fragments.

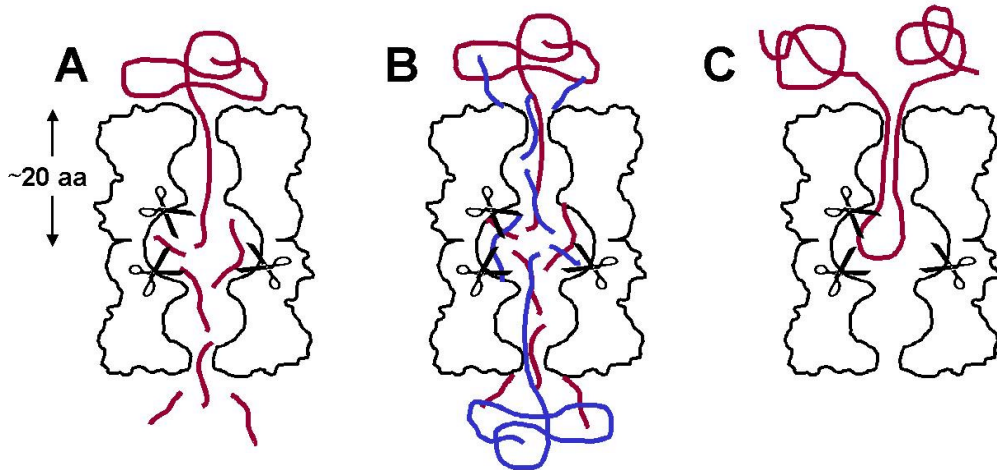
For all of the above, it is difficult to investigate the degradation of native proteins by 20S proteasomes *in vitro*. Up to now, the digestion of proteins has thus almost exclusively been shown for denatured or modified proteins. The degradation of oxidized insulin B-chain was described as 'processive channeling' of degradation intermediates through the human 20S proteasome cylinder (Dick et al., 1991). The fragment lengths found after degradation of oxidized insulin B-chain and denatured hemoglobin by *Thermoplasma* proteasomes led to the above mentioned molecular ruler hypothesis (Wenzel et al., 1994). Also, ovalbumin and  $\beta$ -galactosidase could only be degraded by mammalian 20S after prior denaturation by S-carboxymethylation (Dick et al., 1994). Only bovine  $\beta$ -casein is degraded in an unmodified form by 20S and 26S proteasomes (Pereira et al., 1992; Emmerich et al., 2000). This is probably so because casein is already devoid of tertiary structure in its native form, possibly due to its many proline residues. To speak with the above metaphor, casein and denatured proteins thus rather resemble a snake than a camel when approaching the narrow gate of the proteasomal  $\alpha$ -rings.

#### 1.3.8.2 Processive degradation or fierce chopping?

The degradation of fluorescein isothiocyanate (FITC)- labeled casein yielded final degradation products early on in the digest and accumulating over time, without the appearance of any large degradation intermediates (Akopian et al., 1997). Therefore, a **processive** mechanism was postulated for proteasomal protein degradation. Processivity was defined as complete degradation of an individual substrate "protein to oligopeptides before attacking another protein molecule" (Akopian et al., 1997), requiring continuous channeling of the substrate through the proteasome cylinder and continuous cleaving of the substrate. A very vivid picture, namely that of a **bread-cutter**, was used to illustrate processivity. The lack of large degradation intermediates in proteasomal digests was confirmed by ourselves using yeast 20S and the substrate enolase (Nussbaum et al., 1998) and by others using mammalian 20S and 26S proteasomes and FITC-labeled casein as substrate (Kisselev et al., 1999b). It is unclear, however, whether the observed

early accumulation of final degradation products really reflects processive, i.e. continuous channeling of individual substrate molecules through proteasomes. For, the reported experimental results could also be explained by an **exit size filter** for degradation products, which, according to latest results, could be identical to the openings in the  $\alpha$ -rings (Whitby et al., 2000). Moreover, whereas small, final degradation products accumulate in significant amounts over time, large degradation intermediates might simply not arise in amounts above the detection limits of the analytical systems. Short peptides might be spared from further proteasomal degradation for two reasons: (A) They overcome the exit size filter and diffuse out of the proteolytic chamber. (B) They constitute bad substrates for proteasomes (Dolenc et al., 1998). Proteasomal 'shredding' of proteins into small pieces is thought to protect cells from the potential hazards of biologically active degradation intermediates: For example the separation of a regulatory and an active domain of a transcription factor by partial degradation could lead to uncontrollable induction of the target gene. But still, **nonprocessive** degradation of reduced carboxyamidomethylated lysozyme (RCM-lysozyme) by bovine constitutive and immunoproteasomes was recently reported (Wang et al., 1999). Large degradation intermediates, which were later cleaved to small peptides, could be detected early on in digests. According to these results, proteasomal protein degradation would rather be based on free **diffusion** than on directional **channeling** of substrate, as in the processive model. The authors propose that large cleavage products have the same probability of diffusing out of the proteasome cylinder and of being cleaved again. Nonprocessive or partial degradation of substrates could explain degradation intermediates detected in *in vitro* experiments using lysozyme (Orlowski and Michaud, 1989), insulin B chain (Dick et al., 1991) or naturally occurring peptides as substrates (Niedermann et al., 1996; our own unpublished results). Besides, nonprocessivity could be the prerequisite for the activation of some transcription factors by **partial proteasomal processing**: NF- $\kappa$ B can only be formed after the partial, cotranslational processing of its precursor p105 to p50 (Lin et al., 1998); SPT23 and MGA2 are only released after the partial processing of their ER membrane-bound precursors (Hoppe et al., 2000). Intriguingly, in both cases a loop of a fixed protein needs to be inserted into the proteasome for an initial endoproteolytic cleavage and for the subsequent complete degradation of the C-terminal part of the protein. The two processes differ in that for p105, the N-terminus is fixed by the ribosome and the unfixed, C-terminal part is completely degraded, whereas for the precursors of SPT23 and MGA2, the C-terminal part is at first fixed in the ER membrane but later extracted and degraded. However, nonprocessive degradation *in vitro* has only been

reported for relatively short substrates, the longest of which is lysozyme with 129 aa. Large (and unfixed?) protein substrates, on the contrary, might preferentially be degraded in a processive fashion for their slower diffusion and/or higher affinity to proteasomal active sites (Dolenc et al., 1998).



**Figure 1.3-10: Models for proteasomal protein degradation**

(A) An unfolded substrate aa chain enters the proteasome cylinder and is cleaved by the active sites (scissors) in the central cavity; protein fragments leave via the  $\alpha$ -rings. This model would fit the proposed role for hybrid-proteasomes in which the 19S cap unfolds and inserts the substrate and PA28 regulates the exit of fragments (1.3.2 and 1.3.8.1). The distance from the active sites to the outside of the cylinder approximately equals the length of a 20 aa polypeptide in extended conformation. (B) Substrate proteins could enter proteasomes from both ends. Fragment exit via the  $\alpha$ -rings would probably still be possible in this model, but might be slowed. (C) Cleavage of a polypeptide hairpin loop inserted into the proteasome cylinder. In the proposed examples for this model, one part of the polypeptide chain is completely degraded whereas the other can diffuse away as partial degradation product (Lin et al., 1998; Hoppe et al., 2000). Size estimates are taken from Hoppe et al. (2000). Regulatory particles binding to 20S proteasomes were omitted for clarity.

It was observed that short substrate analogs for the post-acidic activity inhibit the ChT-like one whereas substrate analogs for the ChT-like activity stimulate the ChT-like activity. Based on these findings, a **bite-chew model** for protein degradation was proposed (Kisselev et al., 1999a). According to this model, a protein substrate binds to the ChT-like active sites and thereby stimulates ChT-like cleavages. After one round of ChT-like cleavages ("bite"), intermediate degradation products would move on to post-acidic active sites, thereby inhibiting further ChT-like cleavages, in order to be degraded further ("chew"). The bite-chew model was challenged by the **non-catalytic modifier site model** which proposes a regulatory site distinct from the active sites (Schmidtke et al., 2000).

Both models require allosteric effects which have not been observed yet with any imaging method such as electron microscopy or crystallography.

## 1.4 From Proteasomal Degradation Products to CTL Epitopes

### 1.4.1 What makes proteasomal products good CTL epitopes?

Peptides binding to MHC I molecules must meet two criteria, as pointed out earlier (1.1.3.1): Their lengths must range between 8-10 aa and their sequences have to correlate to the MHC I binding motifs, which often require hydrophobic C-termini (Rammensee et al., 1999). As proteasomes generate peptides with predominantly hydrophobic C-termini (see 1.3.5) and as a large fraction of degradation products is larger than 7 aa (see 1.3.6), proteasomal cleavage products seem to fulfill these requirements. In order to further analyze the relationship between proteasomal cleavage products and MHC I ligands, one study compared the cleavage specificities of proteasomes from different species with sequence signals at the peptide termini of 134 MHC I ligands. It was concluded that (A) the all eucaryotic, but not archaeobacterial proteasomes share cleavage preferences beneficial for the generation of the C-termini of MHC I ligands and (B) that MHC I binding motifs probably evolved to match proteasomal cleavage products (Niedermann et al., 1996). A more recent study more systematically checked 286 MHC I ligands for sequence signals and found that the C-termini, but not the N-termini resemble proteasomal cleavage sites (Altuvia and Margalit, 2000). As the cleavage preferences of immuno-proteasomes augment the output of peptides with hydrophobic C-termini and diminish cleavages after acidic aa (see 1.3.7.2), they are thought to produce more potential MHC I ligands than constitutive proteasomes. In addition to the fact that no known mammalian MHC I allele prefers ligands with acidic C-termini (as an exception, some chicken MHC I alleles do), TAP only poorly translocates such peptides in the ER (see below).

### 1.4.2 Additional selection criteria

As mentioned above, the binding motifs of MHC I molecules represent a high obstacle for amino acid sequences to qualify as CTL epitopes (see 1.1.3.1 and 1.2.2). The MHC binding motifs can be viewed as the last - and probably finest - of a series of filters that peptides have to pass before binding to MHC I molecules. One of the first filters in this chain is proteasomal cleavage specificity (see 1.3.5, 1.3.7.2 and 1.3.7.3). However, some in-between filters probably also affect the selection of CTL epitopes, most notably TAP.

Less specific filters include trimming peptidases in the cytosol and the ER and possibly chaperones that channel peptides from the proteasome to TAP and protect them from degradation into single amino acids.

#### 1.4.2.1 TAP specificity

The peptide transporter TAP signifies an additional selection step for CTL-epitopes. Its translocation specificity is species-dependent and has been investigated in detail for rat, mouse and human TAP (Uebel and Tampe, 1999). Mouse TAP translocates peptides from 9 aa in length and with hydrophobic C-, whereas rat and human TAP translocate peptides with both hydrophobic and basic C-termini (Schumacher et al., 1994; Momburg et al., 1994; van Endert et al., 1995). The preference for C-terminal aa characteristics is in line with binding requirements for MHC I molecules: Human and rat MHC I molecules bind peptides with hydrophobic and basic C-termini; mouse MHC I molecules only bind peptides with hydrophobic C-termini. The three N-terminal residues of peptides are of equal importance for TAP-selectivity: Hydrophobic and charged aa are preferred whereas aromatic and acidic (in P1) and Pro residues (in P2, P3) are strongly disfavored. Peptide-binding to some MHC I alleles depends on aromatic aa in P1 (HLA-A\*0201) or on Pro in P2 (the HLA B7 group) of the peptide, leading to the assumption that the ligands for these MHC alleles need to be transported by TAP as N-terminally elongated precursor peptides (van Endert et al., 1995). The central part of TAP-translocated peptides does not display preferences for specific aa residues, however the peptide backbone seems to be important for binding to TAP (Uebel et al., 1997). The specificity of human TAP is summarized in Figure 1.4-1.

	<b>+</b>	<b>K,N,R</b>	<b>R,I,Q</b>	<b>W,Y</b>		<b>R</b>	<b>I</b>		<b>F,L,R,Y,V</b>
<b>Position</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>
	<b>-</b>	<b>D,E,F</b>	<b>P,L</b>	<b>D,E,G</b>	<b>N</b>	<b>T</b>	<b>D,E</b>		<b>D,E,N,S,G</b>

**Figure 1.4-1: Binding motif of human TAP**

Of a 9mer peptide (shaded rectangle), the N-terminal aa and the C-terminus (dark shading) are most important for translocation by TAP. Please note that TAP can translocate longer peptides than the 9mer given in the example. However, it is always the N-terminal and C-terminal residues that decide about affinity to TAP. Preferred (+) or disliked (-) aa residues at the indicated positions; bold residues show the strongest effects. [The figure was adapted from Uebel and Tampe (1999).]



#### 1.4.2.2 *To trim or not to trim, a nearly philosophical question*

Previously, it was observed that proteasomes cleave out the exact MHC I ligand from protein or peptide sequences, without aa extensions on either end of the fragment (Wenzel et al., 1994; Dick et al., 1994; Dick et al., 1996; Niedermann et al., 1997). However, this contradicts the requirement for ligands of certain MHC alleles to be translocated as N-terminally extended precursors, as imposed by TAP specificity (see 1.4.2.1). Trimming of precursor peptides for MHC I ligands was already suggested 10 years ago (Falk et al., 1990). Later, when the first reports showed proteasome-independent trimming of the extended N-termini of CTL epitopes, this was propagated by some researchers in the field as a general mechanism. Experimental data available to date indicate that proteasomes can contribute to the generation of MHC I ligands in three different ways: A) They excise the final MHC I ligands by dominant cleavages (Lucchiari-Hartz et al., 2000). B) They supply final MHC I ligands as side-products next to more dominant N-terminal extensions (Dick et al., 1994; Groettrup et al., 1995; Niedermann et al., 1995; Dick et al., 1996; Niedermann et al., 1996; Theobald et al., 1998). (C) They only generate N-terminally extended precursor peptides of the MHC I ligand which are subsequently trimmed by non-proteasomal activities in the cytoplasm or ER (Craiu et al., 1997; Stoltze et al., 1998; Mo et al., 1999; Paz et al., 1999). Several cytosolic proteases have been suggested to serve as trim-peptidases: the IFN- $\gamma$ -inducible leucine aminopeptidase (LAP, Beninga et al., 1998), tripeptidyl-peptidase II (TPPII, Geier et al., 1999) and bleomycin hydrolase and puromycin-sensitive aminopeptidase (BH, PSA, Stoltze et al., 2000b). An early report proposing C-terminal trimming in the ER has not been supported by additional data (Elliott et al., 1995).

Taken together, these data strongly indicate that the C-termini of MHC I ligands must be generated by proteasomes.

## 1.5 The Prediction of CTL Epitopes

Many immunologists are searching for CTL epitopes that could be beneficial in prevention and therapy of infectious diseases or cancers. The search usually involves the prediction of protein sequences that are likely to bind to MHC I molecules and therefore to activate the cellular immune response. Different approaches for the prediction of MHC I ligands exist, all of them based on peptides eluted from MHC I molecules. Two methods deduce from the aa sequences of known MHC I ligands either minimal matrices (Rammensee et

al., 1999; [www.syfpeithi.de](http://www.syfpeithi.de)) or more intricate neural networks (Honeyman et al., 1998; Brusic et al., 1998) for prediction; another draws on structural features of known MHC I ligands (Schueler-Furman et al., 2000). The prediction of MHC I ligands is reasonable for two reasons: (A) Binding to MHC I constitutes one of the most downstream and strictest of several selection criteria for CTL epitopes; (B) the binding motifs of MHC I molecules follow comparatively simple rules that can easily serve as the basis for prediction. However, the prediction of MHC I ligands still yields many false-positives. As a consequence, the synthesis of the peptides and subsequent MHC I binding studies are indispensable. But there still remains the possibility that the peptides emerging from this refined testing are not generated by proteasomes *in vivo*. It should thus be possible to narrow down the number of potential CTL epitopes by including proteasomal cleavage preferences into the prediction.

Therefore, several research groups are struggling to develop automated prediction tools for proteasomal cleavages. A group in Berlin published the first prediction algorithm: It was based on the experimental cleavage data of only seven naturally occurring peptides, compiled from different literature sources and thus comprising a small and heterogeneous data set (Holzhütter et al., 1999). Lately, a refined kinetic model was published by the same group (Holzhütter and Kloetzel, 2000). We ourselves have only recently created an algorithm based on proteasomal cleavages in enolase. This algorithm uses simple rules to model proteasomal cleavage behavior and is able to predict cleavages in any protein sequence with some accuracy (Kuttler et al., 2000; for details see 2.2). We have just now incorporated the algorithm into a World Wide Web (www)-site: [www.paproc.de](http://www.paproc.de), the first publicly available proteasomal cleavage predictor (for details see 2.4). At least two more research groups are currently developing proteasomal cleavage predictors: A group in Leiden, The Netherlands, will link the characteristics of proteasomal cleavage sites with those of CTL epitopes for more accurate prediction (F. Ossendorp, C. Melief, personal communications). Researchers in Copenhagen, Denmark, in collaboration with us, have generated neural networks for proteasomal cleavage prediction (C. Kesmir, S. Brunak, personal communications). They are based on both the sequences of MHC I ligands and our proteasomal cleavage data. A group in Berlin will try to link MHC I ligand prediction and proteasomal cleavage prediction, again, in collaboration with us and based on already available programs (H. Mollenkopf, personal communications).

The prediction of TAP-binding peptides has only been attempted ones using neural networks (Daniel et al., 1998). This approach has not been followed up. It is conceivable that the prediction of CTL epitopes might benefit from including TAP specificity into the

selection criteria. However, considering the broad specificity of TAP and the scarcity of experimental data, it might be premature to incorporate TAP in such a prediction method.

## 1.6 Goal and Experimental Outline of the Thesis Project

The **goal** of my thesis project was the **characterization of proteasomal cleavage preferences** in whole protein digests. The large amount of data we hoped to collect was to be used for the **development of a prediction device for proteasomal cleavages**, analogous to the database SYFPEITHI ([www.syfpeithi.de](http://www.syfpeithi.de)), a widely-used public predictor for MHC I binding peptides that was created in our institute (Rammensee et al., 1999).

The basis for my Ph.D. thesis project was already laid during my diploma thesis project (April-Dec. 1997) when I screened for proteins that - despite of the aforementioned experimental obstacles (see 1.3.8) - could be degraded *in vitro* by purified 20S proteasomes from yeast. I found yeast enolase-1 to be a promising candidate. This 436 aa protein turned out to be a sound model substrate: Enolase represents in all aspects (aa composition, structure, charge, etc.) a typical and thus representative protein according to the SAPS database (Brendel et al., 1992). At the same time we received wild-type (wt) and mutant yeast 20S proteasomes through collaborations with the groups of Drs Wolf (Inst. for Biochemistry, University of Stuttgart) and Huber (MPI for Biochemistry, Martinsried). These yeast proteasomes were initially characterized using as substrates fluorogenic peptides and natural peptides harboring MHC I ligands and CTL epitopes (see 2.1, published in Dick et al., 1998). I started the first enolase digestion experiments simultaneously and, at the end of my diploma thesis time (i.e. at the beginning of my Ph.D. thesis time), enolase digestion experiments using all available yeast 20S proteasomes had been performed, but still awaited their biochemical analysis.

The experimental course of events was as follows: Yeast enolase-1, as model substrate, was incubated with purified 20S proteasomes from yeast or human cells. The enolase fragments that were generated by proteasomal degradation were subsequently analyzed by biochemical methods (HPLC fractionation, Edman sequencing and mass spectrometry). The identified enolase **fragments** defined the location of proteasomal **cleavage sites**. Both types of information were compiled graphically into so-called **cleavage maps**. Simple counting and complicated **statistical analysis** (hypergeometric distribution, Student's t-test, Chi<sup>2</sup>-test) of fragments, fragment lengths and, most importantly, cleavage sites allowed us to determine the cleavage preferences for all proteasome species studied. Systematic analysis of the aa sequences flanking cleavage

sites led to the identification of features in the substrate sequence (**cleavage motifs**) that are crucial for cleavage site selection by proteasomes (see 2.2, published in Nussbaum et al., 1998, and 2.5, Toes et al., submitted for publication).

My experimental data were used – in a close and fruitful collaboration with the Department for Biomathematics at the University of Tübingen (Prof. K.P. Hadeler and Dr. C. Kuttler) – as training data for **computer algorithms modeling proteasomal cleavages**. These algorithms were finally tied into a publicly available www-site named **PAProC**, for 'Prediction Algorithm for Proteasomal Cleavages', that is already being accessed by many researchers world-wide (see 2.3, Kuttler et al., 2000, and 2.4, Nussbaum et al., in press).

## 2 Results

### 2.1 Contribution of proteasomal $\beta$ -subunits to the cleavage of peptide substrates analyzed with yeast mutants

#### 2.1.1 Summary

Proteasomes generate peptides that can be presented by MHC class I molecules in vertebrate cells. Using yeast 20S proteasomes carrying different inactivated  $\beta$ -subunits we investigated the specificities and contributions of the different  $\beta$ -subunits to the degradation of polypeptide substrates containing MHC class I ligands and addressed the question of additional proteolytically active sites apart from the active  $\beta$ -subunits. We find a clear correlation between the contribution of the different subunits to the cleavage of fluorogenic and long peptide substrates, with  $\beta 5/Pre2$  cleaving after hydrophobic,  $\beta 2/Pup1$  after basic and  $\beta 1/Pre3$  after acidic residues, but with the exception that  $\beta 2/Pup1$  and  $\beta 1/Pre3$  can also cleave after some hydrophobic residues. All proteolytic activities including the "branched chain amino acid preferring" (BrAAP) component are associated with either  $\beta 5/Pre2$ ,  $\beta 1/Pre3$  or  $\beta 2/Pup1$ , arguing against additional proteolytic sites. Because of the high homology between yeast and mammalian 20S proteasomes in sequence and subunit topology and the conservation of cleavage specificity between mammalian and yeast proteasomes our results can be expected to also describe most of the proteolytic activity of mammalian 20S proteasomes leading to the generation of MHC class I ligands.

#### 2.1.2 Introduction

The 20S proteasome, a threonine protease (Seemüller et al., 1995) and member of the Ntn-amidohydrolase family (Brannigan et al., 1995), is the central component of the non-lysosomal proteolytic system in eukaryotes. It is not only involved in general protein catabolism but also in major regulatory processes such as cell cycle control and signal transduction pathways (Hochstrasser, 1995; Coux et al., 1996; Hilt and Wolf, 1996). In vertebrates the proteasome additionally caters to the immune system: peptides generated as degradation products are delivered to MHC class I molecules for presentation on the cell surface (Lehner and Cresswell, 1996). Although localized in the cytosol and nucleoplasm, the proteasome appears to be responsible for ER-associated protein degradation as well (Hiller et al., 1996; Wiertz et al., 1996; Sommer and Wolf, 1997).

The cylindrical 20S proteasome is composed of four longitudinally stacked heptameric rings. The two outer rings are made up of  $\alpha$ -subunits and connect the 20S particle to supplementary complexes, most notably the 19S cap and the IFN $\gamma$ -inducible PA28 $\alpha\beta$  complex. The 19S cap complex confers specificity for ubiquitinated substrates (Hilt and Wolf, 1996). The PA28 $\alpha\beta$  complex appears to be an immunological adaptation, modulating the cleavage mechanism and improving the yield of antigenic peptides (Dick et al., 1996; Groettrup et al., 1996).

The two inner rings, composed of  $\beta$ -subunits, harbor the catalytically active subunits, displaying their active sites on the inner surface of the central tunnel. Most  $\beta$ -subunits are generated from precursors that undergo N-terminal processing during proteasome assembly (Seemüller et al., 1996). Three of the mature subunits (MB1,  $\delta$  and Z in human;  $\beta$ 5/Pre2,  $\beta$ 1/Pre3 and  $\beta$ 2/Pup1 in yeast) are proteolytically active and carry an amino-terminal threonine residue as the catalytic nucleophile (Fenteany et al., 1995; Groll et al., 1997; Heinemeyer et al., 1997; Arendt and Hochstrasser, 1997). Some of these critical Thr residues can be covalently modified by the proteasome-specific inhibitor lactacystin (Fenteany et al., 1995). In higher vertebrates, the cytokine IFN $\gamma$  induces the expression of three additional active proteasome subunits, LMP7, LMP2 and MECL1, which replace their constitutive counterparts MB1,  $\delta$  and Z, respectively (Savory et al., 1993; Reidlinger et al., 1997; Hisamatsu et al., 1996). However, how these exchanges modulate proteasome function and specificity is still poorly understood.

Biochemical studies on the specificity of the 20S proteasome led to the description of three distinct proteolytic components, designated as the chymotrypsin (CT)-like, trypsin (T)-like and peptidylglutamyl-peptide hydrolyzing (PGPH) activities. Recent studies confirmed that the yeast homologues of MB1, Z and  $\delta$  are the only active subunits with a functional N-terminal Thr nucleophile (Heinemeyer et al., 1997; Arendt and Hochstrasser, 1997) and represent the above mentioned three classical activities, as defined with certain fluorogenic model substrates.

However, some biochemical studies also pointed towards the existence of additional proteolytic components within the proteasome (Orlowski et al., 1993): one of those, the so-called "branched chain amino acid preferring" (BrAAP) component, cleaving after Leu in certain model substrates, was also assumed to represent the major protein degrading activity of the proteasome (Pereira et al., 1992). The existence of such additional activities was concluded from inhibitory studies. While the cleavage of fluorogenic substrates

representing the three classical activities could be inhibited with the general serine protease inhibitor 3,4-dichloroisocoumarin (DCI), the same treatment led to an increased turnover of certain other substrates, defined to represent the BrAAP component (Orlowski et al., 1993; Cardozo et al., 1992). Using more specific inhibitors evidence was provided that the BrAAP and the PGPH-activity might be related to each other (Cardozo et al., 1996; Vinitsky et al., 1994). More recently, the arrangement of the intermediately processed  $\beta 7/\text{Pre}4$  and  $\beta 6/\text{C}5$  propeptides in the crystal structure of the yeast 20S proteasome led to the postulation of an additional catalytic activity, not related to the Thr sites and possibly involved in the processing of the  $\beta$ -subunit propeptides (Groll et al., 1997).

In this study we investigated whether the specificities defined with small fluorogenic substrates do reflect truly distinct catalytic sites and how these specificities relate to the cleavages that can be observed in peptide substrates containing known MHC class I ligands. In addition, we addressed the question of additional proteolytically active sites associated with the yeast 20S proteasome by analysis of the BrAAP component.

Using highly purified wt and mutant yeast proteasomes and several inhibitors of the proteasomal activities, we examined the cleavage specificity of the three active subunits towards short fluorogenic peptides as well as peptide substrates derived from the murine JAK1 tyrosine kinase, the pp89 protein from murine cytomegalovirus (MCMV) and the nucleoprotein (NP) from Influenza A/PR/8/34. We found that all observed cleavages could be correlated to the three known active sites showing that no additional proteolytic activity exists within the yeast 20S proteasome for the given set of substrates. Furthermore, we found that the specificities of the different subunits towards fluorogenic substrates are mainly identical to those observed for peptide substrates with few, but notable exceptions. This allows now the clear association of certain cleavage events within peptide substrates with certain  $\beta$ -subunits and therefore an assessment of the contribution of the different  $\beta$ -subunits to the generation of proteasomal degradation products.

The 20S proteasomes from yeast and higher eukaryotes are highly homologous. As shown in this paper and recently by Niedermann et al. (Niedermann et al., 1997), most cleavage sites in peptides and proteins are conserved between yeast and man. It is therefore very likely that results obtained with the yeast proteasome will improve our understanding of vertebrate proteasome function and specificity and consequently of its role in the generation of CTL epitopes.

## 2.1.3 Materials & Methods

### 2.1.3.1 Generation of yeast 20S proteasome mutants

The generation of mutant strains YUS4 (*pup1-T1A*), YUS1 (*pre3-T1A*) YUS5 (*pup1-T1A pre3-T1A*) and YWH23 (*pre2-K33A*) (numbering refers to amino acid positions in mature subunits) have been described recently (Heinemeyer et al., 1997). Wild-type yeast cells were as described (Groll et al., 1997).

### 2.1.3.2 Purification of proteasomes

The purification of 20S proteasomes from *S. cerevisiae* has been described recently (Groll et al., 1997). The same procedure was applied to the purification of the various mutant proteasomes.

### 2.1.3.3 Measurement of proteasomal activities against substrates with fluorogenic leaving group

Fluorogenic substrates Benzyloxycarbonyl-Gly-Gly-Leu-7-Amino-4-methylcoumarin (Z-GGL-AMC), Succinyl-Leu-Leu-Val-Tyr-7-Amino-4-methylcoumarin (Suc-LLVY-AMC), Benzyloxycarbonyl-Ala-Arg-Arg-7-Amino-4-methylcoumarin (Z-ARR-AMC) and Benzyloxycarbonyl-Leu-Leu-Glu- $\beta$ -Naphthylamide (Z-LLE- $\beta$ NA) (all purchased from Bachem, Heidelberg) were prepared from 10mM stocks in DMSO. Typically 1 $\mu$ g of proteasome (with and without inhibitors) was incubated in a 100 $\mu$ M substrate solution. Fluorescence of the leaving group was determined after incubation times of 1-5 h with a Millipore spectrophotometer at 380nm excitation and 440nm emission for AMC and at 330nm excitation and 410nm emission for  $\beta$ NA.

### 2.1.3.4 Inhibitors

Lactacystin was purchased from E. J. Corey (Harvard University), 3,4-dichloroisocoumarin (DCI) and Calpain Inhibitor I (LLnL) were from Sigma.

### 2.1.3.5 Measurement of non-fluorogenic leaving group by azo coupling

The BrAAP substrate Z-GPALA-p-benzoic acid was a gift from Dr. M. Orlowski (Mount Sinai, New York). It was incubated in 100mM Tris pH 8.0 at a final concentration of 1mM together with 1 $\mu$ g of proteasome, in the presence or absence of 0.01 units Aminopeptidase M (Sigma). The reaction was stopped by addition of one reaction volume 10% trichloroacetic acid and developed by sequential addition of 2 volumes 0.1% NaNO<sub>2</sub>



(in 1N HCl), 2 volumes 0.5% Ammoniumsulfamate (in 1N HCl) and 4 volumes 0.05% N-Naphthylethyldiamine (in 95% ethanol). In the course of the development reaction an azo dye is generated from free p-Amino-Benzoic acid. The extinction of the dye is measured at 545nm.

#### *2.1.3.6 Determination of proteasomal cleavages in unmodified 19-25mer peptides*

Peptides were synthesized on an ABI 432A (Applied Biosystems, Weiterstadt, Germany) automated peptide synthesizer applying Fmoc chemistry and purified by RP-HPLC (Beckman, System Gold). Incubations of 1 $\mu$ g proteasome with peptide substrates at a final concentration of 50 $\mu$ M were performed at 37°C in a total volume of 300 $\mu$ l assay buffer (20mM HEPES/KOH pH 7.8; 2mM MgAc<sub>2</sub>; 1mM DTT).

#### *2.1.3.7 Separation of cleavage products and online analysis by mass spectrometry*

Proteasome digests were analyzed with two different LC-MS systems. In case of the first system, aliquots of 40 $\mu$ l were loaded by autosampler (CTC A200S) and separated by HPLC (Sykam, Gilching, Germany), equipped with a  $\mu$ RPC C2/C18 SC 2.1/10 column (Pharmacia, Freiburg, Germany). Eluent A: 0.05% TFA; Eluent B: 80% Acetonitrile containing 0.055% TFA. Gradient: 23-63% B in 20 min; Flow rate: 50 $\mu$ l/min, gradient started 5 min after sample injection. Analysis was performed on-line by a tandem quadrupole mass spectrometer (TSQ 700; Finnigan MAT, Bremen, Germany) equipped with an ESI-F electrospray ion source. Each scan was acquired in centroid mode over the range m/z 300-2200 in 3 seconds.

In case of the second LC-MS system, aliquots of 5 $\mu$ l were separated by HPLC (ABI 140D solvent delivery system), equipped with a 300 $\mu$ m C18 RP column (Gromsil, Grom, Herrenberg, Germany). Eluent A: 4mM NH<sub>4</sub>Ac, adjusted to pH3 with formic acid; Eluent B: 2mM NH<sub>4</sub>Ac in 70% Acetonitrile, adjusted to pH3 with formic acid. Gradient: 20-75% B in 25 min; Flow rate after pre-column split: 6 $\mu$ l/min. Analysis was performed on-line by a hybrid quadrupole orthogonal acceleration time of flight mass spectrometer (Q-TOF; Micromass, Manchester, UK) equipped with an electrospray ion source. Each scan was acquired over the range m/z 400-1500 in 3 seconds.

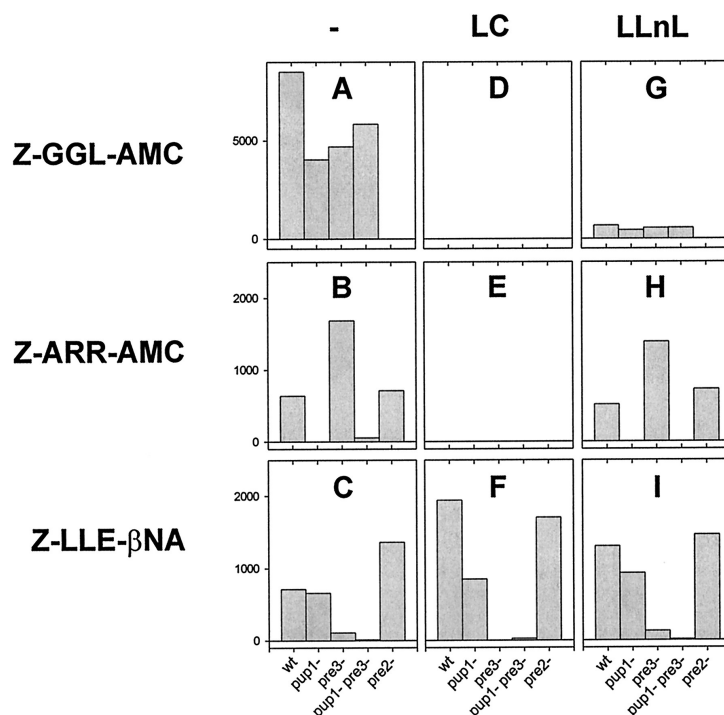
Peptides were identified by their molecular mass. Additionally, the identity of main cleavage products was confirmed by fragmentation with argon atoms in ms/ms experiments. For relative quantitation of single molecular species integrated ion currents were calculated and compared.

## 2.1.4 Results

### 2.1.4.1 Cleavage of small fluorogenic substrates

Cleavage of the standard fluorogenic substrates GGL, ARR and LLE has been shown recently to be subunit-specific (Heinemeyer et al., 1997). Using for the first time yeast proteasome preparations, we repeated the cleavage of fluorogenic peptides to test the reproducibility of the results in our experimental setup. As expected, the GGL-cleaving activity, defined to represent the chymotrypsin-like activity, is clearly correlated with the  $\beta 5$ /Pre2 site: Cleavage of the Z-GGL-AMC substrate is only present in those mutants that retain a functional  $\beta 5$ /Pre2 subunit and absent in *pre2-K33A* mutant proteasomes (Figure 2.1-1A). Identical results were obtained with the Suc-LLVY-AMC substrate (data not shown). The ARR cleaving activity, defined to represent the trypsin-like activity, depends on a functional  $\beta 2$ /Pup1 site: The cleavage of Z-ARR-AMC is not observed in the  $\beta 2$ /Pup1-mutated proteasomes from the *pup1-T1A* and *pup1-T1A pre3-T1A* strains, but in all other mutants (Figure 2.1-1B). Correspondingly,  $\beta 1$ /Pre3 is responsible for the PGPH-like activity: The Z-LLE- $\beta$ NA substrate is not cleaved in the double mutant *pup1-T1A pre3-T1A* and only cleaved weakly in the *pre3-T1A* single mutant (Figure 2.1-1C).

The increased activities of some mutant proteasomes compared to wild-type particles are hard to explain but might be caused by the lack of competition from inactivated subunits for substrate molecules or by an allosteric activation of the remaining active  $\beta$ -subunits by the presence of propeptide remnants still associated with the inactivated  $\beta$ -subunits. In conclusion, the three standard fluorogenic substrates can be used as subunit-specific functional markers. We therefore decided to test the influence of several inhibitors on the proteolytic activity of yeast 20S wt and mutant proteasomes.



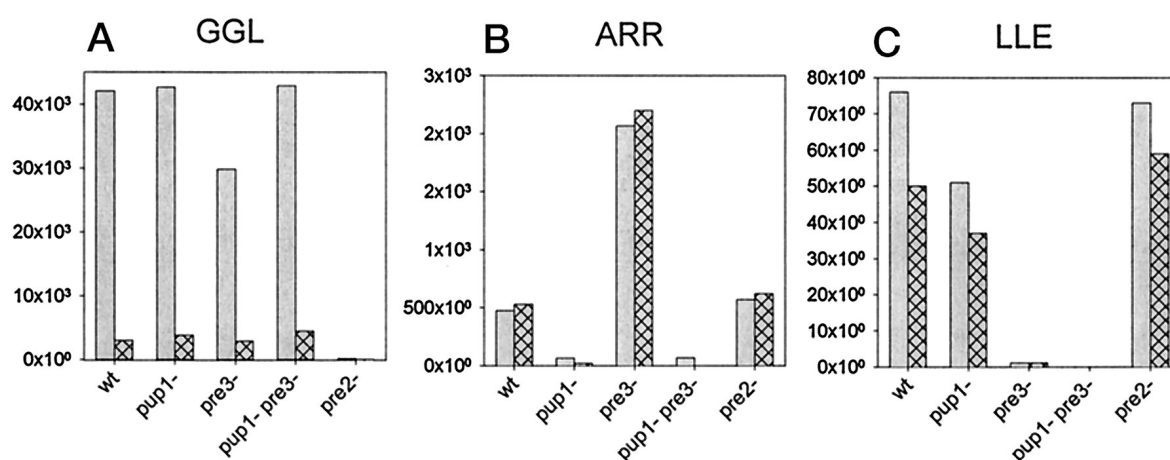
**Figure 2.1-1: Comparison of the ability of wild-type and mutant yeast proteasomes ( $\beta 2/\text{Pup1}^-$ ,  $\beta 1/\text{Pre3}^-$ ,  $\beta 2/\text{Pup1}^- \beta 1/\text{Pre3}^-$  and  $\beta 5/\text{Pre2}^-$ ) to catalyze the release of fluorogenic leaving groups from three different substrates (Z-GGL-AMC, Z-ARR-AMC and Z-LLE- $\beta$ NA) in the absence of inhibitors (A, B, C), in the presence of 1mM lactacystin (LC) (D, E, F) and in the presence of 1mM LLnL (G, H, I).**

The same amount of proteasome (1 $\mu$ g) was used in each incubation. Substrates were added after an inhibitor preincubation time of 45 min. Units of fluorescence are drawn to scale arbitrarily. wt, wild-type yeast proteasome.

#### 2.1.4.2 Proteasome subunits are differentially inhibited by lactacystin, LLnL and DCI

It has been reported previously that the bovine proteasomal subunits corresponding to yeast  $\beta 5/\text{Pre2}$  and  $\beta 2/\text{Pup1}$  (Fenteany et al., 1995) or all six human active  $\beta$ -subunits (Craiu et al., 1997) are inhibited by lactacystin due to covalent binding to the N-terminal threonine. Soaking of yeast 20S proteasome crystals resulted only in a covalent modification of the  $\beta 5/\text{Pre2}$  subunit (Groll et al., 1997). In our hands, pretreatment of proteasomes with 1mM lactacystin for 45 min led to absolute inhibition of GGL and ARR cleaving activity (Figure 2.1-1D and E), but did not have an inhibitory influence on the LLE-cleaving activity (Figure 2.1-1F). Preincubation of proteasomes with 1mM LLnL for 45 min reduced GGL cleavage by ca. 90% (Figure 2.1-1G), but to our surprise did not significantly influence cleavage of the ARR and LLE substrates (Figure 2.1-1H and I),

despite the fact that the same concentration covalently modified  $\beta 5/\text{Pre}2$ ,  $\beta 1/\text{Pre}3$  and  $\beta 2/\text{Pup}1$  in wt 20S yeast proteasome crystals (Groll et al., 1997). Given the established specificities of the different subunits for the fluorogenic substrates, this finding shows that in our setup the  $\beta 5/\text{Pre}2$  catalytic site can be inhibited by both lactacystin and LLnL,  $\beta 2/\text{Pup}1$  only by lactacystin and  $\beta 1/\text{Pre}3$  by neither lactacystin nor LLnL. The vastly different sensitivity of the proteasomal activities towards lactacystin and LLnL was also observed in inhibitor titration experiments (data not shown). Possible explanations for the apparent difference to the crystal structure data, which showed occupation of all active sites by LLnL and occupation of only  $\beta 5/\text{Pre}2$  by lactacystin, will be discussed.



**Figure 2.1-2: Influence of 1mM DCI on wt and mutant yeast proteasomes**

Incubation time: 2h (*open bars*: without inhibitor; *crosshatched bars*: with inhibitor). Substrates were added after an inhibitor preincubation time of 45 min. The influence of the inhibitor on the  $\beta 5/\text{Pre}2$  subunit is shown with GGL (A), on the  $\beta 2/\text{Pup}1$  subunit with ARR (B) and on the  $\beta 1/\text{Pre}3$  subunit with LLE (C).

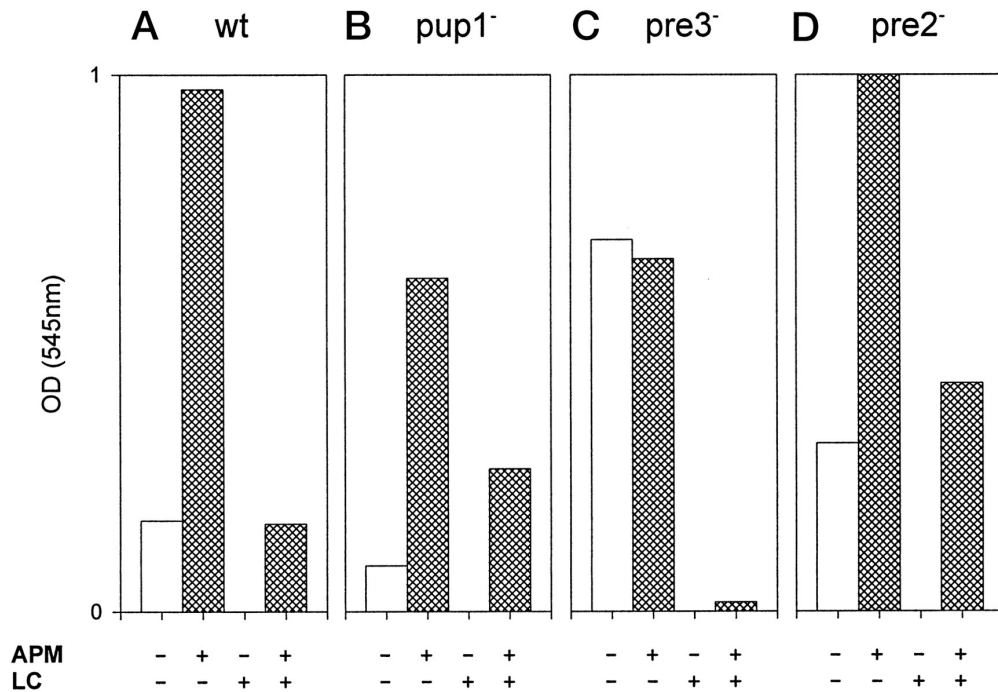
The third inhibitor tested in our study is 3,4-dichloroisocoumarin (DCI). While yeast  $\beta 5/\text{Pre}2$  is clearly sensitive to inactivation by DCI after an 45' incubation,  $\beta 1/\text{Pre}3$  is only weakly affected by DCI concentrations of even 1mM and  $\beta 2/\text{Pup}1$  is not affected at all (Figure 2.1-2). This result is in line with the observation that the three activities in the bovine pituitary proteasome are inactivated with widely differing rate constants (CT-like > PGPH-like > T-like) (Orlowski and Michaud, 1989). However, it has also been reported for the bovine pituitary proteasome that all three classical activities can be inactivated with micromolar concentrations of DCI and short incubation times (Orlowski et al., 1993). This cannot be confirmed in our experiments with yeast wt and mutant proteasomes. To inhibit  $\beta 1/\text{Pre}3$  by 50%, the yeast proteasomes had to be pre-incubated in 100 $\mu\text{M}$  DCI for 90

minutes. Lower DCI concentrations or shorter pre-incubation times did not significantly affect the  $\beta 1/\text{Pre}3$  activity (data not shown).

*2.1.4.3 "BrAAP" activity as measured with the Z-GAPLA-pAB substrate is catalyzed by the  $\beta 1/\text{Pre}3$  active site*

Micromolar concentrations of DCI have been shown to activate an additional proteolytic component of bovine 20S proteasomes, the "branched chain amino acid preferring" (BrAAP) component (Orlowski et al., 1993). We therefore tested whether this activity is also present in yeast 20S proteasomes. The existence of mutant yeast proteasomes gave us the unique possibility to test whether or not this component, if present, can be associated with one of the three active  $\beta$ -subunits or whether it represents an additional, so far unrecognized, active site.

Using the assay introduced by Orlowski et al. to measure the BrAAP activity of the proteasome (Orlowski et al., 1993), the Z-GPALA-pAB substrate was incubated with the different yeast proteasome preparations. Generation of the leaving group (pAB) in the absence of aminopeptidase M (APM) reflects direct cleavage after the C-terminal alanine, while the liberation of additional pAB in the presence of APM was described as cleavage after the penultimate leucine residue, generating Ala-pAB in the first place. The cleavage after Leu is thought to represent the so-called BrAAP activity. Significant BrAAP activity was found in all proteasome preparations but the one with an inactivated  $\beta 1/\text{Pre}3$  subunit. Here, the addition of aminopeptidase M does not lead to release of additional pAB (Figure 2.1-3C) indicating that cleavage after leucine in the GPALA substrate is catalyzed by the  $\beta 1/\text{Pre}3$  subunit. On the other hand, cleavage after alanine is clearly independent of  $\beta 1/\text{Pre}3$  (Figure 2.1-3C) and can be attributed to  $\beta 5/\text{Pre}2$  or  $\beta 2/\text{Pup}1$ . The finding that the BrAAP activity is performed by  $\beta 1/\text{Pre}3$  is also supported by the fact that this activity cannot be inhibited by lactacystin, found previously not to influence  $\beta 1/\text{Pre}3$  activity against LLE (Figure 2.1-1F, Figure 2.1-3A, B, D (+APM, +LC)).

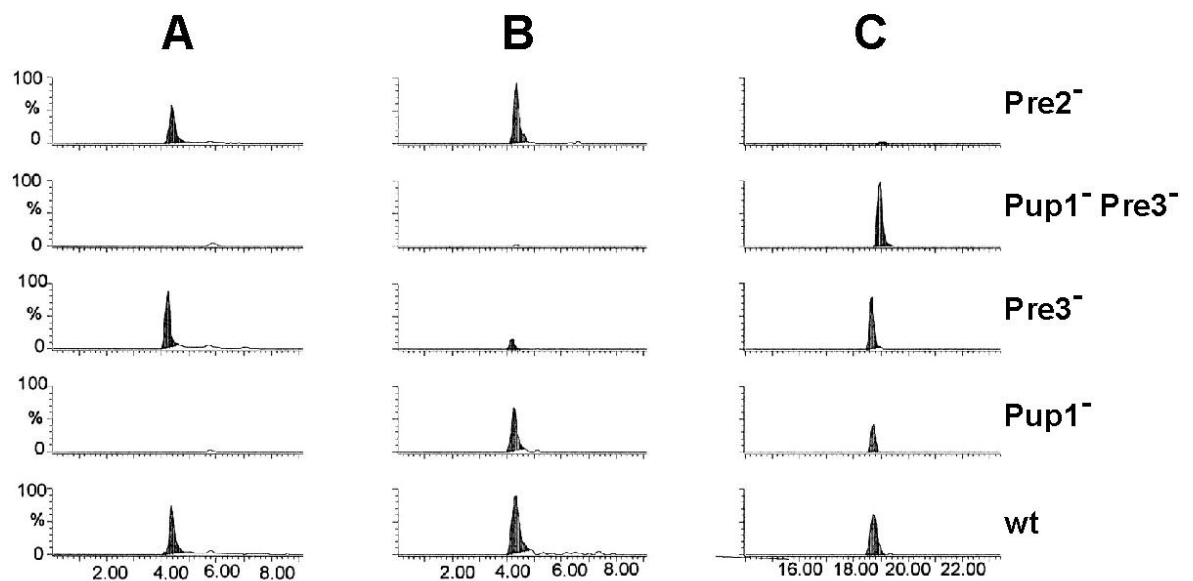


**Figure 2.1-3: BrAAP activity of wt and mutant yeast proteasomes**

Release of p-aminobenzoic acid from Z-GPALA-pAB as measured by azo coupling with wt (A) and mutant (B, C, D) proteasomes, with (+APM, hatched bars) and without (-APM) aminopeptidase M, with (+LC) and without (-LC) lactacystin.

#### 2.1.4.4 Analysis of the degradation of peptide substrates

Most cleavages in long peptide substrates occur after hydrophobic (F, Y, I, L), basic (R, K) and acidic (D, E) residues (see below), a situation similar to the cleavage of fluorogenic substrates, which have been related to chymotrypsin-like, trypsin-like and PGPH-like activities in the proteasome. To investigate whether these cleavages are catalyzed by three non-overlapping and separate activities, we digested three different peptides with yeast wt proteasome and the four different yeast mutant proteasome preparations containing the inactivated  $\beta$ -subunits  $\beta 2/Pup1^-$ ,  $\beta 1/Pre3^-$ ,  $\beta 2/Pup1^- \beta 1/Pre3^-$  and  $\beta 5/Pre2^-$ . Products were separated and analyzed by LC-MS. Three typical sets of HPLC profiles displaying integrated ion currents for the masses of particular degradation products are shown in Figure 2.1-4.



**Figure 2.1-4: Three examples of HPLC profiles displaying integrated ion currents for the masses of particular degradation products.**

(A) Degradation product 14-17 (ALVR) from the Influenza NP peptide is only generated by proteasomes with a functional  $\beta 2$ /Pup1 subunit. (B) Product 1-6 (HSNLND), also from the Influenza peptide, is mostly dependent on a functional  $\beta 1$ /Pre3 site. (C) Product 8-15 (PHFMPTNL) from the pp89 peptide depends on the presence of  $\beta 5$ /Pre2. *wt*, wild-type yeast proteasome.

We limited ourselves to the interpretation of only those cleavages showing a dramatic reduction or complete absence in digestions using mutant proteasomes and avoided to compare relative amounts of fragments generated under the influence of different proteasome mutants. The reason for this is that many peptide fragments generated are intermediate products and subject to further degradation. For quantitative analysis one would have to consider consecutive cleavages, stabilizing or destabilizing the fragment under consideration. This kind of analysis would go far beyond the aim of this study. However, it is possible to recognize a strong requirement for a certain subunit in those cases where corresponding mutants exhibit no or greatly reduced product formation (compare Figure 2.1-4).

#### 2.1.4.5 Degradation of the JAK1-21mer peptide

The JAK1-21mer peptide is part of JAK1 tyrosine kinase and contains the MHC class I ligand SYFPEITHI (Harpur et al., 1993). Incubation of this peptide with mammalian proteasomes leads to efficient generation of the MHC I ligand (Dick et al., 1996; Niedermann et al., 1997), which is flanked by two dominant cleavage sites after F<sub>7</sub> and I<sub>16</sub>,

respectively. The same two cleavages are efficiently catalyzed by the yeast proteasome, in addition to minor cleavages which are also conserved between the mammalian and yeast systems. However, the yeast enzyme additionally cleaves after H<sub>15</sub>, a cleavage not observed in digests with the mammalian proteasome (Figure 2.1–5).

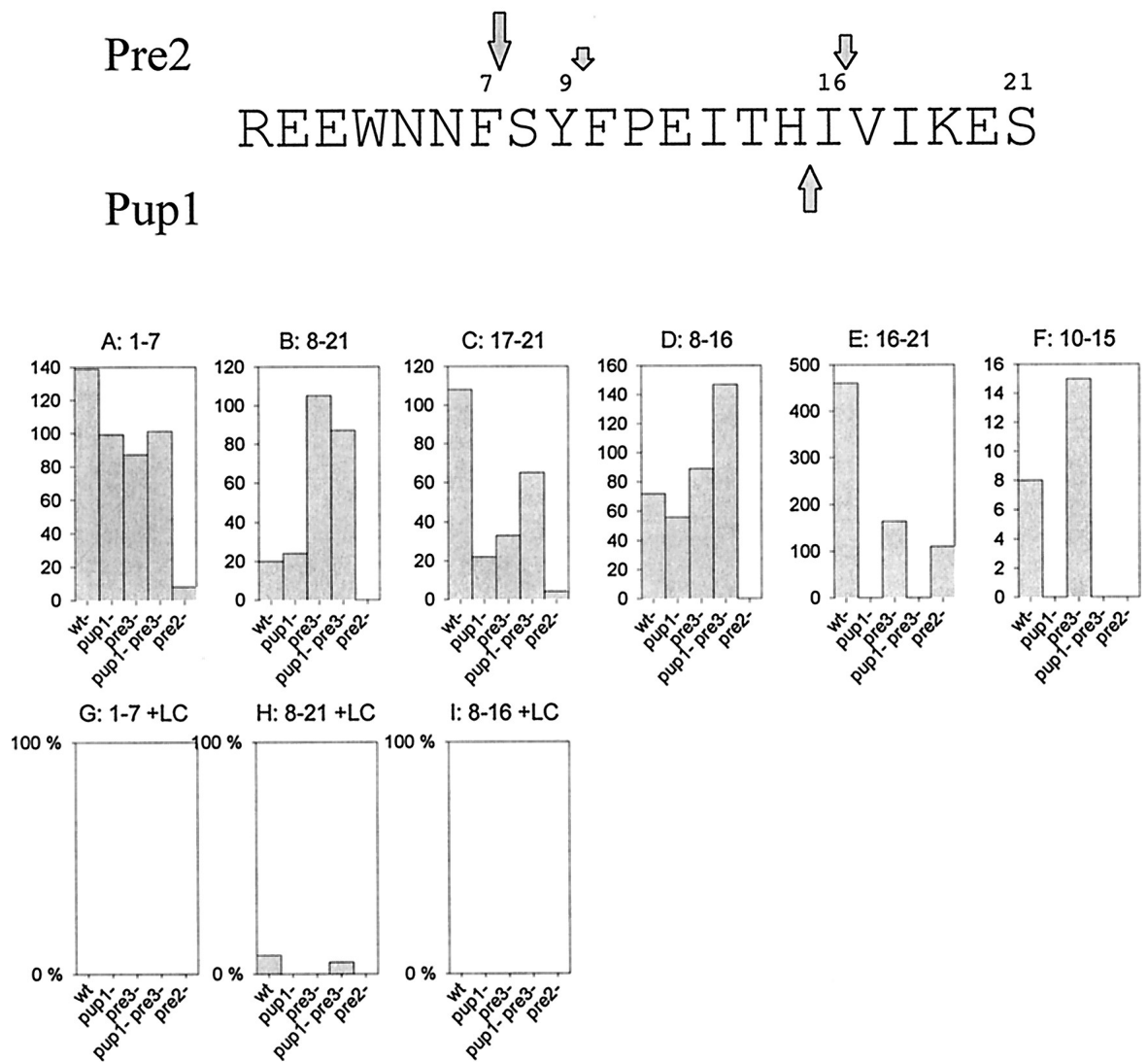
Cleavage after F<sub>7</sub> is  $\beta$ 5/Pre2-dependent, since generation of fragments 1-7 and 8-21 is nearly absent in  $\beta$ 5/Pre2-defective proteasomes (Figure 2.1–5A, B). Cleavage after I<sub>16</sub> is also catalyzed by the  $\beta$ 5/Pre2 subunit: generation of the fragment 17-21 is nearly missing in digests with  $\beta$ 5/Pre2-deficient proteasomes (Figure 2.1–5C). Correspondingly, a functional  $\beta$ 5/Pre2 subunit is necessary and sufficient for the generation of SYFPEITHI (8-16):  $\beta$ 2/Pup1 $\beta$ 1/Pre3<sup>-</sup> double mutants (which only carry functional  $\beta$ 5/Pre2 subunits) efficiently generate SYFPEITHI, while  $\beta$ 5/Pre2<sup>-</sup> mutants do not generate any detectable amount (Figure 2.1–5D). The observation that the formation of fragments 1-7, 8-16 and 8-21 can be completely inhibited with lactacystin (Figure 2.1–5G-I) is in agreement with an essential role of  $\beta$ 5/Pre2 and indicates that  $\beta$ 1/Pre3 does not perform this cleavage.

Cleavage after H<sub>15</sub> is only detected in mutants with a functional  $\beta$ 2/Pup1 subunit, as demonstrated for fragment 16-21 (Figure 2.1–5E). Cleavage after Y<sub>9</sub>, leading to FPEITHI (10-16) in the mammalian system (Dick et al., 1996), also takes place in the yeast system, albeit much weaker. The minor product FPEITH (10-15) is an example for a fragment that depends on two different functional subunits:  $\beta$ 5/Pre2 for the N-terminal cleavage and  $\beta$ 2/Pup1 for the C-terminal cleavage. Correspondingly, fragment 10-15 is only formed in digests with proteasomes that contain both  $\beta$ 5/Pre2 and  $\beta$ 2/Pup1 (Figure 2.1–5F).

#### 2.1.4.6 Degradation of the pp89-25mer peptide

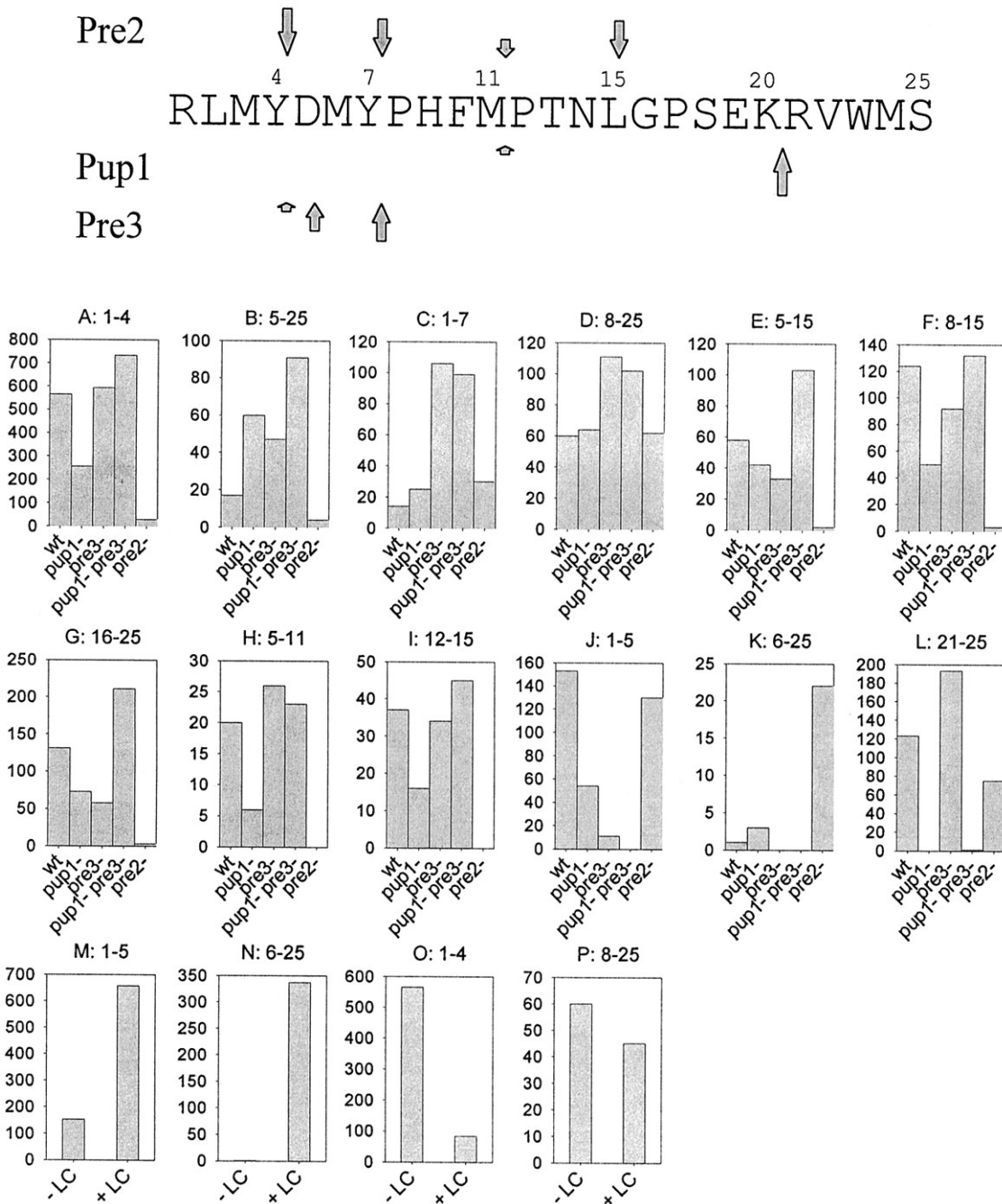
The pp89-25mer peptide is derived from the MCMV pp89 IE-protein and contains the L<sup>d</sup>-presented CTL epitope YPHFMPTNL (Del Val et al., 1991). In digests with the mammalian proteasome two main products are generated: 8-15 and 5-15, the latter one being a candidate precursor peptide for the CTL epitope (Dick et al., 1996). Again, the same dominant cleavage sites and many of the sub-dominant ones are also used by the yeast enzyme (Figure 2.1–7).





**Figure 2.1-5: Summary of cleavage sites in the JAK1-21mer peptide as digested with the yeast 20S proteasome (top) and data explaining the correlation between cleavages and proteasome subunits (bottom).**

A-F: Integrated ion currents from mass spectrometry for the main peptide products generated from the JAK1 peptide after 8h incubation with five different proteasome preparations (wild-type (wt),  $\beta$ 2/Pup1<sup>-</sup>,  $\beta$ 1/Pre3<sup>-</sup>,  $\beta$ 2/Pup1<sup>-</sup> $\beta$ 1/Pre3<sup>-</sup> and  $\beta$ 5/Pre2<sup>-</sup>). G-I: Relative inhibition of product formation by lactacystin (LC) (% of product formed without LC) based on integrated ion currents from LC-MS.



**Figure 2.1-6: Summary of cleavage sites in the pp89-25mer peptide as digested with the yeast 20S proteasome (*top*) and data explaining the correlation between cleavages and proteasome subunits (*bottom*).**

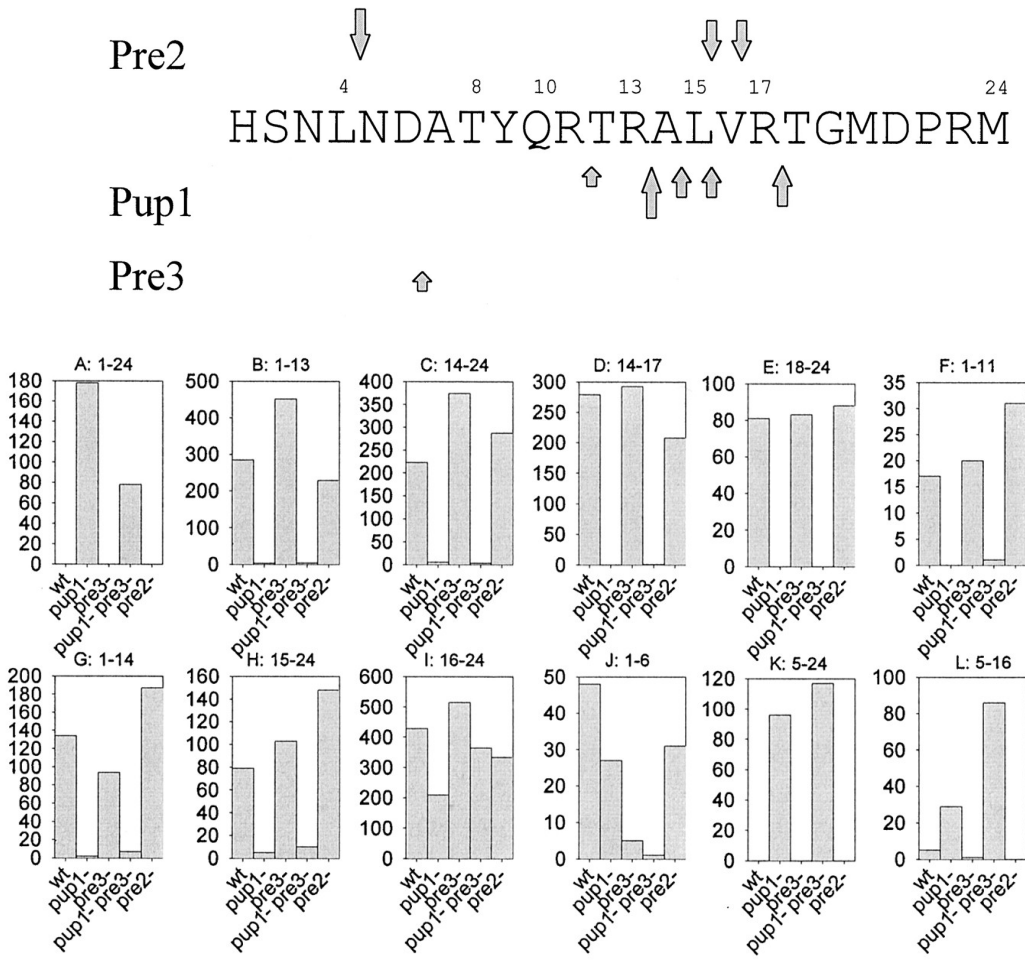
*A-L*: Integrated ion currents for the main peptide products generated from the pp89 peptide after 4h incubation with five different proteasome preparations (wild-type (*wt*),  $\beta 2/Pup1^{-}$ ,  $\beta 1/Pre3^{-}$ ,  $\beta 2/Pup1^{-} \beta 1/Pre3^{-}$  and  $\beta 5/Pre2^{-}$ ). *M-P*: Four examples for the influence of lactacystin (LC) on product formation by the yeast *wt* proteasome.

Cleavage after  $Y_4$  is essentially  $\beta 5/Pre2$ -dependent, since formation of products 1-4 and 5-25 is only present in minimal amounts in  $\beta 5/Pre2$ -defective proteasomes (Figure 2.1-6A and B), and the generation of fragment 1-4 can be inhibited by lactacystin (Figure 2.1-6O). However, the situation is more complicated for the cleavage after  $Y_7$ , which takes place in all mutants tested (Figure 2.1-6C and D). This observation can be explained only if two different subunits effectively catalyze the same cleavage. The  $\beta 5/Pre2$  subunit is clearly capable of performing the cleavage, since it takes place efficiently in the  $\beta 2/Pup1^- \beta 1/Pre3^-$  double mutant. This leads to the question which subunit catalyzes the same cleavage in the  $\beta 5/Pre2$ -defective mutant. Since formation of fragment 8-25 is only weakly inhibited by lactacystin (Figure 2.1-6P), which affects only  $\beta 5/Pre2$  and  $\beta 2/Pup1$ , we conclude that  $\beta 1/Pre3$  contributes to this cleavage. It should be noted that cleavage after  $Y_4$  is also, but to a much lesser extent, catalyzable by  $\beta 1/Pre3$ , explaining the residual activity observed with  $\beta 5/Pre2$ - mutants.

Cleavage after  $L_{15}$  is strictly  $\beta 5/Pre2$ -dependent, since all products that terminate in  $L_{15}$  are only observed in the presence of proteasomes with active  $\beta 5/Pre2$  subunits. Prominent examples are the main products 5-15 and 8-15 (Figure 2.1-6E and F), as well as fragment 16-25 (Figure 2.1-6G). A rather weak cleavage after  $M_{11}$  is also observed and is mainly  $\beta 5/Pre2$ -dependent (Figure 2.1-6H and I). Cleavage after  $D_5$  is almost absolutely  $\beta 1/Pre3$ -dependent (Figure 2.1-6J and K) and therefore not sensitive to lactacystin (Figure 2.1-6M and N). Cleavage after  $K_{20}$  is absolutely  $\beta 2/Pup1$ -dependent: fragment 21-25 only appears when proteasomes with a functional  $\beta 2/Pup1$  subunit are used (Figure 2.1-6L).

#### 2.1.4.7 Degradation of the Influenza NP 24mer peptide

The 24mer peptide from Influenza nucleoprotein contains the  $K^d$ -restricted ligand and CTL epitope TYQRTRALV (147-155) (Röttschke et al., 1990). Digests with the mammalian proteasome (unpublished data) confirmed a strong conservation of preferred cleavage sites with the yeast 20S proteasome. Most dominant cleavage sites are clustered around the C-terminus of the epitope, some of them destroying the MHC ligand (Figure 2.1-8).



**Figure 2.1-7: Summary of cleavage sites in the Influenza nucleoprotein 24mer peptide as digested with the yeast 20S proteasome (top) and data explaining the correlation between cleavages and proteasome subunits (bottom).**

A-L: Integrated ion currents for the main peptide products generated from the influenza nucleoprotein peptide after 2h incubation with five different proteasome preparations (wt, β2/Pup1<sup>-</sup>, β1/Pre3<sup>-</sup>, β2/Pup1β1/Pre3<sup>-</sup> and β5/Pre2<sup>-</sup>).

The overall degradation of the Influenza peptide is dominated by cleavages after arginine residues, catalyzed by the β2/Pup1 subunit. Degradation of the substrate peptide (1-24) is significantly retarded in digests with β2/Pup1-defective proteasomes (Figure 2.1–8A). The most dominant cleavages after R<sub>13</sub> and R<sub>17</sub> are absolutely β2/Pup1-dependent (Figure 2.1–8B-E) and the same is true for the less pronounced cleavage after R<sub>11</sub> (Figure 2.1–8F). Interestingly, cleavage after A<sub>14</sub> is also catalyzed by β2/Pup1, as shown for the fragments 1-14 and 15-24, which are missing in digests using β2/Pup1-defective proteasomes (Figure 2.1–8G and H).

Cleavage after L<sub>15</sub> is a more complicated matter. It is performed by all the mutants; the data shown in Figure 2.1–8I favor the involvement of both  $\beta$ 2/Pup1 and  $\beta$ 5/Pre2 over  $\beta$ 1/Pre3. The fragment 1-6 is generated by a  $\beta$ 1/Pre3 dependent activity cleaving after D (Figure 2.1–8J). The cleavage after L<sub>4</sub> is the most dominant site in the vicinity of the N-terminus of the CTL epitope. Because of the dominant cleavages mediated by the  $\beta$ 2/Pup1 subunit (especially R<sub>13</sub>), fragment 5-24 is usually not observed in digests using proteasomes with active  $\beta$ 1/Pup1 subunits. Only in digests with  $\beta$ 2/Pup1-defective proteasomes this fragment can be detected (Figure 2.1–8K). A similar situation applies to peptide NDATYQRTRALV (5-16) (Figure 2.1–8L), which is a potential precursor of the CTL epitope TYQRTRALV (8-16). The formation of this peptide is strongly enhanced in digests with 20S mutants lacking functional  $\beta$ 2/Pup1. The finding that this fragment is generated when mutants with inactive  $\beta$ 2/Pup1, but active  $\beta$ 5/Pre2 subunit are used argues strongly for the involvement of  $\beta$ 5/Pre2 in this cleavage event. This is also supported by the result that cleavage after L<sub>4</sub> is inhibited by lactacystin (data not shown).

## 2.1.5 Discussion

### 2.1.5.1 Cleavage specificity of the proteasomal subunits

Using short fluorogenic peptide substrates in combination with mutant proteasomes lacking specific subunit activities we found a clear correlation between active subunits and cleavage specificity. This finding supports the conventional belief that the three "classical" fluorogenic substrates are cleaved by three different active sites in the proteasome [ $\beta$ 5/Pre2→ hydrophobic;  $\beta$ 2/Pup1→ basic and  $\beta$ 1/Pre3→ acidic] and can be used as subunit-specific probes. However, the study of longer peptide substrates (21 to 25 aa) in the same system led to more complex results: While the formula [ $\beta$ 5/Pre2→ hydrophobic;  $\beta$ 2/Pup1→ basic and  $\beta$ 1/Pre3→ acidic] can be applied to many of the cleavage sites, there is a certain class of hydrophobic and small residues that does not obey this simple rule. Evidence that the specificities of the different subunits might overlap was provided by the group of Rivett before proteasomes carrying different inactivated subunits have been available. These experiments were performed with the help of inhibitors for the tryptic and chymotryptic activities (Savory et al., 1993; Reidlinger et al., 1997). A special case is the so-called BrAAP activity, cleaving after the hydrophobic residue leucine in the model substrate GPALA, which can now be clearly attributed to  $\beta$ 1/Pre3 (Figure 2.1-3) as suggested previously (Cardozo et al., 1996; Vinitzky et al., 1994). The observed specificities are summarized for the different subunits in the following paragraphs:

$\beta 2/\text{Pup1}$  proved to be responsible for all cleavages after basic residues (R, K) without any exception. This result is in line with the observed specificity towards fluorogenic substrates. However,  $\beta 2/\text{Pup1}$  is not absolutely limited to cleave after basic residues. A cleavage after alanine was clearly identified as  $\beta 2/\text{Pup1}$ -catalyzed (Figure 2.1–8G and H) and there are indications for  $\beta 2/\text{Pup1}$ -catalyzed cleavages after hydrophobic residues (Figure 2.1–8I). Thus,  $\beta 2/\text{Pup1}$  is also (at least partially) responsible for cleavages after small and hydrophobic amino acids.

$\beta 1/\text{Pre3}$  catalyzed all observed cleavages after acidic residues (D, E). In addition,  $\beta 1/\text{Pre3}$  is responsible for the cleavage after leucine in the GPALA substrate, as mentioned above (Figure 2.1-3C), and is involved in the cleavage after  $Y_7$  in the pp89 peptide substrate (Figure 2.1-6C and D). This indicates a broader role of  $\beta 1/\text{Pre3}$  in cleavages after hydrophobic residues.

$\beta 5/\text{Pre2}$  is mainly responsible for cleaving after hydrophobic residues, as expected. However, some of these cleavages are at the same time also catalyzed by  $\beta 1/\text{Pre3}$  or  $\beta 2/\text{Pup1}$ .

In conclusion, while cleavages after basic and acidic residues strictly correlate with subunits  $\beta 2/\text{Pup1}$  and  $\beta 1/\text{Pre3}$ , respectively, cleavages after hydrophobic residues can be the result of  $\beta 5/\text{Pre2}$  alone or with contributions by the  $\beta 1/\text{Pre3}$  or, apparently to a lesser extent, the  $\beta 2/\text{Pup1}$  subunit. Likewise, cleavage after small residues appears not to be restricted to a single subunit.

This overlap might be explained by the contribution of the sequence context around potential cleavage sites as a major factor in determining which subunit will preferentially cleave after a certain hydrophobic or small residue (and also whether there is cleavage at all after a suitable P1 residue). The number of cleavage sites analyzed in this study is too small to make detailed predictions how flanking sequences guide the involvement of the three active subunits in cleavages after certain residues. Recent results, however, namely the extensive characterization of more than 400 degradation products from the 436 aa protein yeast enolase-1 after digestion with wt and mutant yeast proteasomes, support the assumption that the presence of certain amino acids in several positions around the P1 site strongly influences proteasomal cleavage activities (Nussbaum et al., manuscript in preparation; see 2.2).

### 2.1.5.2 Influence of inhibitors on distinct active sites

Inhibition studies with wt and mutant proteasomes provided evidence that under the conditions used in our experiments lactacystin is an inhibitor of both the  $\beta 5/\text{Pre}2$  and  $\beta 2/\text{Pup}1$  catalytic subunits, but not of the  $\beta 1/\text{Pre}3$  subunit, which is only inhibited after longer preincubation times ( $> 4\text{h}$ ) and at lactacystin concentrations  $>1\text{mM}$  (data not shown). This preferential inhibition is in agreement with (Fenteany et al., 1995) who described an irreversible inhibition of only chymotryptic and tryptic activities and (Craiu et al., 1997) who showed that the cleavage after a glutamic acid residue (flanking the N-terminus of the CTL epitope SIINFEKL) cannot be inhibited by lactacystin at a concentration of  $2\mu\text{M}$ . In crystal structural data of the yeast proteasome lactacystin was found covalently associated with only the  $\beta 5/\text{Pre}2$  subunit (Groll et al., 1997). However, the experimental conditions of crystal soaking are very different from those above and this observation cannot exclude a weaker association with  $\beta 2/\text{Pup}1$ , resulting in inhibition of the proteolytic activity of this subunit. The preferential interaction of lactacystin with  $\beta 5/\text{Pre}2$  and  $\beta 2/\text{Pup}1$  under the conditions used in our experiments proved to be very helpful because it provided an additional tool to discriminate catalytic contributions made by the subunits  $\beta 5/\text{Pre}2$  and  $\beta 2/\text{Pup}1$  versus  $\beta 1/\text{Pre}3$ . For example, the use of lactacystin confirmed the association of the BrAAP activity with  $\beta 1/\text{Pre}3$  (Figure 2.1-3) and the contribution of  $\beta 1/\text{Pre}3$  to cleavage after the  $Y_7$  residue in the pp89 peptide substrate (Figure 2.1-6P).

The finding that LLnL inhibits only  $\beta 5/\text{Pre}2$  in our hands, at concentrations where it covalently modifies all three active  $\beta$ -subunits in the yeast crystal structure (Groll et al., 1997), may be explained by the soaking condition of proteasome crystals. The incubation with the inhibitor solution was performed for a much longer period of time (6h vs. 45'). Indeed, increasing the preincubation time of the inhibitor with proteasome to 90 minutes resulted in partial inhibition of the  $\beta 2/\text{Pup}1$  and  $\beta 1/\text{Pre}3$  subunits at LLnL concentrations  $>1\text{mM}$  (data not shown). In addition, hemiacetal formation between peptide aldehyde groups and the active site threonine 1  $O_\gamma$  may be favored under the conditions of crystallization. A similar observation was made for the inhibition by DCI. Preincubations for 45 min with DCI concentrations up to 5mM resulted only in the inhibition of  $\beta 5/\text{Pre}2$ , whereas a 90 min preincubation inhibited also  $\beta 2/\text{Pup}1$  and  $\beta 1/\text{Pre}3$  to 50% at  $100\mu\text{M}$  and 90% at  $1\text{mM}$  (data not shown).

### 2.1.5.3 No additional catalytic sites in the yeast proteasome

Orlowski et al. (1993) biochemically defined catalytic components distinct from the three classical activities (CT-like, T-like and PGPH-like), one of them was called the "branched chain amino acid preferring" (BrAAP) activity. This component was also assumed to represent the "caseinolytic" activity of the proteasome (Pereira et al., 1992). Two other additional, "non-classical" activities were identified and called the "small neutral amino acid preferring" (SNAAP) (Orlowski et al., 1993) and acidic chymotrypsin-like (aCT-like) activities (Figueiredo Pereira et al., 1995).

Testing the wt and mutant yeast 20S proteasomes for cleavage activity against the BrAAP substrate Benzyloxycarbonyl-Gly-Pro-Ala-Leu-Ala-para-Aminobenzoic acid (Z-GPALA-pAB) (Orlowski et al., 1993), we observed that the cleavage after the penultimate leucine residue, defined to represent the BrAAP activity, is performed by  $\beta 1/\text{Pre}3$ . This is evident from the missing BrAAP activity in the  $\beta 1/\text{Pre}3^-$  mutant (Figure 2.1-3C) and in addition from the remaining BrAAP activity of the  $\beta 2/\text{Pup}1^-$  and  $\beta 5/\text{Pre}2^-$  mutants in the presence of lactacystin (Figure 2.1-3B and C), which inhibits  $\beta 5/\text{Pre}2$  and  $\beta 2/\text{Pup}1$ , but not  $\beta 1/\text{Pre}3$  under the applied conditions (Figure 2.1-1F). The cleavage after the ultimate alanine residue is catalyzed by  $\beta 5/\text{Pre}2$  and upon lack of  $\beta 5/\text{Pre}2$  or upon its preferential inhibition by LLnL also by  $\beta 2/\text{Pup}1$  (data not shown). The cleavage after alanine is also lactacystin sensitive (Figure 2.1-3A, B, D), supporting the view that either the  $\beta 2/\text{Pup}1$  or  $\beta 5/\text{Pre}2$  subunit is involved and not  $\beta 1/\text{Pre}3$ , which is not inhibited by lactacystin under the applied conditions (Figure 2.1-1F). The finding that the BrAAP component is not only insensitive to, but even activated by DCI (Orlowski et al., 1993) might be explained by inhibition of the  $\beta 5/\text{Pre}2$  subunit, otherwise competing with  $\beta 1/\text{Pre}3$  for the same substrate molecules. In summary, we conclude that in our system no additional active site apart from the three threonine protease catalytic centers is responsible for cleaving the GPALA substrate and that, as defined by the GPALA substrate,  $\beta 1/\text{Pre}3$  represents the BrAAP component. No evidence for additional activities could be found with any of the other substrates tested in this study. Since the BrAAP or caseinolytic activity has been described as the most significant among the additional activities, we regard it as unlikely that other putative activities (SNAAP and aCT) will turn out to be catalyzed by some unconventional, additional active site.

Since our evidence is based on wt and mutant proteasomes from yeast, we can not definitely rule out that the bovine proteasomes used by Orlowski et al. (1993) do harbour



activities that are distinct from the known threonine-type proteolytic subunits. However, we regard this possibility as unlikely, for the following reasons: (i) proteasomes from yeast and mammals are conserved in subunit structure and topology (Groll et al., 1997; Kopp et al., 1997), as well as in cleavage specificity when tested with long peptide substrates (Niedermann et al., 1997), (ii) yeast proteasomes were highly purified and homogenous in subunit composition, while bovine proteasomes might represent a mixture of different species including inducible subunits. This subunit heterogeneity might lead to experimental complications; (iii) fluorogenic substrates used in our experiments were shown to be cleaved by the yeast proteasome in a subunit-specific manner. It is still to be shown that these and other fluorogenic substrates are cleaved in a strictly subunit-specific manner by mammalian (bovine) proteasomes.

#### *2.1.5.4 Generation of dual cleavage products*

In this study we show that the cleavages generating the SYFPEITHI fragment from the JAK1 peptide and the DMYPHFMPNTL fragment from the pp89 peptide are all performed by the  $\beta$ 5/Pre2 subunit. These results bear on previous experiments performed with mammalian 20S proteasomes. There, we found that the generation of the same peptides is boosted by the presence of PA28 (Dick et al., 1996). Meanwhile, this effect was confirmed (Niedermann et al., 1997) and observed with additional peptide substrates (Dick et al., unpublished data). In the absence of PA28, the N- and C-terminal cleavages that generate an internal fragment occurred independently from each other and could be outcompeted with excess substrate. Kinetics and competition experiments in the presence of PA28 however indicated that the second cleavage did not take place independently from the first cleavage (Dick et al., 1996). These observations led us to propose the model where two concerted cleavages take place between two neighbouring subunits.

However, the crystal structure of the yeast proteasome made clear that the two  $\beta$ 5/Pre2 subunits are not direct neighbours in the 20S particle and are also isolated from the  $\beta$ 1/Pre3 and  $\beta$ 2/Pup1 subunits. An identical subunit topology is expected for the mammalian proteasome (Kopp et al., 1997). Consequently, the positive influence of PA28 on the formation of internal fragments is unlikely to be based on a simultaneous cleavage event performed by the  $\beta$ 5/Pre2-corresponding subunits MB1/LMP7. Furthermore, the distance from MB1/LMP7 to either Z/MECL1 or  $\delta$ /LMP2 is too long to generate peptides of 9-11 amino acids by a simultaneous cleavage event. Whether or not  $\beta$ 2/Pup1 (Z, MECL-1) and  $\beta$ 1/Pre3 ( $\delta$ , LMP2), which are neighbours in the same  $\beta$ -ring, might be involved in the

generation of internal fragments from peptide substrates under the influence of PA28 cannot be tested in the yeast system. No PA28-homologous protein is present in the yeast genome and we could not find significant activation or modulation of the yeast proteasome by human PA28 (data not shown). Studies with mammalian proteasomes and inhibitors at conditions that selectively affect the different proteasomal activities are currently performed to address the involvement of the neighbouring subunits Z/MECL-1 and  $\delta$ /LMP2 in the generation of internal fragments when PA28 is present. Of special interest here is whether PA28 is able to influence the specificity of subunits Z/MECL-1 and  $\delta$ /LMP2 to allow a participation in cleavages after hydrophobic amino acids that might now be able to generate the SYFPEITHI and the DMYPHFMPPTNL fragments by a concerted cleavage event. Our finding that  $\beta$ 2/Pup1 and  $\beta$ 1/Pre3 are able to cleave after hydrophobic amino acids is in favor of this possibility.

Because we used peptide substrates of only up to 25aa in this study, we did not investigate the length of the fragments generated from wild-type or mutant yeast proteasomes in detail. It is however interesting to note that the size of internal fragments generated from synthetic peptide substrates centers around seven amino acids and that the average fragment length does not seem to differ between wild-type and mutant proteasomes with fewer threonine active sites. This is of interest because some of us (Löwe et al., 1995) had suggested that the distance between the subunits carrying the threonine active sites corresponding to an octapeptide in extended conformation, determines the length of fragments generated. The data here do not support this hypothesis. Whether or not this model is correct will be answered by the detailed analysis of proteasomal digestion products of the yeast enolase-1 protein, where more than 400 protein fragments, generated in digestions with yeast wild-type and mutant proteasomes, were characterized (Nussbaum et al., manuscript in preparation; see 2.2 for details).

In addition, in the yeast proteasome, processing intermediates cleaved at residues -8 and -9 of subunits  $\beta$ 7/Pre4 and  $\beta$ 6/C5 were observed in well defined conformations (Groll et al., 1997). This led to the suggestion of the existence of an additional unspecific hydrolytic site at the inner  $\beta$ -annulus of the proteasome. The results presented previously (Heinemeyer et al., 1997) and here argue against this hypothesis. A reconsideration of the hydrolytic activity at the inner  $\beta$ -annulus must await the results of the analysis of processing intermediates in proteasomes with inactivated  $\beta$ -subunits which is on the way (Groll et al., manuscript in preparation).

The analysis of the proteolytic activity of the yeast proteasomes also revealed that the

dominant cleavage activity was performed by the  $\beta 5$ /Pre2 subunit, generating hydrophobic C-termini. This effect is even more enhanced by the contribution of the  $\beta 2$ /Pup1 and  $\beta 1$ /Pre3 subunits to cleavages after hydrophobic residues as well. We are aware of the fact the yeast 20S proteasomes lack the interferon-inducible  $\beta$ -subunits of mammalian proteasomes. But there is no evidence that the presence of these subunits does alter the specificity of the proteasome in a way that will induce the generation of new cleavage sites. The only differences observed so far are changes in the quantity of peptide fragments generated when proteasomes from interferon-induced or LMP2/LMP7 transfected cells were analysed (Niedermann et al., 1997; Groettrup et al., 1995). Therefore, our results will describe most of the cleavage performed in mammalian proteasomes as well and provide an explanation for the dominance of hydrophobic residues at the C-termini of MHC class I ligands. Because the proteasome does not cleave after every hydrophobic, basic or acidic amino acid however, there must be rules that govern the selection of cleavage sites. These rules are not evident in the limited set of data available here but will be looked for in the more than 400 fragments generated from the yeast enolase-1 protein (Nussbaum et al., in preparation; see 2.2 for details). If it is possible to identify rules that guide the interaction of substrate molecules with the proteasome and therefore allow a partial prediction of protein processing, the forecast of CTL epitopes, which so far is based on MHC ligand motifs only, will become much more accurate. The knowledge of subunit contribution to the generation of CTL epitopes will also be extremely useful for the development of subunit-specific proteasome inhibitors that suppress the generation of some CTL epitopes, enhance the production of others but do not affect the entire proteasomal activity.

### 2.1.6 References

Arendt, C.S. and Hochstrasser, M. (1997). Identification of the yeast 20S proteasome catalytic centers and subunit interactions required for active-site formation. *Proc. Natl. Acad. Sci. USA* 94, 7156-7161.

Brannigan, J.A., Dodson, G., Duggleby, H.J., Moody, P.C., Smith, J.L., Tomchick, D.R., and Murzin, A.G. (1995). A protein catalytic framework with an N-terminal nucleophile is capable of self-activation. *Nature* 378, 416-419.

Cardozo, C., Chen, W.E., and Wilk, S. (1996). Cleavage of Pro-X and Glu-X bonds catalyzed by the branched chain amino acid preferring activity of the bovine pituitary multicatalytic proteinase complex (20S proteasome). *Arch. Biochem. Biophys.* 334, 113-120.

Cardozo,C., Vinitzky,A., Hidalgo,M.C., Michaud,C., and Orłowski,M. (1992). A 3,4-dichloroisocoumarin-resistant component of the multicatalytic proteinase complex. *Biochemistry* 31, 7373-7380.

Coux,O., Tanaka,K., and Goldberg,A.L. (1996). Structure and functions of the 20S and 26S proteasomes. *Annu. Rev. Biochem.* 65, 801-847.

Craiu,A., Gaczynska,M., Akopian,T.N., Gramm,C.F., Fenteany,G., Goldberg,A.L., and Rock,K.L. (1997). Lactacystin and clasto-lactacystin beta-lactone modify multiple proteasome beta-subunits and inhibit intracellular protein degradation and major histocompatibility complex class I antigen presentation. *J. Biol. Chem.* 272, 13437-13445.

Del Val,M., Schlicht,H.J., Ruppert,T., Reddehase,M.J., and Koszinowski,U.H. (1991). Efficient processing of an antigenic sequence for presentation by MHC class I molecules depends on its neighboring residues in the protein. *Cell* 66, 1145-1153.

Dick,T.P., Ruppert,T., Groettrup,M., Kloetzel,P.M., Kuehn,L., Koszinowski,U.H., Stevanovic,S., Schild,H., and Rammensee,H.G. (1996). Coordinated dual cleavages induced by the proteasome regulator PA28 lead to dominant MHC ligands. *Cell* 86, 253-262.

Fenteany,G., Standaert,R.F., Lane,W.S., Choi,S., Corey,E.J., and Schreiber,S.L. (1995). Inhibition of proteasome activities and subunit-specific amino-terminal threonine modification by lactacystin. *Science* 268, 726-731.

Figueiredo Pereira,M.E., Chen,W.E., Yuan,H.M., and Wilk,S. (1995). A novel chymotrypsin-like component of the multicatalytic proteinase complex optimally active at acidic pH. *Arch. Biochem. Biophys.* 317, 69-78.

Groettrup,M., Ruppert,T., Kuehn,L., Seeger,M., Standera,S., Koszinowski,U.H., and Kloetzel,P.M. (1995). The interferon-gamma-inducible 11 S regulator (PA28) and the LMP2/ LMP7 subunits govern the peptide production by the 20 S proteasome in vitro. *J. Biol. Chem.* 270, 23808-23815.

Groettrup,M., Soza,A., Eggers,M., Kuehn,L., Dick,T.P., Schild,H., Rammensee,H.G., Koszinowski,U.H., and Kloetzel,P.M. (1996). A role for the proteasome regulator PA28alpha in antigen presentation. *Nature* 381, 166-168.

Groll,M., Ditzel,L., Löwe,J., Stock,D., Bochtler,M., Bartunik,H.D., and Huber,R. (1997). Structure of 20S proteasome from yeast at 2.4 Å resolution. *Nature* 386, 463-471.

Harpur,A.G., Zimiecki,A., Wilks,A.F., Falk,K., Röttschke,O., and Rammensee,H.G. (1993). A

- prominent natural H-2 Kd ligand is derived from protein tyrosine kinase JAK1. *Immunol. Lett.* 35, 235-237.
- Heinemeyer,W., Fischer,M., Krimmer,T., Stachon,U., and Wolf,D.H. (1997). The active sites of the eukaryotic 20 S proteasome and their involvement in subunit precursor processing. *J. Biol. Chem.* 272, 25200-25209.
- Hiller,M.M., Finger,A., Schweiger,M., and Wolf,D.H. (1996). ER degradation of a misfolded luminal protein by the cytosolic ubiquitin-proteasome pathway. *Science* 273, 1725-1728.
- Hilt,W. and Wolf,D.H. (1996). Proteasomes: destruction as a programme. *Trends. Biochem. Sci.* 21, 96-102.
- Hisamatsu,H., Shimbara,N., Saito,Y., Kristensen,P., Hendil,K.B., Fujiwara,T., Takahashi,E., Tanahashi,N., Tamura,T., Ichihara,A., and Tanaka,K. (1996). Newly identified pair of proteasomal subunits regulated reciprocally by interferon gamma. *J. Exp. Med.* 183, 1807-1816.
- Hochstrasser,M. (1995). Ubiquitin, proteasomes, and the regulation of intracellular protein degradation. *Curr. Opin. Cell Biol.* 7, 215-223.
- Kopp,F., Hendil,K.B., Dahlmann,B., Kristensen,P., Sobek,A., and Uerkvitz,W. (1997). Subunit arrangement in the human 20S proteasome. *Proc. Natl. Acad. Sci. USA* 94, 2939-2944.
- Lehner,P.J. and Cresswell,P. (1996). Processing and delivery of peptides presented by MHC class I molecules. *Curr. Opin. Immunol.* 8, 59-67.
- Löwe,J., Stock,D., Jap,B., Zwickl,P., Baumeister,W., and Huber,R. (1995). Crystal structure of the 20S proteasome from the archaeon *T. acidophilum* at 3.4 Å resolution. *Science* 268, 533-539.
- Niedermann,G., Grimm,R., Geier,E., Maurer,M., Realini,C., Gartmann,C., Soll,J., Omura,S., Rechsteiner,M.C., Baumeister,W., and Eichmann,K. (1997). Potential immunocompetence of proteolytic fragments produced by proteasomes before evolution of the vertebrate immune system. *J. Exp. Med.* 186, 209-220.
- Orlowski,M., Cardozo,C., and Michaud,C. (1993). Evidence for the presence of five distinct proteolytic components in the pituitary multicatalytic proteinase complex. Properties of two components cleaving bonds on the carboxyl side of branched chain and small neutral amino acids. *Biochemistry* 32, 1563-1572.
- Orlowski,M. and Michaud,C. (1989). Pituitary multicatalytic proteinase complex. Specificity of

components and aspects of proteolytic activity. *Biochemistry* 28, 9270-9278.

Pereira,M.E., Nguyen,T., Wagner,B.J., Margolis,J.W., Yu,B., and Wilk,S. (1992). 3,4-dichloroisocoumarin-induced activation of the degradation of beta-casein by the bovine pituitary multicatalytic proteinase complex. *J. Biol. Chem.* 267, 7949-7955.

Reidlinger,J., Pike,A.M., Savory,P.J., Murray,R.Z., and Rivett,A.J. (1997). Catalytic properties of 26 S and 20 S proteasomes and radiolabeling of MB1, LMP7, and C7 subunits associated with trypsin-like and chymotrypsin-like activities. *J. Biol. Chem.* 272, 24899-24905.

Röttschke,O., Falk,K., Deres,K., Schild,H., Norda,M., Metzger,J., Jung,G., and Rammensee,H.G. (1990). Isolation and analysis of naturally processed viral peptides as recognized by cytotoxic T cells. *Nature* 348, 252-254.

Savory,P.J., Djaballah,H., Angliker,H., Shaw,E., and Rivett,A.J. (1993). Reaction of proteasomes with peptidylchloromethanes and peptidyl diazomethanes. *Biochem. J.* 296, 601-605.

Seemüller,E., Lupas,A., and Baumeister,W. (1996). Autocatalytic processing of the 20S proteasome. *Nature* 382, 468-471.

Seemüller,E., Lupas,A., Stock,D., Löwe,J., Huber,R., and Baumeister,W. (1995). Proteasome from *Thermoplasma acidophilum*: a threonine protease. *Science* 268, 579-582.

Sommer,T. and Wolf,D.H. (1997). Endoplasmic reticulum degradation: reverse flow of no return. *FESEB J.* 11, 1227-1233.

Vinitzky,A., Cardozo,C., Sepp Lorenzino,L., Michaud,C., and Orłowski,M. (1994). Inhibition of the proteolytic activity of the multicatalytic proteinase complex (proteasome) by substrate-related peptidyl aldehydes. *J. Biol. Chem.* 269, 29860-29866.

Wiertz,E.J., Tortorella,D., Bogyo,M., Yu,J., Mothes,W., Jones,T.R., Rapoport,T.A., and Ploegh,H.L. (1996). Sec61-mediated transfer of a membrane protein from the endoplasmic reticulum to the proteasome for destruction. *Nature* 384, 432-438.

### 2.1.7 Abbreviations

MHC	major histocompatibility complex
PGPH	peptidylglutamylpeptide-hydrolyzing
BrAAP	branched chain amino acid-preferring
DCI	3,4-dichloroisocoumarin
DTT	dithiothreitol

Z-GGL-AMC	benzyl-oxycarbonyl-Gly-Gly-Leu-7-amino-4-methylcoumarin
Suc-LLVY-AMC	succinyl-Leu-Leu-Val-Tyr-7-amino-4-methylcoumarin
Z-ARR-AMC	benzyl-oxycarbonyl-Ala-Arg-Arg-7-amino-4-methylcoumarin
Z-LLE- $\beta$ -NA	benzyl-oxycarbonyl-Leu-Leu-Glu- $\beta$ -naphthylamide
Z-GPALA-pAB	benzyl-oxycarbonyl-Gly-Pro-Ala-Leu-Ala- <i>p</i> -aminobenzoic acid
HPLC	high pressure liquid chromatography
LC-MS	liquid chromatography-mass spectrometry
LLnL	N-acetyl-L-leucyl-L-leucinal-L-norleucinal
CTL	cytotoxic T lymphocyte
wt	wild-type
LC	lactacystin

### 2.1.8 Participating researchers

Tobias P. Dick\*<sup>1</sup>, Martin Deeg<sup>2</sup>, Wolfgang Heinemeyer<sup>3</sup>, Michael Groll<sup>4</sup>, Markus Schirle<sup>1</sup>, Wieland Keilholz<sup>1</sup>, Stefan Stevanović<sup>1</sup>, Dieter H. Wolf<sup>3</sup>, Robert Huber<sup>4</sup>, Hans-Georg Rammensee<sup>1</sup> and Hansjörg Schild<sup>1</sup>

<sup>1</sup> Institute for Cell Biology, Department of Immunology, University of Tübingen, Auf der Morgenstelle 15, D-72076 Tübingen, Germany.

<sup>2</sup> Center for Medical Research, University Hospital, Tübingen, Germany.

<sup>3</sup> Institute for Biochemistry, University of Stuttgart, Pfaffenwaldring 55, D-70569 Stuttgart, Germany.

<sup>4</sup> Max-Planck-Institut für Biochemie, D-82152 Martinsried, Germany.

\*A.K. Nussbaum and T.P. Dick contributed equally to the results summarized in this part.

## 2.2 The cleavage motifs of yeast 20S proteasomes deduced from digests of enolase-1

### 2.2.1 Summary

The 436-amino acid protein enolase-1 from yeast was degraded *in vitro* by purified wild-type and mutant yeast 20S proteasome particles. Analysis of the cleavage products at different time points revealed a processive degradation mechanism and a length distribution of fragments ranging from 3-25 amino acids with an average length of 7 to 8 amino acids. Surprisingly, the average fragment length was very similar between wt and mutant 20S proteasomes with reduced numbers of active sites. This implies that the fragment length is not influenced by the distance between the active sites, as previously postulated. A detailed analysis of the cleavages also allowed the identification of certain amino acid characteristics in positions flanking the cleavage site that guide the selection of the P1 residues by the 3 different active  $\beta$ -subunits. Since yeast and mammalian proteasomes are highly homologous, similar cleavage motifs might be employed by mammalian proteasomes. Therefore, our data provide a first basis for predicting proteasomal degradation products from which peptides are sampled by MHC class I molecules for presentation to cytotoxic T cells.

### 2.2.2 Introduction

The eukaryotic 20S proteasome represents the catalytic core particle of the 26S proteasome which is an essential component of the ubiquitin-dependent protein degradation pathway. The 20S particle is composed of 14 different but related subunits. Two outer disks, containing 7  $\alpha$ -subunits each, guard the two inner heptameric  $\beta$ -subunit rings, which contain three proteolytically active sites each (Groll et al., 1997). How substrate molecules gain access to the active sites at the inner surface of the  $\beta$ -rings is not known. It has been hypothesized that upon association of the 20S particle with regulatory complexes, like the 19S cap structure or PA28, an opening in the middle of the  $\alpha$ -subunit rings is induced allowing unfolded proteins to be fed into the 20S particle.

Probably as a consequence of this closed structure, very few proteins have been found to be degraded by 20S proteasomes *in vitro* and only very few individual fragments and cleavage sites generated during the degradation of proteins by eukaryotic 20S proteasomes have been analyzed until now. This kind of analysis is of interest for several



reasons.

First, it still has to be confirmed that the specificities of the proteasomal  $\beta$ -subunits observed in experiments using peptide substrates (Orlowski and Michaud, 1989; Cardozo et al., 1992; Heinemeyer et al., 1997; Niedermann et al., 1996; Niedermann et al., 1997; Dick et al., 1996), summarized as trypsin (T)-like, chymotrypsin (ChT)-like and peptidylglutamyl-peptide hydrolyzing (PGPH)-activity, correspond to cleavages performed in intact proteins, a situation physiologically more relevant.

Second, the mechanism of protein degradation by the proteasome is not known. For the archebacterial proteasome, which is a less complex structure consisting of 14 identical  $\alpha$ - and  $\beta$ -subunits (Löwe et al., 1995), a processive model has been proposed (Kisselev et al., 1998; Akopian et al., 1997). Because of the high structural homology with the archebacterial particle a similar mechanism can be expected for eukaryotic proteasomes. Furthermore, it has been postulated that the 20S proteasome is equipped with a molecular ruler that determines the length of fragments generated (Wenzel et al., 1994). In the archebacterial proteasomes, the ruler was postulated to reflect the distance between each of the seven active  $\beta$ -subunits (Dick et al., 1994). Therefore, a reduction of active sites is predicted to change the average length of fragments generated. In yeast proteasomes carrying only three active sites, another kind of ruler was proposed to reflect the distance of the 3 active  $\beta$ -subunits to putative proteolytically active sites at the carboxy-terminal ends of  $\alpha$ -helices at the  $\beta$ -annulus (Groll et al., 1997). Two studies recently performed using yeast 20S mutant proteasomes with inactivated  $\beta$ -subunits, however, excluded the latter model and revealed instead that no proteolytically active sites exist within the proteasome besides the  $\beta$ -subunits containing the active site Thr (Dick et al., 1998; Ditzel et al., 1998). The factor determining the length of the degradation products is therefore imposed either by the distance between the three active sites, analogous to the prediction for the *Thermoplasma* model, or by so far unknown mechanisms.

The third reason for an in-depth analysis of fragments resulting from proteasomal protein digests is the lack of knowledge about the selection of cleavage sites during the degradation of substrates. Why the proteasome selects certain cleavage sites and ignores others with identical P1/P1' composition is not known at all. A compilation of cleavage sites from a large panel of fragments should be very helpful in identifying amino acid motifs that explain the selection and allow the prediction of cleavage sites. Because the subunit topology and structure of mammalian proteasomes is analogous to that of yeast

(Kopp et al., 1997), the results will provide crucial information for the general understanding of the proteasomal activity.

We therefore examined in detail peptide fragments generated in digestions of yeast enolase-1 protein by wt and mutant yeast 20S proteasomes containing inactivated  $\beta$ -subunits (Heinemeyer et al., 1997). Degradation products of the individual digests were analyzed for processive substrate degradation, fragment length and the presence of cleavage motifs recognized by the three proteasomal subunits  $\beta$ 2/Pup1,  $\beta$ 1/Pre3 and  $\beta$ 5/Pre2. The data obtained allow us to present proteasomal cleavage maps of a large, unmodified protein for the first time. We find strong evidence for a processive degradation of the protein substrate and demonstrate that the number of active sites does not affect the average fragment length. From this we can exclude that the rules governing the fragment length in yeast 20S proteasomes are the result of the distance between active sites. The detailed analysis of cleavage sites in digests using wt and mutant yeast proteasomes also allows us for the first time to propose distinct cleavage motifs for the three active sites of eukaryotic 20S proteasomes that go far beyond the specificities described using peptide substrates.

## 2.2.3 Materials & Methods

### 2.2.3.1 Generation of mutant proteasomes and purification of yeast 20S proteasomes

Generation of yeast proteasome mutants and the purification of yeast 20S proteasomes has recently been described elsewhere (Groll et al., 1997; Heinemeyer et al., 1997).

### 2.2.3.2 HPLC-Separation of enolase digests

For the separation of degradation products unfractionated enolase digests were subjected to  $\mu$ RP SC 2.1/10-columns (Pharmacia) on a Microbore HPLC-system (SMART-system, Pharmacia). Buffer A: 0,1% TFA; buffer B: 0,081% TFA, 80% acetonitril. Gradients were 10% B for 5 min, in 35 min to 40% B, in 8 min to 75% B and up to 85% in another 7 min at a flow rate of 150  $\mu$ l/min.

### 2.2.3.3 Fragmentation of degradation products

Ms/ms experiments were performed on a hybrid quadrupole orthogonal acceleration tandem mass spectrometer (Q-TOF, Micromass). Fragmentation of the parent ions was achieved by collision with argon atoms. Q1 was set to the mass of interest  $\pm$  0.5 Da and the collision energy optimized for each fragment. Integration time for the TOF analyzer

was 1s with an inter scan delay of 0.1s.

#### 2.2.3.4 Digestions

Digests were incubated at 37°C and stopped by freezing at -20°C. Conditions in all digests were: 0,01% SDS, 10mM EDTA at molar ratios of enolase (10-20 µM) to 20S proteasome (40-60 nM) of 200-400:1. Buffer: 20mM Hepes-KOH, pH 7.6, 2mM MgAc<sub>2</sub>, 1mM DTT.

#### 2.2.3.5 MALDI MS analysis

1 µl of DHAP-matrix (20mg 2,5-dihydroxy-acetophenon, 5mg ammoniumcitrate in 1ml 80% isopropanol) was mixed with 1 µl of each HPLC peak fraction on a gold target. Measurements were executed using a LD-TOF (laser desorption- time of flight)-system (Hewlett-Packard G2025A) at a vacuum of 10<sup>-6</sup> Torr. For signal generation 50 to 150 laser shots were added up in the single shot mode.

#### 2.2.3.6 N-terminal sequencing (Edman degradation)

Routinely, 15 µl of each HPLC peak fraction from digest fractionations were applied to the pulsed-liquid sequencer Procise 494 (Applied Biosystems).

#### 2.2.3.7 Statistical analysis

##### 2.2.3.7.1 Frequencies of amino acids.

The probability  $q(k)$  for finding a given amino acid exactly  $k$  times in a given position by random selection of cleavage sites (considering that each single peptide bond can be randomly selected only once for each enolase molecule) can be calculated according to the hypergeometric distribution.

$q(k)$  values were used to calculate 2-sided tail probabilities  $p$  to indicate deviation of observed frequencies from a random selection of certain amino acids at a given position. Because we considered positions P6 to P6' surrounding a cleavage site the total number of potential cleavage sites ( $N$ ) in enolase-1 was reduced from 435 to 425.

##### 2.2.3.7.2 Comparison of amino acid characteristics.

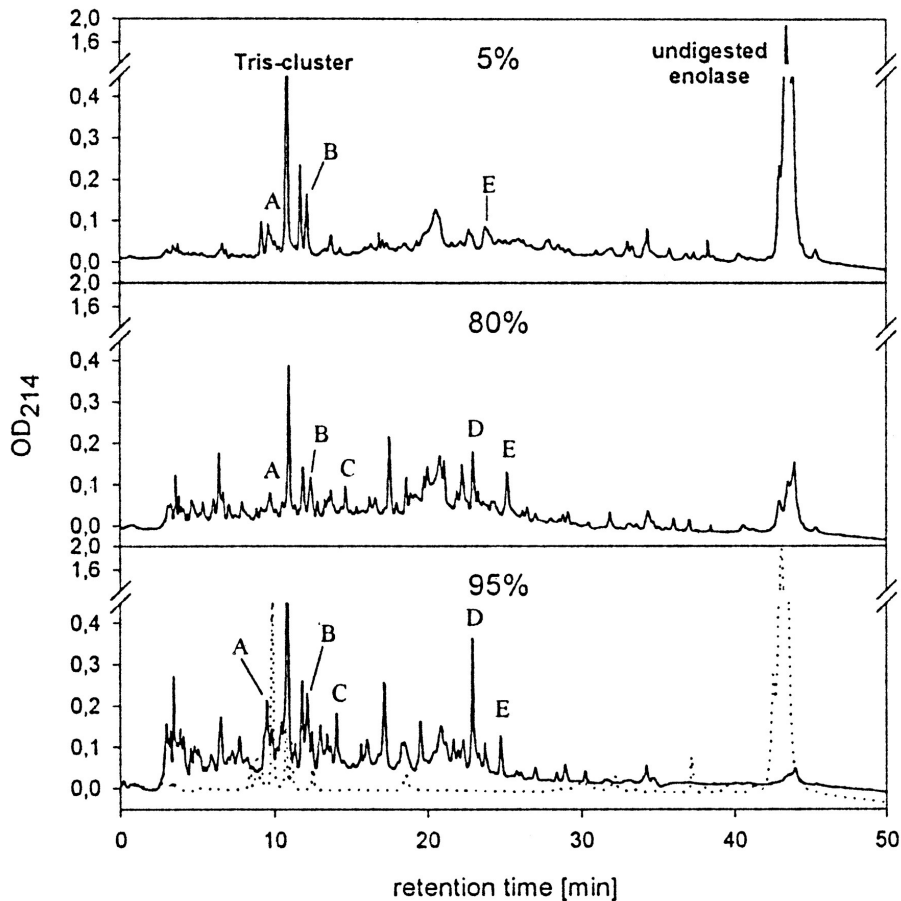
To compare the characteristics of amino acids the 2-sided Student's t-test for two independent data sets was performed. Hydrophobicity parameters were taken from Kyte and Doolittle (Kyte and Doolittle, 1982), bulkiness parameters from Zimmerman et al. (Zimmerman et al., 1968), and normalized frequency parameters for beta-turn from Levitt (Levitt, 1978). Results are shown as  $p$ -values and mean value differences including the

95% confidence limit.

## 2.2.4 Results

### 2.2.4.1 Degradation of enolase-1

Denaturation of proteins is a prerequisite for their digestion by 20S proteasomes. In order to perform digestions of proteins without prior covalent modifications we used protein substrates known for their thermolability, like yeast enolase-1. Enolase-1 was incubated with yeast 20S proteasomes at a molar ratio of about 300:1. Aliquots were taken at different time points and separated by reversed phase HPLC. Peaks containing degradation products increase over time, while the signal for undigested enolase disappears (Figure 2.2-1).



**Figure 2.2-1: Enolase-1 degradation by yeast wt 20S proteasomes**

Percent values represent the degree of enolase degradation as determined by linear regression.

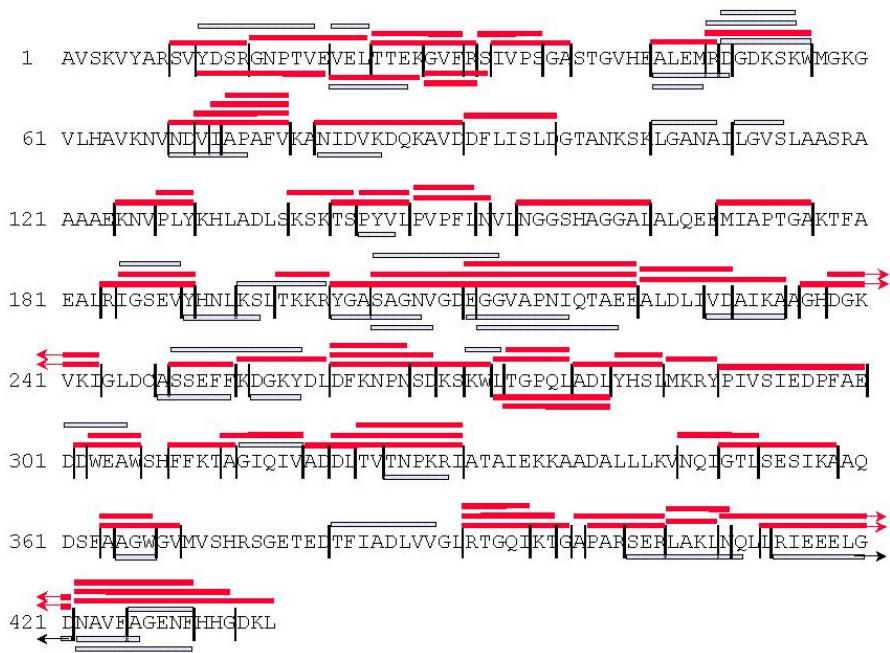
(Bottom) Dotted line in lower panel: Incubation of enolase without proteasome.

Substrate degradation was linear with a degradation velocity of 6 min (data not shown).

The observed stability of the peak pattern over time and the lack of larger enolase-1 fragments, especially at early time points (Figure 2.2-1) argue for a processive substrate degradation mechanism as proposed for the *Thermoplasma* proteasome (Akopian et al., 1997). The peaks A-E observed at different digestion times indeed contained identical fragments as determined by MALDI-MS analysis and Edman degradation (data not shown). The fragments in the range of 12 minutes coelute with digestion buffer components and the peak height therefore does not reflect the amount of peptides.

**2.2.4.2 Digestion map of enolase-1 by wt yeast 20S proteasomes**

All peak fractions obtained during reversed-phase HPLC separation of digests were analyzed by MALDI-MS, Edman sequencing and Q-ToF-MS/MS analysis and compiled into a digestion map (Figure 2.2-2). The digests were reproduced two times, yielding comparable HPLC profiles.



**Figure 2.2-2: Digestion map generated from degradation of enolase-1 by yeast wt 20S proteasomes**

*Vertical lines:* Cleavage sites determined by Edman degradation and mass spectrometry. *Closed bars (█):* Degradation products identified by Edman degradation and mass spectrometry. *Open bars (▭):* Degradation products identified by Edman degradation; C-terminus not yet confirmed by mass spectrometry.

As summarized earlier (Coux et al., 1996), proteasomes are able to perform cleavages

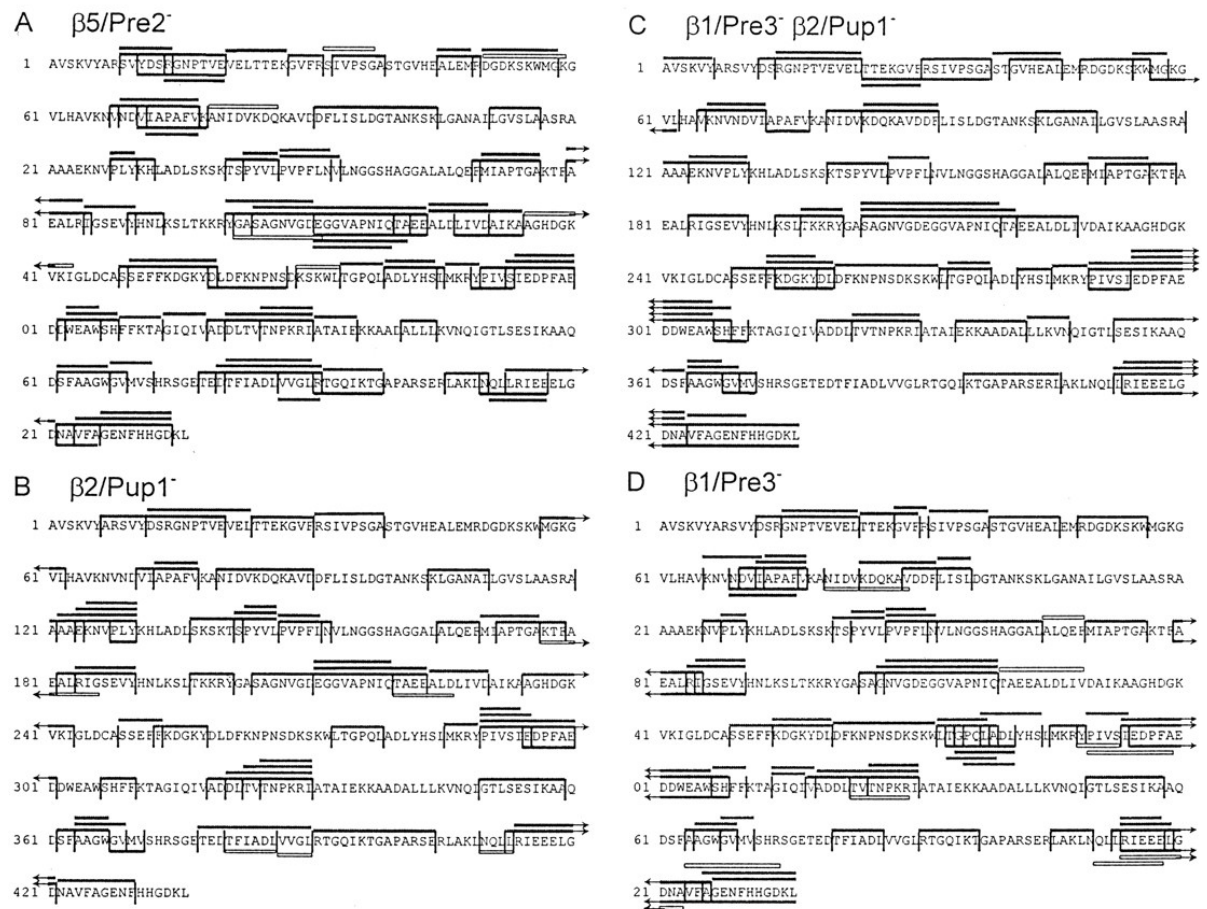
after almost every amino acid, whereby some residues are disfavored (such as Pro, Gln, Asn, Gly, Thr, Ser and Lys) and others favored (Ile, Leu, Trp, Tyr, Asp and Arg) (see Table 2.2-1 for details). The selection of amino acids at positions flanking the cleavage sites will be discussed later.

Three conclusions can be drawn from the observed cleavages: (i) the yeast 20S wt proteasome clearly prefers certain peptide bonds over others as cleavage sites; (ii) not all possible cleavage sites are used in one individual substrate molecule, resulting in the production of overlapping peptides; (iii) some regions of the enolase sequence are characterized by a high number, others by a low number of cleavage sites. This clustering cannot be correlated to charge clusters and repetitive elements which were not detected in the enolase-1 sequence using the SAPS program (Brendel et al., 1992). The fragments identified (n=66) in a single digest of enolase by wt 20S proteasomes cover a range of 3-23 amino acids (Figure 2.2-4). Fragments with a size of 3-9 amino acids clearly dominate with a mean fragment length of 7.5 amino acids (SD = 3.5). The digestion of enolase-1 using human 20S proteasomes resulted in a similar set of fragments, indicating a highly conserved selection of cleavage sites (Nussbaum et al., in preparation; see 2.2.9).

#### 2.2.4.3 Digestion of enolase-1 by mutant yeast 20S proteasomes

To analyze the contribution of the three active  $\beta$ -subunits to the observed cleavages we digested enolase-1 with the following mutant yeast 20S proteasomes carrying inactivated  $\beta$ -subunits, which were purified from the following mutant strains: *pre2-K33A*, *pup1-T1A*, *pup1-T1A pre3-T1A* and *pre3-T1A* (Heinemeyer et al., 1997). Cleavage activity was analyzed as described for wt proteasomes. The  $\beta 2/Pup1^-$  proteasome did not cleave peptide bonds after basic amino acids (Figure 2.2-3A, Table 2.2-1), the  $\beta 1/Pre3^-$  particle performed only one of the 17 cleavages found in wt proteasomes after acidic residues (Figure 2.2-3B, Table 2.2-1) and the  $\beta 2/Pup1^- \beta 1/Pre3^-$  proteasome completely lacked proteolytic activity after basic and acidic residues (Figure 2.2-3C, Table 2.2-1). The  $\beta 5/Pre2^-$  proteasome, completely lacking ChT-like activity as tested with fluorogenic substrates (Heinemeyer et al., 1997; Dick et al., 1998), surprisingly performed a high number of cleavages after hydrophobic amino acids (Figure 2.2-3D, Table 2.2-1). Because yeast 20S proteasomes contain no proteolytically active sites besides the three active  $\beta$ -subunits, the  $\beta 1/Pre3$  and  $\beta 2/Pup1$  subunits, previously thought to cleave only after charged residues, also harbor ChT-like activity. These features were already observed during the analysis of synthetic peptide substrates digested with the same

mutant proteasomes (Dick et al., 1998) and during experiments using inhibitors of the T-like and ChT-like activities (Savory et al., 1993; Reidlinger et al., 1997).

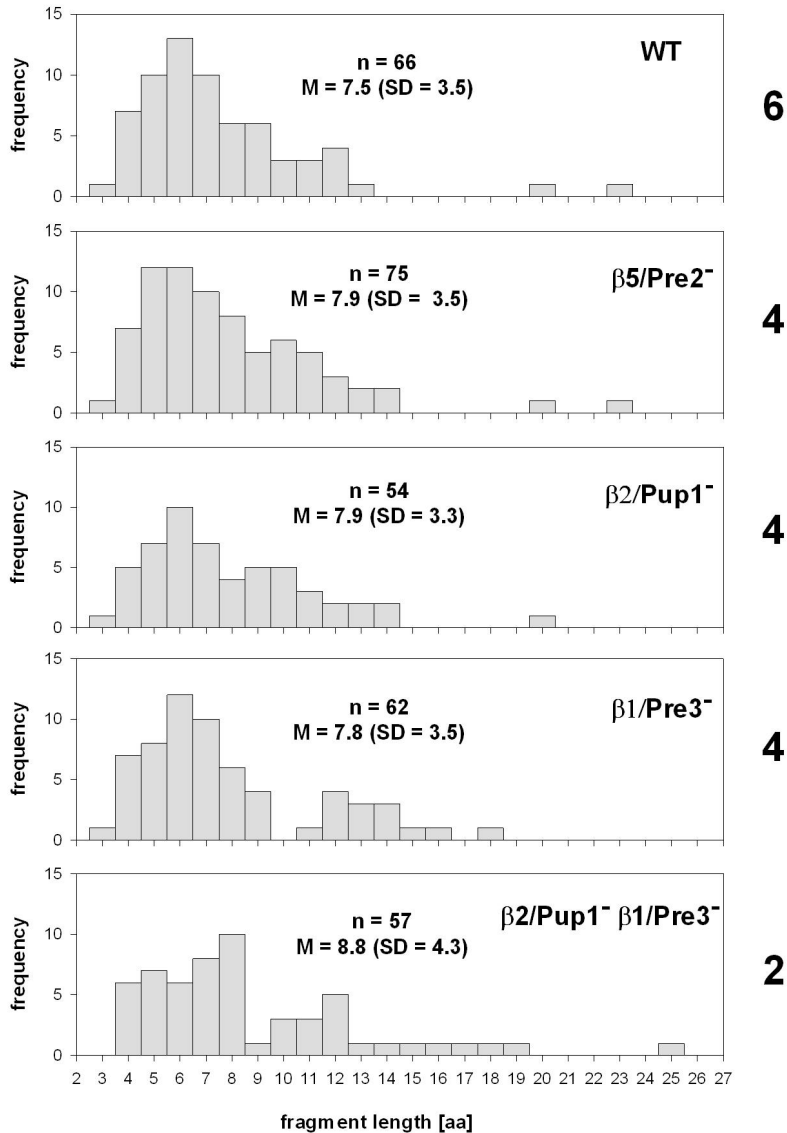


**Figure 2.2-3: Digestion maps generated from degradation of enolase-1 by yeast mutant 20S proteasomes**

The maps include all the degradation products of enolase-1 identified after digestion by (A)  $\beta 5/Pre2^-$ , (B)  $\beta 2/Pup1^-$ , (C)  $\beta 2/Pup1^- \beta 1/Pre3^-$ , (D)  $\beta 1/Pre3^-$  20S proteasomes. Degradation products were identified by Edman degradation and mass spectrometry. Symbols as in Figure 2.2-2.

In addition, the number of cleavages after basic and acidic residues performed by the  $\beta 5/Pre2^-$  particle was increased compared to wt proteasomes, most likely due to lack of competing cuts normally performed by the  $\beta 5/Pre2$  subunit. Even more surprising, all mutant proteasomes generated fragments of an average length of 7-9 amino acids (7.9, SD=3.5 for  $\beta 5/Pre2^-$ ; 7.9, SD=3.3 for  $\beta 2/Pup1^-$ ; 8.8, SD=4.3 for  $\beta 2/Pup1^- \beta 1/Pre3^-$ ; 7.8, SD=3.5 for  $\beta 1/Pre3^-$ ), which is in the range of fragments generated by wild-type proteasomes (Figure 2.2-4). Therefore, the size of fragments generated by yeast proteasomes is not related to the distance between the active  $\beta$ -subunits, as postulated

previously for the *Thermoplasma* proteasome (Wenzel et al., 1994).



**Figure 2.2-4: Distribution of fragment lengths generated by wt and mutant proteasomes from enolase-1**

For each proteasome only internal fragments with known C-terminus and identified from one single digestion experiment were included. M: mean fragment length. SD: standard deviation. n: number of internal fragments identified. Numbers to the right: Number of active  $\beta$ -subunits present in the different proteasome species.

#### 2.2.4.4 P1 amino acids and statistical analysis of flanking residues

The observed frequencies of amino acids in the P1 position of fragments generated by wildtype and mutant proteasomes indicate dominant cleavages after Leu or Arg, which are



used approximately every second time these residues appear in enolase-1. On the other hand, only 2 of 37 Gly residues are used for cleavage (Table 2.2-1). Despite exerting a significant influence, the P1 amino acid is obviously not the only factor determining the location of a cleavage site. By considering only the P1 position for example, it remains unclear why cleavages were detected after certain Leu residues, but not after other Leu residues.

**Table 2.2-1: Frequencies of amino acids in the P1 position**

P1 residue	Total no. in enolase	No. of observed cleavages				
		wt (n=105)	$\beta$ 2/Pup1- (n=80)	$\beta$ 1/Pre3- (n=88)	$\beta$ 2/Pup1- $\beta$ 1/Pre3- (n=73)	$\beta$ 5/Pre2- (n=96)
Ala	55	11	10	13	10	10
Arg	14	8	0	5	0	9
Asn	19	3	1	3	1	3
Asp	31	12	9	1	0	14
Cys	1	1	0	0	0	0
Gln	9	0	1	2	1	1
Glu	25	6	8	1	0	9
Gly	37	2	0	2	1	2
His	11	2	0	1	1	2
Ile	22	7	7	6	8	3
Leu	40	19	16	18	20	10
Lys	37	4	0	3	0	4
Met	5	1	0	1	1	1
Phe	16	5	8	9	8	2
Pro	15	0	0	0	0	0
Ser	31	5	2	3	4	8
Thr	20	2	1	2	1	0
Trp	5	4	4	4	4	2
Tyr	9	3	7	4	7	5
Val	34	10	6	10	6	11

This table only comprises cleavage sites that are not within the 6 N- or C-terminal aa; i.e. it only contains cleavage sites that were also included in the statistical analysis. Therefore, wt and  $\beta$ 2/Pre2<sup>+</sup> have each 1 cleavage site fewer than in the cleavage maps (Figure 2.2-2 and Figure 2.2-3).

In order to statistically evaluate the frequency of all amino acids in positions  $\pm 6$  surrounding cleavable peptide bonds, we compared the observed frequencies with those frequencies expected from a random cleavage distribution. Using the hypergeometric distribution, a value is assigned to each amino acid in every position, representing the probability (p) for the observed frequency to be the result of randomly distributed

cleavages. A statistically significant accumulation or scarcity of particular amino acids in certain positions is taken as an indication for a positive or negative influence on the selection of peptide bonds hydrolyzed by the proteasomal subunits. To justify this approach, a set of cleavages of the same size as the data set was generated by a random number generator and statistically evaluated as a control, resulting in no significant accumulation of amino acids (data not shown).

#### 2.2.4.5 Specificity of the wildtype proteasome

Examination of cleavages performed by the wildtype proteasome (n=105) revealed the following deviations from randomness in the cleavage pattern (p values are two-sided tail probabilities). Leu as a P1 residue has the highest significance (p<0.00082) and reflects pronounced ChT-like activity within wt proteasomes. Weaker preferences for Arg in P1 (p<0.0084), Pro in P4 (p<0.0033) and Gln in P5 (p<0.0011) are also evident. On the other hand, certain residues appear to exert negative influence in certain positions. In the P1 position, Gly is less abundant than expected (p<0.0011). When observed cleavage sites (n=105) are compared to the remaining sites (not observed as cleavage sites, n=320) using the two-sided Student's t-test, a preference for small residues becomes apparent in P1' (p<0.0072;  $1.57 \pm 1.15$ ).

#### 2.2.4.6 Specificity of the $\beta 5$ /Pre2 active site

To evaluate a 'cleavage motif' for a single active  $\beta$ -subunit, in this case  $\beta 5$ /Pre2, cleavages performed by the  $\beta 2$ /Pup1- $\beta 1$ /Pre3<sup>-</sup> 20S particle were analyzed (n=73). The statistical analysis clarifies the order of preference for the different hydrophobic P1 residues relative to their abundance in the substrate: Leu (p<0.0000003) > Tyr (p<0.0001) > Phe (p<0.002) > Trp (p<0.0035) > Ile (p<0.02), but no accumulation is apparent for Val. This is, however, in agreement with recent studies on the cleavage of peptide substrates, indicating that cleavage after Val can be efficiently catalyzed by  $\beta 2$ /Pup1 (Dick et al., 1998). Similar to the wt proteasome, Gly is disfavored in the P1 position. When amino acids are grouped (hydrophobic, basic, acidic and others) the preference for hydrophobic residues in P1 is even more impressive (p<3\*10<sup>-17</sup>) together with a weak preference for basic residues at P1' (p<0.0012). When comparing the observed cleavages (n=73) to the remaining sites (n=352), the following preferences are observed: In positions P3 and P5' polar residues are clearly favored over non-polar ones (p<0.000072;  $1.54 \pm 0.75$  and p<0.0032;  $1.15 \pm 0.76$ , respectively) and a preference for  $\beta$ -turn-promoting residues is seen in position P3 (p<0.0093;  $0.13 \pm 0.10$ ). To determine more precisely why  $\beta 5$ /Pre2

selects certain Leu as P1 residues, but not others, observed cleavage sites with Leu in P1 (n=20) were compared to the corresponding unobserved sites with Leu in P1 (n=19): Again there is a positive correlation between cleavage and the presence of polar ( $p < 0.00076$ ;  $3.2 \pm 1.8$ ) and turn-promoting residues ( $p < 0.0022$ ;  $0.38 \pm 0.24$ ) in the P3 position.

#### 2.2.4.7 *Certain cleavages are not restricted to a single active site*

$\beta 5/Pre2^-$ -catalyzed cleavages as defined by the  $\beta 2/Pup1^- \beta 1/Pre3^-$  double mutant (n=73) should belong to a set complementary to that provided by the  $\beta 5/Pre2^-$  mutant (n=96). However, there is considerable overlap between these two sets of cleavages. Cleavages common to both  $\beta 5/Pre2^-$  and  $\beta 2/Pup1^- \beta 1/Pre3^-$  proteasome digests (n=31) therefore represent features not dependent on a single subunit. We find Tyr preferred at the P1 position ( $p < 0.0002$ ) and Pro at P4 ( $p < 0.0003$ ). When comparing cleavage sites of this intersecting subset (n=31) against the remaining sites (n=394), P1 residues tend to be more hydrophobic ( $p < 0.0013$ ;  $1.8 \pm 1.1$ ) and P3 residues more hydrophilic ( $p < 0.00023$ ;  $2.1 \pm 1.1$ ).

#### 2.2.4.8 *Specificity of the $\beta 2/Pup1$ active site*

Because the respective double mutant proteasomes were not available, the motifs for  $\beta 2/Pup1$  and  $\beta 1/Pre3$  subunits must be approximated. One possibility for approximating  $\beta 2/Pup1$ -specificity is to consider those cleavages catalyzed exclusively by  $\beta 1/Pre3^-$  proteasomes (n=88), but not by  $\beta 1/Pre3^- \beta 2/Pup1^-$  proteasomes (n=73). For these cleavages (n=41) analysis points out weak preferences for Gly in position P1' ( $p < 0.001$ ) and Pro in position P3' ( $p < 0.001$ ). Another possibility of approaching the  $\beta 2/Pup1$  motif is to look at those cleavages common to both  $\beta 1/Pre3^-$  (n=88) and  $\beta 5/Pre2^-$  (n=96) proteasomes, which should mostly originate from the  $\beta 2/Pup1$  subunit (n=46). Analysis of amino acid frequency reveals a strong preference for Pro in P4 ( $p < 0.00004$ ) and P3' and Lys in P3 ( $p < 0.0031$ ). However, this set of 46 cleavages includes 31 cleavages that can be found in either  $\beta 1/Pre3^- \beta 2/Pup1^-$  or  $\beta 2/Pup1^-$  proteasomes and therefore belong to the category of cleavages that are not subunit-specific. Analysis of the remaining 15, strictly  $\beta 2/Pup1$ -specific cleavages reveals Arg in P1 ( $p < 0.000044$ ), Pro in P3', Gly in P1', but not Pro in P4, again showing that the preference for Pro in P4 is strongly associated with those cleavages that can be catalyzed by more than a single subunit.

#### 2.2.4.9 Specificity of the $\beta$ 1/Pre3 active site

Again, there are several possibilities of approximating a  $\beta$ 1/Pre3 cleavage motif. Looking at those cleavages common to both digests with proteasomes lacking active  $\beta$ 5/Pre2 (n=96) and active  $\beta$ 2/Pup1 subunits (n=80), the intersecting set includes 44 cleavages: Pro in P4 is the most prominent feature ( $p < 0.000002$ ), followed by Tyr in P1 ( $p < 0.0009$ ). If those cleavages not specific to a single subunit are eliminated, only 14 cleavages remain as truly  $\beta$ 1/Pre3-specific. Asp now is significant in P1 ( $p < 0.0000007$ ) and Glu in P1 also becomes evident ( $p < 0.0063$ ). Comparing the cleavages common to proteasomes with active  $\beta$ 1/Pre3 subunits (n=44) against the remaining sites of enolase (n=381), the P5 position turns out to prefer hydrophilic residues ( $p < 0.0001$ ;  $3.2 \pm 1.6$ ). The P3 position favors hydrophobic residues ( $p < 0.00061$ ;  $2.82 \pm 1.6$ ) and disfavors turn-promoting residues ( $p < 0.0076$ ;  $0.29 \pm 0.21$ ). If cleavages with Asp or Glu in P1 (n=31) are compared against the remaining Asp and Glu sites (n=24), amino acids at the P4' position tend to be more hydrophobic ( $p < 0.0056$ ;  $2.07 \pm 1.43$ ). The proposed cleavage motifs are summarized in Table 2.2-2.

#### 2.2.5 Discussion

The specificity of 20S proteasomes has been investigated so far either by the use of peptide substrates and inhibitors of the proteasomal activities. To analyze the proteasomal specificity under more physiological conditions we established a protocol that allowed the digestion of an entire protein, yeast enolase-1, without denaturation by covalent modification of amino acids. The identification of more than 400 enolase-1 fragments generated in digests using wild-type and mutated yeast 20S proteasomes carrying different combinations of inactivated  $\beta$ -subunits allowed a detailed analysis of the length of fragments generated, proteolytic specificity,  $\beta$ -subunit contribution to the observed cleavages and a description of motifs that explain the selection of cleavage sites.

Table 2.2-2: Cleavage motifs of yeast 20S active subunits

	Position									
	P5	P4	P3	P2	P1	P1'	P2'	P3'	P4'	P5'
<b>wt</b>		(Pro)*			Leu (Arg) -Gly	$\beta$ -turn small				
<b><math>\beta</math>5/Pre2</b>			polar $\beta$ -turn		Leu Tyr Phe Trp -Gly	basic				$\beta$ -turn polar
<b><math>\beta</math>2/Pup1</b>			(Lys)		Arg (Lys)	(Gly)		(Pro)		
<b><math>\beta</math>1/Pre3</b>	polar		hydro- phobic		Asp (Glu)				hydro- phobic	
<b>Subunit independent</b>		Pro	hydro- phobic		Tyr hydro- phobic					

\*Amino acids in parentheses indicate P values below the significance level after the adjustment for multiple testing according to the Bonferroni method (Bland and Altman, 1995). Significances are, however, supported by amino acid characteristic analysis or crystal structure data. A minus indicates a disfavored amino acid.

### 2.2.5.1 Protein degradation and fragment length

We determined a duration of 6 min for the degradation of one molecule of enolase by one yeast wt 20S proteasome. This value is in the range of activities found so far for degradation of whole proteins by eukaryotic and archaeal 20S proteasomes (Kisselev et al., 1998; Dick et al., 1994; McGuire et al., 1989; Dick et al., 1991), but is likely to be far from physiological levels. In vivo it takes less than 60 min from viral infection to MHC class I restricted presentation of viral peptides to cytotoxic T cells (Hosken et al., 1989). Numerous possibilities can explain this discrepancy, ranging from the lack of crucial components in the digestion buffer to missing in vivo 'helpers of proteasomal proteolysis' such as the 19S cap complex, hsc70 (Bercovich et al., 1997) or activators such as PA28 in mammalian cells (Dick et al., 1996; Dubiel et al., 1992; Ma et al., 1992; Groettrup et al.,

1996).

Our experiments using yeast 20S proteasomes and enolase-1 as a substrate show a very stable pattern of degradation products in HPLC-separated digests over time (Figure 2.2-1) as reported for the *Thermoplasma* 20S proteasome (Kisselev et al., 1998; Akopian et al., 1997). In addition, the same degradation products could be identified at early time points when only 5% of the substrate is degraded and also at late time points when 95% of the substrate molecules are digested (data not shown). This demonstrates that final degradation products are generated without the presence of large degradation intermediates.

The finding that all proteasomes analyzed so far generate peptides with an average length of 7-9 amino acids led to the proposal of an intrinsic molecular ruler determining the fragment size (Groll et al., 1997; Löwe et al., 1995; Wenzel et al., 1994). Our finding that proteasomes carrying 6, 4 or 2 active  $\beta$ -subunits generate fragments of similar length strongly argues against the distance between the active sites influencing the length of proteasomal digestion products (Figure 2.2-4). The small fragment length increase from the wild-type proteasomes to the  $\beta 1/\text{Pre}3\beta 2/\text{Pup}1^-$  proteasomes is most likely due to lack of cleavage activity after charged amino acids which represent 25% of the amino acids in enolase-1. Our observed fragment length distribution is also in agreement with recent results using partially modified protein substrates and archaeal proteasomes (Kisselev et al., 1998). An explanation might be that proteasomes either have an exit size filter preventing the dissociation of products above a certain size or that active  $\beta$ -subunits preferentially bind polypeptide stretches of 7-8 amino acids. The nonsymmetrical shape of the fragment length distributions (Figure 2.2-4) might indicate a log-normal distribution of the fragment size, as postulated (Kisselev et al., 1998; Akopian et al., 1997). The fact that the main part of cleavage products does not fall into the range of MHC-ligands (8-10aa) probably mirrors the predominant contribution of the proteasome to general intracellular proteolysis, which preceded its participation in antigen processing in evolution.

#### 2.2.5.2 Specificity and selectivity of the active $\beta$ -subunits

Analysis of the specificity of the different  $\beta$ -subunits revealed that the  $\beta 5/\text{Pre}2$  subunit is responsible for the ChT-like activity, as expected. Proteasomes containing inactivated  $\beta 2/\text{Pup}1$  or  $\beta 1/\text{Pre}3$  subunits lacked only cleavage activity after basic or acidic residues, respectively. By analyzing the cleavage sites performed by the double mutant containing  $\beta 5/\text{Pre}2$  as the only active subunit we found the strongest selection for Leu in P1 followed

by Tyr, Phe and Trp. By comparing the flanking residues around the  $\beta 5/\text{Pre}2$  cleavage sites with the rest of the observed cleavages we found a strong preference for polar and  $\beta$ -turn promoting amino acids at P3. This observation correlates to characteristics of amino acids at the inner surface of the proteasome surrounding the active site Thr1 of  $\beta 5/\text{Pre}2$ . A polar pocket formed by His98, Asp114, Glu120 and Arg125 of the  $\beta 6/\text{C}5$  subunit might accommodate P3 residues of substrates. A similar situation might exist at the P1' position, where basic residues are enriched and might interact with the negatively charged side-chains of Asp114 and Asp116 of  $\beta 5/\text{Pre}2$ .

The analysis of cleavages performed by  $\beta 2/\text{Pup}1$  revealed a preference for Arg over Lys in P1. The reasons for this selectivity are not known but have already been observed (Ossendorp et al., 1996). Flanking residue analysis revealed also an enrichment of Lys at P3 of the fragments generated by  $\beta 2/\text{Pup}1$ . The basic side-chain might interact with an acidic pocket formed by Asp28 ( $\beta 2/\text{Pup}1$ ) and Asp114, Glu120 and Glu121 of  $\beta 3/\text{Pup}3$ . On the P' side, flexible and  $\beta$ -turn promoting amino acids like Gly at P1' and Pro at P3' are enriched. This feature of  $\beta 2/\text{Pup}1$  substrates might be imposed by amino acids, including His35 of  $\beta 2/\text{Pup}1$ , which block the end of a putative substrate binding cleft and thus require substrate bending.

The  $\beta 1/\text{Pre}3$  subunit favors aspartate over glutamate in P1 and at position P5 polar amino acids, and at position P3 hydrophobic residues are enriched. Whether these features are reflected by the amino acids surrounding the active site Thr1 remains to be determined.

Surprisingly, the  $\beta 2/\text{Pup}1$  and  $\beta 1/\text{Pre}3$  subunits also contribute substantially to the ChT-like activity as evident from the analysis of the  $\beta 5/\text{Pre}2^-$  proteasome, where only a slight reduction in the number of cleavages after hydrophobic amino acids was found. After analyzing the cleavages performed by the  $\beta 5/\text{Pre}2^-$  and the  $\beta 2/\text{Pup}1\beta 1/\text{Pre}3^-$  proteasomes, expected to contain complementary sets of cleavage sites, 31 identical cleavages were found in both groups. By examining the residues flanking these cleavage sites, we found Pro at P4 significantly enriched possibly favoring a turn at this position. When we extended the analysis of amino acid frequencies and characteristics to 10 amino acids flanking observed cleavage sites in both directions, we did not observe any additional significant amino acid preferences. This finding might indicate that a stretch of about 10 amino acids (P5 to P5') is critical for the selection of a cleavage site by the  $\beta$ -subunits.

The cleavage motifs for proteasomal cleavages described here represent the first tools for

developing an algorithm for the prediction of proteasomal cleavage sites. These motifs will be relevant for the prediction of peptides generated by mammalian proteasomes since the cleavages in enolase-1 generated by yeast or human 20S proteasomes overlap to a large extent (Nussbaum et al., unpublished observations; see 2.2.9). Nevertheless, detailed studies using human 20S proteasomes, lacking or containing interferon-inducible  $\beta$ -subunits, will be performed to assess their contribution to the generation of epitopes for cytotoxic T cells and to refine the cleavage motifs.

## 2.2.6 References

- Akopian, T.N., Kisselev, A.F., and Goldberg, A.L. (1997). Processive degradation of proteins and other catalytic properties of the proteasome from *Thermoplasma acidophilum*. *J. Biol. Chem.* 272, 1791-1798.
- Bercovich, B., Stancovski, I., Mayer, A., Blumenfeld, N., Laszlo, A., Schwartz, A.L., and Ciechanover, A. (1997). Ubiquitin dependent degradation of certain protein substrates in vitro requires the molecular chaperone hsc70. *J. Biol. Chem.* 272, 9002-9010.
- Bland, J.M. and Altman, D.G. (1995). Multiple significance tests: the Bonferroni method. *BMJ* 310, 170.
- Brendel, V., Bucher, P., Nourbakhsh, I.R., Blaisdell, B.E., and Karlin, S. (1992). Methods and algorithms for statistical analysis of protein sequences. *Proc. Natl. Acad. Sci. U. S. A* 89, 2002-2006.
- Cardozo, C., Vinitsky, A., Hidalgo, M.C., Michaud, C., and Orłowski, M. (1992). A 3,4-dichloroisocoumarin-resistant component of the multicatalytic proteinase complex. *Biochemistry* 31, 7373-7380.
- Coux, O., Tanaka, K., and Goldberg, A.L. (1996). Structure and functions of the 20S and 26S proteasomes. *Annu. Rev. Biochem.* 65, 801-847.
- Dick, L.R., Aldrich, C., Jameson, S.C., Moomaw, C.R., Pramanik, B.C., Doyle, C.K., DeMartino, G.N., Bevan, M.J., Forman, J.M., and Slaughter, C.A. (1994). Proteolytic processing of ovalbumin and beta-galactosidase by the proteasome to a yield antigenic peptides. *J. Immunol.* 152, 3884-3894.
- Dick, L.R., Moomaw, C.R., DeMartino, G.N., and Slaughter, C.A. (1991). Degradation of oxidized insulin B chain by the multiproteinase complex macropain (proteasome). *Biochemistry* 30, 2725-2734.



- Dick, T.P., Nussbaum, A.K., Deeg, M., Heinemeyer, W., Groll, M., Schirle, M., Keilholz, W., Stevanovic, S., Wolf, D.H., Huber, R., Rammensee, H.G., and Schild, H. (1998). Contribution of proteasomal beta-subunits to the cleavage of peptide substrates analyzed with yeast mutants. *J. Biol. Chem.* 273, 25637-25646.
- Dick, T.P., Ruppert, T., Groettrup, M., Kloetzel, P.M., Kuehn, L., Koszinowski, U.H., Stevanovic, S., Schild, H., and Rammensee, H.G. (1996). Coordinated dual cleavages induced by the proteasome regulator PA28 lead to dominant MHC ligands. *Cell* 86, 253-262.
- Ditzel, L., Huber, R., Mann, K., Heinemeyer, W., Wolf, D.H., and Groll, M. (1998). Conformational constraints for protein self-cleavage in the proteasome. *J. Mol. Biol.* 279, 1187-1191.
- Dubiel, W., Pratt, G., Ferrell, K., and Rechsteiner, M.C. (1992). Purification of an 11 S regulator of the multicatalytic protease. *J. Biol. Chem.* 267, 22369-22377.
- Groettrup, M., Soza, A., Eggers, M., Kuehn, L., Dick, T.P., Schild, H., Rammensee, H.G., Koszinowski, U.H., and Kloetzel, P.M. (1996). A role for the proteasome regulator PA28alpha in antigen presentation. *Nature* 381, 166-168.
- Groll, M., Ditzel, L., Löwe, J., Stock, D., Bochtler, M., Bartunik, H.D., and Huber, R. (1997). Structure of 20S proteasome from yeast at 2.4 Å resolution. *Nature* 386, 463-471.
- Heinemeyer, W., Fischer, M., Krimmer, T., Stachon, U., and Wolf, D.H. (1997). The active sites of the eukaryotic 20 S proteasome and their involvement in subunit precursor processing. *J. Biol. Chem.* 272, 25200-25209.
- Hosken, N.A., Bevan, M.J., and Carbone, F.R. (1989). Class I-restricted presentation occurs without internalization or processing of exogenous antigenic peptides. *J. Immunol.* 142, 1079-1083.
- Kisselev, A.F., Akopian, T.N., and Goldberg, A.L. (1998). Range of sizes of peptide products generated during degradation of different proteins by archaeal proteasomes. *J. Biol. Chem.* 273, 1982-1989.
- Kopp, F., Hendil, K.B., Dahlmann, B., Kristensen, P., Sobek, A., and Uerkevitz, W. (1997). Subunit arrangement in the human 20S proteasome. *Proc. Natl. Acad. Sci. USA* 94, 2939-2944.
- Kyte, J. and Doolittle, R.F. (1982). A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157, 105-132.
- Levitt, M. (1978). Conformational preferences of amino acids in globular proteins. *Biochemistry* 17,

4277-4285.

Löwe, J., Stock, D., Jap, B., Zwickl, P., Baumeister, W., and Huber, R. (1995). Crystal structure of the 20S proteasome from the archaeon *T. acidophilum* at 3.4 Å resolution. *Science* 268, 533-539.

Ma, C.P., Slaughter, C.A., and DeMartino, G.N. (1992). Identification, purification, and characterization of a protein activator (PA28) of the 20 S proteasome (macropain). *J. Biol. Chem.* 267, 10515-10523.

McGuire, M.J., McCullough, M.L., Croall, D.E., and DeMartino, G.N. (1989). The high molecular weight multicatalytic proteinase, macropain, exists in a latent form in human erythrocytes. *Biochim. Biophys. Acta* 995, 181-186.

Niedermann, G., Grimm, R., Geier, E., Maurer, M., Realini, C., Gartmann, C., Soll, J., Omura, S., Rechsteiner, M.C., Baumeister, W., and Eichmann, K. (1997). Potential immunocompetence of proteolytic fragments produced by proteasomes before evolution of the vertebrate immune system. *J. Exp. Med.* 186, 209-220.

Niedermann, G., King, G., Butz, S., Birsner, U., Grimm, R., Shabanowitz, J., Hunt, D.F., and Eichmann, K. (1996). The proteolytic fragments generated by vertebrate proteasomes: structural relationships to major histocompatibility complex class I binding peptides. *Proc. Natl. Acad. Sci. USA* 93, 8572-8577.

Orlowski, M. and Michaud, C. (1989). Pituitary multicatalytic proteinase complex. Specificity of components and aspects of proteolytic activity. *Biochemistry* 28, 9270-9278.

Ossendorp, F.A., Eggers, M., Neisig, A., Ruppert, T., Groettrup, M., Sijts, A.J., Mengede, E., Kloetzel, P.M., Neefjes, J., Koszinowski, U.H., and Melief, C.J. (1996). A single residue exchange within a viral CTL epitope alters proteasome-mediated degradation resulting in lack of antigen presentation. *Immunity* 5, 115-124.

Reidlinger, J., Pike, A.M., Savory, P.J., Murray, R.Z., and Rivett, A.J. (1997). Catalytic properties of 26 S and 20 S proteasomes and radiolabeling of MB1, LMP7, and C7 subunits associated with trypsin-like and chymotrypsin-like activities. *J. Biol. Chem.* 272, 24899-24905.

Savory, P.J., Djaballah, H., Angliker, H., Shaw, E., and Rivett, A.J. (1993). Reaction of proteasomes with peptidylchloromethanes and peptidyl diazomethanes. *Biochem. J.* 296, 601-605.

Wenzel, T., Eckerskorn, C., Lottspeich, F., and Baumeister, W. (1994). Existence of a molecular ruler in proteasomes suggested by analysis of degradation products. *FEBS Lett.* 349, 205-209.

Zimmerman, J.M., Eliezer, N., and Simha, R. (1968). The characterization of amino acid sequences in proteins by statistical methods. *J. Theor. Biol.* 21, 170-201.

### 2.2.7 Abbreviations

T	trypsin
ChT	chymotrypsin
PGPH	peptidylglutamyl-peptide hydrolyzing
MHC	major histocompatibility complex
wt	wild type
MS	mass spectrometry
MALDI	matrix associated laser desorption ionisation
Q-Tof	Quadrupole Time of Flight

### 2.2.8 Participating researchers

Tobias P. Dick\*<sup>1</sup>, Wieland Keilholz<sup>1</sup>, Markus Schirle<sup>1</sup>, Stefan Stevanović<sup>1</sup>, Klaus Dietz<sup>2</sup>, Wolfgang Heinemeyer<sup>3</sup>, Michael Groll<sup>4</sup>, Dieter H. Wolf<sup>3</sup>, Robert Huber<sup>4</sup>, Hans-Georg Rammensee<sup>1</sup>, and Hansjörg Schild<sup>1</sup>

<sup>1</sup> Institut für Zellbiologie, Abteilung Immunologie, Universität Tübingen, Auf der Morgenstelle 15, D-72076 Tübingen, Germany.

<sup>2</sup> Institut für Medizinische Biometrie, Universität Tübingen, Westbahnhofstrasse 55, D-72070 Tübingen, Germany.

<sup>3</sup> Institut für Biochemie, Universität Stuttgart, Pfaffenwaldring 55, D-70569 Stuttgart, Germany.

<sup>4</sup> Max-Planck-Institut für Biochemie, D-82152 Martinsried, Germany.

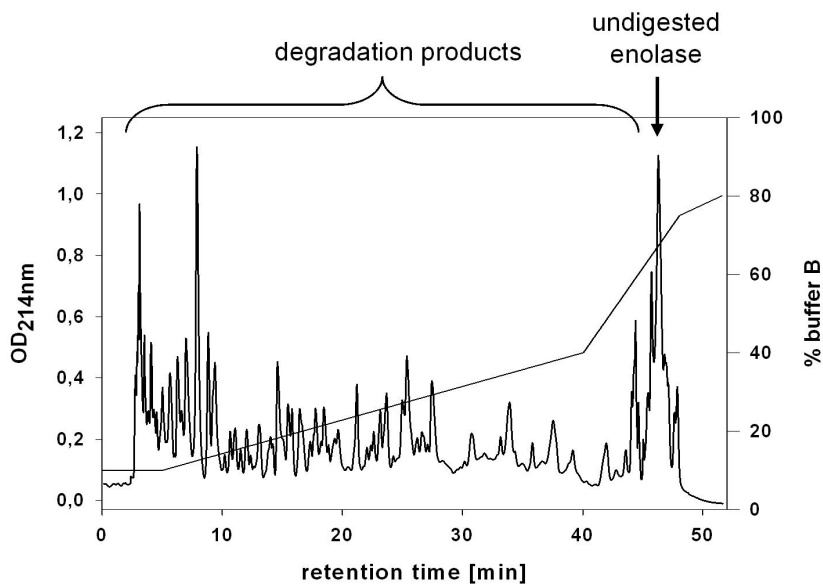
\*A.K. Nussbaum and T.P. Dick contributed equally to the results summarized in this part.

## 2.2.9 Additional data: Cleavage motif of human erythrocyte 20S proteasomes

Soon after analysing the yeast 20S cleavage motifs, I finished the analysis of the digest of enolase by human erythrocyte 20S proteasomes, with the help of members of the analytics group in the lab (Dr. Stevanović and M. Schirle).

The results have only been published in detail as poster contributions at several conferences (e.g. DGfI Jahrestagung 1999, Hannover) and are summarized below:

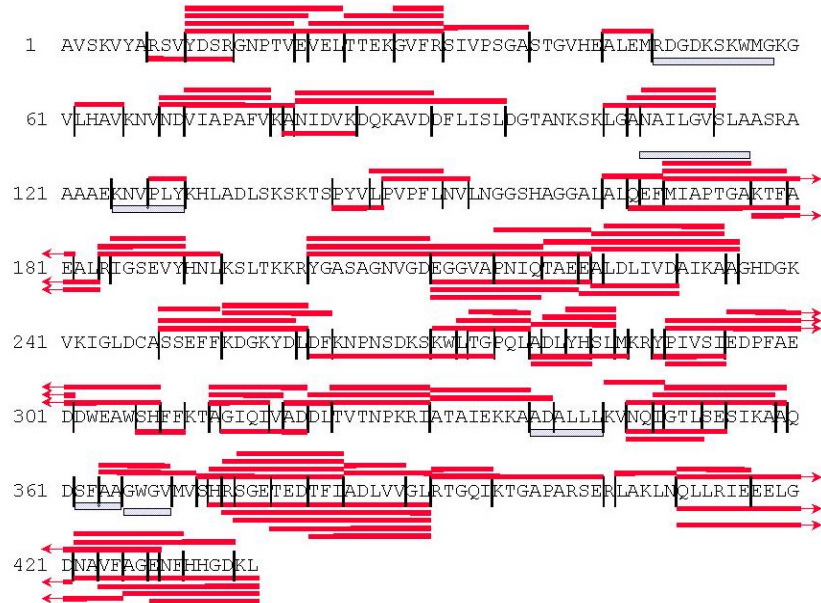
### 2.2.9.1 Fractionation of the degradation products by HPLC



**Figure 2.2-5: HPLC-chromatogram of digest of enolase-1 by human erythrocyte 20S proteasomes**

The digest was stopped by freezing at  $-20^{\circ}\text{C}$  after 5 days of incubation at  $37^{\circ}\text{C}$ . At this time, enolase was degraded to about 80%. No enolase disintegration was observed in a control without proteasomes. For buffer conditions, check 2.2.3.2.

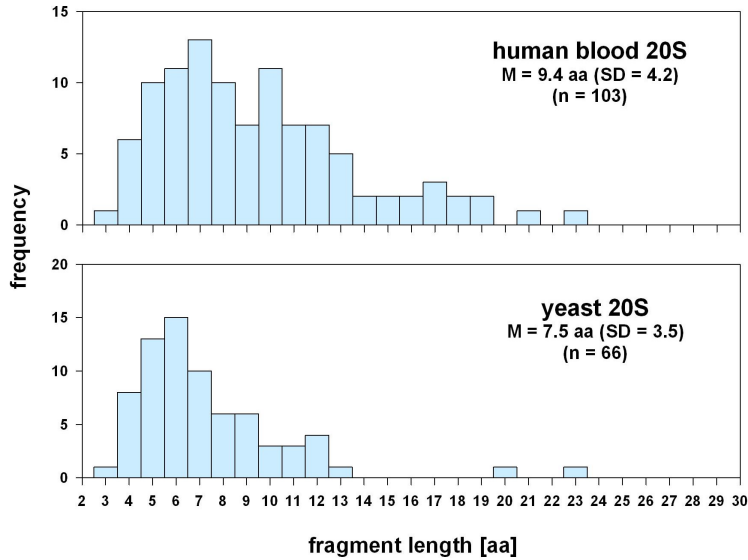
### 2.2.9.2 Cleavage map



**Figure 2.2-6: Cleavage map of the degradation of enolase-1 by human erythrocyte 20S proteasomes**

The map comprises the 107 full fragments (*horizontal bars*) and the 118 cleavage sites (*vertical lines*) identified in digests of enolase by human erythrocyte 20S proteasomes. (*Solid bars*) full fragments of which both N- and C-termini are known. (*Open bars*) fragments for which only the N-termini were identified.

### 2.2.9.3 Fragment length distribution



**Figure 2.2-7: Human and yeast proteasomes generate fragments of similar length in digests of enolase**

The general fragment length distribution is similar for yeast and human 20S proteasomes, with a range from 3-23 aa. The slightly longer average length for human 20S proteasomes can be explained by improved analytical methods which made it possible to identify more (especially longer) fragments. Only internal fragments with known C-terminus and identified from one single digestion experiment were included. M: mean fragment length. SD: standard deviation. n: number of internal fragments identified.

Cleavage motif

**Table 2.2-3: The cleavage motifs of yeast (wt) and human erythrocyte 20S proteasomes overlap largely**



	P5	P4	P3	P2	P1	P'1	P'2	P'3	P'4	P'5
yeast 20S		P			L R -G hydrophobic	b-turn small				
human 20S		(P)			L -G bulky hydrophobic	small polar	-K	bulky hydrophobic		

Both proteasome species prefer large, hydrophobic and aa in P1, whereas they dislike the small aa Gly at this position. They both like small aa in P1' and Pro in P4. For the human 20S proteasome, the statistical analysis yielded p-values of barely significant numbers.

In summary, the cleavage preferences of yeast and human 20S proteasomes are similar.

We used the yeast and human 20S proteasome cleavage data to generate algorithms for the prediction of proteasomal cleavages. This is shown in the next section (2.3) of the Results part.

## **2.3 An algorithm for the prediction of proteasomal cleavages**

### **2.3.1 Summary**

Proteasomes, major proteolytic sites in eukaryotic cells, play an important part in major histocompatibility class I (MHC I) ligand generation and thus, for the regulation of specific immune responses. Their cleavage specificity is of outstanding interest for this process.

In order to generalize previously determined cleavage motifs of 20S proteasomes, we developed network-based model proteasomes trained by an evolutionary algorithm with experimental cleavage data of yeast and human 20S proteasomes. A window of 10 flanking amino acids (AAs) proved sufficient for the model proteasomes to reproduce the experimental results with 98-100% accuracy. Actual experimental data were reproduced significantly better than randomly selected cleavage sites, suggesting that our model proteasomes were able to extract rules inherent to proteasomal cleavage data. The affinity parameters of the model, which decide for or against cleavage, correspond with the cleavage motifs determined experimentally. The predictive power of the model was verified for unknown (to the program) test conditions: the prediction of cleavage numbers in proteins and the generation of MHC I ligands from short peptides.

In summary, our model proteasomes reproduce and predict proteasomal cleavages with high accuracy. They present a promising approach for predicting proteasomal cleavage products in future attempts and, in combination with existing algorithms for MHC I ligand prediction, will be tested to improve cytotoxic T lymphocyte (CTL) epitope prediction.

### **2.3.2 Introduction**

Proteasomes are cytosolic multisubunit proteases which are involved in cell cycle control, transcription factor activation and the generation of peptide ligands for MHC I molecules (for reviews, see Baumeister et al. (1998), Rock and Goldberg (1999), Uebel and Tampe (1999)). They exist in several forms: The proteolytically active core complexes, or 20S proteasomes, and, when associated with the ATP-dependent 19S cap complexes, the larger 26S proteasomes that are able to recognize proteins marked by ubiquitin for proteasomal degradation (Jentsch and Schlenker, 1995; Hershko and Ciechanover, 1998). Another protein complex known to associate with the 20S core particle is PA28, the 11S regulator (Ahn et al., 1995), which was shown to improve the yield of antigenic



peptides (Groettrup et al., 1996; Dick et al., 1996).

Eukaryotic 20S proteasomes consist of four stacked rings (overall stoichiometry  $\alpha_7\beta_7\beta_7\alpha_7$ ), each consisting of 7 different subunits (Groll et al., 1997). Each of the two inner  $\beta$ -rings carries three catalytically active sites on its inner surface. Their proteolytic specificities have been described as chymotrypsin-like (cleaving after large, hydrophobic AAs), trypsin-like (cleaving after basic AAs) and peptidyl-glutamyl-peptide-hydrolyzing (cleaving after acidic AAs) (for review, see Uebel and Tampe (1999)). Strings of unfolded proteins are thought to be inserted into the cylinder and to be cut into pieces by the active sites; the resulting peptide fragments are then released into the cytosol. Functionally, proteasomal protein degradation is believed to proceed from one substrate end to the other (“processively”), without the release of large degradation intermediates (Akopian et al., 1997; Nussbaum et al., 1998; Kisselev et al., 1999a).

In vertebrate cells, some of the proteolytic fragments produced by the proteasome are fed into the antigen processing machinery. Since peptide presentation by MHC I molecules at the cell surface is an intrinsic requirement for the ability of the immune system to eradicate virus-infected or transformed cells (Rammensee et al., 1993; Pamer and Cresswell, 1998), it is of general interest to know exactly how the proteasome is involved in this process. Proteasomal cleavage specificity has been assessed by *in vitro* digestion experiments using either tri- or tetrapeptides with fluorogenic leaving groups (Kuckelkorn et al., 1995; Heinemeyer et al., 1997; Arendt and Hochstrasser, 1997), peptides of 15-40 AAs (Boes et al., 1994; Niedermann et al., 1995; Niedermann et al., 1996; Dick et al., 1998), or denatured proteins (Dick et al., 1991; Dick et al., 1994; Kisselev et al., 1998, Kisselev et al., 1999a) as substrates. We analyzed the cleavage preferences of yeast wild-type and mutant proteasomes in a non-modified protein (Nussbaum et al., 1998)<sup>3</sup>. Using statistical analysis of cut sites, it was possible for the first time to determine so-called cleavage motifs, i.e. the preferred sequences around cleavage sites, for the three active  $\beta$ -subunits of yeast proteasomes.

In order to apply this cleavage site information to any possible proteasome substrate, one needs an automated prediction device. Such devices already exist for the binding of

---

<sup>3</sup> Mutant yeast proteasomes:  $\beta_5/Pre2^-$ ,  $\beta_2/Pup1^-$ ,  $\beta_1/Pre3^-$  carrying one inactivated  $\beta$ -subunit are denoted here as “single mutants”;  $\beta_2/Pup1^- \beta_1/Pre3^-$  carrying two inactivated  $\beta$ -subunits is denoted here as “double mutant”.

peptides to MHC I molecules (Rammensee et al., 1997) and have been described for peptide transport by the transporter associated with antigen processing (TAP) (Daniel et al., 1998). However, devices for the prediction of proteasomal cleavages are only at the beginning of their development. Recently, published peptide cleavage data were used to develop a prediction algorithm (Holzhütter et al., 1999), which reproduced its training data with 93% and predicted non-training cleavages in one peptide substrate with 80% accuracy. For different AAs in the P1 position<sup>4</sup>, cleavage motifs spanning up to 13 AAs were calculated by the algorithm. However, it has recently been shown that the three different proteolytic activities of eukaryotic proteasomes exhibit overlapping specificities (Dick et al., 1998; Nussbaum et al., 1998). We therefore planned to generate a prediction device that does not rely on various motifs for different P1-AAs, but reflects a combination of the cutting of three active sites. This should mirror the experimental situation more efficiently where observed cleavages arise from a mixture of three different, partly overlapping proteasomal cleavage specificities. More importantly, we wanted our approach to be based on a more homogeneous set of training data, possibly generated under identical conditions. A device for proteasomal cleavage prediction would take us one step further in predicting the selection of CTL epitopes presented on MHC I by the three well-described “funnels of specificity”: proteasome cleavages, TAP transport, MHC I binding.

To this end, we developed a network-based model for proteasome cleavages, trained by an evolutionary algorithm on cleavages in protein and some peptide substrates (Nussbaum et al., 1998; Nussbaum et al., in preparation; Niedermann et al., 1995; Niedermann et al., 1996; Niedermann et al., 1997). Our program performed significantly better for experimental data than for randomly positioned cuts, suggesting that it extracts rules inherent to proteasomal cleavages from training data. Besides, it reproduced the training data with very high accuracy (98-100%). The parameters of the proteasome model largely reflect the roles of particular AAs in the cleavage motifs determined experimentally. Prediction of non-training cleavages was tested for some peptide substrates containing known MHC I ligands. The results indicate our approach to represent a promising starting point for refined algorithms for proteasomal cleavage and

---

<sup>4</sup> The positions relative to a cleavage site (↓) are numbered as described by Schechter and Berger (1967):

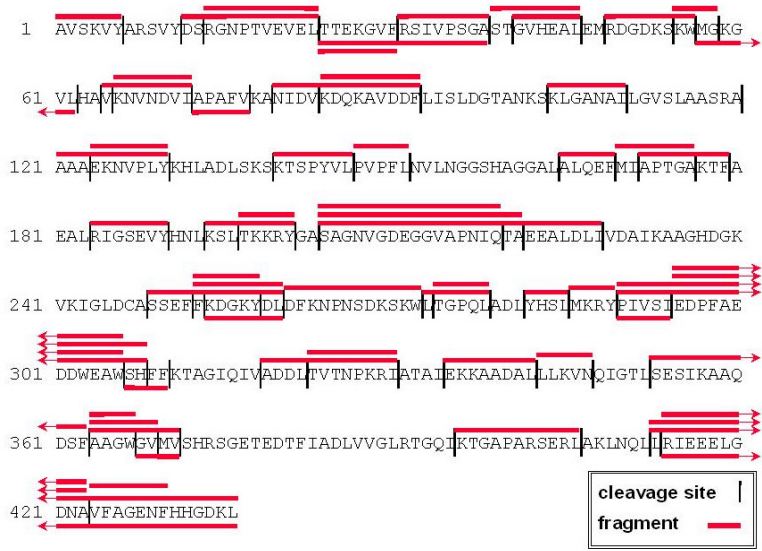
Pk-Pk-1-...-P2-P1 ↓ P1'-P2'-...-Pm-1'-Pm'.

fragment prediction.

### 2.3.3 Materials & Methods/Procedures

#### 2.3.3.1 The experimental training data

Before designing a prediction tool, it was necessary to carefully inspect the experimental training data<sup>5</sup>. In brief, the experimental setup was as follows: purified 20S proteasomes were coincubated with a protein substrate (enolase, 436 AAs), enolase fragments were generated by the proteolytic action of the proteasome, and these fragments were identified by biochemical methods, allowing to locate proteasomal cleavage sites. Thus, the experimental data consist of fragments and cleavage sites. The cleavage sites are not randomly distributed. Hence, the question is: What are the rules for cutting the protein?



**Figure 2.3-1: Typical experimental distribution of cuts and fragments**

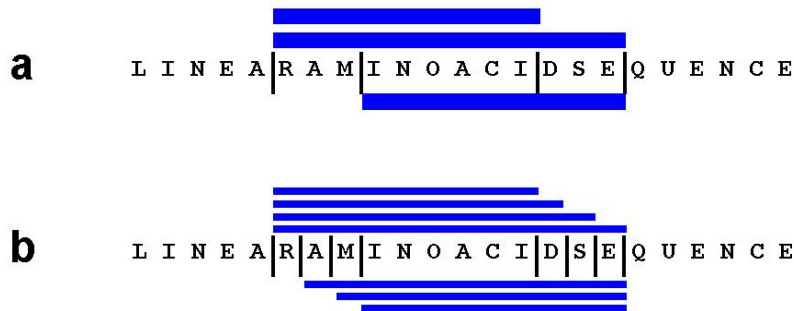
Vertical bars indicate proteasomal cut sites; horizontal bars depict the detected cleavage products. The cut positions represent the training data for the model proteasomes. The example shown here illustrates the fragments produced by the yeast double mutant in enolase (taken from Nussbaum et al. (1998); see 2.2).

In each experiment, a large set ( $10^{10}$ ) of proteasome molecules of a given type (wild-type,

<sup>5</sup> The training data for the program and some working hypotheses used herein are largely taken from Nussbaum et al., 1998 (see 2.2). If not stated otherwise, this source is referred to throughout this section.

single mutant, double mutants) was exposed to enolase molecules, typically 300 enolase molecules to one proteasome. The experiments show overlapping fragments (Figure 2.3-1), which can only be explained by assuming that cutting is a probabilistic event on the level of individual molecules. However, the sample is so large that the probability of a cut occurring at a given position in enolase can be viewed as a deterministic parameter.

Due to the biochemical nature of the detection process, some fragments are more easily detected than others, and there may be undetected fragments and cleavage sites. Stretches of AA with few detected cleavage sites can be explained in two ways: Either very few cleavages are made in this region or many cleavages (fragments) of very low frequency are generated and are hence not detectable (Figure 2.3-2). We thus assume that the experiment provides information on the most frequent cuts.



**Figure 2.3-2: Two ways in which substrate regions with few detected cleavage sites can be explained**

(A) Few cleavages are made, generating few fragments of high abundance. (B) Many cleavages are performed, thus producing many different fragments of low abundance each, possibly below the detection limit.

Experiments using short peptide substrates have already demonstrated that cleavage sites are used with different frequencies (Boes et al., 1994; Niedermann et al., 1995; Dick et al., 1996). Our training data behave similarly, as suggested by the existence of overlapping fragments, but relative frequencies of fragments were not determined. This will affect our modeling approach: There are underlying statistical parameters (cleavage probabilities at given positions) that until now have not been experimentally determined for the training data.

Due to these limitations in fragment detection, of the two fragments meeting at one cleavage site only one may be found. Therefore, the present modeling approach is

directed towards cleavage sites rather than fragments. We did not to consider fragment structure, although overlapping fragments exist. Instead we simulated the observed cut patterns. We shall return to the problem of overlapping fragments below.

### 2.3.3.2 *The model*

A (possible) cleavage always occurs between two positions (two AAs). A “cut after position  $i$ ” denotes a cut between positions  $i$  (e.g. Y190) and  $i+1$  (e.g. H191). With respect to a (possible) cleavage site, these positions are also denoted by  $P_k, \dots, P_1, P_1', \dots, P_m'$ . If a peptide and cut positions are presented to us at random, then the AA in position  $P_i$  ( $P_i$  ranging over  $P_k, \dots, P_1, P_1', \dots, P_m'$ ) is a random variable assuming the values A, C, D, ..., Y, representing the twenty AAs in the one-letter code. Thus,  $X_{3'} = G$  is the third position after the cut when occupied by the AA G.

A simple random cutting device would inspect the positions successively (or simultaneously) and would cut with a certain probability  $P \in (0,1)$ , regardless of which of the possible twenty AAs occupy the positions and independently of any other cuts; the lengths of the fragments would follow a geometric distribution. This is clearly not the case in the experimental situation (Nussbaum et al., 1998; Kisselev et al., 1999a). More sophisticated devices would either A) use distance information, e.g. the distance from the previous cut, or B) neighborhood information, e.g. the AAs adjacent to the prospective cleavage position. Neighborhood information could be limited to the pair  $P_1/P_1'$  or comprise a wider window,  $P_k-P_m$ . For these more sophisticated devices, the proteasome would either A) work on the positions successively and cut with probability  $P_v$ , if  $v$  is the distance from the previous cut, or B) cut at a position with a probability depending on the AAs in  $P_1$  and  $P_1'$ , but not on any previous cuts. In this case  $P(X_1, X_{1'})$  is a measure for the affinity of the adjacent AAs to the active site of the proteasome. In even more complex devices, the probability of cutting would depend on previous cuts plus neighborhood information. Many of these suggested cleavage mechanisms will be rather artificial and will not provide insight because enolase is just one given protein of finite length: The total experimental data on cuts and fragments can be encoded in many different ways in terms of distance and neighborhood information. Thus, one encounters the usual problem of biological modeling: A detailed model with many parameters can reproduce a finite data set exactly but does not provide any information on the underlying biochemical mechanism. Consequently, simple and plausible hypotheses on proteasomal cleaving should be specified. These hypotheses should be minimal (i.e., simpler models cannot explain the observed features) and there should be a nontrivial test for the model.

Therefore we proceeded as follows: First, a hypothesis was formulated for the mechanism. Then a model proteasome was designed in the form of a “neural network” and trained on the experimental data in order to yield an optimal parameter set. The training procedure can also be viewed as an evolutionary algorithm to determine an appropriate model proteasome. Finally the performance of this optimized model proteasome was tested.

As a working model, the following hypothetical mechanism was chosen: The enolase molecule is inserted into the proteasome starting with its N-terminus (for notational convenience, the results do not depend on the direction) and is then threaded along the proteolytic sites. According to the findings that there are no statistically significant effects of positions  $P_i$  or  $P_i'$  with  $i > 6$  or  $i' > 6$  (Nussbaum et al., 1998), it is suggested that the decision for a cleavage is based on local properties of the substrate molecule. Obviously, the  $P_1/P_1'$ -pair plays a prominent role in this process (e.g. combination LT appears four times and is cleaved four times; double mutant). However, if a cutting algorithm is based on these positions alone, the performance will be rather poor (e.g. combination SK appears six times, but is only cleaved three times; double mutant).

In detail, our proteasome model is based on the following hypotheses:

In determining whether to cut or not, the proteasome inspects only a neighborhood  $P_k \dots P_1 \mid P_1' \dots P_m'$  of the prospective cleavage site;  $k$  and  $m$  are fixed small numbers  $\leq 6$ .

The main effect results from the affinity of the pair of AAs in the  $P_1$  and  $P_1'$  positions to the active subunits in the proteasome. This effect is modeled by an affinity parameter  $\alpha_1(X_1, X_{1'})$ . This parameter is not an affinity of the two AAs per se but affinity when exposed to the active sites of proteasome. The value  $-\alpha_1(X_1, X_{1'})$  could be interpreted as an affinity of the pair of AAs to the active sites of the proteasome, i.e., a negative  $\alpha_1$  indicates a positive influence on cleavage.

Each of the positions  $P_i$ ,  $i=2, \dots, k$ , (or  $P_i'$ ,  $i=2, \dots, m$ ,) exerts an affinity  $\alpha_i(X_i)$  ( $\alpha_i(X_{i'})$ ) towards the prospective cut which depends on  $X_i$  ( $X_{i'}$ ) but not on the AAs at the other positions. The affinities can be positive, negative, or zero.

The model is additive: The total affinity at the position considered is

$$\delta = \alpha_1(X_1, X_{1'}) + \sum_{i=2}^k \alpha_i(X_i) + \sum_{i=2}^m \alpha_{i'}(X_{i'}).$$

Notice that the two sums start with  $i=2$ , i.e. there is no additional affinity parameter for

positions P1, P1'.

The probability of a cut depends only on the total affinity  $\delta$ . The cutting process is deterministic: The probability of a cut is equal to 1 if  $\delta$  is negative or zero and it is equal to 0 if  $\delta$  is positive (choosing the threshold zero is a mere normalization). Of course, assumption 5 excludes the occurrence of overlapping fragments (in contrast to the experiments).

### 2.3.3.3 *The learning process*

For a given set of affinities, the model proteasome will work along the enolase and apply the local rule consecutively to each position from  $k$  to  $n-m$ , where  $n$  is the length of the substrate. Thus, following assumption 5, it will always perform certain (the same) cuts in enolase. Then, comparing the results with the experimental data, for each position there are four possibilities: A) There is no cut in the experimental data and the model proteasome did not cut. B) There is a cut in the data and the model did indeed cut. C) There is a cut in the data but the model did not cut (missing cut). D) There is no cut in the data but the model did cut (superfluous cut).

A stochastic hill-climbing algorithm (evolutionary algorithm) was used to train the network, and for this purpose a measure was chosen to "tell" the network how good or bad its present performance was. The choice of the measure was somewhat arbitrary (for the lack of biochemical or mathematical guidelines for judging the performance), but we felt that a superfluous cut is less significant than a missing cut. Hence the performance measure was defined as:

$$F = K \times \text{the number of missing cuts} + \text{the number of superfluous cuts},$$

where typically  $K = 2$ . Choosing  $K > 1$  amounts to penalizing missing cuts. A good performance correlates with a small number  $F$ , and the data are reproduced if  $F=0$  has been achieved. The training algorithm works as follows, starting from an arbitrary or suitably chosen initial affinity configuration and the corresponding number  $F$ , called  $F_0$ .

1. Store the present affinity configuration and compute a new affinity configuration by applying a random perturbation.
2. Compute the value  $F$  for the new affinity configuration and call it  $F_1$ .
3. If  $F_1 \leq F_0$  then store the new affinities and go to 1.
4. If  $F_1 > F_0$  then keep the old affinities and go to 1.

The training process obviously depends on the choice of the random variations. If  $k=5$  and  $m=1$ , for example, then we have  $20 \times 20$  (for the AAs in pair P1P1') +  $4 \times 20$  (for positions P5-P2) = 480 parameter values. Thus, a hill-climbing algorithm is applied to a landscape on a space with dimension 480; this space is very large and there is no way to explore it by a deterministic screening procedure. We decided to vary only one parameter (affinity) at a time (i.e., in the parameter space, each variation is parallel to a coordinate axis). Varying several parameters at a time did not accelerate the optimization process (data not shown). Thus, in the algorithm a parameter is chosen at random, and then a number from the uniform distribution on some interval  $[-\varepsilon, \varepsilon]$  is selected, where  $\varepsilon$  may even depend on the affinity in question. The length  $\varepsilon$ , measuring the variance of the perturbations, has to be carefully selected. If  $\varepsilon$  is too small or too large, the process will take very long to approach a local optimum or will often skip over it, respectively. In the actual program, non-constant variance was adapted to the temporal increase of the goal function  $F$ .

Since the unknown parameters vary over a continuum and the goal function  $F$  assumes entire values, the goal function is constant on certain sets and a unique optimal parameter configuration cannot be expected. In general there is a whole set in the parameter space such that all configurations in this set produce the same result. This observation corresponds to biochemical reality, where slight variations in non-essential parts of the proteasome would probably not lead to malfunction of the enzyme. In principle, there could be parameter configurations in quite distant regions of the parameter space which also produce  $F=0$ .

As with any hill-climbing optimization procedure, good initial guesses do provide assistance. We chose the following approach based on the relative frequencies of AA pairs around cleavage sites: For two AAs  $X_1$  and  $X_{1'}$ , let  $S = S(X_1, X_{1'})$  be the frequency of the pair  $X_1 X_{1'}$  in the substrate and  $C = C(X_1, X_{1'})$  be the frequency of cleaved pairs. If  $S > 0$  then  $\alpha_1(X_1, X_{1'}) = 2(1-2C/S)$ , and otherwise equals zero. Similarly, let  $S_i = S_i(X_i)$  be the frequency of  $X_i$  (up to an error near the terminus, this number is the frequency of  $X_i$  in the substrate) and let  $C_i = C_i(X_i)$  be the frequency of  $X_i$  in position  $P_i$  such that a cut is performed at  $P_1$ . Then the initial guess is  $\alpha_4(X_4) = 2(1-2C_4/S_4)$  and  $\alpha_5(X_5) = 0.5(1-2C_5/S_5)$ . We chose  $\alpha_i = 0$  for all other positions.

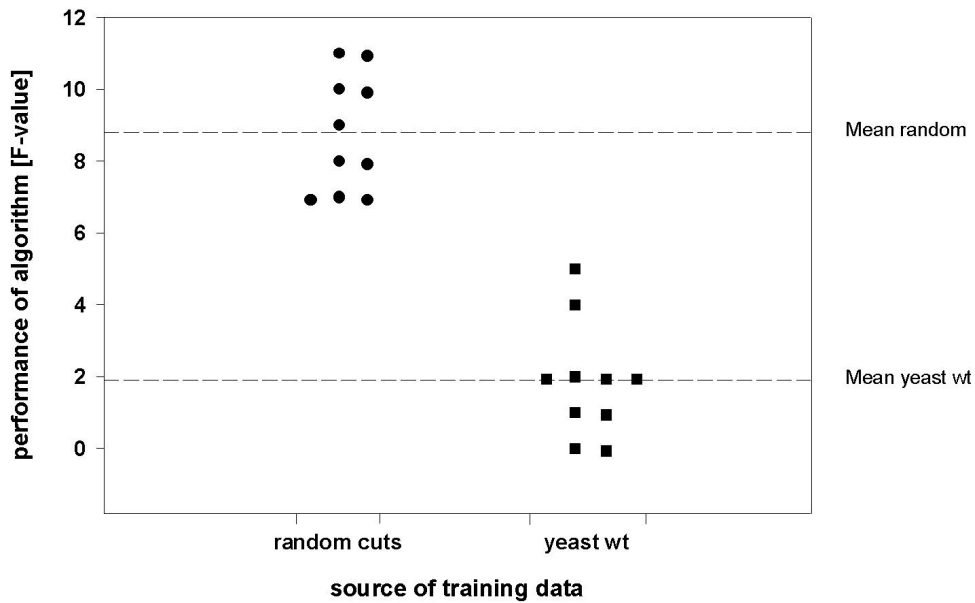
#### 2.3.3.4 Testing the model

As explained earlier, our goal is to obtain a proteasome model rather than a mere



computational device or statistical analysis. There are about  $2.5 \times 10^{10}$  potential octameric peptide sequences of which only about 500 occur in enolase (later, the cutting algorithm is based on such octamers, see next section). If the model proteasome just extracted information and reproduced the cuts, then it should perform similarly on experimental and random data sets. However, if the model proteasome recognized inherent rules in the experimental data, it should perform better on experimental training data.

To test this hypothesis, we compared the performance of the model on artificial, randomly distributed cleavage sites as opposed to cleavages generated by yeast wt 20S proteasomes in enolase, in each case 105 cleavages. Ten independent runs were performed for each and the F value was checked after 200,000 steps. The algorithm reproduced the experimental data significantly better than the random cuts ( $p=0.0002^6$ ): The mean F-value for the experimental cleavages was 1.9 versus 8.8 for the random cuts (Figure 2.3-3).



**Figure 2.3-3: Performance of algorithm when fed with random cleavage data (circles) or yeast wt proteasome cleavages in enolase (squares).**

The training data comprised 105 cleavages, both for the experimental data generated with yeast wt proteasomes and for the randomly selected data. Dashed lines: mean values of ten independent runs. The results shown were reached after a constant number of 200,000 iteration steps.

Interestingly, for the experimental data only superfluous cuts contributed to  $F$ , i.e., the model recovered all training cleavages and made some additional cuts (data not shown). On the other hand, for random cleavage data, the high value of  $F$  resulted from both superfluous and missing cuts (with respect to the training data) although the latter were strongly penalized during the learning process. Similar results were obtained for two more sets of randomly selected cuts (data not shown). Thus, the model performs better on experimental data and for these it recovers all cuts. These findings demonstrate on the one hand that our concept of cleavage affinities leads to a working model, even with relatively few parameters. On the other hand, they show that the experimental cleavage data (proteasome cuts in enolase) contain a large amount of hidden regularity that can be extracted by the present algorithm.

## 2.3.4 Results

### 2.3.4.1 *Reproduction performance*

Our method was applied to experimental cleavage data generated by different 20S proteasomes (yeast wild-type, single mutants, double mutant, human wild-type) in the substrate enolase (Nussbaum et al., 1998; Nussbaum et al., in preparation), yielding different model proteasome “species” ## 1-8 (Table 2.3-1).

In initial attempts, the double mutant data were utilized because of their “cleavage homogeneity”. The decision to cut was based on information from the octameric interval  $P4, \dots, P4'$ , i.e.  $k = m = 4$ . An initial parameter configuration was based on the frequencies of AAs in  $P1$  and  $P1'$  around cleavage sites in the experiment. This rather crude approximation led to a value  $F = 72$  (with  $K = 2$ ), which appears rather weak, but is far from the performance of an arbitrary parameter configuration. After approximately 194.000 steps of the training algorithm, an optimal solution with  $F = 0$  was reached. When the search for a good initial choice was automated as outlined above, the same result was obtained after 176.000 steps. In all subsequent runs, automated initial guesses were used (data not shown).

In an attempt to gradually reduce the number of parameters, both shorter intervals and intervals with “holes” were used. These attempts were made by trial and error, guided by the statistical analysis of cleavage sites. The intervals  $P6, \dots, P1', P4'$  (65.000 steps),

---

<sup>6</sup> The p-value was determined by a two-sided, paired Student's t-test.

P5,...,P1',P4' (71.000) and P4,...,P1',P4' (80.000) indicated that shorter AA windows contain sufficient information for the correct reproduction of the training data, and that P5 seems to play an important role for  $\beta 5/\text{Pre}2$ . Choosing the interval P5,...,P1' reduced the number of steps further to 36.000 (Table 2.3-1), suggesting that positions P2'-P4' contain information that obscure the data and hence decelerate the process. It also shows that the positions N-terminal ("to the left") of the cleavage site seem more important than the ones C-terminal ("to the right") of the cleavage. (It should be emphasized that the number of iteration steps is not entirely relevant for the biological problem or interpretation, since it depends on the quality of the initial guess. These numbers are quoted to give an impression of how a training algorithm behaves in spaces of very high dimensions.) Similarly, for all model proteasomes other than the double mutant, the interval P6,...P1',P4' turned out to be the most suitable for the reproduction of the experimental data. (Note that positions P2', P3' were left out.) Particularly, the data generated with human proteasomes are badly reproduced when P5 and especially P4 are not included in the interval (data not shown). The results of the reproduction performances are summarized in Table 2.3-1.

A substrate molecule of finite length only contains a finite amount of information as far as cleavage data are concerned. A set of affinity parameters reproducing data for one sequence might not be suited for other sequences. A finite set of training data might "teach" the program rules specific only for the training data. The capacity for generalizing the model proteasomes should thus improve if the set of training data is extended. Therefore, for two model proteasomes (#7 and #8), the cleavage data of two short ovalbumin-peptides (Niedermann et al., 1995; Niedermann et al., 1996; Niedermann et al., 1997) were added to the enolase cleavage data generated by human proteasomes (for the reproduction performances see (Table 2.3-1).

The different proteasome models reproduce their training data with nearly perfect scores (98-100%). The majority of the wrongly assigned cuts are superfluous cuts and could theoretically represent cuts that remained undetected in the experiment. The system simulating the yeast double mutant 20S proteasome ( $\beta 5/\text{Pre}2^+$ ) uses the shortest interval (P5,...,P1') to assess the cleavages, probably reflecting the fact that these cleavages are performed by one single proteolytic activity, thus rendering the training data rather homogeneous. For wild-type and single mutant proteasomes the cleavages are generated by three and two different proteolytic activities, respectively. The training data are therefore expected to be less homogeneous, which could be a reason why a wider interval

(P6,...,P1',P4') is needed. This inhomogeneity might be brought about by the need for the active sites  $\beta 1/\text{Pre}3$  and/or  $\beta 2/\text{Pup}1$  to inspect a wider environment of the substrate.

**Table 2.3-1: The reproduction performance of the model proteasome**

No.	proteasome species*	experi- mental cleavages	positions considered	# of steps ‡	performance §			
					correct	not found	super- fluous	% correctly reproduced
#1	yeast wt	105	P6-P1', P4'	116,000	105	none	none	100
#2	yeast $\beta 5/\text{Pre}2^-$	95	P6-P1', P4'	297,000	95	"	"	100
#3	yeast $\beta 2/\text{Pup}1^-$	80	P6-P1', P4'	270,000	80	"	"	100
#4	yeast $\beta 1/\text{Pre}3^-$	88	P6-P1', P4'	500,000	88	"	"	100
#5	yeast $\beta 1/\text{Pre}3^-$ $\beta 2/\text{Pup}1^-$	73	P5-P1'	36,000	73	"	"	100
#6	human wt	117	P6-P1', P4'	275,000	117	"	2	99.5
#7	human wt+	136	P6-P1', P4'	500,000	136	"	4	99.2
#8	human wt+ ¶¶	151	P6-P1', P4'	445,000	150	1	4	98.9

\* 20S species used to generate cleavages in enolase; + additional training data;

‡ number of iteration steps until the reported performance was reached

§ reproduction performance reached after so many steps of the algorithm; the three columns on the left refer to the cleavages made by the algorithm, the %-value in the right column refers to all possible cleavage sites (theoretical number of cleavage sites in enolase:  $427 = 435 - (k+m) + 2$ )

|| learning data: cleavages in ovalbumin peptides 37-77 and Y249-269 [human 20S] (Niedermann et al., 1997) + enolase data [20S from human erythrocytes] (Nussbaum, in preparation; see 2.2.9)

¶¶ learning data: the most dominant cleavages in ovalbumin peptides 37-77 and 239-282 [human and mouse 20S] (Niedermann et al., 1995; Niedermann et al., 1996; Niedermann et al., 1997) + enolase data [20S from human erythrocytes] (Nussbaum, in preparation; see 2.2.9)

In brief, theoretical proteasomes, which do not produce overlapping fragments, were able to reproduce the experimental cleavage data with 98-100% accuracy. A minimal window size of 6-10 AAs, containing 6-8 individual AA positions, provided sufficient information for highly accurate reproduction. Using experimental data from proteasomes bearing more than one active site seems to complicate the cutting rules and therefore requires a wider window for the reproduction, probably reflecting the fact that three different proteolytic specificities operate in eukaryotic proteasomes.

### 2.3.4.2 Computed affinities

Will the affinities of an optimal solution ( $F=0$ ) yield information on the “physical” nature of the cleaving process? As stressed previously, there is no unique optimal affinity parameter configuration. Indeed, for the same problem with the same initial guess, different optimal solutions are found due to the stochastic nature of the search process. Although these optimal solutions vary in detail, some characteristic affinity patterns do emerge. In order to “extract” these characteristics, mean values of affinity parameter configurations were computed, and the results were compared to the experimental findings.

First  $\bar{\alpha}_1$ , a measure for the influence of the P1 position on cleavage, based on the P1,P1' pair affinity in the optimal solution, was introduced:

$$\bar{\alpha}_1(X_1) = \sum_{X_{1'}} \alpha_1(X_1, X_{1'}) \chi(X_1, X_{1'}),$$

where  $\chi(X_1, X_{1'})$  is equal to 1 if the pair  $X_1, X_{1'}$  occurs somewhere in enolase and zero otherwise. The affinity parameters were normalized in such a way that the sum of the absolute values of all twenty AAs was equal to 1,

$$\frac{\bar{\alpha}_1(X_1)}{\sum_{X_1} |\bar{\alpha}_1(X_1)|}.$$

These values were then averaged over the set of the runs of evolutionary algorithm. For the yeast wt proteasome, the results are presented in Table 2.3-2, similarly for positions P1' and P4.

For position P1, large hydrophobic AAs (F, L, W) and the basic AA R were found to promote a cleavage by the program. This is in line with the experimental results (L and R promoting cleavage). A similar agreement was found for the AAs inhibiting a cleavage when appearing in P1: For both the program and the experimental data, G was the most cut-inhibiting AA in P1. For position P1', the program found P, R, Y, and F to be cut-promoting, whereas the large branched AAs V, L, and I, as well as the acidic ones E and D, were cut-inhibiting. This is in agreement with the experiments, where  $\beta$ -turn-promoting AAs (e.g. P) are enriched at P1' and large AAs are rare at P1' (e.g. V, L, I). The cut-promoting influence of R in P1' in the program could reflect the preference of the chymotryptic site Pre2 for basic AAs at this position. The reason for the cut-promoting

influence of the large aromatic AAs Y and F at P1' is currently unclear. As our group, as well as others, (Nussbaum et al., 1998; Shimbara et al., 1998) have already reported on the influence of P at P4, it was interesting to find the cleavage-promoting power of P in P4 in the program. On the other hand, H in P4 seems to strongly inhibit cleavage, a result not observed in the statistical analysis of the experimental data. We are currently examining experimentally whether there is a cut-inhibiting influence of H in P4.

**Table 2.3-2: Cleavage characteristics of yeast wild-type proteasomes deduced from computed affinity parameters**

position	positive effect on cleavage		negative effect on cleavage	
	computed affinities	experimental results*	computed affinities	experimental results*
P1	F, L, W‡, R (I, Y)	L, R	G, K, P, Q, N (A, T, M‡, S)	G
P1'	P, R, Y, F, M‡ (K, T)	small, $\beta$ -turn	V, L, I, E, D (W‡, S, G)	---
P4	P	P	H	---

Strength of AA influence in descending order from left to right. AAs in brackets only show a minor influence. The affinity parameters for the AA cysteine, which only appears once in enolase, are prone to a large variance and were not included in the table.

\* determined by statistical analysis of cleavage data (hypergeometric distribution)

‡ low frequency in enolase (n=5); large variance of computed affinity parameters possible

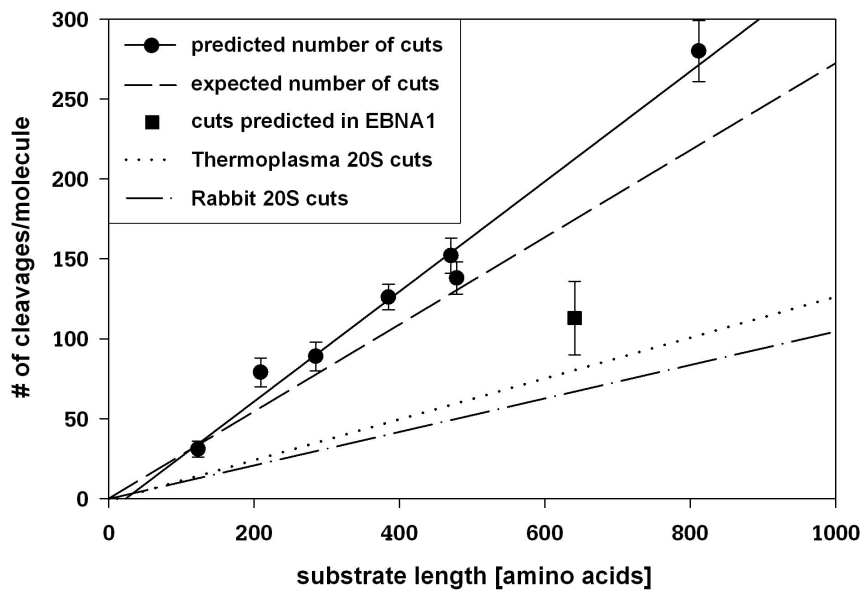
In conclusion, the cleavage motifs of the model proteasome, i.e., the rules the program has learned to reproduce the training data, can be extracted from the computed affinity parameters in order to gain insight into the biochemical mechanism. For the yeast wild-type proteasome, they largely reflect the results obtained from statistical analysis of the experimental data. Some differences between computed affinity parameters and experimental results were detected. Such variances can serve as working hypotheses for experimental investigations.

#### 2.3.4.3 Number of cleavages predicted in proteins

The number of proteasomal cleavages in proteins has been estimated by a calculation using the experimentally determined mean fragment length (Kisselev et al., 1998; Kisselev

et al., 1999a). However, this indirect approach does not account for the presence of overlapping fragments. We found many more cleavages by yeast and human 20S proteasomes in enolase than predicted by the above method (Nussbaum et al., 1998; Nussbaum et al., in preparation; Toes et al., in preparation). We therefore had our program theoretically cleave protein substrates of different lengths (results depicted in Figure 2.3-4). The program trained on cleavages in enolase generated by human 20S proteasomes “cleaved” the test substrates with a frequency of about 33 cuts per 100 AAs in the substrate molecule, 2.5 times more cuts than proposed before (Kisselev et al., 1999a). According to the above mentioned method, 22 cuts are made in casein by mammalian 20S proteasomes (Kisselev et al., 1999a), whereas we found 50-60 cuts by individual fragment detection (Emmerich et al., in preparation). The program predicted about 64 cuts. We therefore conclude that our program mirrors the number of 20S cleavages in protein substrates more realistically than previous methods.

The regression line for the number of cleavages versus protein length does not go through zero. If this effect is relevant at all, then it possibly indicates that short peptides are bad substrates, as reported before (Dolenc et al., 1998). The number of predicted cleavages clearly depends on the AA composition of the test substrate, as suggested by the strong deviation of the number of cleavages predicted in Epstein-Barr-virus nuclear antigen 1 (EBNA1), a viral protein of 641 AAs, from what would be expected (Figure 2.3-4). A glycine-alanine repeat region (GARR) of approximately 240 AAs in EBNA1 is the reason why EBNA1 is not degraded by proteasomes in EBV-infected cells (Levitskaya et al., 1995; Levitskaya et al., 1997; Sharipo et al., 1998). Our model proteasomes on average only perform 7 cleavages in the 240 AAs of this repeat. This can be explained by the influence of the AA G, which is known to be cut-preventing in P1 (Table 2.3-2). Our experimental data agree with this finding: peptides containing short GARR's are hardly ever cleaved within the GARR's (Nussbaum, unpublished observations). Also the deviation of the number of expected cuts from the regression line of predicted cuts could stem from the program using actual sequence information, as opposed to length information, to decide about cuts. Moreover, it could reflect the fact that it is easier to find cleavages in short proteins in the experimental setup (because there are fewer fragments), while the program is not limited by the number of fragments.



**Figure 2.3-4: Number of cleavages predicted in proteins of different lengths.**

The results for the predicted cuts are a mean value of 7 programs (error bars indicate standard deviations). The number of expected cuts (dashed line) was calculated by applying the enolase-specific ratio of cuts per AAs (117 cuts/436 AAs) to other substrates. The number of cuts produced by Thermoplasma 20S (dotted line) and rabbit 20S (dash-dot) are taken from Kisselev et al. (1998) and Kisselev et al. (1999a), respectively. Test substrates: lactalbumin (123 AAs), bovine  $\beta$ -casein (209), influenza A matrixprotein (285), ovalbumin (385), *E.coli* alkaline phosphatase (471), influenza A nucleoprotein (479), and EBNA3 (812). EBNA1 (641) was not included in the regression of the predicted number of cuts.

In summary, based only on cuts made in enolase, our program predicts the number of cleavages in other test substrates more realistically than previous methods, as indicated by the comparison with experimental cleavage numbers in casein. This information is useful for the estimation of the number of potential MHC I binding peptides generated during protein degradation.

#### 2.3.4.4 Predicting the generation of MHC I ligands

The true challenge for a prediction device should be the correct identification of individual cleavages in sequences not included in the training data. For this purpose, a 25mer peptide of pp89 from murine cytomegalovirus (corresponding to positions 162-186) and a 21mer peptide from murine JAK1 tyrosine kinase (positions 348-368) were chosen, because they contain MHC I ligands and are common experimental proteasome substrates. They were cleaved by model proteasomes #6-8 (described in Table 2.3-1) and



the performance was assessed (Table 2.3-3).

**Table 2.3-3: Predictive power of the algorithm in peptides containing MHC I ligands**

data source	cleavage map	performance			
		correct	not found	superfluous	% correctly reproduced
<b>a</b>	lit. 1 R L M Y   5 D   M Y   10 P H F M   15 P T N L   20 G P S E K   25 R V W M S			---	
	# 6 R L M Y D M Y   P H F M   P T N L   G P S E K   R V W M S	3	1	3	75
	# 7 R L M Y D M Y   P H F M   P T N L   G P S E K   R V W M S	4	0	2	87.5
	# 8 R L M Y D M Y   P H F M   P T N L   G   P S E K   R V W M S	4	0	5	68.8
<b>b</b>	lit. 1 R E E W N N   5 F   10 S Y F P E I   15 T H I   20 V I K E S			---	
	# 6 R E E W N N F   S Y F P E I T H I   V I K E S	3	2	2	66.6
	# 7 R E E W N N F   S Y F P E I T H I   V I K E S	2	3	3	50
	# 8 R E E W N N F   S Y F P E I T H I   V I K E S	5	0	3	75

Cleavage performance in peptide sequences (pp89 (a) and JAK1 (b)) containing known MHC I ligands. lit.: extract of experimental findings as described in Boes et al. (1994), Kuckelkorn et al. (1995), Groettrup et al. (1996), Dick et al. (1996), Dick et al. (1998). #6-8: model proteasomes (compare to Table 2.3-1). Short vertical bars indicate cuts; gray coloring indicates experimental cuts in regions close to the termini that the program cannot predict. Long dotted lines indicate the position of experimentally observed cuts. The underlined sequences represent MHC I ligands. Calculations as in Table 2.3-1.

The prediction performance reached values of 75-88% (counting both missing and superfluous cuts, the latter of which we consider less critical). At first glance, it appears contradictory that for the pp89-peptide program #7 performs better than program #8 (Table 2.3-3 A), but vice versa for the JAK1-peptide (Table 2.3-3 B). The training data for program #7 seem to contain information more suitable for the prediction of cleavages in pp89 than those for program #8, and vice versa for the JAK1-peptide. Presumably, adding data to the training set improves the prediction performance. A good example is the cut between L15 and G16 in the pp89-peptide. This cut is not generated by program #6 (trained on cuts in enolase only), probably because the pair LG appears three times in enolase but is never cut. However, the cut is generated by programs #7 and #8 (trained on cuts in enolase and ovalbumin-peptides), which can be explained by the appearance of a cut between L and G in the ovalbumin-peptides. The situation for the JAK1-peptide is similar. Here, positions other than P1 play the predominant role. One could therefore

conclude that the training data do not yet contain enough information to predict every single possible cleavage.

In summary, the predictive power of the program is high for certain short test sequences. Shortcomings of the model proteasomes are likely to be caused by the scarceness of training data and could be influenced by not considering the effects of in vivo regulators of proteasome function, e.g. PA28, the 19S cap or interferon (IFN)- $\gamma$ -inducible subunits. We therefore assume that by including such data and generally by enlarging the set of training data the performance of the program will improve.

#### 2.3.4.5 Refined model

As described above, proteasomes use the cleavage sites with distinct preferences, which can be determined by quantifying cleavage products. However, the cleavage products (fragments) in enolase, the data our model proteasomes are based on, had not been quantified. Nevertheless we wanted to refine the model in such a way that a cut is performed with some probability depending on the flanking AAs. These probabilities were therefore included as parameters of the model. Again, we chose numbers  $k$  and  $m$  to define the neighborhood of the prospective cleavage site. Hypotheses 1-4 were imposed as before, but property 5 was replaced by a more subtle procedure.

In accordance with the additivity requirement we assume that the probability of a cut does not depend on all the particular AAs in the neighborhood (as the affinities do) but only on the total affinity  $\delta$ . Since the variable  $\delta$  ranges over the real numbers, we must specify the probability of a cut in terms of the total affinity  $\delta$ . Thus a function  $P(\delta)$  depending on few parameters was determined by training. It is evident what the properties of  $P(\delta)$  should be:  $P(\delta)$  is a decreasing function,  $P(\delta) \rightarrow 1$  for  $\delta \rightarrow -\infty$ , and  $P(\delta) \rightarrow 0$  for  $\delta \rightarrow +\infty$ . In the previous case of "deterministic cutting" the function was just  $P(\delta)=1$  for  $\delta \leq 0$  and  $P(\delta)=0$  for  $\delta > 0$ . For

an initial test one can choose, e.g.  $P(\delta) = \frac{1}{1 + e^{\delta/\kappa}}$ , which depends on a positive parameter  $\kappa$ . In the limit of  $\kappa \rightarrow 0$  the deterministic cutting can be recovered.

We wanted to choose a simple model to avoid too much arbitrariness when assigning probabilities to cuts. Thus we defined only two cases: A) A high probability cut is a cut not overlapped by any fragment, e.g. the cuts after L183, Y190, L193, Y201 and A203 (Figure 2.3-1). B) A low probability cut is a cut overlapped by some fragment (i.e. not every fragment is cleaved at this site), e.g. the cuts after L196, F252, F253 and Y258 (Figure 2.3-1).

A goal function for measuring the prediction was introduced as follows: If  $\delta > 0$ , the probability of cleaving is so low that cleavages will not be observed. Then there is a critical level  $\delta_1 < 0$ . If  $\delta < \delta_1$ , the probability of cleaving is high. That means fragments containing (i.e. overlapping) the cleavage site in question will not be observed, because this cleavage site is always used. If  $\delta_1 < \delta \leq 0$ , the probability of cleaving is so far from 0 and 1 that fragments resulting from cleaving at this site but also fragments overlapping the site can be found. In view of the homogeneity of our model we can assume that  $\delta_1 = -1$ . An energy function was chosen to estimate the correctness of the prediction:

$$F = 2 \times \# \text{ of missing cuts} + \# \text{ of superfluous cuts} + \# \text{ of wrongly assigned cuts}$$

The refined model was applied to the experimental cleavage data of the yeast double mutant because of their "homogeneous consistency". The data were classified according to the rules described above. First an optimal solution for the mere reproduction of the cleavages was generated using the positions P6,...,P1',P4'. Starting from this point, a parameter set was found which reproduced exactly ( $F = 0$ ) the training data with their assigned probabilities (data not shown). This finding demonstrates that, at least in the case of the double mutant, the network system has sufficient "freedom" to determine, in addition to cleavage positions, probabilities for cuts to occur. Thus there should be sufficient freedom to accommodate information on experimental fragment frequencies in the future. The refined model has not been applied to other experimental data yet.

### 2.3.5 Discussion

Here we present a simple (one-layer) network capable of generating proteasomal cleavages in any given AA sequence. For training the network, goal functions counting missing and superfluous cuts (and cleavage probability in case of overlapping fragments in a refined model) were used together with a stochastic, hill-climbing optimization process. Thus, parameter sets were found which reproduced the training data (cuts performed by 20S proteasomes in enolase plus some cuts in ovalbumin peptides) with almost perfect scores (98-100% exactness, Table 2.3-1). More interestingly, they predicted cleavages in two non-training sequences containing MHC I ligands with an accuracy of 75-87% (Table 2.3-3). Thus, the theoretical proteasomes described here represent the most successful prediction tool for proteasomal cleavages to date.

The precision with which the training data were reproduced (98-100%) and the much poorer reproduction performance when applied to a set of random cuts (Figure 2.3-3) imply that the selected approach is able to extract actual proteasomal cleavage rules

inherent to experimental training data. A window spanning 10 AAs (positions P6-P4') contains sufficient information for this excellent reproduction performance, probably indicating that only a short stretch of about 10 AAs is "seen" by the inner surface of the proteasome at the site of cleavage. This result is in line with the experimental data, where no statistically significant AA effects were observed beyond positions P6 and P6'. It is also in accordance with Holzhütter et al. (1999), who found no significant influences beyond P8 and P6'. Whether there are long-distance effects on cleavage site selection, as suggested by Kisselev et al. (1999b), beyond the intervals checked so far, remains to be determined. The fact that the P-side (AAs P6-P1 needed for the program) is more critical than the P'-side (P1', P4') was detected earlier (Nussbaum et al., 1998; Holzhütter et al., 1999). It could indicate that the P-side of prospective cleavage sites is more strongly bound by the inner surface of the proteasome near its active sites than the P'-side. The theoretical proteasomes relied on parameters similar to the cleavage motifs determined experimentally, as shown by the correspondence of computed affinity parameters with the statistical analysis of the experimental data for yeast wt proteasomes (Table 2.3-2). This is especially true for the preference for large hydrophobic AAs (F, L, W) in P1, the cut-inhibiting influence of G in P1, and the accumulation of P in P4. Thus, the program has not reproduced the training data "by chance", but by applying supposedly proteasome-specific rules. Certain affinity parameters the program uses to decide for or against a cleavage differ from the ones observed experimentally. We cannot exclude that artifacts imposed by the nature of the training algorithm are to blame in such cases. However, in these cases, the influence of the AAs in question can be tested experimentally. An example for this is the cut-inhibiting effect of the AA H at P4, which we are now examining by *in vitro* peptide digestions. Thus, the program can provide working hypotheses to be explored experimentally and thereby possibly accelerate the investigation process.

Other important aspects of our approach are the homogeneity and the size of the set of training data. In a previous approach, published cleavage data from different literature sources were used as training data of an algorithm (Holzhütter et al., 1999). The data encompassed roughly 200 scissile peptide bonds and had been generated under dissimilar experimental conditions (source of 20S, IFN- $\gamma$ -treatment, incubation times). It cannot be excluded that these differences influenced the proteasomal cleavage pattern. We thus based our model largely on cuts generated in one single experiment, forming a set of around 400-500 scissile peptide bonds. The likelihood of experimental variations and statistical artifacts should therefore be low.

The number of cleavages in proteins predicted by the program (Figure 2.3-4) very accurately reflects the occurrence of overlapping fragments, as suggested by the comparison with the number of cuts we found in casein (Emmerich et al., in preparation). Previous attempts to determine cleavage frequencies in protein substrates yielded considerably lower numbers of cleavages, probably due to the fact that they had been determined indirectly (by calculation) instead of directly by fragment identification, thereby not giving full credit to the presence of overlapping fragments (Kisselev et al., 1998; Kisselev et al., 1999a). As the number of possible cleavage sites will determine the number of different peptides generated, our program reflects this point of interest for immunological questions more realistically.

Typically, a limited window containing 8 AAs (positions P6,...,P1',P4'), or altogether about 500 parameters, was sufficient to reproduce the training data. Similar networks usually contain an enormous set of free parameters (Chou et al., 1996; Thompson et al., 1995) which do not reflect the biochemical nature of the process under investigation. Keeping the number of free parameters to a minimum prevents the program from putting too much weight on minor characteristics and makes it easier to extract rules and to interpret their meaning for the biological process.

Our program predicts cleavages on the basis of AA "names", not AA properties such as volume or hydrophobicity, as in Holzhütter et al. (1999). The significance of such properties might hence be underestimated by our approach. However, AAs can be classified according to many additional characteristics such as polarity, tendency to promote  $\beta$ -turns, positive or negative charge, etc. Thus we felt that a sequence of AAs in the one-letter code possibly contained important information beyond selected traits. In order to avoid overlooking this information, our training data were in the form of AA sequences in the one-letter code. This code can be translated into AA characteristics to test for the influence of selected side chain properties (e.g. the proline-effect at P4 is probably caused by the kink-inducing properties of this AA, not by its size or hydrophobicity). In the above mentioned approach, two AA properties were sufficient to yield satisfactory results, suggesting that proteasomes only check a limited amount of AA features to decide for or against a cleavage. However, before this point is resolved, we consider it premature to base an automated prediction device on selected properties only.

The predictive power of the program is not strong for all test data, e.g. the cuts at the C-termini of MHC I ligands (data not shown). It is possible that test data consisting of selected MHC I ligands do not fit our program because the involvement of trimming

proteases different from proteasomes in the generation of the C-termini of MHC I ligands cannot be excluded (Elliott et al., 1995). However, from examining falsely predicted cuts in different types of test sequences, it is more likely that the lack of a larger set of training data is responsible (data not shown). Many “cases”, i.e. AA sequences, do not exist in the training data. For example, only 214 AA pairs, i.e. 54% of all possible 400 (theoretical P1,P1'-) pairs, occur in enolase. Therefore, in judging a prospective cut site, the program will base its decision partly on chance. It should consequently be possible to improve the predictive power of the program by increasing the set of training data. For this purpose we are currently generating more cleavage data in new protein substrates, using the same conditions as for the enolase data in order to guarantee the homogeneity of the training data.

Studies on in vivo antigen processing indicate that, for certain CTL-epitopes, proteasomes only participate in the generation of their C-termini and that their N-termini can be trimmed to the correct size by cytosolic or ER-resident aminopeptidases (Craiu et al., 1997; Stoltze et al., 1998; Paz et al., 1999; Mo et al., 1999). However, a recent report claims that proteasomes do generate both termini of CTL-epitopes (Lucchiari-Hartz et al., 2000). It is therefore possible that antigen processing contains an epitope-specific component and that the prediction of proteasomal cleavages will turn out to be valuable only for the C-terminus of certain CTL-epitopes.

At present the program is only able to predict cleavage sites, not individual fragments, and does not “know” about overlapping fragments. However, fragment prediction will be more useful for the application to immunological questions than just cut prediction. Also, the program only generates deterministic cuts (in the basic model) or cuts based on estimated probabilities (in the refined model) at the moment. Both shortcomings should be overcome by using quantified cleavage fragments as training data. Quantitation will provide the network with probabilities for cleavages and for fragments to occur, a necessity for probabilistic cut and fragment prediction. As predicted cuts and fragments will be coupled to defined probabilities for them to occur, the presence of overlapping fragments will no longer pose a problem. The new cleavage data currently being generated are therefore quantified on a routine basis.

In summary, we have created automated prediction devices for cuts by yeast and human 20S proteasomes. They reproduce their training data nearly perfectly and show strong predictive power in selected test substrates. The program provides information on the functioning of “real” proteasomes, which mirror experimental findings (e.g. number of

cleavage sites in proteins, cleavage motifs) better than previous approaches. Shortcomings in the prediction of individual cuts are likely to be caused by the limited amount of training data. However, as network-based prediction programs can continuously be fed with new training data, the predictive power of the program should improve as new experimental data are incorporated. The program therefore represents a suitable format for the development of fragment prediction programs. It can be tested in the future in combination with existing programs for the prediction of MHC I ligands to improve the forecast of CTL-epitopes.

### 2.3.6 References

- Ahn, J. Y., Tanahashi, N., Akiyama, K., Hisamatsu, H., Noda, C., Tanaka, K., Chung, C. H., Shimbara, N., Willy, P. J., Mott, J. D., Slaughter, C. A. and DeMartino, G. N. (1995). Primary structures of two homologous subunits of PA28, a gamma- interferon-inducible protein activator of the 20S proteasome. *FEBS Lett.* 366, 37-42.
- Akopian, T. N., Kisselev, A. F. and Goldberg, A. L. (1997). Processive degradation of proteins and other catalytic properties of the proteasome from *Thermoplasma acidophilum*. *J. Biol. Chem.* 272, 1791-1798.
- Arendt, C. S. and Hochstrasser, M. (1997). Identification of the yeast 20S proteasome catalytic centers and subunit interactions required for active-site formation. *Proc. Natl. Acad. Sci. USA*, 94, 7156-7161.
- Baumeister, W., Walz, J., Zuhl, F. and Seemüller, E. (1998). The proteasome: paradigm of a self-compartmentalizing protease. *Cell* 92, 367-380.
- Boes, B., Hengel, H., Ruppert, T., Multhaup, G., Koszinowski, U. H. and Kloetzel, P. M. (1994). Interferon gamma stimulation modulates the proteolytic activity and cleavage site preference of 20S mouse proteasomes. *J. Exp. Med.* 179, 901-909.
- Chou, K. C., Tomasselli, A. G., Reardon, I. M. and Henrikson, R. L. (1996). Predicting human immunodeficiency virus protease cleavage sites in proteins by a discriminant function method. *Proteins* 24, 51-72.
- Craiu, A., Akopian, T., Goldberg, A. and Rock, K. L. (1997). Two distinct proteolytic processes in the generation of a major histocompatibility complex class I-presented peptide. *Proc. Natl. Acad. Sci. USA* 94, 10850-10855.
- Daniel, S., Brusica, V., Caillat-Zucman, S., Petrovsky, N., Harrison, L., Riganelli, D., Sinigaglia, F.,

Gallazzi, F., Hammer, J. and van Endert, P. M. (1998). Relationship between peptide selectivities of human transporters associated with antigen processing and HLA class I molecules. *J. Immunol.* 161, 617-624.

Dick, L. R., Moomaw, C. R., DeMartino, G. N. and Slaughter, C. A. (1991). Degradation of oxidized insulin B chain by the multiproteinase complex macropain (proteasome). *Biochemistry* 30, 2725-2734.

Dick, L. R., Aldrich, C., Jameson, S. C., Moomaw, C. R., Pramanik, B. C., Doyle, C. K., DeMartino, G. N., Bevan, M. J., Forman, J. M. and Slaughter, C. A. (1994). Proteolytic processing of ovalbumin and beta-galactosidase by the proteasome to a yield antigenic peptides. *J. Immunol.* 152, 3884-3894.

Dick, T. P., Ruppert, T., Groettrup, M., Kloetzel, P. M., Kuehn, L., Koszinowski, U., Stevanovic, S., Schild, H. and Rammensee, H. G. (1996). Coordinated dual cleavages induced by the proteasome regulator PA28 lead to dominant MHC ligands. *Cell* 86, 253-262.

Dick, T. P., Nussbaum, A. K., Deeg, M., Heinemeyer, W., Groll, M., Schirle, M., Keilholz, W., Stevanovic, S., Wolf, D. H., Huber, R., Rammensee, H. G. and Schild, H. (1998). Contribution of proteasomal beta-subunits to the cleavage of peptide substrates analyzed with yeast mutants. *J. Biol. Chem.* 273, 25637-25646.

Dolenc, I., Seemüller, E. and Baumeister, W. (1998). Decelerated degradation of short peptides by the 20S proteasome. *FEBS Lett.* 434, 357-361.

Elliott, T., Willis, A., Cerundolo, V. and Townsend, A. (1995). Processing of major histocompatibility class I-restricted antigens in the endoplasmic reticulum. *J. Exp. Med.* 181, 1481-1491.

Groettrup, M., Soza, A., Eggers, M., Kuehn, L., Dick, T. P., Schild, H., Rammensee, H. G., Koszinowski, U. H. and Kloetzel, P. M. (1996). A role for the proteasome regulator PA28alpha in antigen presentation. *Nature* 381, 166-168.

Groll, M., Ditzel, L., Lowe, J., Stock, D., Bochtler, M., Bartunik, H. D. and Huber, R. (1997). Structure of 20S proteasome from yeast at 2.4 Å resolution. *Nature* 386, 463-471.

Heinemeyer, W., Fischer, M., Krimmer, T., Stachon, U. and Wolf, D. H. (1997). The active sites of the eukaryotic 20 S proteasome and their involvement in subunit precursor processing. *J. Biol. Chem.* 272, 25200-25209.

Hershko, A. and Ciechanover, A. (1998). The ubiquitin system. *Annu. Rev. Biochem.* 67, 425-479.



Holzhütter, H. G., Frommel, C. and Kloetzel, P. M. (1999). A theoretical approach towards the identification of cleavage- determining amino acid motifs of the 20 S proteasome. *J. Mol. Biol.* 286, 1251-1265.

Jentsch, S. and Schlenker, S. (1995). Selective protein degradation: a journey's end within the proteasome. *Cell* 82, 881-884.

Kisselev, A. F., Akopian, T. N. and Goldberg, A. L. (1998). Range of sizes of peptide products generated during degradation of different proteins by archaeal proteasomes. *J. Biol. Chem.* 273, 1982-1989.

Kisselev, A. F., Akopian, T. N., Woo, K. M. and Goldberg, A. L. (1999a). The sizes of peptides generated from protein by mammalian 26 and 20 S proteasomes. Implications for understanding the degradative mechanism and antigen presentation. *J. Biol. Chem.* 274, 3363-3371.

Kisselev, A.F., Akopian, T.N., Castillo, V. and Goldberg, A.L. (1999b). Proteasome active sites allosterically regulate each other, suggesting a cyclical bite-chew mechanism for protein breakdown. *Mol. Cell* 4, 395-402.

Kuckelkorn, U., Frentzel, S., Kraft, R., Kostka, S., Groettrup, M. and Kloetzel, P. (1995). Incorporation of major histocompatibility complex--encoded subunits LMP2 and LMP7 changes the quality of the 20S proteasome polypeptide processing products independent of interferon-gamma. *Eur. J. Immunol.* 25, 2605-2611.

Levitskaya, J., Coram, M., Levitsky, V., Imreh, S., Steigerwald-Mullen, P. M., Klein, G., Kurilla, M. G. and Masucci, M. G. (1995). Inhibition of antigen processing by the internal repeat region of the Epstein-Barr virus nuclear antigen-1. *Nature* 375, 685-688.

Levitskaya, J., Sharipo, A., Leonchiks, A., Ciechanover, A. and Masucci, M. G. (1997). Inhibition of ubiquitin/proteasome-dependent protein degradation by the Gly-Ala repeat domain of the Epstein-Barr virus nuclear antigen 1. *Proc. Natl. Acad. Sci. USA* 94, 12616-12621.

Lucchiari-Hartz, M., van Endert, P.M., Lauvau, G., Maier, R., Meyerhans, A., Mann, D., Eichmann, K. and Niedermann, G. (2000). Cytotoxic T lymphocyte epitopes of HIV-1 nef. Generation Of multiple definitive major histocompatibility complex class I ligands by proteasomes. *J. Exp. Med.* 191, 239-252.

Mo, X.Y., Cascio, P., Lemerise, K., Goldberg, A.L. and Rock, K. (1999). Distinct proteolytic processes generate the C and N termini of MHC class I-binding peptides. *J. Immunol.* 163, 5851-5859.

Niedermann, G., Butz, S., Ihlenfeldt, H. G., Grimm, R., Lucchiari, M., Hoschutzky, H., Jung, G., Maier, B. and Eichmann, K. (1995). Contribution of proteasome-mediated proteolysis to the hierarchy of epitopes presented by major histocompatibility complex class I molecules. *Immunity* 2, 289-299.

Niedermann, G., King, G., Butz, S., Birsner, U., Grimm, R., Shabanowitz, J., Hunt, D. F. and Eichmann, K. (1996). The proteolytic fragments generated by vertebrate proteasomes: structural relationships to major histocompatibility complex class I binding peptides. *Proc. Natl. Acad. Sci. USA* 93, 8572-8577.

Niedermann, G., Grimm, R., Geier, E., Maurer, M., Realine, C., Gartmann, C., Soll, J., Omura, S., Rechsteiner, M. C., Baumeister, W. and Eichmann, K. (1997). Potential immunocompetence of proteolytic fragments produced by proteasomes before evolution of the vertebrate immune system. *J. Exp. Med.* 186, 209-220.

Nussbaum, A. K., Dick, T. P., Keilholz, W., Schirle, M., Stevanovic, S., Dietz, K., Heinemeyer, W., Groll, M., Wolf, D. H., Huber, R., Rammensee, H. G. and Schild, H. (1998). Cleavage motifs of the yeast 20S proteasome beta subunits deduced from digests of enolase 1. *Proc. Natl. Acad. Sci. USA* 95, 12504-12509.

Pamer, E. and Cresswell, P. (1998). Mechanisms of MHC class I-restricted antigen processing. *Annu. Rev. Immunol.* 16, 323-358.

Paz, P., Brouwenstijn, N., Perry, R. and Shastri, N. (1999). Discrete proteolytic intermediates in the MHC class I antigen processing pathway and MHC I-dependent peptide trimming in the ER. *Immunity* 11, 241-251.

Rammensee, H. G., Falk, K. and Rotzschke, O. (1993). Peptides naturally presented by MHC class I molecules. *Annu. Rev. Immunol.* 11, 213-244.

Rammensee, H. G., Bachmann, J. and Stevanovic, S. (1997). MHC ligands and peptide motifs. Landes Bioscience, Austin, Texas, USA.

Rock, K. L. and Goldberg, A. L. (1999). Degradation of cell proteins and the generation of MHC class I-presented peptides. *Annu. Rev. Immunol.* 17, 739-779.

Schechter, I. and Berger, A. (1967). On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* 27, 157-162.

Sharipo, A., Imreh, M., Leonchiks, A., Imreh, S. and Masucci, M. G. (1998). A minimal glycine-

alanine repeat prevents the interaction of ubiquitinated I kappaB alpha with the proteasome: a new mechanism for selective inhibition of proteolysis. *Nat. Med.* 4, 939-944.

Shimbara, N., Ogawa, K., Hidaka, Y., Nakajima, H., Yamasaki, N., Niwa, S., Tanahashi, N. and Tanaka, K. (1998). Contribution of proline residue for efficient production of MHC class I ligands by proteasomes. *J. Biol. Chem.* 273, 23062-23071.

Stoltze, L., Dick, T. P., Deeg, M., Pommerl, B., Rammensee, H. G. and Schild, H. (1998). Generation of the vesicular stomatitis virus nucleoprotein cytotoxic T lymphocyte epitope requires proteasome-dependent and -independent proteolytic activities. *Eur. J. Immunol.* 28, 4029-4036.

Thompson, T. B., Chou, K. C. and Zheng, C. (1995). Neural network prediction of the HIV-1 protease cleavage sites. *J. Theor. Biol.* 177, 369-379.

Uebel, S. and Tampe, R. (1999). Specificity of the proteasome and the TAP transporter. *Curr. Opin. Immunol.* 11, 203-208.

### 2.3.7 Abbreviations

AA(s)	amino acid(s)
CTL	cytotoxic T-lymphocyte
EBNA1	Epstein-Barr-virus nuclear antigen 1
GARR	glycine-alanine repeat region
IFN	interferon
MHC I	major histocompatibility complex class I

### 2.3.8 Participating researchers

Christina Kuttler<sup>\*1</sup>, Tobias P. Dick<sup>3</sup>, Hans-Georg Rammensee<sup>2</sup>, Hansjörg Schild<sup>2</sup>, and Karl-Peter Haderl<sup>1</sup>.

<sup>1</sup> Biomathematik, University of Tübingen, Auf der Morgenstelle 10, D-72076 Tübingen, Germany.

<sup>2</sup> Institute for Cell Biology, Department of Immunology, University of Tübingen, Auf der Morgenstelle 15, D-72076 Tübingen, Germany.

<sup>3</sup> present address: Section of Immunobiology, Howard Hughes Medical Institute, Yale University School of Medicine, 310 Cedar Street, PO Box 208011, New Haven, CT 06520-8011, USA.

\* A. K. Nussbaum and C. Kuttler contributed equally to the results presented in this part.

## 2.4 PProC: A Prediction Algorithm for Proteasomal Cleavages available on the WWW

### 2.4.1 Summary

The first version of PProC (Prediction Algorithm for Proteasomal Cleavages) is now available to the general public. PProC is a prediction tool for cleavages by human and yeast proteasomes, based on experimental cleavage data. It will be particularly useful for immunologists working on antigen processing and the prediction of major histocompatibility class I molecule (MHC I) ligands and cytotoxic T-lymphocyte (CTL) epitopes. Likewise, in cases in which proteasomal protein degradation has been indicated in disease, PProC can be used to assess the general cleavability of disease-linked proteins. On its web site (<http://www.paproc.de>), background information and hyperlinks are provided for the user (e.g. to SYFPEITHI, the database for the prediction of MHC class I ligands).

### 2.4.2 Introduction

Proteasomes are the key players in cytosolic protein degradation. Some of the “waste” products of this process, peptides of 3-40 amino acids in length, are spared from further degradation to single amino acids and are translocated by the transporter associated with antigen processing (TAP) into the endoplasmic reticulum (ER) (Rock and Goldberg, 1999; Pamer and Cresswell, 1998; Uebel and Tampe, 1999; van Endert, 1999). Whereas the C-termini of MHC I ligands most likely have to be generated by proteasomes (Craiu et al., 1997; Stoltze et al., 1998; Mo et al., 1999), cytosolic or ER-resident trimpeptidases can reduce longer fragments to the correct size of 8-10 aa for them to bind to MHC I and be presented on the cell surface (Paz et al., 1999; Stoltze et al., 2000). The recognition of MHC I-bound peptides is responsible for the induction of cytotoxic T-lymphocytes (CTL) against intracellular pathogens, such as viruses, and against mutated proteins, e.g. in tumor cells. Thus, it is conceivable that proteasomal cleavage specificity regulates immune responses, albeit indirectly. It would therefore be beneficial to be able to predict proteasomal cleavages. In combination with MHC I ligand predictors such as SYFPEITHI (Rammensee et al., 1999), in particular, proteasomal cleavage prediction should help to improve the forecast of MHC class I restricted CTL epitopes. More specifically, it could support researchers seeking CTL epitopes by limiting the number of predicted MHC class I ligands from protein antigens. In addition, the effect of amino acid mutations in viral or

tumor-specific proteins on antigen presentation could also be assessed. Proteasomal cleavage prediction can thus play an important part in rational vaccine design.

It is well-established that altered proteasomal protein degradation can be a factor in disease processes. A change in the general cleavability of a protein could be one possible mechanism by which disease-linked proteins are spared from degradation, thus increasing their intracellular levels and possibly killing cells. EBNA-1, a viral protein from EBV is spared from proteasomal degradation due to a Gly-Ala repeat region (GARR) (Levitskaya et al., 1995; Levitskaya et al., 1997). The length of the GARR is critical for this effect (Dantuma et al., 2000). Our proteasomal cleavage algorithm predicts hardly any cleavages within the GARR (Kuttler et al., 2000), and short GARRs are not cleaved by proteasomes *in vitro* (Nussbaum, unpublished observations). Poly-glutamine (polyQ) – containing proteins are involved in neurodegenerative diseases such as Huntington's disease (HD) and spinocerebellar ataxias (SCA). Conformational changes in mutant forms of polyQ-proteins possibly result in altered proteasomal degradation and lead to the formation of protein deposits (Wanker, 2000; Evert et al., 2000; Zoghbi and Orr, 2000; Cummings et al., 1998). Again, longer polyQ-sequences more readily cause disease (Krobitsch and Lindquist, 2000), possibly by more efficiently inhibiting proteasomal degradation. For disease-linked proteins such as the polyQ-proteins, PAProC can predict whether proteasomal cleavage specificity might be involved in protein accumulation and the initiation of disease.

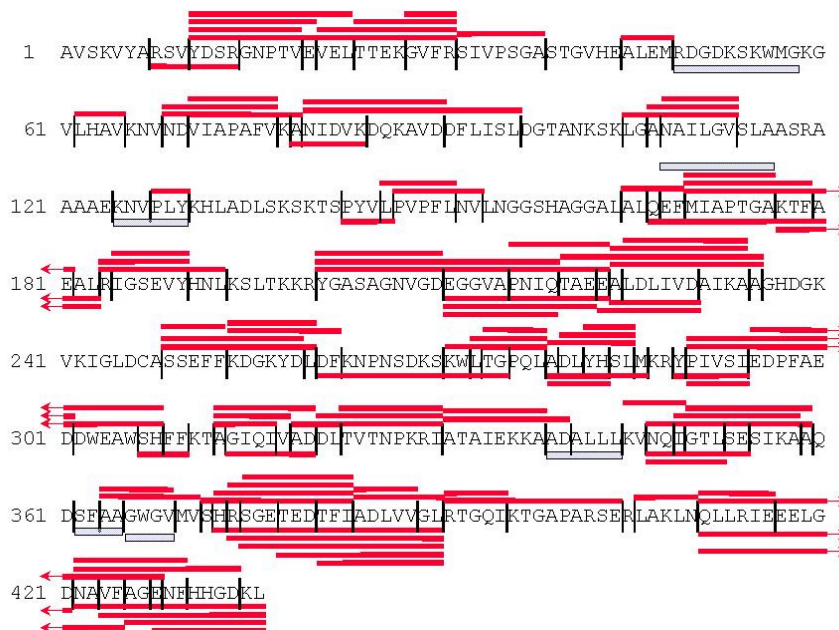
Recently, the first prediction algorithms for proteasomal cleavages were published on paper (Holzhütter et al., 1999; Kuttler et al., 2000), and a third, neural network-based approach, has just been completed (Kesmir et al, submitted). In addition, several other groups are currently working on this intriguing topic.

We now introduce the first publicly available tool for proteasomal cleavage prediction, PAProC, and present its mode of function. PAProC is a prediction tool for cleavages by human as well as wild-type (wt) and mutant yeast proteasomes that is based on experimental cleavage data (Nussbaum et al., 1998; Kuttler et al., 2000). On the PAProC-web site (<http://www.paproc.de>), easy-to-handle predictions of proteasomal cleavages in amino acid sequences can be performed. In addition, background information on proteasomes and on PAProC and hyperlinks (e.g. to SYFPEITHI, the database for the prediction of MHC class I ligands) are available.

## 2.4.3 Materials & Methods

### 2.4.3.1 Experimental training data and algorithm

The experimental training data for PAProC have already been described in detail (Nussbaum et al., 1998; Kuttler et al., 2000). In brief, a non-modified protein (yeast enolase-1) was incubated with purified yeast or human erythrocyte 20S proteasomes, resulting in the proteasomal fragmentation of enolase. Enolase fragments were identified by biochemical methods (fractionation, sequencing, mass spectrometry) and led to the compilation of a cleavage map (see Figure 2.4-1). All the detected cleavage sites were then used to train an algorithm, which has recently been described in detail (Kuttler et al. 2000). The algorithm “sees” a window of 10 amino acids around cleavage sites (6 to the “left”, 4 to the “right”). Based on this amino acid window, the algorithm can reproduce its training data and, more importantly, is able to predict proteasomal cleavages in sequences not included in the training data. All the PAProC-predictions mentioned in this paper are based on algorithms trained with human 20S proteasome data.



**Figure 2.4-1: Cleavage map of the degradation of yeast enolase-1 by human erythrocyte proteasomes**

The figure shows the 107 full fragments (*horizontal bars*) and the 118 cleavage sites (*vertical lines*) identified in digests of enolase by human erythrocyte 20S proteasomes. (*Solid bars*) full fragments of which both N- and C-termini are known. (*Open bars*) fragments for which only the N-termini were identified. See 2.2.9 for details.

### 2.4.3.2 *Server requirements and programming language*

PAProC runs on any common internet browser (successfully tested for Netscape versions 3.0, 4.05, 4.5, 4.7, Microsoft Internet Explorer 5.0 and Lynx). The PAProC algorithm was written in PERL 5, a standard programming language that guarantees functioning of the program independent of the users' computer systems.

### 2.4.3.3 *Comparison of MHC I binding prediction and proteasomal cleavage prediction of CTL epitopes*

The SYFPEITHI database (<http://www.syfpeithi.de>) (Rammensee et al., 1999) was used to predict MHC I ligands from the tumor associated proteins HER-2/neu, pmel/gp100, p53 and the influenza (strain A/Japan/305/57) protein hemagglutinin (HA). We then checked in PAProC if any proteasomal cleavages were predicted to generate the exact C-termini of the predicted MHC I ligands. This was carried out for all the predicted ligands down to the SYFPEITHI prediction score of the known CTL epitope with the weakest predicted binding ability (at the most 75 ligands), or for at least the 20 best predicted binders. Example: For HER-2/neu, two decameric HLA-A\*0201-restricted CTL epitopes are known. Their predicted MHC I binding scores are 24 and 20, respectively. We therefore checked all the predicted binders, from the highest score (in this case: score 29) down to the score for the known CTL epitope with the weakest predicted binding ability (here: score 20), for C-terminal cleavage in PAProC and came up with 38 peptides. For influenza HA, where all known CTL epitopes fell into the first few predicted ligands, we checked the predicted cleavages for the 20 MHC I ligands with the highest predicted binding scores.


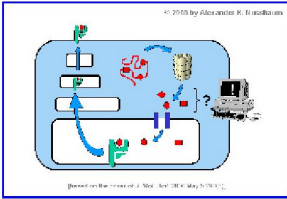
## 2.4.4 Results

### 2.4.4.1 *Sequence submission*

We have recently described the development of an algorithm for the prediction of proteasomal cleavages (Kuttler et al., 2000) and have now incorporated it into a publicly available web site, the Prediction Algorithm for Proteasomal Cleavages (PAProC), available at <http://www.paproc.de>. The homepage of PAProC is also the submission page for amino acid sequences (Figure 2.4-2). From this page, information on the functions of PAProC and on proteasomes can be accessed.

Amino acid sequences must be entered in the one-letter code, e.g. as FASTA- or original SWISS-PROT-format. With the exception of letters coding for aa, spaces, numbers and carriage returns, PAProC will not tolerate additional characters in the sequence. A

sequence name can be entered that will appear on the result output form for user information only. A proteasome species (human or yeast) must be selected before submission. For the human proteasome, it is possible to choose between three predictors based on slightly different training data (type I: based only on cleavages in enolase; types II + III: based on cleavages both in enolase and in different ovalbumin peptides). For the yeast proteasome, it is possible to choose between predictors trained with wt or different mutant proteasome data. Please refer to (Kuttler et al., 2000), for detailed information. The default setting for the proteasome species is human proteasome type III.

## Prediction Algorithm for Proteasomal Cleavages

[User information](#) on PAPROC functions  
[Background information](#) on proteasomes and PAPROC

Name of the Amino acid sequence:

Type in your [amino acid sequence](#) (in 1-letter-code)

Which [proteasome species](#) should be used ?

<p>Human proteasome</p> <p><input type="radio"/> wild type I</p> <p><input type="radio"/> wild type II</p> <p><input checked="" type="radio"/> wild type III</p>	<p>Yeast proteasome</p> <p><input type="radio"/> wild type</p> <p><input type="radio"/> beta5/Pre2- single mutant</p> <p><input type="radio"/> beta1/Pre3- single mutant</p> <p><input type="radio"/> beta2/Pup1- single mutant</p> <p><input type="radio"/> beta2/Pup1- beta1/Pre3- double mutant</p>
--	--

Please choose the [output style](#) :

short form  long form

**Figure 2.4-2: Entry and submission web site of PAPROC**

The entry and submission web site of PAPROC displays spaces for sequence entry and click buttons for prediction and result output options. It additionally provides links to background information for the first-time user.

As PAPROC needs a window of at least 6 aa N-terminal (“to the left”) and 4 aa C-terminal (“to the right”) of the cleavage site of interest, the minimum length for submitted sequences is 10 amino acids. Within a 10 aa sequence, PAPROC can only predict the cleavage between positions 6 and 7.

In order to fill in blanks in submitted sequences or to prolong short sequences to reach the



minimal length of 10 aa, the letter “X” (or “x”) can be used as a wildcard. Please note that “X” has a neutral (zero) effect on the cleavage decision by PAProC.

#### 2.4.4.2 Results output



## Results

### Short output form

PAProC predicts the following (9) proteasomal cleavages (made by **human proteasome type III**) in **pp89** (25 amino acids):

1	RLMYDMY   PHF   M   P   TN   L   G   PSE   K
21	RVVMS

PAProC predicts the following (60) proteasomal cleavages (made by **human proteasome type III**) in **presenilin\_1** (175 amino acids):

1	VVLVNM   AEG   DPEAQR   RV   SKN
21	SKYNAE   STE   RE   S   QD   TVAE   ND
41	DGG   F   SE   E   WEA   QRDS   HLG   P   HR
61	ST   PES   RAA   VQ   EL   SS   SI   LAGE
81	DPEER   GV   KLG   LGDFI   FY   S   VL
101	VGKASATA   SG   DU   NTTIA   C   F   V
121	AILIG   LC   L   TLLLLAI   FFK   A   L
141	PAL   PISI   TF   G   LVF   Y   F   AT   DY   L
161	V   Q   PFMDQL   AF   H   QFYI

### Long output form

The program predicts the following (9) proteasomal cleavages (made by **human proteasome type III**) in **pp\_89** (25 amino acids):

Position	Amino acid	Cleavage prediction
1	R	non-cleavable area
2	L	non-cleavable area
3	M	non-cleavable area
4	Y	non-cleavable area
5	D	non-cleavable area
6	M	-
7	Y	++
8	P	-
9	H	-
10	F	+++
11	M	+++
12	P	+
13	T	-
14	N	++
15	L	+++
16	G	+
17	P	-
18	S	-
19	E	+
20	K	+
21	R	-
22	V	non-cleavable area
23	W	non-cleavable area
24	M	non-cleavable area
25	S	non-cleavable area

- : no cleavage behind this position, +, ++, +++ : cleavage behind this position, with a hint for its strength (number of "+")  
Non-cleavable area: the flanking window of this position reaches to the boundary of the sequence

### Figure 2.4-3: Two formats for result output

There are two output formats for results: Short form (A, C) and long form (B). The short format is more graphical and displays predicted cleavages as red, vertical bars, inserted into the submitted sequence. The long format gives in a tabular form detailed information on cleavage positions and approximate cleavage strength. Examples: (A, B) cleavages predicted in a CTL epitope-containing peptide from murine cytomegalovirus pp89 IE-protein; (C) cleavages predicted in the C-terminal 175 aa of Presenilin-1, a protein whose mutated forms are involved in Alzheimer's disease (Checler et al., 2000).

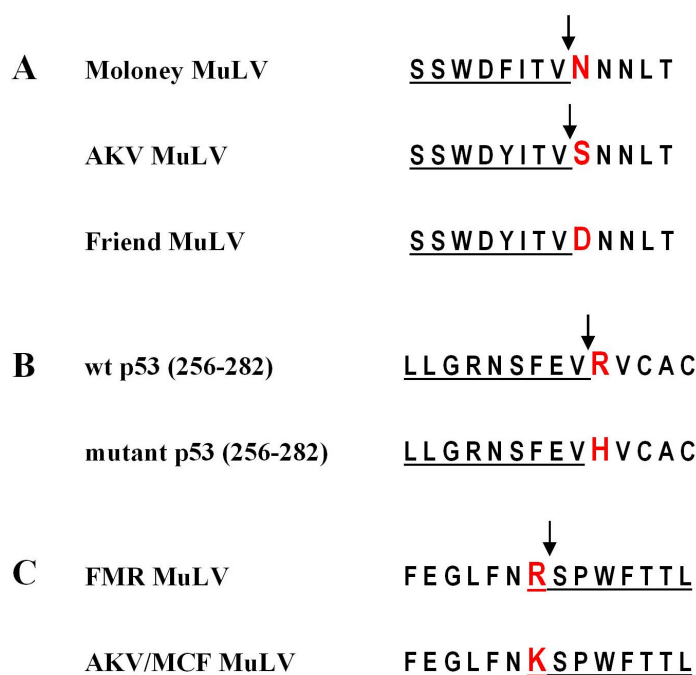
For the results, a short and a long output format can be selected. The short format is more graphical and shows the cleavages inserted into the submitted sequence as red, vertical bars (Figure 2.4-3 A, C). This format is recommended for a quick overview and serves as

the default setting. The long format gives more detailed tabular information on cleavage positions and estimated cleavage strength (Figure 2.4-3 B). It is useful, for example, for testing effects of mutations on the cleavage pattern and cleavage strength. (Please note that the assessment of cleavage strength is not based on quantified cleavage data, but purely on the qualitative nature of the aa flanking the prospective cleavage site.)

#### 2.4.4.3 Prediction results – mutated peptides containing CTL epitopes

PAProC can rapidly assess the changes of proteasomal cleavage imposed by mutations in an amino acid sequence containing CTL epitopes. In the case of a Moloney murine leukemia virus (MuLV) epitope, it was reported that a mutation N→D (as in the related Friend MuLV), but not N→S (as in the endogenous AKV-type MuLV), just C-terminal of the CTL epitope abrogates the generation of the C-terminus of the epitope by inducing a new cleavage site after D (Beekman et al., 2000). PAProC predicts these experimentally verified cleavages correctly (Figure 2.4-4 A). In another case, the R→H point mutation at position 273 of the tumor suppressor p53 was shown to abrogate the cleavage generating the C-terminus of the HLA-A\*0201-epitope 264-272 (Theobald et al., 1998). Again, PAProC correctly predicts this change of the proteasomal cleavage pattern (Figure 2.4-4 B). It should be stressed that PAProC predicts the C-terminal cleavages in the viral and in the p53-peptides to be very strong cleavages, i.e. the mutations in the peptides turn strong cleavage sites into non-cleavage sites in PAProC.

In yet another example, the exchange of the N-terminal residue of a viral CTL epitope rescues it from proteasomal destruction. Whereas the residue R promotes an epitope-destroying cleavage in the FMR type MuLV sequence, the R→K exchange in the AKV/MCF type of MuLV weakens this cleavage considerably, thus conserving the AKV/MCF CTL epitope (Ossendorp et al., 1996). PAProC correctly predicts the epitope-destroying cleavage behind R; however, PAProC does not predict a cleavage behind K any more (Figure 2.4-4 C). The latter is in contrast to the experimental results, where a very weak cleavage was found behind K. However, the result is in line with the experimental finding that the R→K mutation rescues the CTL epitope.



**Figure 2.4-4: PProC predictions of changes in the proteasomal cleavage pattern in sequences containing CTL epitopes, imposed by aa mutations within the epitope or at the flanking residue**

(A) PProC prediction of the C-terminal cleavage in related viral sequences (experimental results in Beekman et al., 2000); (B) PProC prediction of the C-terminal cleavage in a wt and mutant p53-epitope (experimental results in Theobald et al., 1998); (C) PProC prediction of an epitope-destroying cleavage at the N-terminus of related viral sequences (experimental results in Ossendorp et al., 1996). See text for details. Arrows indicate predicted cleavages. The mutated position is indicated by large letters. The underlined sequence corresponds to the CTL epitope. Please note that mouse 20S proteasomes were used to generate the experimental results for the CTL epitopes from MuLV (A and C). The predictions were done using human proteasome option type III.

#### 2.4.4.4 Prediction results – proteins linked to disease

For polyQ-diseases, altered proteasomal degradation of disease-linked proteins probably contributes to their intraneural aggregation, a crucial step in the pathogenesis of these diseases (Wanker, 2000). PProC can serve to assess the general cleavability and the proteasomal cleavage pattern in such proteins. For both huntingtin in Huntington's disease (HD) and ataxin in spinocerebellar ataxias (SCA), altered proteasomal degradation has been indicated in development of disease (Evert et al., 2000; Wyttenbach et al., 2000). We now report that two out of three PProC-algorithms trained by human 20S proteasome cleavage data (types II and III) predict the polyQ-regions of wt human ataxin-7 (10 aa) and of wt huntingtin (repeat length: 23 aa) to be resistant to proteasomal

cleavage (not shown). As the algorithm behind PAProC only inspects a window of 10 aa, the same is true for mutated huntingtin and ataxin with longer polyQ-sequences. It is conceivable that especially pathologically elongated polyQ-sequences could protect polyQ-containing proteins from proteasomal degradation, thus leading to protein accumulation within neurons and to cytosolic and nuclear inclusions. It must be mentioned that it is not an intrinsic feature of our prediction algorithm to leave poly-aa-sequences uncleaved: One-third of the 20 possible poly-aa-sequences (like AAAA..., CCCC..., DDDD..., etc.) are still predicted to be cleaved by proteasomes (not shown).

#### 2.4.4.5 Prediction results – combination of SYFPEITHI and PAProC

For immunologists, PAProC will be most interesting when used in combination with predictors of MHC I binding. To assess whether PAProC-predictions correlate with known CTL epitopes, we tested whether PAProC predicts cleavages at the C-termini of CTL epitopes in the melanoma-associated proteins MAGE-1 and MAGE-3. For MAGE-1, the C-terminal cleavage was predicted for 8 of 10 CTL epitopes (human proteasome type III, not shown). In 6 of the 8 cases, cleavages at both the N- and C-terminus of the CTL epitope were predicted. For MAGE-3, 4 of 7 C-termini were predicted to be cleaved correctly. In 3 of the 4 cases, PAProC predicted cleavages at both the N- and C-terminus of the epitope (data not shown). Known CTL epitopes in MAGE-1 include peptides binding to different HLA-alleles, e.g. to A2 (Pascolo S, submitted), A3, A24, A28, B7, B37, B53, Cw2 and two epitopes binding to both A1/B35 and Cw3/Cw16, respectively. Known CTL epitopes in MAGE-3 include peptides binding to A1, A2, A24, B37 and B44 (for references see Rammensee et al., 1999).

For the combination of MHC I ligand prediction and proteasomal cleavage prediction, one would probably first predict MHC I binders from a protein sequence of interest. In the second step, one would check the predicted MHC I binders for correct C- (and perhaps N-) terminal processing by proteasomes. To mimic this approach, we a) asked SYFPEITHI to predict MHC I-binding peptides from the tumor-associated proteins HER-2/neu, pmel/gp100 and p53 and from the influenza HA protein, b) asked PAProC to predict the cleavages in these proteins and c) compared the outcome to that of known CTL epitopes from these proteins (Table 2.4-1).

**Table 2.4-1: PProC predictions of the C-terminal cleavages of CTL epitopes**

Protein	MHC <sup>#</sup>	PProC / CTL epitopes (proteasome type) <sup>§</sup>	PProC / predicted MHC-ligands <sup>§</sup>
HER-2/neu	HLA-A*0201 (decamers)	2/2 (III)	12/38 (III <sup>§</sup> )
		1/2 (II, I)	
pmel/gp100	HLA-A*0201 (nonamers)	4/5 (I)	29/75 (I)
		3/5 (II, III)	
p53	HLA-A*0201 (nonamers)	2/2 (III)	12/25 (III)
Infl. Jap. HA	H2-Kd (nonamers)	2/3 (I/II)	6/20 (I), 7/20 (II)

# MHC class I molecule for which ligands were predicted; HLA: Human leukocyte antigen; H2: mouse MHC I; in brackets: length of predicted MHC I ligands

§ PProC predictions of the C-terminal cleavage for CTL epitopes (Number of correctly predicted C-termini / number of CTL epitopes tested). Proteasome type: algorithms used for PProC predictions. Types I-III, all based on human 20S proteasome cleavage data, slightly differ in parts of their training data (Refer to Kuttler et al. (2000) for detail; see 2.3).

§ PProC predictions of the C-terminal cleavage for predicted MHC I ligands (Number of predicted cleavage sites corresponding to the C-termini of predicted MHC I ligands / number of predicted MHC I ligands tested). In brackets: proteasome type.

For HER-2/neu decamers, the C-termini of the two known HLA-A\*0201-restricted CTL epitopes were predicted to be cleaved correctly by PProC's human proteasome type III (one of two for types I + II). However, PProC's overall hit rate was only 12 of 38 for the C-termini of predicted MHC I ligands from HER-2/neu, including the two correctly predicted CTL epitopes (see Material & Methods for details). This discrepancy reflects the influence of antigen processing on actual MHC I ligand generation. Similarly, for the three other proteins tested, fitting PProC predictions were "enriched" among the CTL epitopes as compared to predicted MHC I ligands (Table 2.4–1). These results suggest that PProC can indeed be used in combination with the prediction of MHC I binding to narrow down the search for CTL epitopes from any protein sequence.

## 2.4.5 Discussion

Here we have described the first publicly available prediction tool for proteasomal

cleavages, PAPProC (Prediction Algorithm for Proteasomal Cleavages). This internet-based program (<http://www.paproc.de>) allows the rapid submission of aa sequences and provides an output of possible proteasomal cleavages sites within these sequences (Figure 2.4-2, Figure 2.4-3). PAPProC offers information on both the general cleavability of amino acid sequences (cuts per amino acids) and the individual cleavages (positions and estimated strength) (Figure 2.4-3). Our results suggest that PAPProC can be used to assess the influence of aa mutations on CTL epitope generation (Figure 2.4-4). More importantly, PAPProC will support the search for CTL epitopes by making it more efficient. The number of candidates (i.e. predicted MHC I ligands) will be reduced by excluding those not predicted to be cleaved correctly (Table 2.4–1).

#### 2.4.5.1 Benefits of PAPProC

Proteasomes are involved in many essential cellular processes, such as cell cycle control and transcription factor activation (Rock and Goldberg, 1999; Pamer and Cresswell, 1998). They have also been linked to disease, e.g. Alzheimer's, Parkinson's and Huntington's disease (Checler et al., 2000; Wanker, 2000). Thus, PAPProC might be attractive for researchers in a wide range of fields. However, as proteasomal cleavage specificity can regulate immune responses (see 2.4.2 Introduction), immunologists studying antigen processing will most likely constitute the largest group of future PAPProC-users. Rational vaccine-design in particular could profit from proteasomal cleavage prediction. For multi-epitope vaccines, PAPProC could for example be used to determine epitope-linking regions that most efficiently support correct processing of the amino acid sequence. Moreover, proteasomal cleavage predictors such as PAPProC could be used for the design of peptide-based proteasome inhibitors.

#### 2.4.5.2 Prediction success rate of PAPProC

There are examples in which PAPProC does not predict proteasomal cleavages at the C-termini of known CTL epitopes (hit rate in HER-2/neu HLA-A\*0201-restricted nonamers: 3 of 7 CTL-epitopes, 15 of 40 predicted MHC I ligands; pmel17/gp100 HLA-A\*0201-restricted decamers: 1 of 6 CTL-epitopes; EBV EBNA-6 HLA-B\*2705-restricted nonamers: 1 of 2; EBV EBNA-3 HLA-B\*08-restricted nonamers: 0 of 4 CTL-epitopes; not shown). Although C-terminal trimming cannot be excluded to play a role *in vivo* (Elliott et al., 1995), it is more likely that the limited amount of training data for the PAPProC algorithm is responsible for this (Kuttler et al., 2000).

PAPProC also predicts epitope-destroying cleavages within known CTL epitopes (not

shown). We do not consider this a problem because there are many examples of proteasomal cleavage sites within CTL epitopes that do not abrogate presentation of the epitope *in vivo* (Ossendorp et al., 1996; Theobald et al., 1998; Morel et al., 2000; Lucchiari-Hartz et al., 2000, to name a few). Moreover, epitope-destroying cleavages might be weak ones, whereas cleavages at the C-termini of CTL epitopes might be dominant ones. Our preliminary prediction data suggest this notion (Kuttler, unpublished observation). However, only predictions based on quantified cleavage data will solve this problem adequately. The assessment of cleavage strength from non-quantified data, as executed in PAProC at present, is only a rough estimate.

The PAProC algorithms trained with human 20S proteasome data correctly predict cleavages that were generated experimentally or *in vivo* by mouse proteasomes (Figure 2.4-4, Table 2.4-1). This result suggests that PAProC's human proteasome algorithms can be used to predict cleavages generated by any other mammalian proteasome. This concept is supported by the finding that the cleavage preferences of mammalian proteasomes are similar (Bjorkman et al., 1987). The successful prediction results for CTL epitopes binding to different human HLA- (text and Table 2.4-1) and the mouse H2-K<sup>d</sup> MHC I molecules (Table 2.4-1) confirm that the success rate of PAProC predictions does not depend on factors imposed by distinct MHC I binding motifs.

#### 2.4.5.3 Future developments

PAProC was developed by antigen processing and proteasome specialists in close collaboration with biomathematicians, and has thus profited from competent biological and mathematical expertise. However, we are aware of the fact that PAProC is still in its "childhood years". The precision of PAProC predictions varies depending on the substrate and, for human proteasome, on the proteasome type chosen. The lack of training data is probably responsible for this (Kuttler et al., 2000). For this reason, we are continuously working on improving PAProC both by the generation of additional training data and by the fine tuning of the prediction model itself. For example, in the future we will use quantified cleavage data for greater accuracy. We will use cleavages by 26S proteasomes (Emmerich et al., 2000) and cleavages by immuno- or constitutive human 20S proteasomes (Toes et al., in preparation; see 2.5 for details) as training data. The latter approach in particular will be of great interest for immunologists. Moreover, we are developing an algorithm based on fragments, not cleavage sites, hence allowing the prediction of fragments, not just cleavage sites. This approach will give more insight into whether cleavages within MHC-ligands will still allow the generation of the uncut ligand.

However, PAPProC will profit most efficiently from working experience, which is one of the main reasons for making it public at this point. We therefore encourage PAPProC users to let us know how the program performed for them. We will be happy to include experimental data generated outside our lab into our training data. However, strict quality measures will be applied.

#### 2.4.5.4 Distant goal

For immunologists, proteasomal cleavage prediction will be most useful in combination with the prediction of MHC I ligands. Several groups are preparing such combined approaches. A group in Berlin is combining the previously reported proteasomal cleavage algorithm (Holzhütter et al., 1999) and PAPProC to available predictors of MHC I binding, such as SYFPEITHI (H. Mollenkopf, personal communications). In Copenhagen, NetChop, a neural network for proteasomal cleavage prediction (C. Kesmir, submitted), and PAPProC will be combined to different available MHC I ligand predictors (C. Kesmir and S. Brunak, personal communications). A group in Leiden is preparing an approach by which proteasomal cleavage data will be combined with characteristics of known CTL epitopes (F. Ossendorp, C. Melief, personal communications).

#### 2.4.6 References

Beekman, N.J., van Veelen, P.A., van Hall, T., Neisig, A., Sijts, A.J., Camps, M., Kloetzel, P.M., Neefjes, J., Melief, C.J., and Ossendorp, F.A. (2000). Abrogation of CTL epitope processing by single amino acid substitution flanking the C-terminal proteasome cleavage site. *J. Immunol.* 164, 1898-1905.

Bjorkman, P.J., Saper, M.A., Samraoui, B., Bennett, W.S., Strominger, J.L., and Wiley, D.C. (1987). The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. *Nature* 329, 512-518.

Checler, F., da Costa, C.A., Ancolio, K., Chevallier, N., Lopez-Perez, E., and Marambaud, P. (2000). Role of the proteasome in Alzheimer's disease. *Biochim. Biophys. Acta* 1502, 133-138.

Craiu, A., Akopian, T.N., Goldberg, A.L., and Rock, K.L. (1997). Two distinct proteolytic processes in the generation of a major histocompatibility complex class I-presented peptide. *Proc. Natl. Acad. Sci. USA* 94, 10850-10855.

Cummings, C.J., Mancini, M.A., Antalffy, B., DeFranco, D.B., Orr, H.T., and Zoghbi, H.Y. (1998). Chaperone suppression of aggregation and altered subcellular proteasome localization imply



protein misfolding in SCA1. *Nat. Genet.* 19, 148-154.

Dantuma, N.P., Heessen, S., Lindsten, K., Jellne, M., and Masucci, M.G. (2000). Inhibition of proteasomal degradation by the Gly-Ala repeat of Epstein-Barr virus is influenced by the length of the repeat and the strength of the degradation signal. *Proc. Natl. Acad. Sci. USA* 97, 8381-8385.

Elliott, T., Willis, A., Cerundolo, V., and Townsend, A. (1995). Processing of major histocompatibility class I-restricted antigens in the endoplasmic reticulum. *J. Exp. Med.* 181, 1481-1491.

Emmerich, N.P.N., Nussbaum, A.K., Stevanovic, S., Priemer, M., Toes, R.E.M., Rammensee, H.G., and Schild, H. (2000). The human 26S and 20S proteasomes generate overlapping but different sets of peptide fragments from a model protein substrate. *J. Biol. Chem.* 275, 21140-21148.

Evert, B.O., Wullner, U., and Klockgether, T. (2000). Cell death in polyglutamine diseases. *Cell Tissue Res.* 301, 189-204.

Holzthütter, H.G., Frommel, C., and Kloetzel, P.M. (1999). A theoretical approach towards the identification of cleavage-determining amino acid motifs of the 20 S proteasome. *J. Mol. Biol.* 286, 1251-1265.

Krobitsch, S. and Lindquist, S. (2000). Aggregation of huntingtin in yeast varies with the length of the polyglutamine expansion and the expression of chaperone proteins. *Proc. Natl. Acad. Sci. USA* 97, 1589-1594.

Kuttler, C., Nussbaum, A.K., Dick, T.P., Rammensee, H.G., Schild, H., and Haderl, K.P. (2000). An algorithm for the prediction of proteasomal cleavages. *J. Mol. Biol.* 298, 417-429.

Levitskaya, J., Coram, M., Levitsky, V., Imreh, S., Steigerwald-Mullen, P.M., Klein, G., Kurilla, M.G., and Masucci, M.G. (1995). Inhibition of antigen processing by the internal repeat region of the Epstein-Barr virus nuclear antigen-1. *Nature* 375, 685-688.

Levitskaya, J., Sharipo, A., Leonchiks, A., Ciechanover, A., and Masucci, M.G. (1997). Inhibition of ubiquitin/proteasome-dependent protein degradation by the Gly-Ala repeat domain of the Epstein-Barr virus nuclear antigen 1. *Proc. Natl. Acad. Sci. USA* 94, 12616-12621.

Lucchiari-Hartz, M., van Endert, P.M., Lauvau, G., Maier, R., Meyerhans, A., Mann, D., Eichmann, K., and Niedermann, G. (2000). Cytotoxic T lymphocyte epitopes of HIV-1 nef. Generation Of multiple definitive major histocompatibility complex class I ligands by proteasomes. *J. Exp. Med.*

191, 239-252.

Mo, X.Y., Cascio, P., Lemerise, K., Goldberg, A.L., and Rock, K.L. (1999). Distinct proteolytic processes generate the C and N termini of MHC class I-binding peptides. *J. Immunol.* 163, 5851-5859.

Morel, S., Levy, F., Burlet-Schiltz, O., Brasseur, F., Probst-Kepper, M., Peitrequin, A.L., Monsarrat, B., Van Velthoven, R., Cerottini, J.C., Boon, T., Gairin, J.E., and Van den Eynde, B.J. (2000). Processing of some antigens by the standard proteasome but not by the immunoproteasome results in poor presentation by dendritic cells. *Immunity* 12, 107-117.

Nussbaum, A.K., Dick, T.P., Keilholz, W., Schirle, M., Stevanovic, S., Dietz, K., Heinemeyer, W., Groll, M., Wolf, D.H., Huber, R., Rammensee, H.G., and Schild, H. (1998). Cleavage motifs of the yeast 20S proteasome beta subunits deduced from digests of enolase 1. *Proc. Natl. Acad. Sci. USA* 95, 12504-12509.

Ossendorp, F.A., Eggers, M., Neisig, A., Ruppert, T., Groettrup, M., Sijts, A.J., Mengede, E., Kloetzel, P.M., Neefjes, J., Koszinowski, U.H., and Melief, C.J. (1996). A single residue exchange within a viral CTL epitope alters proteasome-mediated degradation resulting in lack of antigen presentation. *Immunity* 5, 115-124.

Pamer, E. and Cresswell, P. (1998). Mechanisms of MHC class I-restricted antigen processing. *Annu. Rev. Immunol.* 16, 323-358.

Paz, P., Brouwenstijn, N., Perry, R., and Shastri, N. (1999). Discrete proteolytic intermediates in the MHC class I antigen processing pathway and MHC I-dependent peptide trimming in the ER. *Immunity* 11, 241-251.

Rammensee, H.G., Bachmann, J., Emmerich, N.P.N., Bachor, O.A., and Stevanovic, S. (1999). SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics* 50, 213-219.

Rock, K.L. and Goldberg, A.L. (1999). Degradation of cell proteins and the generation of MHC class I-presented peptides. *Annu. Rev. Immunol.* 17, 739-779.

Stoltze, L., Dick, T.P., Deeg, M., Pömmel, B., Rammensee, H.G., and Schild, H. (1998). Generation of the vesicular stomatitis virus nucleoprotein cytotoxic T lymphocyte epitope requires proteasome-dependent and -independent proteolytic activities. *Eur. J. Immunol.* 28, 4029-4036.

Stoltze, L., Schirle, M., Schwarz, G., Schröter, C., Thompson, M.W., Hersh, L.B., Kalbacher, H., Stevanovic, S., Rammensee, H.G., and Schild, H. (2000). Two new players in the MHC class I

antigen processing pathway. *Nat. Immunol.* 1, 413-418.

Theobald, M., Ruppert, T., Kuckelkorn, U., Hernandez, J., Haussler, A., Ferreira, E.A., Liewer, U., Biggs, J., Levine, A.J., Huber, C., Koszinowski, U.H., Kloetzel, P.M., and Sherman, L.A. (1998). The sequence alteration associated with a mutational hotspot in p53 protects cells from lysis by cytotoxic T lymphocytes specific for a flanking peptide epitope. *J. Exp. Med.* 188, 1017-1028.

Uebel, S. and Tampe, R. (1999). Specificity of the proteasome and the TAP transporter. *Curr. Opin. Immunol.* 11, 203-208.

van Endert, P.M. (1999). Genes regulating MHC class I processing of antigen. *Curr. Opin. Immunol.* 11, 82-88.

Wanker, E.E. (2000). Protein aggregation in Huntington's and Parkinson's disease: implications for therapy. *Mol. Med. Today* 6, 387-391.

Wytenbach, A., Carmichael, J., Swartz, J., Furlong, R.A., Narain, Y., Rankin, J., and Rubinsztein, D.C. (2000). Effects of heat shock, heat shock protein 40 (HDJ-2), and proteasome inhibition on protein aggregation in cellular models of Huntington's disease. *Proc. Natl. Acad. Sci. USA* 97, 2898-2903.

Zoghbi, H.Y. and Orr, H.T. (2000). Glutamine repeats and neurodegeneration. *Annu. Rev. Neurosci.* 23, 217-247.

## 2.4.7 Abbreviations

CTL	cytotoxic T lymphocytes
EBNA	Ebstein-Barr virus nuclear antigen
EBV	Epstein-Barr virus
ER	endoplasmic reticulum
GARR	Gly-Ala-rich region
H2	mouse MHC locus
HD	Huntington's disease
HLA	human leukocyte antigen (human MHC locus)
MHC I	major histocompatibility complex class I
MuLV	Moloney murine leukemia virus
PAProC	Prediction Algorithm for Proteasomal Cleavages
polyQ	poly-glutamine
SCA	spinocerebellar ataxia
TAP	transporter associated with antigen processing
wt	wild-type

WWW

world-wide web

### 2.4.8 Participating researchers

Christina Kuttler<sup>2</sup>, Karl-Peter Haderer<sup>2</sup>, Hans-Georg Rammensee<sup>1</sup>, Hansjörg Schild<sup>1</sup>

<sup>1</sup> Universität Tübingen, Interfakultäres Institut für Zellbiologie, Abteilung Immunologie, Auf der Morgenstelle 15, D-72076 Tübingen, Germany.

<sup>2</sup> Universität Tübingen, Biomathematik, Auf der Morgenstelle 10, D-72076 Tübingen, Germany.

## **2.5 Discrete cleavage motifs of constitutive and immuno-proteasomes revealed by quantitative analysis of cleavage products**

### **2.5.1 Summary**

Proteasomes are the main proteases responsible for cytosolic protein degradation and the production of MHC class I-ligands. Incorporation of the IFN- $\gamma$ -inducible subunits LMP2, LMP7 and MECL-1 leads to the formation of immuno-proteasomes which have been associated with more efficient MHC class I antigen processing. Although differences in cleavage specificities of constitutive- and immuno-proteasomes have been observed frequently, cleavage motifs have not been described.

Using the 436-amino acid protein enolase-1 as unmodified model substrate and a novel quantitative approach, we analyzed a large collection of peptides generated by either set of proteasomes. Inspection of the amino acids flanking proteasomal cleavage sites allowed the description of two different cleavage motifs. These motifs finally explain recent findings describing differential processing of CTL-epitopes by constitutive- and immuno-proteasomes and are important to the understanding of peripheral T cell tolerization/activation as well as for effective vaccine development.

### **2.5.2 Introduction**

Cytotoxic T lymphocytes (CTL) are crucial for the defense against many invading organisms and certain tumors. Presentation of antigenic peptides bound to major histocompatibility complex (MHC) class I molecules is a prerequisite for stimulation of a CTL-response, and therefore plays a pivotal role in providing CTL with the capacity to respond to foreign antigens (Rammensee et al., 1993).

Peptides that meet the restrictive binding characteristics of MHC class I molecules for presentation to CTL are generated after intracellular protein degradation by cytosolic proteases. The central enzyme responsible for protein degradation is the proteasome (Rock and Goldberg, 1999). Because of their intimate involvement in antigen processing and presentation (Pamer and Cresswell, 1998), detailed knowledge on the cleavage preferences of proteasomes will be crucial for understanding CTL-epitope generation and thus for the regulation of specific immune responses.

The 20S proteasome represents the proteolytic core of the larger 26S proteasome complex that encompasses either one or two regulatory particles of at least 18 subunits (Coux et al., 1996). The eukaryotic 20S particle is composed of 14 different but related subunits organized in a barrel-shaped complex with the stoichiometry  $\alpha_7\beta_7\beta_7\alpha_7$ . Three subunits of the two inner  $\beta$ -rings ( $\beta_1$ ,  $\beta_2$  and  $\beta_5$ ) participate directly in peptide bond cleavage. They represent three distinct proteolytic activities, designated as the chymotrypsin (ChT)-like, trypsin-like and peptidylglutamylpeptide-hydrolyzing (PGPH) activities (Heinemeyer et al., 1997; Dick et al., 1998). As the  $\text{NH}_2$ -terminal threonine residues responsible for peptide bond cleavage do not directly prefer certain peptide bonds over others, the basis for the three distinct proteolytic activities most likely resides in the characteristics of the amino acids in the vicinity (pockets) of each active  $\text{NH}_2$ -terminal threonine (Groll et al., 1997).

Upon interferon- $\gamma$  (IFN- $\gamma$ ) exposure of cells, the three active  $\beta$ -subunits that are constitutively expressed in 20S-proteasomes can be replaced by three IFN- $\gamma$ -inducible homologues, LMP-2 (=  $\beta_{1i}$ ) (for Y ( $\beta_1$ )), MECL-1 ( $\beta_{2i}$ ) (for Z ( $\beta_2$ )) and LMP-7 ( $\beta_{5i}$ ) (for X ( $\beta_5$ )). Although there is extensive sequence homology, these replacements alter the nature of peptides that are generated by proteasomes (Boes et al., 1994; Groettrup et al., 1995; Gaczynska et al., 1993; Eleuteri et al., 1997; Cardozo and Kohanski, 1998).

Proteasomes harboring these IFN- $\gamma$ -inducible subunits are also called immuno-proteasomes, as opposed to the constitutively expressed "constitutive" proteasomes, because immuno-proteasomes were found to process a number of viral epitopes with greater efficacy *in vitro* (Sijts et al., 2000a; Sijts et al., 2000b; van Hall et al., 2000; Schwarz et al., 2000). Using several artificial fluorogenic substrates *in vitro*, it was found that immuno-proteasomes display a better capacity to cleave after hydrophobic and basic residues, but are less well equipped for cleavage after acidic amino acids (Gaczynska et al., 1994). The finding that proteasomes are responsible for the generation of the correct COOH-terminus of several CTL-epitopes (Craiu et al., 1997; Stoltze et al., 1998), and the notion that hydrophobic or positively charged amino acids serve in most cases as COOH-terminal anchor-residues of MHC class I-ligands, led to the concept that immuno-proteasomes contribute to more efficient MHC class I antigen-processing. Nonetheless, more recent studies have shown that some antigenic peptides are efficiently produced by constitutive proteasomes but cannot be produced by immuno-proteasomes (Morel et al., 2000). This clearly contradicts the concept that the immuno-proteasome is generally better suited for the processing of MHC class I-ligands.

To better understand the reasons why certain MHC class I-ligands are destroyed and others generated with greater efficiency in cells expressing different sets of proteasomes, we have performed an in-depth analysis of peptide fragments generated after proteasomal cleavage.

We have employed, for the first time, a strictly quantitative method to analyse a large collection of peptide-fragments produced by either set of proteasome. Our observations allowed the identification of certain amino acids (or their characteristics) in positions distant, or directly flanking the cleavage sites selected by either set of proteasomes. The (quantified) mapping of cleavage sites using a large protein substrate provides the basis for a better understanding of proteasomal cleavage specificity, allowing a refined proteasomal cleavage prediction, which will be helpful for the identification of new CTL epitopes, the design of new (recombinant) vaccines, and for better insight into immunity against infection.

### 2.5.3 Materials & Methods

#### 2.5.3.1 Purification of 20S proteasomes

20S proteasomes were isolated as described before (Groettrup et al., 1995). Frozen pellets of LCL-721 cells or LCL-721.174 cells were lysed in a buffer containing 0.1% Triton-X-100 on ice and homogenized in a Dounce homogenizer. The 40.000xg supernatant of the lysate was bound to DEAE-Sephacel. After elution, the protein fraction was concentrated and loaded onto a 10 - 40% sucrose gradient. After centrifugation, gradient fractions were tested for protease activity using the fluorogenic substrates Suc-LLVY-AMC and Suc-YVAD-AMC. Active fractions were pooled and further purified by anion exchange chromatography on a MonoQ HR5/5 FPLC column (Pharmacia). The purity of the proteasome preparates, checked by SDS-PAGE, was >95%. Quantification of native proteasome protein was determined by a variation of the Lowry Method (Bio-Rad Protein Assay, Bio-Rad) and bovine serum albumine as a standard.

#### 2.5.3.2 Immunoblotting

5 µg of purified proteasome polypeptides were separated by 12% SDS-PAGE, and transferred to polyvinylidene difluoride (PVDF) (DuPont) with a semidry transfer system. Human LMP-7 was detected using a rabbit polyclonal antiserum by chemoluminescence (PW8200, Affiniti research Products Ltd. Mamhead, UK).

### 2.5.3.3 Measurement of proteasomal activities against substrates with fluorogenic leaving group

The fluorogenic substrates benzyloxycarbonyl-Leu-Leu-Glu- $\beta$ -naphthylamide (Z-LLE- $\beta$ -NA), succinyl-Tyr-Val-Ala-Asp-7-amino-4-methylcoumarin (Suc-YVAD-AMC), succinyl-Leu-Leu-Val-Tyr-7-amino-4-methylcoumarin (Suc-LLVY-AMC), and benzyloxycarbonyl-Ala-Arg-Arg-7-amino-4-methylcoumarin (Z-ARR-AMC) (all from Bachem, Heidelberg, Germany) were prepared from 10 mM stocks in Me<sub>2</sub>SO. 1  $\mu$ g proteasome was incubated with 100  $\mu$ M substrate solution. Fluorescence of the leaving group was determined after incubation for 0 - 6 hours with a Tecan spectrophotometer (Tecan, Crailsheim, Germany) at 360 nm excitation and 450 nm emission for AMC and at 330 nm excitation and 410 nm emission for  $\beta$ -NA. Fluorescence readings of released AMC or  $\beta$ -NA were recorded as arbitrary units.

### 2.5.3.4 In vitro degradation of enolase 1

150  $\mu$ g of yeast enolase 1 was incubated in digestion buffer (20mM HEPES/KOH (pH 7.6), 2 mM MgAc<sub>2</sub>, and 0.01% SDS) with proteasomes in a molar ratio of 150:1. Digestions were stopped by freezing the samples at -80°C when approximately 50% of the substrate was digested (usually after approximately 48 hours).

### 2.5.3.5 Separation and analysis of cleavage products

For the separation of degradation products, unfractionated enolase digests were subjected to  $\mu$ RP SC 2.1/10 columns (Pharmacia) on a Microbore HPLC-system (SMART system, Pharmacia). Buffer A contained 0.1% trifluoroacetic acid (TFA); buffer B contained 0.081% TFA and 80% acetonitrile. Gradients were 0% for 5min, in 40 min to 40% B, in 8 min to 75% B, and up to 85% in another 7 min at a flow rate of 150  $\mu$ l/min. Fractions were collected by peak fractionation with a maximal volume of 500  $\mu$ l/peak.

Peak fractions were dried and redissolved in 25 $\mu$ l of 40% methanol, 1% formic acid and subsequently analysed by Matrix-associated Laser desorption ionization-time of flight (MALDI-TOF) mass spectrometry (G2025A, Hewlett Packard, Waldbronn, Germany) and N-terminal sequencing (Edman degradation) (pulsed liquid protein sequencer procise 494A, Applied Biosystems, Weiterstadt, Germany). Alternatively, peptides were analysed on a hybrid quadrupole orthogonal acceleration tandem mass spectrometer (Micromass). All these techniques were applied as described recently (Nussbaum et al., 1998). Pmol amounts for each peptide detected in the HPLC fraction were determined by Edman



sequencing and used for the quantitative analysis of the data.

#### *2.5.3.6 Statistical analysis - frequencies of amino acids*

To detect statistically significant features in the amino acid distributions flanking the cleavage sites, we compared percent values using a classical Chi<sup>2</sup>-test for four tables (variance assumed due to counting). This method was used to compare constitutive and immuno-proteasomes with each other and with enolase. For a more thorough comparison of the absolute pmol-amounts of constitutive and immuno-proteasomes, we accounted for the experimental variability and, according to the quasi-likelihood approach of Wedderburn, assumed a mean-variance structure. We assumed the variance to be proportional to the mean and fitted the proportionality constant  $\alpha$  from all the data. Then, the usual Chi<sup>2</sup>-test variable was scaled by  $1/\alpha$ , which led asymptotically to a test variable that is Chi<sup>2</sup> distributed with one degree of freedom. – The results of the latter, more thorough approach correlated to the approach neglecting experimental variability and using percent values. Only Chi<sup>2</sup>-values above 3.841 are considered to be significant.

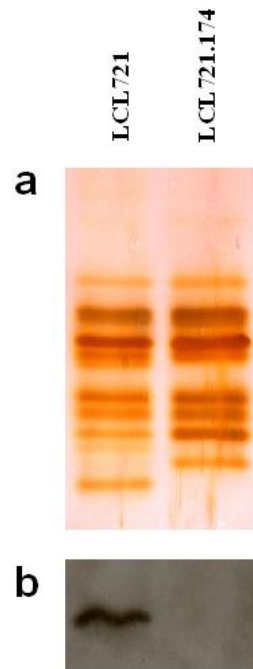
#### *2.5.3.7 Statistical analysis - comparison of amino acid characteristics*

To compare the characteristics of amino acids, The observed frequencies of amino acids at P6 to P6' around cleavage sites in both proteasomes were compared to each other using the Chi<sup>2</sup>-test. Hydrophobicity, bulkiness, and flexibility characteristics (values as in 24) of both proteasomes, and enolase were compared by translating the percentage amino acids found to the corresponding hydrophobicity, bulkiness, and flexibility scales. This resulted in spectra per cleavage site, and these were compared by means of regression analysis.

## 2.5.4 Results

### *2.5.4.1 Isolation of proteasomes*

Proteasomes were isolated from EBV-transformed B cells. Constitutive proteasomes were purified from cells (LCL721.174) lacking LMP2 and LMP7 due to a chromosomal deletion in the MHC-locus (Spies et al., 1990; DeMars et al., 1985). The lack of LMP2 or LMP7 results in inefficient maturation of MECL-1 in 20S proteasomes. Therefore, this cell line contains only proteasomes carrying active constitutive subunits (Griffin et al., 1998). The immuno-proteasome preparation was isolated from the parental line (LCL721) that served for the generation of LCL 721.174.

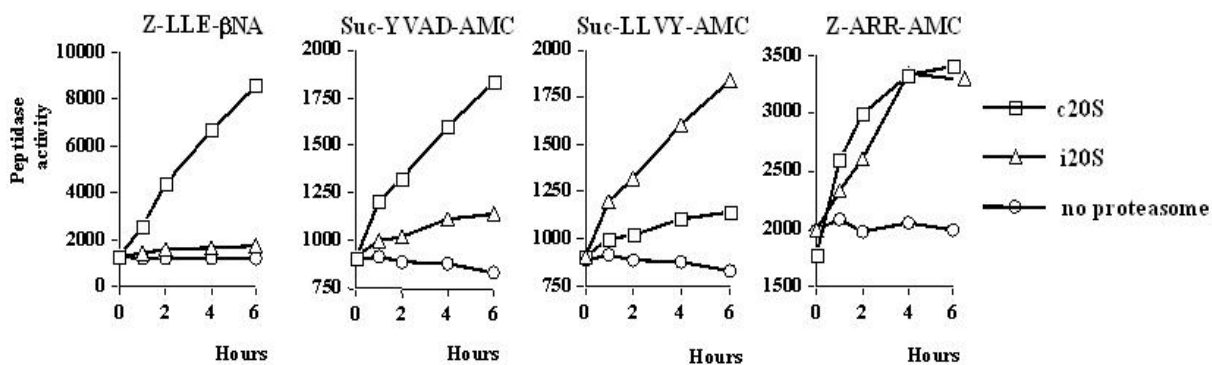


**Figure 2.5-1: Immuno-subunit-incorporation into 20S proteasomes purified from LCL-721 cells (left), but not into 20S-proteasomes derived from LCL-721.174 cells (right).**

20S proteasomes were isolated from LCL-721 and LCL-721.174 cells as described in materials and methods. Proteasome subunits were separated by SDS-PAGE (a) and probed with an LMP-7-specific antiserum (b). LMP-7 was only detected in 721-proteasomes.

As expected, only proteasomes isolated from LCL721 cells expressed immuno-subunits, as exemplified by the presence of LMP7 (Figure 2.5-1a,b). Moreover, these proteasomes showed an increased cleavage activity after the hydrophobic amino acid tyrosine and a reduced ability to release fluorogenic groups linked to acidic amino acids, compared to proteasomes isolated from LCL721.174 cells (Figure 2.5-2). Both proteasome preparations harboured the constitutive subunit Z (not shown). This indicates that not all constitutive subunits were exchanged in the proteasome pool of the LCL721 cells, which most likely gives a good reflection of the intracellular proteasome pool present in IFN- $\gamma$  treated cells.

Together, these data indicate that proteasomes purified from LCL721, although containing some constitutive proteasomes, behave like immuno-proteasomes with very little PGPH-activity, whereas LCL721.174-derived proteasomes can be classified as constitutive proteasomes with high PGPH-activity.

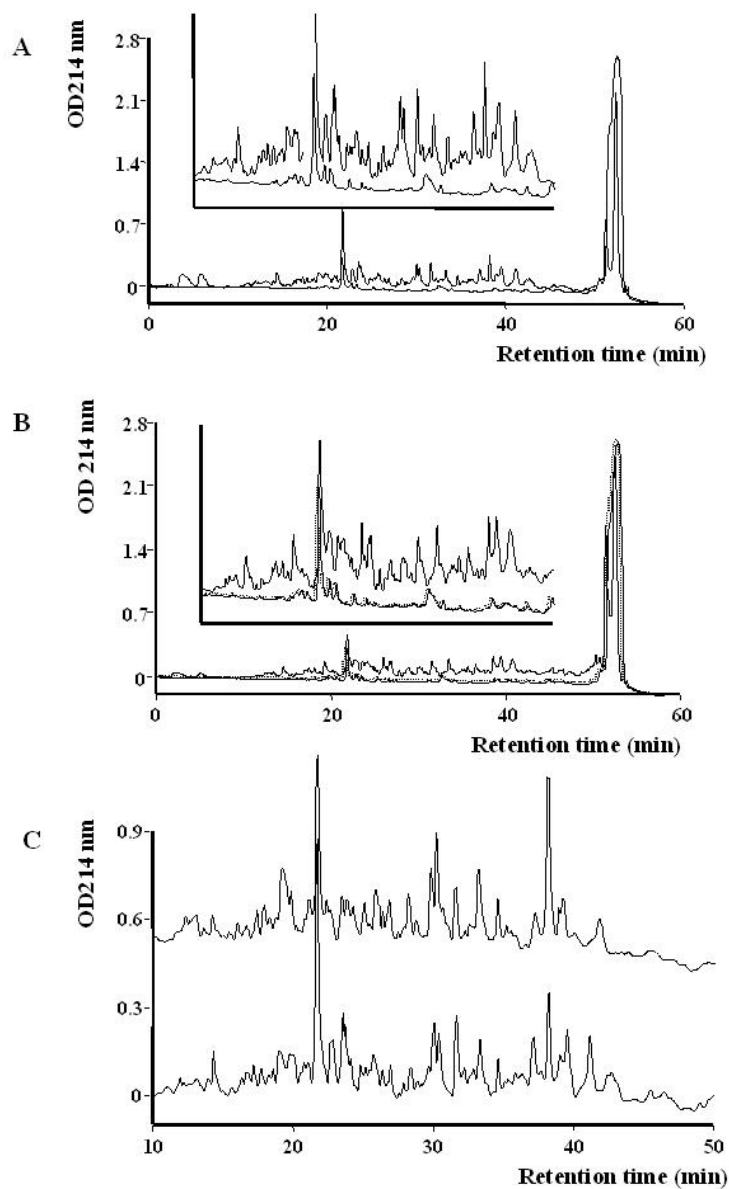


**Figure 2.5-2: Comparison of the ability of 20S proteasomes isolated from LCL-721 and LCL-721.174 cells to catalyze the release of fluorogenic groups from three different substrates.**

Peptidylglutamylpeptide-hydrolyzing activity was analyzed using Z-LLE- $\beta$ NA and Suc-YVAD-AMC as substrates. The substrates Suc-LLVY-AMC and Z-ARR-AMC were used to analyze chymotrypsin-like and trypsin-like activities, respectively. One representative experiment of three is shown.

#### 2.5.4.2 Digestion of Enolase

Although the ability of constitutive- and immuno-proteasomes to cleave a set of standard fluorogenic substrates or some CTL-epitope containing peptides has been well documented, little is known about the selection of cleavage sites during the degradation of proteins, especially on a quantitative basis. To obtain further insight into these cleavage preferences, we used the complete protein enolase-1 from yeast as substrate, which can be digested by proteasomes *in vitro* without prior modifications. Enolase is a 436 amino acid long protein in which the frequency of amino acids resembles the average amino acid frequency in proteins (Brendel et al., 1992). Digestion of enolase was performed by incubation with constitutive- or immuno-proteasomes at a molar ratio of 150 : 1 (enolase : proteasome). The reaction was stopped when approximately 50% of the substrate was degraded, and subsequently separated by reversed phase HPLC (Figure 2.5-3a,b). Comparison of two independent digests obtained after incubation with two independent constitutive proteasome batches, revealed that highly comparable degradation profiles were obtained (Figure 2.5-3c). These data indicate that different incubations by the same proteasome type yields a similar set of degradation products, as was confirmed by MALDI-MS analyses of several fractions that eluted at the same time (not shown).



**Figure 2.5-3: Enolase degradation by proteasomes purified from LCL-721 and LCL-721.174 cells.**

Enolase was incubated with constitutive proteasomes (a), immuno-proteasomes (b) or was left untreated (dotted lines). A magnification of the degradation-profile is shown in the inserts. The comparison of two independent digestion-profiles (in c) acquired after incubating enolase with two different constitutive proteasome-batches purified from LCL-721.174 cells shows high reproducibility, indicating that a similar set of peptides is generated by different "174-proteasome"-batches. Optical density at 214 nm of eluted digestion products from reversed phase HPLC is shown.

After having established the reproducibility of the digestion profiles, the peptide fragments in all fractions were analysed by Edman sequencing, in combination with MALDI-MS, and

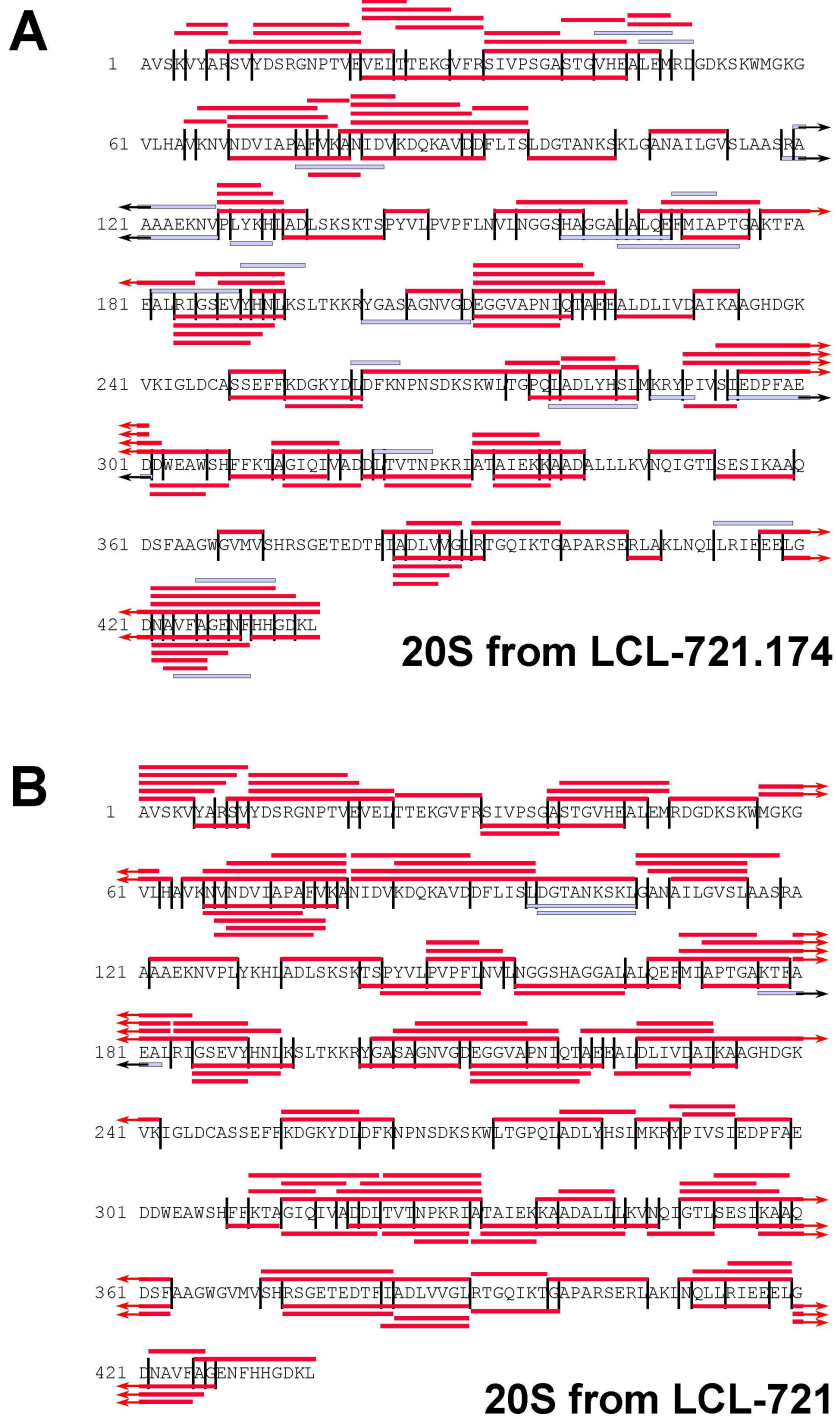
compiled in a digestion map (Figure 2.5-4). Approximately 50% of cleavages generated by constitutive proteasomes were not produced by immuno-proteasomes. Indeed, when the identity of fragments produced by either set of proteasome was compared on a qualitative basis, only around 25% of peptides produced by immuno-proteasomes were also found in constitutive proteasome digests.

Therefore, the pool of peptides generated by cells expressing immuno-proteasomes is very different from the peptide-pool generated by cells harbouring constitutive proteasomes only.

#### *2.5.4.3 Quantification of digestion profiles*

For careful examination of proteasomal cleavage preferences, it is important to know the quantity of each fragment to calculate how often particular cleavage sites are selected. In contrast to MS-data, data acquired by Edman-sequencing are quantitative, and can thus be used to determine the amount of peptide liberated. The combination of MS-analysis and the quantified Edman-sequencing data identified the absolute amount of each peptide detected in HPLC-fractions. In the constitutive proteasome digests, a total of 136 fragments was detected, representing 6135 pmol of peptide (Table 2.5-1). By adding the pmol of all fragments starting or ending at a particular cleavage site (and then choosing the higher one of the two sums), pmol amounts of peptide generated from a given cleavage site, and thus the frequency of cleavage site utilization was determined. The most frequently used cleavage sites were found at amino acid position 278 and 404, resulting in the liberation of 265 pmol peptide each (Table 2.5-1 top). As usage of many other cleavage sites resulted in only 5 pmol of peptide, these data indicate that the relative usage of cleavage sites within one protein can differ substantially.

Similar data were obtained for the immuno-proteasome digest (123 peptides representing 6370 pmol of peptide) (Table 2.5-2). The most abundant cleavage site was found at amino acid position 419, resulting in the generation of 250 pmol peptide (Table 2.5-2 top). As for constitutive proteasomes, many other cleavage sites were used less often, indicating that both proteasome species prefer certain peptide bonds over others as cleavage sites. The almost identical amounts of peptides generated by both proteasomes (6135 pmol versus 6370 pmol) demonstrate a comparable substrate turnover, allowing a direct comparison of pmol amounts of amino acids at different positions around cleavage sites as well as a comparative statistical analysis of both digests.



**Figure 2.5-4: Digestion map generated from degradation of enolase by constitutive proteasomes (A) and immuno-proteasomes (B).**

Vertical lines: cleavage sites determined by Edman degradation and/or MS; solid bars: degradation products identified by Edman degradation in combination with MS; open bars: degradation products identified by Edman degradation only (COOH- terminus of peptide not identified).

**Table 2.5-1: Absolute amounts of amino acids found in positions P6 to P1 and P1' to P6' of peptides generated by constitutive proteasomes.**

	Position		pmol		P6-P1 ↓ P1'-P6'	
	278	265	LTGPQL	ADLYHS		
	404	265	APARSE	RLAKLN		
	383	185	TEDTFI	ADLVVG		
	31	175	EKGVFR	SIVPSG		
	142	165	SKSKTS	PYVLPV		
	183	165	TFAEAL	RIGSEV		
	230	165	LDLIVD	AIKAAG		
	330	155	TNPKRI	ATAIEK		
	146	150	TSPYVL	PVPFLN		
	79	145	PAFVKA	NIDVKD		

	P6	P5	P4	P3	P2	P1	↓	P1'	P2'	P3'	P4'	P5'	P6'
<b>A</b>	865	595	1020	485	620	825	1545	330	710	590	930	365	
<b>C</b>	0	0	0	0	130	0	0	0	0	0	0	0	
<b>D</b>	580	460	255	215	185	805	300	1185	315	235	170	400	
<b>E</b>	530	465	385	455	250	605	300	275	460	150	465	440	
<b>F</b>	50	275	430	165	430	145	190	250	165	575	220	50	
<b>G</b>	300	575	750	230	545	175	185	415	760	185	320	780	
<b>H</b>	175	125	10	150	280	65	165	110	65	35	385	90	
<b>I</b>	505	125	385	335	180	510	60	845	220	450	405	170	
<b>K</b>	220	580	240	945	455	80	365	115	370	675	600	655	
<b>L</b>	540	550	535	525	230	1185	345	900	1120	435	670	565	
<b>M</b>	10	0	95	10	15	35	70	40	75	65	10	60	
<b>N</b>	120	285	235	140	240	95	390	215	195	160	270	730	
<b>P</b>	365	475	550	365	95	40	480	20	255	240	330	315	
<b>Q</b>	45	260	25	20	340	60	75	60	155	305	35	75	
<b>R</b>	125	45	45	425	290	275	595	95	155	145	80	60	
<b>S</b>	575	365	380	90	535	395	500	215	170	530	490	555	
<b>T</b>	735	460	140	365	365	140	150	430	140	360	250	105	
<b>V</b>	270	380	365	920	840	450	275	410	695	705	355	695	
<b>W</b>	40	15	35	30	20	30	0	50	0	0	50	10	
<b>Y</b>	85	100	255	265	90	220	145	175	110	295	100	15	
<b>sum</b>	6135	6135	6135	6135	6135	6135	6135	6135	6135	6135	6135	6135	

**(Top)** 10 most abundant cleavage sites from P6-P1 and P1'-P6', the position of the cleavage site (pos) and the calculated amount of enolase that has been cleaved at this position (pmol) (see also Materials and Methods).

**(Bottom)** Based upon quantified data (as exemplified in **Top**), the absolute amounts of amino acids (in pmol) around all identified cleavage sites are shown.

**Table 2.5-2: Absolute amounts of amino acids found in positions P6 to P1 and P1' to P6' of peptides generated by immuno-proteasomes.**

Position	pmol	P6-P1 ↓	P1'-P6'
419	250	RIEEEL	GDNAVF
390	200	DLVVGL	RTGQIK
183	155	TFAEAL	RIGSEV
313	155	HFFKTA	GIQIVA
133	150	PLYKHL	ADLSKS
142	150	SKSKTS	PYVLPV
253	150	ASSEFF	KDGKYD
285	150	DLYHSL	MKRYPI
289	150	SLMKRY	PIVSIE
383	150	TEDTFI	ADLVVG

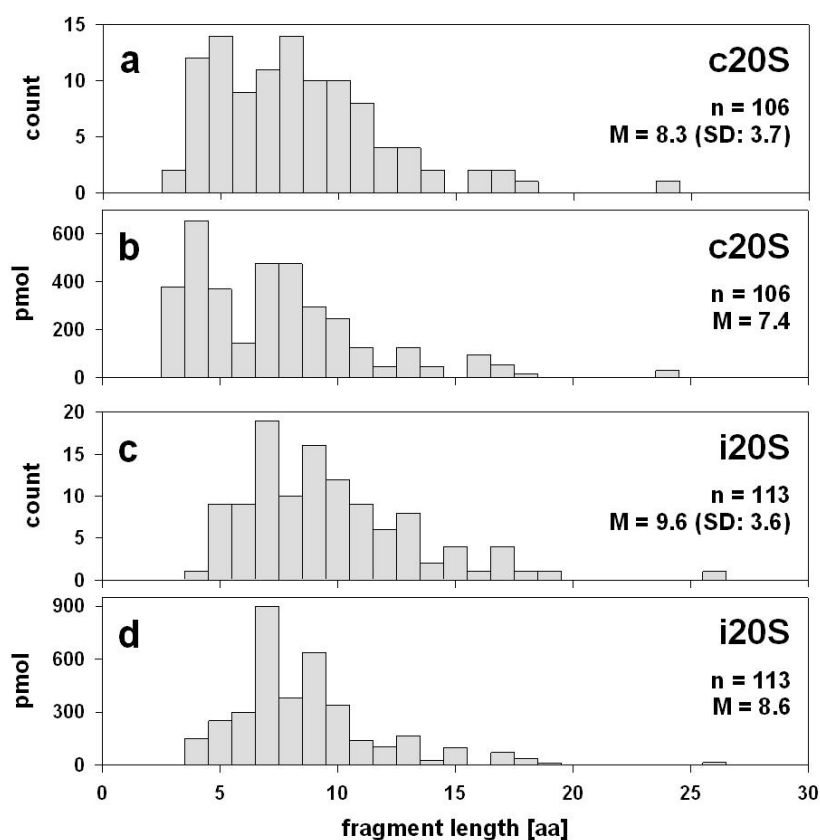
	P6	P5	P4	P3	P2	P1	↓	P1'	P2'	P3'	P4'	P5'	P6'
<b>A</b>	680	615	940	420	590	745	1440	665	560	775	480	810	
<b>C</b>	0	0	0	0	0	0	0	0	0	0	0	0	
<b>D</b>	725	360	190	225	160	150	335	1210	255	325	220	375	
<b>E</b>	300	370	515	870	580	155	255	320	410	270	400	515	
<b>F</b>	0	495	280	105	665	550	50	100	120	260	190	250	
<b>G</b>	500	550	485	180	360	170	880	475	970	265	465	675	
<b>H</b>	300	35	80	285	205	130	140	40	50	40	340	285	
<b>I</b>	440	345	290	470	145	585	20	800	295	420	730	395	
<b>K</b>	90	475	400	1065	620	175	450	275	290	405	590	830	
<b>L</b>	560	975	450	235	160	2305	65	370	1025	410	105	510	
<b>M</b>	20	0	215	10	80	50	295	0	0	100	0	50	
<b>N</b>	245	285	290	165	205	95	520	290	560	575	190	290	
<b>P</b>	325	370	255	380	35	0	375	175	305	180	440	80	
<b>Q</b>	10	260	210	105	170	10	175	75	225	250	100	20	
<b>R</b>	365	100	65	215	395	200	480	0	190	290	320	100	
<b>S</b>	545	250	610	225	735	185	345	320	265	695	380	520	
<b>T</b>	890	200	85	235	570	195	310	560	275	195	275	55	
<b>V</b>	255	485	585	940	695	320	85	535	495	555	925	610	
<b>W</b>	25	60	0	25	0	70	0	0	0	75	0	0	
<b>Y</b>	95	140	425	215	0	280	150	160	80	285	220	0	
<b>sum</b>	6370	6370	6370	6370	6370	6370	6370	6370	6370	6370	6370	6370	

**(Top)** 10 most abundant cleavage sites from P6-P1 and P1'-P6', the position of the cleavage site (pos) and the calculated amount of enolase that has been cleaved at this position (pmol) (see also Materials and Methods).

**(Bottom)** Based upon quantified data (as exemplified in **Top**), the absolute amounts of amino acids (in pmol) around all identified cleavage sites are shown.



Comparison -on a qualitative and quantitative basis- of the fragment length distributions revealed that both proteasome species generated peptides in the size range from 4-24 (c20S) and 3-26 (i20S) amino acids with an average length of 7 - 9 amino acids (Figure 2.5-5). As observed after qualitative analysis, quantitative comparison of the data sets showed that only 30% of the peptides produced by constitutive fragments were also produced by immuno-proteasomes (data not shown). Together, these findings strongly indicate that in cells harbouring constitutive- or immuno-proteasomes, respectively, the pool of peptides generated from proteasomal protein turnover will differ substantially.



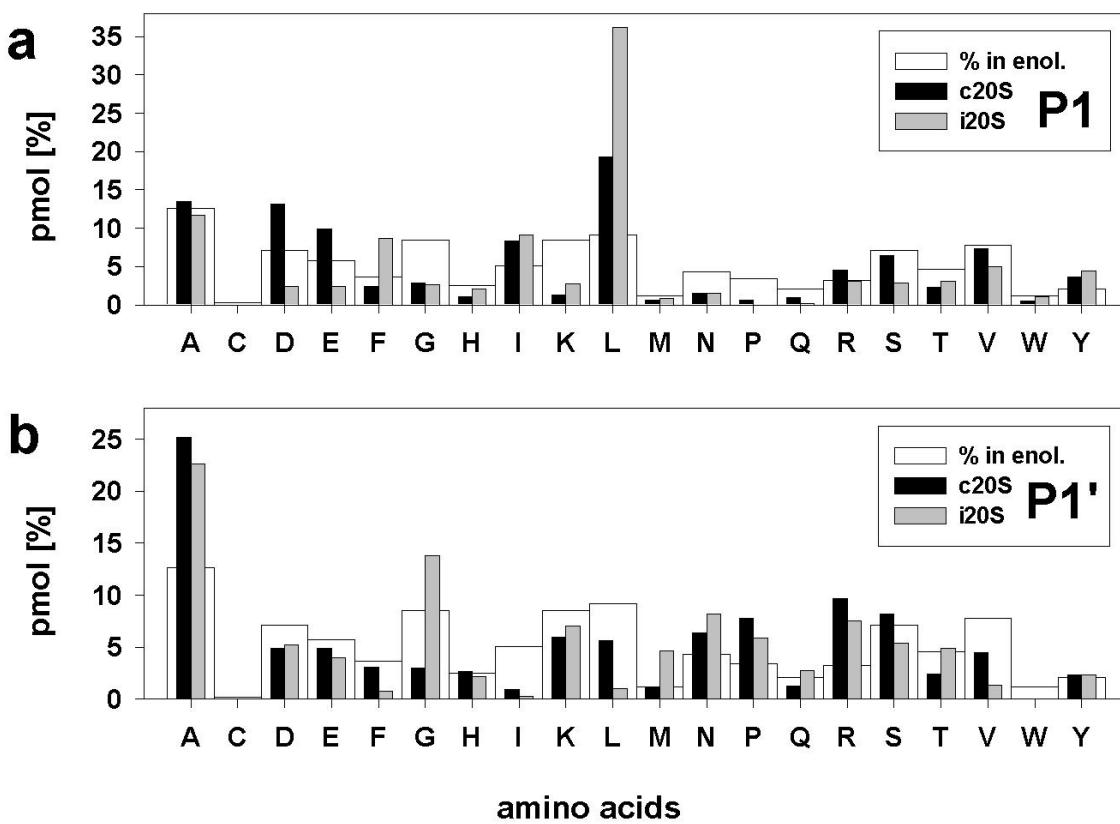
**Figure 2.5-5: Distribution of fragment lengths generated by constitutive proteasomes and immuno-proteasomes.**

The frequencies (a and c) and absolute amounts (b and d) of different fragments with the same size obtained from enolase after digestion with constitutive proteasomes (c and d) and immuno-proteasomes (a and b) is depicted.

#### 2.5.4.4 Analysis of cleavage site-usage

As outlined above, some cleavage sites are used often, whereas others are used less frequently or not at all by one or both proteasome species. Until now, most studies

addressing the specificity of proteasomal peptide/protein degradation did not take the frequency of cleavages into account. To study more accurately the influence of all 20 amino acids flanking proteasomal cleavage sites, we determined the frequencies of amino acids around cleavage sites (P6 to P1 NH<sub>2</sub>-terminal of cleavage site; P1' to P6' COOH-terminal of cleavage site) using the quantified data set described above (Table 2.5-1, Table 2.5-2 and Figure 2.5-6). No statistically significant differences in cleavage site selection were observed between the quantified cleavage data of two independent digests performed by two independent immuno-proteasome batches, indicating that their cleavage preferences were highly reproducible (not shown).



**Figure 2.5-6: Relative frequencies of amino acids in position P1 (a) P1' (b).**

The absolute amount of amino acids found in a defined positions (P6 to P6') around cleavage sites used by constitutive proteasomes and immuoproteasomes was divided by the total amount of peptides detected in the digests resulting in the relative frequency of amino acid-usage in that position. Big white bars: relative frequency of amino acids found in enolase; black bars: relative frequency of amino acids at P1 or P1' positions in peptide-fragments generated by constitutive proteasomes; grey bars: relative frequency of amino acids at P1 or P1' in peptide-fragments generated by immuno-proteasomes.

Examination of cleavages performed by immuno-proteasomes revealed several deviations

from randomness at the amino acid level (i.e. the background in enolase) as determined by  $\text{Chi}^2$  analysis (not shown). L as P1 residue has the highest preference ( $\text{Chi}^2$ : 21) and reflects a pronounced ChT-like activity within immuno-proteasomes. Weaker preferences for T at P6 ( $\text{Chi}^2$ : 5.2) and D at P2' ( $\text{Chi}^2$ : 6.2) are also evident. On the other side, several residues appear to exert a negative influence on cleavage site selection, with the hydrophobic residues I, L and V ( $\text{Chi}^2$ : >4.3) at P1' being the most prominent. Apart from these amino acids at the P1' position, L at P2 and P5', as well as K at P6 (all  $\text{Chi}^2$ : 4) were found less abundantly as expected. When amino acids are grouped according to their characteristics, it was found that there was a positive correlation between cleavage and the presence of hydrophobic, and bulky amino acids in P1 ( $p < 0.0002$ ).

An analogous examination for the cleavages selected by constitutive proteasomes revealed that the frequency of S as P3- and the positively charged amino acid K as P1- and P2'-residue is reduced (all  $\text{Chi}^2$ : 3.9), indicating that these amino acids are disfavored by constitutive proteasomes at these positions. Constitutive proteasomes, like immuno-proteasomes, prefer A at P1' and L as P1 residue (albeit to a lesser extent;  $\text{Chi}^2$ : 4.2). Similarly, an enrichment for D at P2' was noted ( $\text{Chi}^2$ : 8.5), indicating that both proteasome species have a preference for negatively charged amino acids at the P2' position. When comparing characteristics of amino acids at these positions, a preference for hydrophobic, bulky amino acids in P1 ( $p < 0.0002$ ) was observed. There is a negative correlation between cleavage site selection and the presence of hydrophobic, bulky amino acids in P1' ( $p < 0.0002$ ).

#### *2.5.4.5 Effect of immuno-subunit incorporation on cleavage site selection.*

The data described above reveal that both proteasome species exhibit a different, but partially overlapping (L in P1, A in P1' and D in P2') cleavage preference. However, they do not address directly the influence of immuno-subunit incorporation on cleavage site selection. Therefore, we compared absolute and relative amino acid frequencies around cleavage sites used by immuno-proteasomes to those used by constitutive proteasomes. As control, we evaluated cleavage site selection of two independent immuno-proteasome batches, yielding no statistically significant differences (not shown). When comparing the two proteasome species for individual amino acid usage, a strong enrichment of L ( $\text{Chi}^2$ : 7.1 and to a lesser extent F,  $\text{Chi}^2$ : 3.8) was detected at P1 in peptides generated by immuno-proteasomes. The strong enrichment of L became most prominent when we analysed in detail those cleavages exclusively generated by immuno-proteasomes, as

43% of these cleavages were performed after L (data not shown). Acidic amino acids D and E in P1 exert a negative influence on cleavage site selection by immuno-proteasomes ( $\text{Chi}^2$ : 4.5 and 8.1 respectively). G in P1' is enriched in peptides generated by immuno-proteasomes ( $\text{Chi}^2$ : 7.8), pointing to a preference of immuno-proteasomes for this small nonpolar amino acid in P1'. Analysis of amino acid frequencies at positions P2 to P6 and P2' to P6' showed a preference of immuno-proteasomes for V at P5 ( $\text{Chi}^2$ : 4.2), whereas L at P5' and P2' is disfavored by immuno-proteasomes ( $\text{Chi}^2$ : 7.3 and 4.3 respectively). When testing for amino acid characteristics, we found that immuno-proteasomes prefer bulky, hydrophobic amino acids at P1 ( $P < 0.003$ ), but dislike flexible amino acids at this position ( $p < 0.001$ ). Similarly, immuno-proteasomes prefer hydrophobic amino acids in P5 ( $p < 0.03$ ), but disfavor flexible amino acids in this location ( $p < 0.04$ ). The opposite was found in P1': hydrophobic, bulky amino acids were favored more by constitutive proteasomes ( $p < 0.04$ ). Thus, comparison of peptide bonds used as cleavage sites by immuno-proteasomes to the ones used by constitutive proteasomes revealed a pronounced ChT-like activity of immuno-proteasomes, as is reflected by a strong enrichment of hydrophobic amino acids in P1. The results mentioned above are summarized in Table 2.5-3.

### 2.5.5 Discussion

The proteasome is the key enzyme responsible for cytosolic protein degradation and the generation of peptides that are presented by MHC class I molecules by liberation of the appropriate COOH- or  $\text{NH}_2$ - and COOH-terminus. However, proteasomes also can prevent MHC class I presentation by destruction of potential MHC-ligands or their inability to select the proper C-terminal cleavage site. Cleavage site selection might change substantially after incorporation of IFN- $\gamma$ -inducible subunits, as indicated by experiments using small fluorogenic substrates or short synthetic peptides. Therefore, detailed knowledge on the specificity of protein degradation by either constitutive or immuno-proteasomes is vital to the comprehension and prediction of CTL-epitope generation.

To analyse alterations of cleavage preference upon immuno-subunit incorporation, we used an entire protein as substrate without modification of amino acids or bias to known CTL epitopes. The quantitative identification of more than 120 peptide fragments generated by either constitutive- or immuno-proteasomes allowed a detailed analysis of the cleavage specificities. Our data reveal that especially the amino acids in P3 to P3' have a strong influence on cleavage site selection as the observed frequencies of amino acids in these locations showed the largest discrepancy with their background frequencies

in enolase. Nonetheless, amino acids further away most likely influence cleavage site selection as well, as was reported earlier for the effect of P at P4 on peptide bond cleavage by yeast and constitutive proteasomes (Nussbaum et al., 1998; Shimbara et al., 1998; Kuttler et al., 2000; Miconnet et al., 2000). The average length of fragments produced by either set of proteasome is almost the same. The slight increase in the length of fragments generated by immuno-proteasomes can be explained by the lack of cleavages after acidic amino acids. As a consequence more fragments with MHC class I ligand potential are generated.

**Table 2.5-3: Cleavage motifs of constitutive and immuno-proteasomes**

		P6	P5	P4	P3	P2	P1	P1'	P2'	P3'	P4'	P5'	P6'
<b>c20S</b>	+			(P) <i>h</i>	(K) <i>b</i>		L <i>h,b</i>	A	D,I <i>h,b</i>				
	-				S <i>f,h</i>		K <i>f</i>	<i>h,b,f</i>	K				
<b>i20S</b>	+				(K) <i>b,f</i>	<i>b</i>	L <i>h,b</i>	A <i>f</i>	D,I				
	-				<i>h</i>	L <i>h</i>	P <i>f</i>	I,L,V <i>h,b</i>					
<b>i20S/c20S</b>	+						L,F <i>h,b</i>	G <i>f</i>					
	-						D,E <i>f</i>	<i>h</i>	L				

Preferred (+) or disliked (-) aa residues flanking cleavage sites (↓). c20S (i20S): Frequencies of aa surrounding cleavages performed by constitutive (immuno-) 20S proteasomes as compared to the aa frequencies in enolase. i20S/c20S: direct comparison of the aa frequencies around immuno-proteasome cleavages to constitutive proteasome cleavages. The results shown here are based on pmol-quantities. Letters: aa in one-letter code. Brackets: results only slightly significant. Italic: aa characteristics: *f*, flexible, *b*: bulky, *h*, hydrophobic.

By comparing flanking residues around cleavage sites, we found that both proteasomes display a partially overlapping, but different cleavage specificity (Figure 2.5-6 and Table 2.5-3). They both prefer L as P1-residue, whereas K in P3, A in P1' and D in P2' were also favored. Nonetheless, immuno-proteasomes have a much stronger preference for L at P1, as well as other hydrophobic amino acids in this position. These findings point to a

pronounced ChT-like activity after immuno-subunit incorporation. In contrast, the acidic amino acids D and E were clearly disfavored by immuno-proteasomes, whereas these amino acids were enriched at P1 in peptides generated by constitutive proteasomes. Some of the different characteristics regarding amino acid preferences at the P1 position were observed previously in a small number of fragments generated from covalently modified lysozyme using proteasomes from bovine spleen and pituitaries (Cardozo and Kohanski, 1998).

Our observations most likely correlate with the characteristics of amino acids at the inner surface of the proteasome. Analysis of the contribution of individual active  $\beta$ -subunits to cleavage site selection in yeast 20S proteasomes (Heinemeyer et al., 1997; Nussbaum et al., 1998), revealed that the active site of  $\beta 1$  prefers to cleave after acidic residues,  $\beta 2$  after basic residues and  $\beta 5$  after hydrophobic residues. Structural analysis of proteasomes has shown that the pockets around the active site threonine of  $\beta 5$  and  $\beta 2$  do not change after immuno-subunit incorporation (Groll et al., 1997). Therefore, it is most likely that the cleavage site-flanking amino acids that are preferred by both proteasomes (K in P3, L in P1, A in P1' and D in P2') correlate with the characteristics of amino acids in vicinity of the S-1 pockets of  $\beta 5/\beta 5i$  and/or  $\beta 2/\beta 2i$ .

The enhanced preference for hydrophobic amino acids at P1 of immunoproteasomal cleavages does most probably not result from exchange of  $\beta 5$ , which mediates the ChT-like activity, for  $\beta 5i$  (LMP7), as replacement of  $\beta 5$  for  $\beta 5i$  does not alter the pocket surrounding the active site threonine. Moreover, functional data indicate that  $\beta 5i$  influences the structural features of 20S proteasomes, thereby enhancing the activity of  $\beta 1i$  (LMP2) (Sijts et al., 2000a; Gileadi et al., 1999). The stronger preference of immuno-proteasomes for hydrophobic amino acids at P1, however, correlates with the characteristics at the inner surface of the proteasome when  $\beta 1$  (Y) is replaced for  $\beta 1i$ . Replacement of  $\beta 1$  for  $\beta 1i$  is predicted to cause the formation of a more apolar pocket around the active site threonine. This most likely results in the accommodation of hydrophobic amino acids at P1, leading to an enrichment of peptides with hydrophobic COOH-termini, and a reduction of fragments with charged COOH-termini (Groll et al., 1997). Thus, it is conceivable that exchange of  $\beta 1$  for  $\beta 1i$  causes a reduction of peptides with charged COOH-termini (amino acids like D and E), and enhanced liberation of peptides with hydrophobic COOH-termini.

Structural analyses of proteasomes also predict that the active S-1 pocket of  $\beta 5i$  is more

constricted than the pocket of  $\beta 5$  (Groll et al., 1997). As most hydrophobic amino acids are rather bulky, it is conceivable that small, flexible, nonpolar amino acids, like G, in close vicinity to hydrophobic amino acids, support accommodation of these amino acids in the S1 pocket of  $\beta 5i$  most efficiently. Therefore, we anticipate that the strong enrichment of G in P1' of cleavage sites selected by immuno-proteasomes is a direct consequence of these altered characteristics of the  $\beta 5$ -immuno-subunit.



The observation that the P1 preference of immuno-proteasomes closely resembles the preferences of the F-pocket of most MHC class I-molecules is thought to explain why immuno-proteasomes are associated with more efficient processing of MHC class I-ligands. As some CTL-epitopes are more efficiently generated by immuno-proteasomes (Stoltze et al., 2000), it is postulated that immuno-proteasome expression is generally associated with more efficient CTL-epitope generation. However, the correlation between epitope production and expression of immuno-proteasomes is more subtle. Some MHC class I-alleles harbour F-pockets that accommodate amino acids with charged polar side chains (e.g. K and R). These MHC alleles are unlikely to "profit" from immuno-proteasome expression, as the trypsin-like activity is not enhanced by immuno-subunit incorporation. More importantly, current literature indicates that immuno-subunit expression does not exert a positive effect on all CTL-epitopes (Morel et al., 2000). In view of our results, increased CTL-epitope generation can now be explained and also predicted by several different mechanisms. They include enhanced liberation of the proper  $\text{NH}_2$ - and/or  $\text{COOH}$ -terminus resulting from the combination hydrophobic amino acid-small amino acid at P1 and P1' (Table 2.5-4: HbcAg 141-151, LCMV pp89, Adeno E1B), enhanced generation of transporter-associated with antigen processing (TAP)-compatible CTL epitope precursors (Table 2.5-4: LCMV pp89, LCMV NP) or reduced PGPH-activity leaving the CTL-epitope intact (Table 2.5-4: Influenza A NP). However, the presence of a hydrophobic amino acid within a CTL-epitope can also result in its destruction as shown for a CTL epitope derived from the ubiquitous self-protein RU1 (Table 2.5-4, Morel et al., 2000).

The finding that the peptide pool generated by immuno-proteasomes differs from the one produced by constitutive proteasomes leads to important biological consequences. Recently, it was found that dendritic cells (DC) and other professional antigen presenting cells (APC) express immuno-proteasomes (Morel et al., 2000). In contrast, most cells outside the lymphoid system harbour mainly constitutive proteasomes, unless they are exposed to inflammatory stimuli. These findings indicate that the peptide pool available to MHC molecules, and thus for presentation to the immune system, differs substantially

between professional APC and nonlymphoid cells. Professional APC are not only involved in T cell priming, but are also playing a pivotal role in central and peripheral T cell tolerization (Sauter et al., 2000; Huang et al., 2000; Steinman et al., 2000). Professional APC not only tolerize for endogenously expressed antigens, but also for antigens acquired from exogenous sources (cross-tolerization) (Kurts et al., 1997). Although these antigens are derived from nonlymphoid cells that only express constitutive proteasomes, tolerance will be induced against self-peptides that are generated by cells expressing immuno-proteasomes. During viral infection, virus-infected cells but also neighbouring cells will be exposed to inflammatory cytokines, leading to immuno-proteasome expression. Because tolerance has been generated to the peptide pool expressed by cells harbouring immuno-proteasomes, self-reactive immune attack is thus reduced to a minimum.

**Table 2.5-4: Correlation between observed and predicted cleavages preferentially performed by immuno- (i20S) or constitutive (c20S) proteasomes**

Source	MHC-restriction	Influence of IFN- $\gamma$	Sequence	Reference
HbcAg 141-151	HLA-Aw68	+	NAPIL <b>STLPETT</b> VRRRGRSEPR	Sijts et al. (2000), <i>J. Exp. Med.</i>
Influenza A NP	H2-D <sup>b</sup>	+	RGVQI <b>ASNENMD</b> <b>AM</b> DSRTLELR	Dick et al, unpublished
LCMV pp89	H2-L <sup>d</sup>	+	LMY <b>DM</b> <b>YPHFMP</b> TNL GPSEKR	Boes et al. (1994), <i>J. Exp. Med.</i>
LCMV NP	H2-L <sup>d</sup>	+	KIM <b>RTE</b> <b>RPOASGVY</b> MGNLTAQ	Schwarz et al. (2000), <i>J. Immunol.</i>
Adeno E1B	H2-K <sup>b</sup>	+	YKISKL <b>VNIRNCCYI</b> SGNLAE	Sijts et al., (2000), <i>J. Immunol.</i>
RUI	HLA-B51	-	TGSTAV <b>VPYGSFK</b> <b>HV</b> DTRLQ	Morel et al. (2000), <i>Immunity</i>

 i20S preference
  c20S preference

These considerations might also be relevant for induction of CTL-immunity against ubiquitously expressed self proteins. If tolerance to self proteins was restricted to peptides generated by immuno-proteasomes, nonlymphoid cells expressing constitutive proteasomes would display peptides for which no CTL tolerance has been induced. In this respect, it is noteworthy that all CTL-epitopes derived from ubiquitous self-proteins (like RU1, Mage-1, gp100, Melan-A; Table 2.5-4 and Morel et al., 2000) identified thusfar are generated less efficiently by immuno-proteasomes, whereas presentation of viral epitopes is, in general, enhanced after immuno-proteasome expression (Table 2.5-4).



The combination of MHC ligand motifs and constitutive-/immunoproteasomal cleavage motifs for T cell epitope prediction will enhance progress in precise intervention of specific immune responses. It should, for example, now be possible to identify new CTL epitopes preferentially generated by constitutive proteasomes (like the RUI epitope). Such peptides are attractive candidates to be used as cancer vaccines, especially if they are derived from antigens that are, for example, overexpressed in tumors.

## 2.5.6 References

Boes, B., Hengel, H., Ruppert, T., Multhaup, G., Koszinowski, U.H., and Kloetzel, P.M. (1994). Interferon gamma stimulation modulates the proteolytic activity and cleavage site preference of 20S mouse proteasomes. *J. Exp. Med.* 179, 901-909.

Brendel, V., Bucher, P., Nourbakhsh, I.R., Blaisdell, B.E., and Karlin, S. (1992). Methods and algorithms for statistical analysis of protein sequences. *Proc. Natl. Acad. Sci. U. S. A* 89, 2002-2006.

Cardozo, C. and Kohanski, R.A. (1998). Altered properties of the branched chain amino acid-preferring activity contribute to increased cleavages after branched chain residues by the "immunoproteasome". *J. Biol. Chem.* 273, 16764-16770.

Coux, O., Tanaka, K., and Goldberg, A.L. (1996). Structure and functions of the 20S and 26S proteasomes. *Annu. Rev. Biochem.* 65, 801-847.

Craiu, A., Akopian, T.N., Goldberg, A.L., and Rock, K.L. (1997). Two distinct proteolytic processes in the generation of a major histocompatibility complex class I-presented peptide. *Proc. Natl. Acad. Sci. USA* 94, 10850-10855.

DeMars, R., Rudersdorf, R., Chang, C., Petersen, J., Strandtmann, J., Korn, N., Sidwell, B., and Orr, H.T. (1985). Mutations that impair a posttranscriptional step in expression of HLA-A and -B antigens. *Proc. Natl. Acad. Sci. U. S. A* 82, 8183-8187.

Dick, T.P., Nussbaum, A.K., Deeg, M., Heinemeyer, W., Groll, M., Schirle, M., Keilholz, W., Stevanovic, S., Wolf, D.H., Huber, R., Rammensee, H.G., and Schild, H. (1998). Contribution of proteasomal beta-subunits to the cleavage of peptide substrates analyzed with yeast mutants. *J. Biol. Chem.* 273, 25637-25646.

Eleuteri, A.M., Kohanski, R.A., Cardozo, C., and Orlowski, M. (1997). Bovine spleen multicatalytic proteinase complex (proteasome). Replacement of X, Y, and Z subunits by LMP7, LMP2, and MECL1 and changes in properties and specificity. *J. Biol. Chem.* 272, 11824-11831.

Gaczynska, M., Rock, K.L., and Goldberg, A.L. (1993). Gamma-interferon and expression of MHC genes regulate peptide hydrolysis by proteasomes. *Nature* 365, 264-267.

Gaczynska, M., Rock, K.L., Spies, T., and Goldberg, A.L. (1994). Peptidase activities of proteasomes are differentially regulated by the major histocompatibility complex-encoded genes for LMP2 and LMP7. *Proc. Natl. Acad. Sci. USA* 91, 9213-9217.

Gileadi, U., Moins-Teisserenc, H.T., Correa, I., Booth, B.L.J., Dunbar, P.R., Sewell, A.K., Trowsdale, J., Phillips, R.E., and Cerundolo, V. (1999). Generation of an immunodominant CTL epitope is affected by proteasome subunit composition and stability of the antigenic protein. *J. Immunol.* 163, 6045-6052.

Griffin, T.A., Nandi, D., Cruz, M., Fehling, H.J., Kaer, L.V., Monaco, J.J., and Colbert, R.A. (1998). Immunoproteasome assembly: cooperative incorporation of interferon gamma (IFN-gamma)-inducible subunits. *J. Exp. Med.* 187, 97-104.

Groettrup, M., Ruppert, T., Kuehn, L., Seeger, M., Standera, S., Koszinowski, U.H., and Kloetzel, P.M. (1995). The interferon-gamma-inducible 11 S regulator (PA28) and the LMP2/ LMP7 subunits govern the peptide production by the 20 S proteasome in vitro. *J. Biol. Chem.* 270, 23808-23815.

Groll, M., Ditzel, L., Löwe, J., Stock, D., Bochtler, M., Bartunik, H.D., and Huber, R. (1997). Structure of 20S proteasome from yeast at 2.4 Å resolution. *Nature* 386, 463-471.

Heinemeyer, W., Fischer, M., Krimmer, T., Stachon, U., and Wolf, D.H. (1997). The active sites of the eukaryotic 20 S proteasome and their involvement in subunit precursor processing. *J. Biol. Chem.* 272, 25200-25209.

Huang, F.P., Platt, N., Wykes, M., Major, J.R., Powell, T.J., Jenkins, C.D., and MacPherson, G.G. (2000). A discrete subpopulation of dendritic cells transports apoptotic intestinal epithelial cells to T cell areas of mesenteric lymph nodes. *J. Exp. Med.* 191, 435-444.

Kurts, C., Kosaka, H., Carbone, F.R., Miller, J.F., and Heath, W.R. (1997). Class I-restricted cross-presentation of exogenous self-antigens leads to deletion of autoreactive CD8(+) T cells. *J. Exp. Med.* 186, 239-245.

Kuttler, C., Nussbaum, A.K., Dick, T.P., Rammensee, H.G., Schild, H., and Haderl, K.P. (2000). An algorithm for the prediction of proteasomal cleavages. *J. Mol. Biol.* 298, 417-429.

Miconnet, I., Servis, C., Cerottini, J.C., Romero, P., and Levy, F. (2000). Amino acid identity and/or position determines the proteasomal cleavage of the HLA-A\*0201-restricted peptide tumor antigen

MAGE-3271-279. *J. Biol. Chem.* 275, 26892-26897.

Morel, S., Levy, F., Burlet-Schiltz, O., Brasseur, F., Probst-Kepper, M., Peitrequin, A.L., Monsarrat, B., Van Velthoven, R., Cerottini, J.C., Boon, T., Gairin, J.E., and Van den Eynde, B.J. (2000). Processing of some antigens by the standard proteasome but not by the immunoproteasome results in poor presentation by dendritic cells. *Immunity* 12, 107-117.

Nussbaum, A.K., Dick, T.P., Keilholz, W., Schirle, M., Stevanovic, S., Dietz, K., Heinemeyer, W., Groll, M., Wolf, D.H., Huber, R., Rammensee, H.G., and Schild, H. (1998). Cleavage motifs of the yeast 20S proteasome beta subunits deduced from digests of enolase 1. *Proc. Natl. Acad. Sci. USA* 95, 12504-12509.

Pamer, E. and Cresswell, P. (1998). Mechanisms of MHC class I-restricted antigen processing. *Annu. Rev. Immunol.* 16, 323-358.

Rammensee, H.G., Falk, K., and Rötzschke, O. (1993). Peptides naturally presented by MHC class I molecules. *Annu. Rev. Immunol.* 11, 213-244.

Rock, K.L. and Goldberg, A.L. (1999). Degradation of cell proteins and the generation of MHC class I-presented peptides. *Annu. Rev. Immunol.* 17, 739-779.

Sauter, B., Albert, M.L., Francisco, L., Larsson, M., Somersan, S., and Bhardwaj, N. (2000). Consequences of cell death: exposure to necrotic tumor cells, but not primary tissue cells or apoptotic cells, induces the maturation of immunostimulatory dendritic cells. *J. Exp. Med.* 191, 423-434.

Schwarz, K., van Den, B.M., Kostka, S., Kraft, R., Soza, A., Schmidtke, G., Kloetzel, P.M., and Groettrup, M. (2000). Overexpression of the proteasome subunits LMP2, LMP7, and MECL-1, but not PA28 alpha/beta, enhances the presentation of an immunodominant lymphocytic choriomeningitis virus T cell epitope. *J. Immunol.* 165, 768-778.

Shimbara, N., Ogawa, K., Hidaka, Y., Nakajima, H., Yamasaki, N., Niwa, S., Tanahashi, N., and Tanaka, K. (1998). Contribution of proline residue for efficient production of MHC class I ligands by proteasomes. *J. Biol. Chem.* 273, 23062-23071.

Sijts, A.J., Ruppert, T., Rehmann, B., Schmidt, M., Koszinowski, U.H., and Kloetzel, P.M. (2000a). Efficient generation of a hepatitis B virus cytotoxic T lymphocyte epitope requires the structural features of immunoproteasomes. *J. Exp. Med.* 191, 503-514.

Sijts, A.J., Standera, S., Toes, R.E.M., Ruppert, T., Beekman, N.J., van Veelen, P.A., Ossendorp,

F.A., Melief, C.J., and Kloetzel, P.M. (2000b). MHC class I antigen processing of an adenovirus CTL epitope is linked to the levels of immunoproteasomes in infected cells. *J. Immunol.* 164, 4500-4506.

Spies, T., Bresnahan, M., Bahram, S., Arnold, D., Blanck, G., Mellins, E., Pious, D., and DeMars, R. (1990). A gene in the human major histocompatibility complex class II region controlling the class I antigen presentation pathway. *Nature* 348, 744-747.

Steinman, R.M., Turley, S., Mellman, I., and Inaba, K. (2000). The induction of tolerance by dendritic cells that have captured apoptotic cells. *J. Exp. Med.* 191, 411-416.

Stoltze, L., Dick, T.P., Deeg, M., Pömmmerl, B., Rammensee, H.G., and Schild, H. (1998). Generation of the vesicular stomatitis virus nucleoprotein cytotoxic T lymphocyte epitope requires proteasome-dependent and -independent proteolytic activities. *Eur. J. Immunol.* 28, 4029-4036.

Stoltze, L., Nussbaum, A.K., Sijts, A.J., Emmerich, N.P.N., Kloetzel, P.M., and Schild, H. (2000). The function of the proteasome system in MHC class I antigen processing. *Immunol. Today* 21, 317-319.

van Hall, T., Sijts, A.J., Camps, M., Offringa, R., Melief, C.J., Kloetzel, P.M., and Ossendorp, F.A. (2000). Differential Influence on Cytotoxic T Lymphocyte Epitope Presentation by Controlled Expression of Either Proteasome Immunosubunits or PA28. *J. Exp. Med.* 192, 483-494.

## 2.5.7 Abbreviations

aa	amino acid(s)
APC	Antigen presenting cell
DC	Dendritic cell
LCL	lymphoblastoid cell
MS	mass spectrometry / mass spectrometric
TAP	Transporter associated with antigen processing

## 2.5.8 Participating researchers

R. E. M. Toes\*<sup>2</sup>, M. Schirle<sup>1</sup>, M. Kraft<sup>1</sup>, N. P. N. Emmerich<sup>1</sup>, A. Zwinderman<sup>3</sup>, T. P. Dick<sup>4</sup>, J. Müller<sup>5</sup>, B. Schoenfisch<sup>5</sup>, S. Stevanovic<sup>1</sup>, H. G. Rammensee<sup>1</sup> and H. Schild<sup>1</sup>.

<sup>1</sup> Institute for Cell Biology, Department of Immunology, University of Tübingen, Auf der Morgenstelle 15, D-72076 Tübingen, Germany.

<sup>2</sup> Department of Immunohematology and Blood transfusion, Leiden University Medical Center, Box 9600, 2333 AZ Leiden, The Netherlands.

<sup>3</sup> Department of Biostatistics, Leiden University Medical Center, Box 9600, 2333 AZ Leiden, The Netherlands

<sup>4</sup> Section of Immunobiology, Howard Hughes Medical Institute, Yale University School of Medicine, 310 Cedar Street, PO Box 208011, New Haven, CT 06520-8011, USA.

<sup>5</sup> Biomathematik, University of Tübingen, Auf der Morgenstelle 10, D-72076 Tübingen, Germany.

\*R.E.M. Toes and A.K. Nussbaum contributed equally to the results presented in this part.

## 3 Summary

### 3.1 Summary & Perspectives

In my Ph.D. research, I characterized in detail the cleavage specificities of yeast and human proteasomes by *in vitro* (i.e. "in a test tube") experiments and their biochemical and statistical analysis. In collaboration with biomathematicians, our data were used to train computer models of the proteasome in order to apply proteasomal cleavage rules to any aa sequence. Recently, an internet version of the resulting prediction program for proteasomal cleavages PProC has been instated. Thus, I have truly made a research journey "From the test tube to the World Wide Web" during the course of my Ph.D. work. The landmarks of the trip are summarized below.

The way for my Ph.D. thesis work was already paved by two events during my Diploma thesis work: 1) Our lab obtained highly purified wild-type and mutant yeast 20S proteasomes through a collaboration with the groups of Prof. Wolf (Dept. of Biochemistry, University of Stuttgart) and Prof. Huber (MPI for Biochemistry, Martinsried). 2) I found yeast enolase-1 to be a substrate for proteasomal degradation *in vitro*, without prior covalent modification of enolase for its unfolding.

#### 3.1.1 Summary of Results section 2.1

The cleavage specificities of the aforementioned wild-type and mutant yeast proteasomes were initially characterized by digests of peptides containing MHC I ligands (see 2.1; published in Dick et al., 1998). Using inhibitors differentially affecting the active sites of 20S proteasomes, we found that the so-called BrAAP-activity of the proteasome was executed by  $\beta 1$ /Pre3, the active site subunit generally indicated exclusively in post-acidic cleavages. In addition, we found  $\beta 2$ /Pup1, the subunit with T-like activity, to contribute to cleavages after some hydrophobic and small aa. Our results confirmed the assignment of the "classical" activities ChT-, T-like and post-acidic to the subunits  $\beta 5$ ,  $\beta 2$  and  $\beta 1$ , respectively, to be correct. In particular,  $\beta 1$  and  $\beta 2$  exclusively perform post-acidic and T-like cleavages, respectively. The outcome of the specificity analysis is summarized in Table 3.1-1. Moreover, our results for the first time revealed the contribution of individual active site subunits to the generation of MHC I ligands and directly lead to the assumption that it should be possible to "tailor" MHC I processing by active site-specific inhibitors. Unfortunately, such strictly active site-specific inhibitors (in addition to being proteasome-

specific) are not yet available.

**Table 3.1-1: P1-residues used by yeast 20S proteasome active subunits, as deduced from the results in 2.1.**

active $\beta$ -subunit	P1-specificity	
$\beta$ 5/Pre2	large hydrophobic:	L, Y, V, W, F, I, (M)
$\beta$ 1/Pre3	exclusive for:	D, E (post-acidic)
	additional:	L, Y
$\beta$ 2/Pup1	exclusive for:	R, K, H (T-like)
	additional:	A, (L, M)

### 3.1.2 Summary of Results section 2.2

Almost simultaneous to the characterization of the different yeast 20S proteasomes, I managed to generate nice HPLC digestion profiles of enolase using the wild-type and mutant yeast proteasomes. Their biochemical analysis for enolase fragments took many months to be finished. The results of this analysis was presented in part 2.2 (published in Nussbaum et al., 1998). We saw processive degradation of enolase (i.e. no generation of large degradation intermediates) into overlapping fragments, indicating that individual substrate molecules are cleaved slightly differently. If however the observed processivity is brought about by an exit size filter (see 1.3.8.2), individual substrate molecules might all be degraded in exactly the same fashion in a first round of cleavages. Overlapping fragments might then be generated in a second round, a stochastic process influenced by diffusion to active sites or through the exit filter. However, the distribution of cleavage sites along the enolase sequence suggests mutually exclusive cleavages that are difficult to reconcile with the latter hypothesis. As the truth often lies somewhere in between, it is possible that individual substrate molecules go through a stochastic **"get chopped or get away"** process from the beginning of their proteasomal degradation. Diffusion of the substrate aa chain to different active sites (**"get chopped"**) would be controlled by chance much in the same way as the egress of fragments from the 20S proteasome cylinder (**"get away"**). The "affinity" of the substrate sequence towards the active sites as well as the size of the exit filter would be regulators of the process. It is tempting to postulate that the likelihood of cleavage increases with the time a substrate sequence remains in close proximity to the nucleophilic threonines, the time being regulated by how

well the substrate aa sequence fits into the active site pocket. In other words, the flanking sequences of a potential cleavage site would determine the probability for a cleavage. There is ample evidence for this notion, and all the different models explaining the processivity of degradation can be brought to terms with the selection of cleavage sites on the basis of flanking sequences.

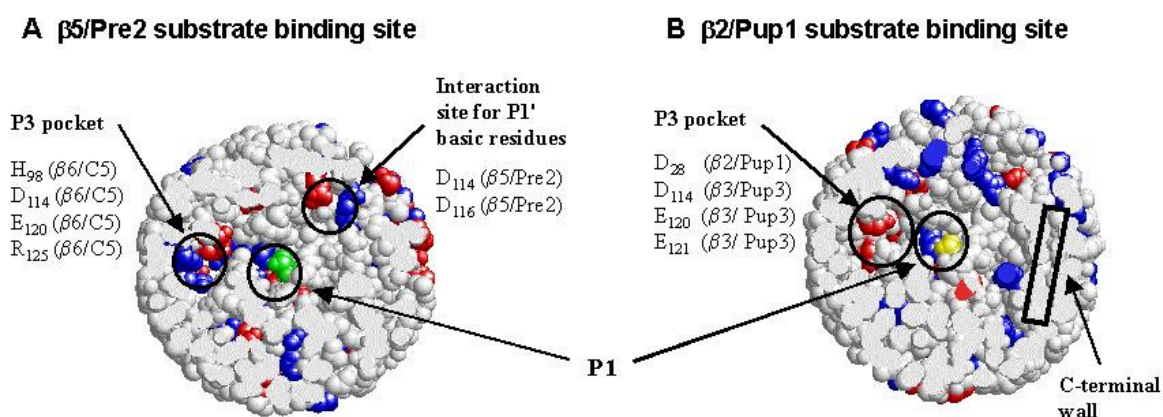
A surprising result of the enolase digestion using wild-type and mutant yeast 20S proteasomes with partly "toothless" active sites was the fact that the distribution of fragment lengths did not significantly vary between the proteasomes. This finding was not in agreement with the classical view of the molecular ruler hypothesis, according to which the distance of active sites in proteasomes determines the fragment length (see 1.3.6). Our results therefore represent the strongest evidence against the molecular ruler hypothesis to date.

The main reason, however, for which we had attempted the time-consuming analysis of so many enolase fragments was the search for the rules governing cleavage site selection by eucaryotic proteasomes. We found that all three active site subunits contributed to the degradation of enolase, as judged by missing cleavage sites in proteasome mutants lacking particular activities. There seems to exist some redundancy in the system, as the cleavages of missing activities were replaced by alternative cleavages of present activities, resulting in similar fragment lengths. This was even true for proteasomes lacking the activity of  $\beta 5/\text{Pre}2$ , a mutation that is lethal in yeast. Similar to the findings in 2.1, we saw that  $\beta 1$  and  $\beta 2$  exclusively perform post-acidic and T-like cleavages, respectively, and that they contribute to cleavages after small and hydrophobic aa, such as A, F, G, I, L, M, N, Q, S, V, W, Y (Table 2.2-1). The most prominent P1-residues found in digests by wt yeast proteasomes were L (selected by  $\beta 5/\text{Pre}2$ ), R ( $\beta 2/\text{Pup}1$ ) and D ( $\beta 1/\text{Pre}3$ ). We could deduce cleavage motifs, i.e. the preferred flanking aa around cleavage sites, for each of the three activities by combinatorial and subtractive analysis of the results obtained using the different mutant proteasomes (Table 2.2-2). Interestingly, the cleavage motifs for  $\beta 5/\text{Pre}2$  and  $\beta 2/\text{Pup}1$  matched the inner surface of the substrate binding pockets of these subunits (Figure 3.1-1).

Unfortunately, there was only one proteasome available in which two activities had been silenced. This proteasome only carried  $\beta 5/\text{Pre}2$  as active subunit. Such 'single-activity' proteasomes would be best to determine the cleavage preferences of the individual active site subunits. However, they seem to be difficult to generate.



Soon after the analysis of the experiments using yeast proteasomes I finished the analysis of an enolase digest using human erythrocyte 20S proteasomes (see 2.2.9). Their cleavage preferences were less pronounced than those of the wild-type yeast proteasomes, but largely overlapped for the size of fragments (Figure 2.2-7) and the cleavage motif (Table 2.2-3).



**Figure 3.1-1: The inner surfaces of the substrate binding pockets of  $\beta 5$ /Pre2 (A) and  $\beta 2$ /Pup1 (B) correlate to the cleavage motifs of these subunits (see Table 2.2-2)**

(A) A pocket lined by polar aa residues could accommodate polar P3-residues found in the cleavage motif of  $\beta 5$ /Pre2. On the P1' side, two Asp residues might be responsible for the enrichment of basic P1' residues in the cleavage motif. (B) An acidic P3 pocket in  $\beta 2$ /Pup1 could interact with Lys residues at P3 of the substrate chain. The enrichment of the bend-promoting aa Gly and Pro on the P'-side of  $\beta 2$ /Pup1-substrates might be imposed by a steric obstacle ("C-terminal wall") that blocks the way. P1: location of the nucleophilic Thr residue. The location of the substrate chain was fitted to the known position of the propeptides found in the crystal structure of mutant yeast proteasomes.

### 3.1.3 Summary of Results sections 2.3 and 2.4

In order to turn our experimental results into a prediction tool for proteasomal cleavages, we sought the help of people at the Department for Biomathematics of Prof. Haderler at the University of Tübingen. To obtain a simple model with few parameters, we constructed a proteasome model inspecting a flanking window of several aa around the cleavage sites. This model was then trained with our different experimental data in the fashion of an evolutionary algorithm, aiming at the reproduction of the training data. Trial and error led to an optimal flanking window P6-P1',P4' (P2' and P3' omitted) around the cleavage site for the model to decide about a cleavage. Together with the aa-pair at P1/P1', especially the residues at P5 and P4 were of great importance for the model. For all sets of experimental data, the algorithms arrived at nearly perfect reproduction of the training

data. A data set of randomly selected cuts was not reproduced as efficiently as actual proteasomal cleavage data, suggesting that there were regularities/rules present in the experimental data and that our mathematical proteasome model could detect them. In fact, the theoretical proteasomes used basically the same kind of rules ("affinity parameters") for cleavage site selection as the rules we had previously determined statistically from our experimental data. The algorithm trained with human proteasome data reached almost 90% precision in the prediction of cleavages that had not been included in the training data (see 2.3, published in Kuttler et al., 2000).

The predictive power of the algorithms was not satisfactory for all test sequences. We found evidence that mainly the lack of training data is to blame for that. It is also possible that our proteasome model is not accurate enough. The current model, for example, weighs the influence of aa residues at certain positions within the inspected window independently of the aa at other positions. However, experimental results indicate that this might be too simple an approach. In order to improve the performance of the prediction it might therefore be necessary to come up with a model that considers the identity of all other aa in the inspected window when judging the influence of a particular aa at a fixed position. Moreover, it might be necessary to include positions P2' and P3' into the inspected window for future sets of experimental data. These positions seem important for cleavage site selection by human 20S proteasomes from LCL cells (see 2.5). We are therefore continuously struggling to refine the model and to generate and gather more training and test data. Refined models that can accommodate quantified experimental cleavage data have already been developed for the experimental data introduced in 2.5.

We have only recently made our prediction algorithms available via World Wide Web under the name PProC (for Prediction Algorithm for Proteasomal Cleavages, [www.paproc.de](http://www.paproc.de); see 2.4; Nussbaum et al., in press), a user-friendly program that can easily be handled by researchers unfamiliar with proteasomes. In addition we have provided background information for users to expand their knowledge about proteasomes. It might be risky to make PProC available at a time when the performance of the predictions is still suffering from the aforementioned teething troubles: Users could lose faith in PProC before it reaches higher predictive power. However, we specifically ask for and hope to get feedback from users that should allow us to improve PProC. In particular, PProC could turn into a platform for the collection of experimental cleavage data. This would probably most efficiently improve the predictions. During the first four months of its existence, PProC has already been accessed more than 1000 times by researchers from all over the world, although it will only be announced in a peer-reviewed

journal in a few months (Nussbaum et al., in press).

In collaborations with S. Brunak/C. Kesmir (Copenhagen), C. Melief/F. Ossendorp (Leiden) and G. Holzhütter/H. Mollenkopf (Berlin) we are trying to establish more reliable proteasomal cleavage prediction. This group of people combines the know-how of experienced proteasome researchers and mathematical and computational skills. The neural networks by C. Kesmir or the algorithms by G. Holzhütter might ultimately perform better than PAMProC, although first comparative results do not support this view. The experimental data generated in the group of F. Ossendorp could complement the training data for PAMProC and other prediction tools. Eventually, all groups working on the prediction of proteasomal cleavages want to create a link between the prediction of MHC I ligands and proteasomal cleavages. For the above-mentioned reasons (see 1.5) such combined prediction could be valuable in the search for CTL epitopes. The "manual" combination of SYFPEITHI and PAMProC has led to a superior prediction of candidates for CTL epitopes from different protein sequences (see Table 2.4-1). However, this did not hold true for all test cases. Again, the lack of training data for the PAMProC algorithms seems to be the reason for weaknesses in the predictive power.

#### 3.1.4 Summary of Results section 2.5

The differential effects of immuno-proteasomes on the generation of CTL epitopes was pointed out above (1.3.7.3). In order to come up with general rules for cleavages by immuno- versus constitutive proteasomes, we compared their cleavage preferences in digests of enolase (part 2.5, Toes et al., submitted for publication). The analysis of enolase fragments was now performed in a strictly quantitative way. No differences were found for the lengths of fragments produced by the two proteasome species; their sizes ranged from 3-26 aa with an average length of 7-9 aa. However, individual fragments and thus cleavage sites only partially overlapped. We found that immuno-proteasomes, as compared to constitutive proteasomes, have a more pronounced preference for cleavages after large, hydrophobic aa (L, F), but do not cleave after acidic residues any more. In addition to this confirmation of earlier results, we could extend the knowledge on cleavage site selection by immuno- and constitutive proteasomes to several residues flanking the cleavage sites (see Table 2.5-3). Interestingly, the cleavage motifs we determined properly explain many effects of immune-proteasomes on CTL epitope processing and presentation (Table 2.5-4). Our quantified data are already being used to train refined algorithms for the prediction of proteasomal cleavages. We are confident that the application of quantified cleavage data for cleavage site prediction and the differentiation

between the cleavage preferences of immuno- and constitutive proteasome will substantially improve the predictive power of PAPProC. Preliminary results show that an algorithm trained with immuno-proteasome cleavages is better at predicting the cleavages at the C-termini of MHC I ligands. However, it is too early to draw final conclusions about the performance of the new algorithms.

## 3.2 Unmentioned and Ongoing Projects

### 3.2.1 Insensitivity of $\beta$ -peptides to proteasomal cleavage<sup>7</sup>

**Peptide-analogs** are amino acid chains with unusual structural features, such as D- instead of L-aa or more stable bonds than peptide bonds to connect the single aa. Despite of their structural differences, peptide analogs can be biologically active. On the other hand, they are usually resistant to enzymatic degradation and therefore stable once brought into a target organism. Long-lived peptide-analogs with biological activity are thus pharmacologically interesting. One group of peptide analogs are  **$\beta$ -peptides**, which are built from aa with an additional  $\text{CH}_2$ -group between the  $\alpha$ -C-atom and the terminal carboxy group. In a collaboration with Jürg Schreiber from the ETH Zürich, I was able to show that neither the multicatalytic proteasome nor very vigorous proteases or protease mixes such as Pronase and Proteinase K were able to degrade  $\beta$ -peptides *in vitro* (Seebach et al., 1998). These results suggest that  $\beta$ -peptides might be long-lived in the cytosol of cells and hence have important implications for their use as pharmacological agents.

### 3.2.2 Directionality – a story with an end?

If proteins are degraded by proteasomes in a processive way (see 1.3.8), then it might be required that the substrate enters the proteolytic chamber in a defined direction, either starting with the N- or with the C-terminus of the aa chain. To find the 'end' of the substrate that is inserted/degraded first, we constructed several modified versions of enolase. By making use of the single Cys residue approximately in the middle of the enolase sequence (at position 247 of 436), we coupled gold-beads, agarose-beads and rhodamine to enolase. We hoped that the bulky obstacle would inhibit full entry of enolase

---

<sup>7</sup> The publication summarized here is part of the prospective Ph.D. theses of Stefan Abele and Jürg V. Schreiber, ETH Zürich.

into the proteasome cylinder, but still expected to see partial enolase degradation. The left-over half of enolase and the generated fragments from the other, degraded half should have made it possible to identify the enolase terminus degraded first. Unfortunately, none of these experiments yielded any result as proteasome activity was in all cases completely blocked, probably by adsorption to the beads or inhibitory effects of the chemical groups coupled to enolase. A different try using miniature dialysis chambers to "pull out" early degradation products was too insensitive to detect casein or enolase fragments before single substrate molecules were completely degraded.

This project, which was originally started by Tobias P. Dick, was continued by David Gatfield, who constructed recombinantly a Casein-derived Artificial Proteasome Substrate (CAPROS) for radioactive labeling during his Diploma thesis project. The project is currently carried on by Ph.D. student Stefan Tenzer who will perform the radioactive labeling and the actual digestion experiments. If it is possible to stop the experiment after one protein terminus has been degraded and before the other terminus has been touched, then it should theoretically be possible to determine a degradation direction for CAPROS. The use of a radioactively labeled protein substrate will hopefully overcome the problem of sensitivity in the previous approaches.

### 3.2.3 The Gly-Ala story

When we were looking for a method to block enolase degradation in order to determine the directionality of degradation, we remembered the observation that the Gly-Ala-repeat region (GARR) from Epstein-Barr virus nuclear antigen 1 (EBNA1) inhibits proteasomal degradation of proteins harboring GARRs (Levitskaya et al., 1995; Levitskaya et al., 1997; Sharipo et al., 1998). The mechanism for this inhibition not being known, we assumed that the GARR physically inhibited proteasomal digestion on the level of the 20S proteasome core. (Later it was reported that a 22 aa long GARR inserted into I $\kappa$ B $\alpha$  exhibits random coil structure, thus eliminating structural stabilization as reason for protection from proteasomal degradation; Leonchiks et al., 1998). It therefore came as a surprise to us when 20S proteasomes *in vitro* were perfectly able to degrade enolase harboring a GARR of about 20 aa. This observation meant we could not use the GARR to determine directionality, but at the same time it raised our interest in the mechanism by which GARRs block proteasomal degradation. Our proteasomal cleavage algorithms predicted GARRs to be spared from proteasomal cleavages (2.3). This hint proved right in digestion experiments using various (roughly 30mer) peptides harboring an 8 aa GARR: Whereas the rest of the peptide was extensively cleaved, the GARR stayed untouched.

Interestingly, the presence of the GARR prolonged the half-life of the GARR-containing peptide substantially, confirming a stabilizing effect for the whole polypeptide, not just for the GARR itself, as already observed with GARR-containing proteins. Unfortunately, we have up to now not been able to investigate the degradation of GARR-containing enolase in cells. Despite the valiant efforts of David Gatfield during his Diploma thesis, we continuously had problems with the immunoprecipitation of cmyc-tagged enolase from the cells in pulse-chase experiments.

In the near future, we would like to examine the *in vitro* digestion of enolase containing longer GARRs, as it was recently reported that the length of the GARR determines protein stability (Dantuma et al., 2000). Besides, we will test whether the related Gly-Ser-Ala-regions from monkey viruses (Blake et al., 1999) also inhibit proteasomal cleavages in *in vitro* peptide substrates. We are currently analyzing whether GARR-containing peptides and proteins interfere with the functions of the 26S proteasome, for example by the binding and therefore competition of the hydrophobic GARRs to ubiquitin-recognition sites in the 19S cap.

The work with GARRs has brought other proteins with unusual aa repeat sequences to my attention, most notably poly-glutamine (polyQ)-containing proteins implicated in the generation of neurodegenerative diseases such as Huntington's disease and several forms of spinocerebellar ataxias (see 2.4 for more information). We have just initiated a collaboration with G.B. Landwehrmeyer and K.S. Lindenberg from the Department for Neurology, Experimental Neurology, University of Ulm, Germany, in order to investigate the role of proteasomes in the accumulation of polyQ-proteins in neurons, a possibly critical factor in the onset of disease.

### 3.2.4 When does the PA28-effect kick in?

Despite earlier reports denying a role for the proteasome activator PA28 in protein degradation (Dubiel et al., 1992), I assumed that PA28 not only participates in the generation of peptides, but must also have a function in the degradation of proteins. Therefore I tried to show that PA28 can accelerate the degradation of FITC-labeled bovine  $\beta$ -casein by human 20S proteasomes *in vitro*, which was not successful: PA28-activated proteasomes with increased activity towards short fluorogenic model substrates did not degrade FITC-casein faster than just 20S proteasomes alone. However, I observed in HPLC-chromatograms that the signals for (what looked like) intermediate degradation products of casein were shifted faster to early retention times (i.e. smaller degradation

products) in the presence of PA28. My interpretation of this result was that the activating effect of PA28 only "kicks in" when the degradation products have reached a certain "smallness". This hypothesis is in line with the latest results that PA28 induces the  $\alpha$ -rings of 20S proteasomes to open (Whitby et al., 2000; see Introduction 1.3.7.1). The diffusion of peptides in and out of the proteolytic chamber, that is most likely regulated by the  $\alpha$ -ring opening, is probably most effective for small peptides. Therefore, once degradation products have reached a threshold "tininess", their increased diffusion could accelerate their degradation into even smaller fragments.

In order to test the hypothesis that PA28-mediated proteasome activation only affects protein fragments of a threshold size, in future experiments substrates of different sizes will be incubated with 20S proteasomes in the presence and absence of PA28. If a threshold size exists, differences in the initial velocity of degradation should be detected for the substrate of the critical size.

## 4 References

- Ahn, J.Y., Tanahashi, N., Akiyama, K., Hisamatsu, H., Noda, C., Tanaka, K., Chung, C.H., Shimbara, N., Willy, P.J., Mott, J.D., Slaughter, C.A., and DeMartino, G.N. (1995). Primary structures of two homologous subunits of PA28, a gamma- interferon-inducible protein activator of the 20S proteasome. *FEBS Lett.* 366, 37-42.
- Akopian, T.N., Kisselev, A.F., and Goldberg, A.L. (1997). Processive degradation of proteins and other catalytic properties of the proteasome from *Thermoplasma acidophilum*. *J. Biol. Chem.* 272, 1791-1798.
- Altuvia, Y. and Margalit, H. (2000). Sequence signals for generation of antigenic peptides by the proteasome: implications for proteasomal cleavage mechanism. *J. Mol. Biol.* 295, 879-890.
- Arendt, C.S. and Hochstrasser, M. (1997). Identification of the yeast 20S proteasome catalytic centers and subunit interactions required for active-site formation. *Proc. Natl. Acad. Sci. USA* 94, 7156-7161.
- Arnold, D., Driscoll, J., Androlewicz, M.J., Hughes, E., Cresswell, P., and Spies, T. (1992). Proteasome subunits encoded in the MHC are not generally required for the processing of peptides bound by MHC class I molecules. *Nature* 360, 171-174.
- Bachmair, A., Finley, D., and Varshavsky, A. (1986). In vivo half-life of a protein is a function of its amino-terminal residue. *Science* 234, 179-186.
- Baumeister, W., Walz, J., Zuhl, F., and Seemüller, E. (1998). The proteasome: paradigm of a self-compartmentalizing protease. *Cell* 92, 367-380.
- Ben-Shahar, S., Komlosch, A., Nadav, E., Shaked, I., Ziv, T., Admon, A., DeMartino, G.N., and Reiss, Y. (1999). 26 S proteasome-mediated production of an authentic major histocompatibility class I-restricted epitope from an intact protein substrate. *J. Biol. Chem.* 274, 21963-21972.
- Benaroudj, N. and Goldberg, A.L. (2000). PAN, the proteasome-activating nucleotidase from archaeobacteria, is a protein-unfolding molecular chaperone. *Nat. Cell Biol.* 2, 833-839.
- Beninga, J., Rock, K.L., and Goldberg, A.L. (1998). Interferon-gamma can stimulate post-proteasomal trimming of the N terminus of an antigenic peptide by inducing leucine aminopeptidase. *J. Biol. Chem.* 273, 18734-18742.
- Bjorkman, P.J., Saper, M.A., Samraoui, B., Bennett, W.S., Strominger, J.L., and Wiley, D.C.



- (1987). Structure of the human class I histocompatibility antigen, HLA-A2. *Nature* 329, 506-512.
- Blake, N.W., Moghaddam, A., Rao, P., Kaur, A., Glickman, R., Cho, Y.G., Marchini, A., Haigh, T., Johnson, R.P., Rickinson, A.B., and Wang, F. (1999). Inhibition of antigen presentation by the glycine/alanine repeat domain is not conserved in simian homologues of Epstein-Barr virus nuclear antigen 1. *J. Virol.* 73, 7381-7389.
- Boes, B., Hengel, H., Ruppert, T., Multhaup, G., Koszinowski, U.H., and Kloetzel, P.M. (1994). Interferon gamma stimulation modulates the proteolytic activity and cleavage site preference of 20S mouse proteasomes. *J. Exp. Med.* 179, 901-909.
- Braun, B.C., Glickman, M., Kraft, R., Dahlmann, B., Kloetzel, P.M., Finley, D., and Schmidt, M. (1999). The base of the proteasome regulatory particle exhibits chaperone-like activity. *Nat. Cell Biol.* 1, 221-226.
- Brendel, V., Bucher, P., Nourbakhsh, I.R., Blaisdell, B.E., and Karlin, S. (1992). Methods and algorithms for statistical analysis of protein sequences. *Proc. Natl. Acad. Sci. U. S. A* 89, 2002-2006.
- Brooks, P., Fuentes, G., Murray, R.Z., Bose, S., Knecht, E., Rechsteiner, M.C., Hendil, K.B., Tanaka, K., Dyson, J., and Rivett, A.J. (2000). Subcellular localization of proteasomes and their regulatory complexes in mammalian cells. *Biochem. J.* 346, 155-161.
- Brown, J.H., Jardetzky, T.S., Gorga, J.C., Stern, L.J., Urban, R.G., Strominger, J.L., and Wiley, D.C. (1993). Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature* 364, 33-39.
- Brusic, V., Rudy, G., Honeyman, G., Hammer, J., and Harrison, L. (1998). Prediction of MHC class II-binding peptides using an evolutionary algorithm and artificial neural network. *Bioinformatics* 14, 121-130.
- Cardozo, C. and Kohanski, R.A. (1998). Altered properties of the branched chain amino acid-preferring activity contribute to increased cleavages after branched chain residues by the "immunoproteasome". *J. Biol. Chem.* 273, 16764-16770.
- Cardozo, C., Michaud, C., and Orlowski, M. (1999). Components of the bovine pituitary multicatalytic proteinase complex (proteasome) cleaving bonds after hydrophobic residues. *Biochemistry* 38, 9768-9777.

Cardozo, C., Vinitsky, A., Michaud, C., and Orlowski, M. (1994). Evidence that the nature of amino acid residues in the P3 position directs substrates to distinct catalytic sites of the pituitary multicatalytic proteinase complex (proteasome). *Biochemistry* 33, 6483-6489.

Cerundolo, V., Benham, A., Braud, V., Mukherjee, S., Gould, K.G., Macino, B., Neefjes, J., and Townsend, A. (1997). The proteasome-specific inhibitor lactacystin blocks presentation of cytotoxic T lymphocyte epitopes in human and murine cells. *Eur. J. Immunol.* 27, 336-341.

Ciechanover, A. (1998). The ubiquitin-proteasome pathway: on protein death and cell life. *EMBO J.* 17, 7151-7160.

Craiu, A., Akopian, T.N., Goldberg, A.L., and Rock, K.L. (1997). Two distinct proteolytic processes in the generation of a major histocompatibility complex class I-presented peptide. *Proc. Natl. Acad. Sci. USA* 94, 10850-10855.

Cresswell, P. (1994). Assembly, transport, and function of MHC class II molecules. *Annu. Rev. Immunol.* 12, 259-293.

Dahlmann, B., Ruppert, T., Kuehn, L., Merforth, S., and Kloetzel, P.M. (2000). Different Proteasome Subtypes in a Single Tissue Exhibit Different Enzymatic Properties. *J. Mol. Biol.* 303, 643-653.

Daniel, S., Brusic, V., Caillat-Zucman, S., Petrovsky, N., Harrison, L., Riganelli, D., Sinigaglia, F., Gallazzi, F., Hammer, J., and van Endert, P.M. (1998). Relationship between peptide selectivities of human transporters associated with antigen processing and HLA class I molecules. *J. Immunol.* 161, 617-624.

Dantuma, N.P., Heessen, S., Lindsten, K., Jellne, M., and Masucci, M.G. (2000). Inhibition of proteasomal degradation by the Gly-Ala repeat of Epstein-Barr virus is influenced by the length of the repeat and the strength of the degradation signal. *Proc. Natl. Acad. Sci. USA* 97, 8381-8385.

DeMartino, G.N. and Slaughter, C.A. (1999). The proteasome, a novel protease regulated by multiple mechanisms. *J. Biol. Chem.* 274, 22123-22126.

Deveraux, Q., Ustrell, V., Pickart, C., and Rechsteiner, M.C. (1994). A 26 S protease subunit that binds ubiquitin conjugates. *J. Biol. Chem.* 269, 7059-7061.

Dick, L.R., Aldrich, C., Jameson, S.C., Moomaw, C.R., Pramanik, B.C., Doyle, C.K., DeMartino, G.N., Bevan, M.J., Forman, J.M., and Slaughter, C.A. (1994). Proteolytic processing of ovalbumin and beta-galactosidase by the proteasome to a yield antigenic peptides. *J. Immunol.* 152, 3884-

3894.

Dick, L.R., Moomaw, C.R., DeMartino, G.N., and Slaughter, C.A. (1991). Degradation of oxidized insulin B chain by the multiproteinase complex macropain (proteasome). *Biochemistry* 30, 2725-2734.

Dick, T.P., Nussbaum, A.K., Deeg, M., Heinemeyer, W., Groll, M., Schirle, M., Keilholz, W., Stevanovic, S., Wolf, D.H., Huber, R., Rammensee, H.G., and Schild, H. (1998). Contribution of proteasomal beta-subunits to the cleavage of peptide substrates analyzed with yeast mutants. *J. Biol. Chem.* 273, 25637-25646.

Dick, T.P., Ruppert, T., Groettrup, M., Kloetzel, P.M., Kuehn, L., Koszinowski, U.H., Stevanovic, S., Schild, H., and Rammensee, H.G. (1996). Coordinated dual cleavages induced by the proteasome regulator PA28 lead to dominant MHC ligands. *Cell* 86, 253-262.

Dolenc, I., Seemüller, E., and Baumeister, W. (1998). Decelerated degradation of short peptides by the 20S proteasome. *FEBS Lett.* 434, 357-361.

Dubiel, W., Pratt, G., Ferrell, K., and Rechsteiner, M.C. (1992). Purification of an 11 S regulator of the multicatalytic protease. *J. Biol. Chem.* 267, 22369-22377.

Ehring, B., Meyer, T.H., Eckerskorn, C., Lottspeich, F., and Tampe, R. (1996). Effects of major-histocompatibility-complex-encoded subunits on the peptidase and proteolytic activities of human 20S proteasomes. Cleavage of proteins and antigenic peptides. *Eur. J. Biochem.* 235, 404-415.

Eleuteri, A.M., Kohanski, R.A., Cardozo, C., and Orlowski, M. (1997). Bovine spleen multicatalytic proteinase complex (proteasome). Replacement of X, Y, and Z subunits by LMP7, LMP2, and MECL1 and changes in properties and specificity. *J. Biol. Chem.* 272, 11824-11831.

Elliott, T., Willis, A., Cerundolo, V., and Townsend, A. (1995). Processing of major histocompatibility class I-restricted antigens in the endoplasmic reticulum. *J. Exp. Med.* 181, 1481-1491.

Emmerich, N.P.N., Nussbaum, A.K., Stevanovic, S., Priemer, M., Toes, R.E.M., Rammensee, H.G., and Schild, H. (2000). The human 26S and 20S proteasomes generate overlapping but different sets of peptide fragments from a model protein substrate. *J. Biol. Chem.* 275, 21140-21148.

Enenkel, C., Lehmann, A., and Kloetzel, P.M. (1998). Subcellular distribution of proteasomes implicates a major location of protein degradation in the nuclear envelope-ER network in yeast.

*EMBO J.* 17, 6144-6154.

Enenkel, C., Lehmann, H., Kipper, J., Guckel, R., Hilt, W., and Wolf, D.H. (1994). PRE3, highly homologous to the human major histocompatibility complex-linked LMP2 (RING12) gene, codes for a yeast proteasome subunit necessary for the peptidylglutamyl-peptide hydrolyzing activity. *FEBS Lett.* 341, 193-196.

Falk, K., Rötzschke, O., and Rammensee, H.G. (1990). Cellular peptide composition governed by major histocompatibility complex class I molecules. *Nature* 348, 248-251.

Falk, K., Rötzschke, O., Stevanovic, S., Jung, G., and Rammensee, H.G. (1991). Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. *Nature* 351, 290-296.

Fehling, H.J., Swat, W., Laplace, C., Kuhn, R., Rajewsky, K., Muller, U., and von Boehmer, H. (1994). MHC class I expression in mice lacking the proteasome subunit LMP-7. *Science* 265, 1234-1237.

Ferrell, K., Deveraux, Q., van Nocker, S., and Rechsteiner, M.C. (1996). Molecular cloning and expression of a multiubiquitin chain binding subunit of the human 26S protease. *FEBS Lett.* 381, 143-148.

Ferrell, K., Wilkinson, C.R., Dubiel, W., and Gordon, C. (2000). Regulatory subunit interactions of the 26S proteasome, a complex problem. *Trends Biochem. Sci.* 25, 83-88.

Früh, K. and Yang, Y. (1999). Antigen presentation by MHC class I and its regulation by interferon gamma. *Curr. Opin. Immunol.* 11, 76-81.

Gaczynska, M., Rock, K.L., and Goldberg, A.L. (1993). Gamma-interferon and expression of MHC genes regulate peptide hydrolysis by proteasomes. *Nature* 365, 264-267.

Gaczynska, M., Rock, K.L., Spies, T., and Goldberg, A.L. (1994). Peptidase activities of proteasomes are differentially regulated by the major histocompatibility complex-encoded genes for LMP2 and LMP7. *Proc. Natl. Acad. Sci. USA* 91, 9213-9217.

Garboczi, D.N., Ghosh, P., Utz, U., Fan, Q.R., Biddison, W.E., and Wiley, D.C. (1996). Structure of the complex between human T-cell receptor, viral peptide and HLA-A2. *Nature* 384, 134-141.

Garcia, K.C., Degano, M., Stanfield, R.L., Brunmark, A., Jackson, M.R., Peterson, P.A., Teyton, L., and Wilson, I.A. (1996). An alphabeta T cell receptor structure at 2.5 Å and its orientation in the TCR-MHC complex. *Science* 274, 209-219.

Geier, E., Pfeifer, G., Wilm, M., Lucchiari-Hartz, M., Baumeister, W., Eichmann, K., and Niedermann, G. (1999). A giant protease with potential to substitute for some functions of the proteasome. *Science* 283, 978-981.

Glas, R., Bogyo, M., McMaster, J.S., Gaczynska, M., and Ploegh, H.L. (1998). A proteolytic system that compensates for loss of proteasome function. *Nature* 392, 618-622.

Glickman, M.H., Rubin, D.M., Coux, O., Wefes, I., Pfeifer, G., Cjeka, Z., Baumeister, W., Fried, V.A., and Finley, D. (1998). A subcomplex of the proteasome regulatory particle required for ubiquitin-conjugate degradation and related to the COP9-signalosome and eIF3. *Cell* 94, 615-623.

Glotzer, M., Murray, A.W., and Kirschner, M.W. (1991). Cyclin is degraded by the ubiquitin pathway. *Nature* 349, 132-138.

Grant, E.P., Michalek, M.T., Goldberg, A.L., and Rock, K.L. (1995). Rate of antigen degradation by the ubiquitin-proteasome pathway influences MHC class I presentation. *J. Immunol.* 155, 3750-3758.

Gray, C.W., Slaughter, C.A., and DeMartino, G.N. (1994). PA28 activator protein forms regulatory caps on proteasome stacked rings. *J. Mol. Biol.* 236, 7-15.

Griffin, T.A., Nandi, D., Cruz, M., Fehling, H.J., Kaer, L.V., Monaco, J.J., and Colbert, R.A. (1998). Immunoproteasome assembly: cooperative incorporation of interferon gamma (IFN-gamma)-inducible subunits. *J. Exp. Med.* 187, 97-104.

Groettrup, M., Ruppert, T., Kuehn, L., Seeger, M., Standera, S., Koszinowski, U.H., and Kloetzel, P.M. (1995). The interferon-gamma-inducible 11 S regulator (PA28) and the LMP2/ LMP7 subunits govern the peptide production by the 20 S proteasome in vitro. *J. Biol. Chem.* 270, 23808-23815.

Groettrup, M., Soza, A., Eggers, M., Kuehn, L., Dick, T.P., Schild, H., Rammensee, H.G., Koszinowski, U.H., and Kloetzel, P.M. (1996). A role for the proteasome regulator PA28alpha in antigen presentation. *Nature* 381, 166-168.

Groll, M., Bajorek, M., Kohler, A., Moroder, L., Rubin, D.M., Huber, R., Glickman, M.H., and Finley, D. (2000). A gated channel into the proteasome core particle. *Nat. Struct. Biol.* 7, 1062-1067.

Groll, M., Ditzel, L., Löwe, J., Stock, D., Bochtler, M., Bartunik, H.D., and Huber, R. (1997). Structure of 20S proteasome from yeast at 2.4 Å resolution. *Nature* 386, 463-471.

Groll, M., Heinemeyer, W., Jager, S., Ullrich, T., Bochtler, M., Wolf, D.H., and Huber, R. (1999).

The catalytic sites of 20S proteasomes and their role in subunit maturation: a mutational and crystallographic study. *Proc. Natl. Acad. Sci. U. S. A* 96, 10976-10983.

Harding, C.V., France, J., Song, R., Farah, J.M., Chatterjee, S., Iqbal, M., and Siman, R. (1995). Novel dipeptide aldehydes are proteasome inhibitors and block the MHC-I antigen-processing pathway. *J. Immunol.* 155, 1767-1775.

Heinemeyer, W., Fischer, M., Krimmer, T., Stachon, U., and Wolf, D.H. (1997). The active sites of the eukaryotic 20 S proteasome and their involvement in subunit precursor processing. *J. Biol. Chem.* 272, 25200-25209.

Heinemeyer, W., Gruhler, A., Mohrle, V., Mahe, Y., and Wolf, D.H. (1993). PRE2, highly homologous to the human major histocompatibility complex-linked RING10 gene, codes for a yeast proteasome subunit necessary for chymotryptic activity and degradation of ubiquitinated proteins. *J. Biol. Chem.* 268, 5115-5120.

Hendil, K.B., Khan, S., and Tanaka, K. (1998). Simultaneous binding of PA28 and PA700 activators to 20 S proteasomes. *Biochem. J.* 332, 749-754.

Hershko, A. and Ciechanover, A. (1998). The ubiquitin system. *Annu. Rev. Biochem.* 67, 425-479.

Hershko, A., Ciechanover, A., and Varshavsky, A. (2000). The ubiquitin system. *Nat. Med.* 6, 1073-1081.

Hilt, W., Enenkel, C., Gruhler, A., Singer, T., and Wolf, D.H. (1993). The PRE4 gene codes for a subunit of the yeast proteasome necessary for peptidylglutamyl-peptide-hydrolyzing activity. Mutations link the proteasome to stress- and ubiquitin-dependent proteolysis. *J. Biol. Chem.* 268, 3479-3486.

Holzhütter, H.G., Frommel, C., and Kloetzel, P.M. (1999). A theoretical approach towards the identification of cleavage-determining amino acid motifs of the 20 S proteasome. *J. Mol. Biol.* 286, 1251-1265.

Holzhütter, H.G. and Kloetzel, P.M. (2000). A kinetic model of vertebrate 20S proteasome accounting for the generation of major proteolytic fragments from oligomeric peptide substrates. *Biophys. J.* 79, 1196-1205.

Honeyman, M.C., Brusica, V., Stone, N.L., and Harrison, L.C. (1998). Neural network-based prediction of candidate T-cell epitopes. *Nat. Biotechnol.* 16, 966-969.

- Hoppe, T., Matuschewski, K., Rape, M., Schlenker, S., Ulrich, H.D., and Jentsch, S. (2000). Activation of a membrane-bound transcription factor by regulated ubiquitin/proteasome-dependent processing. *Cell* 102, 577-586.
- Jiang, H. and Monaco, J.J. (1997). Sequence and expression of mouse proteasome activator PA28 and the related autoantigen Ki. *Immunogenetics* 46, 93-98.
- Kingsbury, D.J., Griffin, T.A., and Colbert, R.A. (2000). Novel propeptide function in 20 S proteasome assembly influences beta subunit composition. *J. Biol. Chem.* 275, 24156-24162.
- Kisselev, A.F., Akopian, T.N., Castillo, V., and Goldberg, A.L. (1999a). Proteasome active sites allosterically regulate each other, suggesting a cyclical bite-chew mechanism for protein breakdown. *Mol. Cell* 4, 395-402.
- Kisselev, A.F., Akopian, T.N., and Goldberg, A.L. (1998). Range of sizes of peptide products generated during degradation of different proteins by archaeal proteasomes. *J. Biol. Chem.* 273, 1982-1989.
- Kisselev, A.F., Akopian, T.N., Woo, K.M., and Goldberg, A.L. (1999b). The sizes of peptides generated from protein by mammalian 26 and 20 S proteasomes. Implications for understanding the degradative mechanism and antigen presentation. *J. Biol. Chem.* 274, 3363-3371.
- Kisselev, A.F., Songyang, Z., and Goldberg, A.L. (2000). Why does threonine, and not serine, function as the active site nucleophile in proteasomes? *J. Biol. Chem.* 275, 14831-14837.
- Koopmann, J.O., Albring, J., Huter, E., Bulbuc, N., Spee, P., Neefjes, J., Hammerling, G.J., and Momburg, F. (2000). Export of antigenic peptides from the endoplasmic reticulum intersects with retrograde protein translocation through the Sec61p channel. *Immunity* 13, 117-127.
- Kopp, F., Hendil, K.B., Dahmann, B., Kristensen, P., Sobek, A., and Uerkvitz, W. (1997). Subunit arrangement in the human 20S proteasome. *Proc. Natl. Acad. Sci. USA* 94, 2939-2944.
- Kuckelkorn, U., Frentzel, S., Kraft, R., Kostka, S., Groettrup, M., and Kloetzel, P.M. (1995). Incorporation of major histocompatibility complex-encoded subunits LMP2 and LMP7 changes the quality of the 20S proteasome polypeptide processing products independent of interferon-gamma. *Eur. J. Immunol.* 25, 2605-2611.
- Kuttler, C., Nussbaum, A.K., Dick, T.P., Rammensee, H.G., Schild, H., and Haderl, K.P. (2000). An algorithm for the prediction of proteasomal cleavages. *J. Mol. Biol.* 298, 417-429.

Lammert, E., Arnold, D., Nijenhuis, M., Momburg, F., Hammerling, G.J., Brunner, J., Stevanovic, S., Rammensee, H.G., and Schild, H. (1997a). The endoplasmic reticulum-resident stress protein gp96 binds peptides translocated by TAP. *Eur. J. Immunol.* 27, 923-927.

Lammert, E., Stevanovic, S., Brunner, J., Rammensee, H.G., and Schild, H. (1997b). Protein disulfide isomerase is the dominant acceptor for peptides translocated into the endoplasmic reticulum. *Eur. J. Immunol.* 27, 1685-1690.

Leonchiks, A., Liepinsh, E., Barishev, M., Sharipo, A., Masucci, M.G., and Otting, G. (1998). Random coil conformation of a Gly/Ala-rich insert in I $\kappa$ B $\alpha$  excludes structural stabilization as the mechanism for protection against proteasomal degradation. *FEBS Lett.* 440, 365-369.

Levitskaya, J., Coram, M., Levitsky, V., Imreh, S., Steigerwald-Mullen, P.M., Klein, G., Kurilla, M.G., and Masucci, M.G. (1995). Inhibition of antigen processing by the internal repeat region of the Epstein-Barr virus nuclear antigen-1. *Nature* 375, 685-688.

Levitskaya, J., Sharipo, A., Leonchiks, A., Ciechanover, A., and Masucci, M.G. (1997). Inhibition of ubiquitin/proteasome-dependent protein degradation by the Gly-Ala repeat domain of the Epstein-Barr virus nuclear antigen 1. *Proc. Natl. Acad. Sci. USA* 94, 12616-12621.

Lin, L., DeMartino, G.N., and Greene, W.C. (1998). Cotranslational biogenesis of NF- $\kappa$ B p50 by the 26S proteasome. *Cell* 92, 819-828.

Loidl, G., Groll, M., Musiol, H.J., Huber, R., and Moroder, L. (1999). Bivalency as a principle for proteasome inhibition. *Proc. Natl. Acad. Sci. USA* 96, 5418-5422.

Löwe, J., Stock, D., Jap, B., Zwickl, P., Baumeister, W., and Huber, R. (1995). Crystal structure of the 20S proteasome from the archaeon *T. acidophilum* at 3.4 Å resolution. *Science* 268, 533-539.

Lucchiari-Hartz, M., van Endert, P.M., Lauvau, G., Maier, R., Meyerhans, A., Mann, D., Eichmann, K., and Niedermann, G. (2000). Cytotoxic T lymphocyte epitopes of HIV-1 nef. Generation Of multiple definitive major histocompatibility complex class I ligands by proteasomes. *J. Exp. Med.* 191, 239-252.

Ma, C.P., Slaughter, C.A., and DeMartino, G.N. (1992). Identification, purification, and characterization of a protein activator (PA28) of the 20 S proteasome (macropain). *J. Biol. Chem.* 267, 10515-10523.

Macagno, A., Gilliet, M., Sallusto, F., Lanzavecchia, A., Nestle, F.O., and Groettrup, M. (1999). Dendritic cells up-regulate immunoproteasomes and the proteasome regulator PA28 during



maturation. *Eur. J. Immunol.* 29, 4037-4042.

Madden, D.R., Gorga, J.C., Strominger, J.L., and Wiley, D.C. (1991). The structure of HLA-B27 reveals nonamer self-peptides bound in an extended conformation. *Nature* 353, 321-325.

McCormack, T.A., Cruikshank, A.A., Grenier, L., Melandri, F.D., Nunes, S.L., Plamondon, L., Stein, R.L., and Dick, L.R. (1998). Kinetic studies of the branched chain amino acid preferring peptidase activity of the 20S proteasome: development of a continuous assay and inhibition by tripeptide aldehydes and clasto-lactacystin beta-lactone. *Biochemistry* 37, 7792-7800.

Michalek, M.T., Grant, E.P., Gramm, C., Goldberg, A.L., and Rock, K.L. (1993). A role for the ubiquitin-dependent proteolytic pathway in MHC class I-restricted antigen presentation. *Nature* 363, 552-554.

Miconnet, I., Servis, C., Cerottini, J.C., Romero, P., and Levy, F. (2000). Amino acid identity and/or position determines the proteasomal cleavage of the HLA-A\*0201-restricted peptide tumor antigen MAGE-3271-279. *J. Biol. Chem.* 275, 26892-26897.

Mo, A.X.Y., Cascio, P., Lemerise, K., Goldberg, A.L., and Rock, K.L. (1999). Distinct proteolytic processes generate the C and N termini of MHC class I-binding peptides. *J. Immunol.* 163, 5851-5859.

Momburg, F., Ortiz-Navarrete, V., Neefjes, J., Goulmy, E., van de Wal, Y., Spits, H., Powis, S.J., Butcher, G.W., Howard, J.C., and Walden, P. (1992). Proteasome subunits encoded by the major histocompatibility complex are not essential for antigen presentation. *Nature* 360, 174-177.

Momburg, F., Roelse, J., Howard, J.C., Butcher, G.W., Hammerling, G.J., and Neefjes, J. (1994). Selectivity of MHC-encoded peptide transporters from human, mouse and rat. *Nature* 367, 648-651.

Morel, S., Levy, F., Burlet-Schiltz, O., Brasseur, F., Probst-Kepper, M., Peitrequin, A.L., Monsarrat, B., Van Velthoven, R., Cerottini, J.C., Boon, T., Gairin, J.E., and Van den Eynde, B.J. (2000). Processing of some antigens by the standard proteasome but not by the immunoproteasome results in poor presentation by dendritic cells. *Immunity* 12, 107-117.

Mott, J.D., Pramanik, B.C., Moomaw, C.R., Afendis, S.J., DeMartino, G.N., and Slaughter, C.A. (1994). PA28, an activator of the 20 S proteasome, is composed of two nonidentical but homologous subunits. *J. Biol. Chem.* 269, 31466-31471.

Murakami, Y., Matsufuji, S., Kameji, T., Hayashi, S.I., Igarashi, K., Tamura, T., Tanaka, K., and

Ichihara, A. (1992). Ornithine decarboxylase is degraded by the 26S proteasome without ubiquitination. *Nature* 360, 597-599.

Murata, S., Kawahara, H., Tohma, S., Yamamoto, K., Kasahara, M., Nabeshima, Y., Tanaka, K., and Chiba, T. (1999). Growth retardation in mice lacking the proteasome activator PA28gamma. *J. Biol. Chem.* 274, 38211-38215.

Newman, R.H., Whitehead, P., Lally, J., Coffey, A., and Freemont, P. (1996). 20S human proteasomes bind with a specific orientation to lipid monolayers in vitro. *Biochim. Biophys. Acta* 1281, 111-116.

Niedermann, G., Butz, S., Ihlenfeldt, H.G., Grimm, R., Lucchiari, M., Hoschutzky, H., Jung, G., Maier, B., and Eichmann, K. (1995). Contribution of proteasome-mediated proteolysis to the hierarchy of epitopes presented by major histocompatibility complex class I molecules. *Immunity* 2, 289-299.

Niedermann, G., Grimm, R., Geier, E., Maurer, M., Realini, C., Gartmann, C., Soll, J., Omura, S., Rechsteiner, M.C., Baumeister, W., and Eichmann, K. (1997). Potential immunocompetence of proteolytic fragments produced by proteasomes before evolution of the vertebrate immune system. *J. Exp. Med.* 186, 209-220.

Niedermann, G., King, G., Butz, S., Birsner, U., Grimm, R., Shabanowitz, J., Hunt, D.F., and Eichmann, K. (1996). The proteolytic fragments generated by vertebrate proteasomes: structural relationships to major histocompatibility complex class I binding peptides. *Proc. Natl. Acad. Sci. USA* 93, 8572-8577.

Nussbaum, A.K., Dick, T.P., Keilholz, W., Schirle, M., Stevanovic, S., Dietz, K., Heinemeyer, W., Groll, M., Wolf, D.H., Huber, R., Rammensee, H.G., and Schild, H. (1998). Cleavage motifs of the yeast 20S proteasome beta subunits deduced from digests of enolase 1. *Proc. Natl. Acad. Sci. USA* 95, 12504-12509.

Orlowski, M. (1990). The multicatalytic proteinase complex, a major extralysosomal proteolytic system. *Biochemistry* 29, 10289-10297.

Orlowski, M., Cardozo, C., Eleuteri, A.M., Kohanski, R.A., Kam, C.M., and Powers, J.C. (1997). Reactions of [<sup>14</sup>C]-3,4-dichloroisocoumarin with subunits of pituitary and spleen multicatalytic proteinase complexes (proteasomes). *Biochemistry* 36, 13946-13953.

Orlowski, M., Cardozo, C., and Michaud, C. (1993). Evidence for the presence of five distinct proteolytic components in the pituitary multicatalytic proteinase complex. Properties of two

- components cleaving bonds on the carboxyl side of branched chain and small neutral amino acids. *Biochemistry* 32, 1563-1572.
- Orlowski, M. and Michaud, C. (1989). Pituitary multicatalytic proteinase complex. Specificity of components and aspects of proteolytic activity. *Biochemistry* 28, 9270-9278.
- Osmulski, P.A. and Gaczynska, M. (2000). Atomic force microscopy reveals two conformations of the 20 S proteasome from fission yeast. *J. Biol. Chem.* 275, 13171-13174.
- Palombella, V.J., Rando, O.J., Goldberg, A.L., and Maniatis, T. (1994). The ubiquitin-proteasome pathway is required for processing the NF- $\kappa$ B1 precursor protein and the activation of NF- $\kappa$ B. *Cell* 78, 773-785.
- Pamer, E. and Cresswell, P. (1998). Mechanisms of MHC class I-restricted antigen processing. *Annu. Rev. Immunol.* 16, 323-358.
- Parham, P., Adams, E.J., and Arnett, K.L. (1995). The origins of HLA-A,B,C polymorphism. *Immunol. Rev.* 143, 141-180.
- Paz, P., Brouwenstijn, N., Perry, R., and Shastri, N. (1999). Discrete proteolytic intermediates in the MHC class I antigen processing pathway and MHC I-dependent peptide trimming in the ER. *Immunity* 11, 241-251.
- Pereira, M.E., Nguyen, T., Wagner, B.J., Margolis, J.W., Yu, B., and Wilk, S. (1992). 3,4-dichloroisocoumarin-induced activation of the degradation of beta-casein by the bovine pituitary multicatalytic proteinase complex. *J. Biol. Chem.* 267, 7949-7955.
- Preckel, T., Fung-Leung, W.P., Cai, Z., Vitiello, A., Salter-Cid, L., Winqvist, O., Wolfe, T.G., Von Herrath, M., Angulo, A., Ghazal, P., Lee, J.D., Fourie, A.M., Wu, Y., Pang, J., Ngo, K., Peterson, P.A., Früh, K., and Yang, Y. (1999). Impaired immunoproteasome assembly and immune responses in PA28<sup>-/-</sup> mice. *Science* 286, 2162-2165.
- Rammensee, H.G., Bachmann, J., Emmerich, N.P.N., Bachor, O.A., and Stevanovic, S. (1999). SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics* 50, 213-219.
- Rammensee, H.G., Falk, K., and Rötzschke, O. (1993). Peptides naturally presented by MHC class I molecules. *Annu. Rev. Immunol.* 11, 213-244.
- Reits, E.A., Vos, J.C., Gromme, M., and Neefjes, J. (2000). The major substrates for TAP in vivo are derived from newly synthesized proteins. *Nature* 404, 774-778.

Rivett, A.J. (1989). The multicatalytic proteinase. Multiple proteolytic activities. *J. Biol. Chem.* 264, 12215-12219.

Rivett, A.J., Palmer, A., and Knecht, E. (1992). Electron microscopic localization of the multicatalytic proteinase complex in rat liver and in cultured cells. *J. Histochem. Cytochem.* 40, 1165-1172.

Rock, K.L., Gramm, C., Rothstein, L., Clark, K., Stein, R., Dick, L., Hwang, D., and Goldberg, A.L. (1994). Inhibitors of the proteasome block the degradation of most cell proteins and the generation of peptides presented on MHC class I molecules. *Cell* 78, 761-771.

Sadasivan, B., Lehner, P.J., Ortmann, B., Spies, T., and Cresswell, P. (1996). Roles for calreticulin and a novel glycoprotein, tapasin, in the interaction of MHC class I molecules with TAP. *Immunity* 5, 103-114.

Schechter, I. and Berger, A. (1967). On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* 27, 157-162.

Scheffner, M., Huibregtse, J.M., Vierstra, R.D., and Howley, P.M. (1993). The HPV-16 E6 and E6-AP complex functions as a ubiquitin-protein ligase in the ubiquitination of p53. *Cell* 75, 495-505.

Scherrer, K. and Bey, F. (1994). The prosomes (multicatalytic proteinases; proteasomes) and their relationship to the untranslated messenger ribonucleoproteins, the cytoskeleton, and cell differentiation. *Prog. Nucleic Acid Res. Mol. Biol.* 49, 1-64.

Schmidtke, G., Emch, S., Groettrup, M., and Holzhütter, H.G. (2000). Evidence for the existence of a non-catalytic modifier site of peptide hydrolysis by the 20 S proteasome. *J. Biol. Chem.* 275, 22056-22063.

Schmidtke, G., Holzhütter, H.G., Bogyo, M., Kairies, N., Groll, M., de Giuli, R., Emch, S., and Groettrup, M. (1999). How an inhibitor of the HIV-1 protease modulates proteasome activity. *J. Biol. Chem.* 274, 35734-35740.

Schubert, U., Anton, L.C., Gibbs, J., Norbury, C.C., Yewdell, J.W., and Bennink, J.R. (2000). Rapid degradation of a large fraction of newly synthesized proteins by proteasomes. *Nature* 404, 770-774.

Schueler-Furman, O., Altuvia, Y., Sette, A., and Margalit, H. (2000). Structure-based prediction of binding peptides to MHC class I molecules: application to a broad range of MHC alleles. *Protein Sci.* 9, 1838-1846.

Schumacher, T.N., Kantesaria, D.V., Heemels, M.T., Ashton Rickardt, P.G., Shepherd, J.C., Früh, K., Yang, Y., Peterson, P.A., Tonegawa, S., and Ploegh, H.L. (1994). Peptide length and sequence specificity of the mouse TAP1/TAP2 translocator. *J. Exp. Med.* 179, 533-540.

Schwarz, K., van Den, B.M., Kostka, S., Kraft, R., Soza, A., Schmidtke, G., Kloetzel, P.M., and Groettrup, M. (2000). Overexpression of the proteasome subunits LMP2, LMP7, and MECL-1, but not PA28 alpha/beta, enhances the presentation of an immunodominant lymphocytic choriomeningitis virus T cell epitope. *J. Immunol.* 165, 768-778.

Seebach, D., Abele, S., Schreiber, J.V., Martinoni, B., Nussbaum, A.K., Schild, H., Schulz, H., Hennecke, H., Woessner, R., and Bitsch, F. (1998). Biological and pharmacokinetic studies with  $\beta$ -peptides. *Chimia* 52, 734-739.

Seemüller, E., Lupas, A., and Baumeister, W. (1996). Autocatalytic processing of the 20S proteasome. *Nature* 382, 468-471.

Seemüller, E., Lupas, A., Stock, D., Löwe, J., Huber, R., and Baumeister, W. (1995). Proteasome from *Thermoplasma acidophilum*: a threonine protease. *Science* 268, 579-582.

Seufert, W., Futcher, B., and Jentsch, S. (1995). Role of a ubiquitin-conjugating enzyme in degradation of S- and M-phase cyclins. *Nature* 373, 78-81.

Seufert, W. and Jentsch, S. (1990). Ubiquitin-conjugating enzymes UBC4 and UBC5 mediate selective degradation of short-lived and abnormal proteins. *EMBO J.* 9, 543-550.

Sharipo, A., Imreh, M., Leonchiks, A., Imreh, S., and Masucci, M.G. (1998). A minimal glycine-alanine repeat prevents the interaction of ubiquitinated I kappaB alpha with the proteasome: a new mechanism for selective inhibition of proteolysis. *Nat. Med.* 4, 939-944.

Sheaff, R.J., Singer, J.D., Swanger, J., Smitherman, M., Roberts, J.M., and Clurman, B.E. (2000). Proteasomal turnover of p21Cip1 does not require p21Cip1 ubiquitination. *Mol. Cell* 5, 403-410.

Shimbara, N., Ogawa, K., Hidaka, Y., Nakajima, H., Yamasaki, N., Niwa, S., Tanahashi, N., and Tanaka, K. (1998). Contribution of proline residue for efficient production of MHC class I ligands by proteasomes. *J. Biol. Chem.* 273, 23062-23071.

Sijts, A.J., Ruppert, T., Rehmann, B., Schmidt, M., Koszinowski, U.H., and Kloetzel, P.M. (2000a). Efficient generation of a hepatitis B virus cytotoxic T lymphocyte epitope requires the structural features of immunoproteasomes. *J. Exp. Med.* 191, 503-514.

Sijts, A.J., Standera, S., Toes, R.E.M., Ruppert, T., Beekman, N.J., van Veelen, P.A., Ossendorp, F.A., Melief, C.J., and Kloetzel, P.M. (2000b). MHC class I antigen processing of an adenovirus CTL epitope is linked to the levels of immunoproteasomes in infected cells. *J. Immunol.* 164, 4500-4506.

Song, X., Mott, J.D., von Kampen, J., Pramanik, B.C., Tanaka, K., Slaughter, C.A., and DeMartino, G.N. (1996). A model for the quaternary structure of the proteasome activator PA28. *J. Biol. Chem.* 271, 26410-26417.

Stohwasser, R., Salzmann, U., Giesebrecht, J., Kloetzel, P.M., and Holzhütter, H.G. (2000). Kinetic evidences for facilitation of peptide channelling by the proteasome activator PA28. *Eur. J. Biochem.* 267, 6221-6230.

Stohwasser, R., Standera, S., Peters, I., Kloetzel, P.M., and Groettrup, M. (1997). Molecular cloning of the mouse proteasome subunits MC14 and MECL- 1: reciprocally regulated tissue expression of interferon-gamma- modulated proteasome subunits. *Eur. J. Immunol.* 27, 1182-1187.

Stoltze, L., Dick, T.P., Deeg, M., Pömmel, B., Rammensee, H.G., and Schild, H. (1998). Generation of the vesicular stomatitis virus nucleoprotein cytotoxic T lymphocyte epitope requires proteasome-dependent and -independent proteolytic activities. *Eur. J. Immunol.* 28, 4029-4036.

Stoltze, L., Nussbaum, A.K., Sijts, A.J., Emmerich, N.P.N., Kloetzel, P.M., and Schild, H. (2000a). The function of the proteasome system in MHC class I antigen processing. *Immunol. Today* 21, 317-319.

Stoltze, L., Schirle, M., Schwarz, G., Schröter, C., Thompson, M.W., Hersh, L.B., Kalbacher, H., Stevanovic, S., Rammensee, H.G., and Schild, H. (2000b). Two new players in the MHC class I antigen processing pathway. *Nat. Immunol.* 1, 413-418.

Strickland, E., Hakala, K., Thomas, P.J., and DeMartino, G.N. (2000). Recognition of misfolding proteins by PA700, the regulatory subcomplex of the 26 S proteasome. *J. Biol. Chem.* 275, 5565-5572.

Tamura, T., Nagy, I., Lupas, A., Lottspeich, F., Cejka, Z., Schoofs, G., Tanaka, K., De Mot, R., and Baumeister, W. (1995). The first characterization of a eubacterial proteasome: the 20S complex of *Rhodococcus*. *Curr. Biol.* 5, 766-774.

Tanahashi, N., Murakami, Y., Minami, Y., Shimbara, N., Hendil, K.B., and Tanaka, K. (2000). Hybrid proteasomes. Induction by interferon-gamma and contribution to ATP-dependent proteolysis. *J. Biol. Chem.* 275, 14336-14345.

Tanahashi, N., Yokota, K., Ahn, J.Y., Chung, C.H., Fujiwara, T., Takahashi, E., DeMartino, G.N., Slaughter, C.A., Toyonaga, T., Yamamura, K., Shimbara, N., and Tanaka, K. (1997). Molecular properties of the proteasome activator PA28 family proteins and gamma-interferon regulation. *Genes Cells* 2, 195-211.

Tarcsa, E., Szymanska, G., Lecker, S., O'Connor, C.M., and Goldberg, A.L. (2000). Ca<sup>2+</sup>-free calmodulin and calmodulin damaged by in vitro aging are selectively degraded by 26 S proteasomes without ubiquitination. *J. Biol. Chem.* 275, 20295-20301.

Theobald, M., Ruppert, T., Kuckelkorn, U., Hernandez, J., Haussler, A., Ferreira, E.A., Liewer, U., Biggs, J., Levine, A.J., Huber, C., Koszinowski, U.H., Kloetzel, P.M., and Sherman, L.A. (1998). The sequence alteration associated with a mutational hotspot in p53 protects cells from lysis by cytotoxic T lymphocytes specific for a flanking peptide epitope. *J. Exp. Med.* 188, 1017-1028.

Townsend, A., Bastin, J., Gould, K.G., Brownlee, G., Andrew, M., Coupar, B., Boyle, D., Chan, S., and Smith, G. (1988). Defective presentation to class I-restricted cytotoxic T lymphocytes in vaccinia-infected cells is overcome by enhanced degradation of antigen. *J. Exp. Med.* 168, 1211-1224.

Uebel, S., Kraas, W., Kienle, S., Wiesmuller, K.H., Jung, G., and Tampe, R. (1997). Recognition principle of the TAP transporter disclosed by combinatorial peptide libraries. *Proc. Natl. Acad. Sci. USA* 94, 8976-8981.

Uebel, S. and Tampe, R. (1999). Specificity of the proteasome and the TAP transporter. *Curr. Opin. Immunol.* 11, 203-208.

Ustrell, V., Pratt, G., and Rechsteiner, M.C. (1995a). Effects of interferon gamma and major histocompatibility complex- encoded subunits on peptidase activities of human multicatalytic proteases. *Proc. Natl. Acad. Sci. USA* 92, 584-588.

Ustrell, V., Realini, C., Pratt, G., and Rechsteiner, M.C. (1995b). Human lymphoblast and erythrocyte multicatalytic proteases: differential peptidase activities and responses to the 11S regulator. *FEBS Lett.* 376, 155-158.

van Endert, P.M. (1999). Genes regulating MHC class I processing of antigen. *Curr. Opin. Immunol.* 11, 82-88.

van Endert, P.M., Riganeli, D., Greco, G., Fleischhauer, K., Sidney, J., Sette, A., and Bach, J.F. (1995). The peptide-binding motif for the human transporter associated with antigen processing. *J. Exp. Med.* 182, 1883-1895.

van Hall, T., Sijts, A.J., Camps, M., Offringa, R., Melief, C.J., Kloetzel, P.M., and Ossendorp, F.A. (2000). Differential Influence on Cytotoxic T Lymphocyte Epitope Presentation by Controlled Expression of Either Proteasome Immunoproteasome Subunits or PA28. *J. Exp. Med.* 192, 483-494.

Van Kaer, L., Ashton Rickardt, P.G., Eichelberger, M., Gaczynska, M., Nagashima, K., Rock, K.L., Goldberg, A.L., Doherty, P.C., and Tonegawa, S. (1994). Altered peptidase and viral-specific T cell response in LMP2 mutant mice. *Immunity* 1, 533-541.

Varshavsky, A. (1996). The N-end rule: functions, mysteries, uses. *Proc. Natl. Acad. Sci. USA* 93, 12142-12149.

Varshavsky, A., Turner, G., Du, F., and Xie, Y. (2000). The ubiquitin system and the N-end rule pathway. *Biol. Chem.* 381, 779-789.

Verma, R. and Deshaies, R.J. (2000). A proteasome howdunit: the case of the missing signal. *Cell* 101, 341-344.

Wang, R., Chait, B.T., Wolf, I., Kohanski, R.A., and Cardozo, C. (1999). Lysozyme degradation by the bovine multicatalytic proteinase complex (proteasome): evidence for a nonprocessive mode of degradation. *Biochemistry* 38, 14573-14581.

Watts, C. (1997). Capture and processing of exogenous antigens for presentation on MHC molecules. *Annu. Rev. Immunol.* 15, 821-850.

Wenzel, T., Eckerskorn, C., Lottspeich, F., and Baumeister, W. (1994). Existence of a molecular ruler in proteasomes suggested by analysis of degradation products. *FEBS Lett.* 349, 205-209.

Whitby, F.G., Masters, E.I., Kramer, L., Knowlton, J.R., Yao, Y., Wang, C.C., and Hill, C.P. (2000). Structural basis for the activation of 20S proteasomes by 11S regulators. *Nature* 408, 115-120.

Yao, Y., Huang, L., Krutchinsky, A., Wong, M.L., Standing, K.G., Burlingame, A.L., and Wang, C.C. (1999). Structural and functional characterizations of the proteasome-activating protein PA26 from *Trypanosoma brucei*. *J. Biol. Chem.* 274, 33921-33930.

Yaron, A., Gonen, H., Alkalay, I., Hatzubai, A., Jung, S., Beyth, S., Mercurio, F., Manning, A.M., Ciechanover, A., and Ben-Neriah, Y. (1997). Inhibition of NF- $\kappa$ B cellular function via specific targeting of the I- $\kappa$ B-ubiquitin ligase. *EMBO J.* 16, 6486-6494.

Yukawa, M., Sakon, M., Kambayashi, J., Shiba, E., Kawasaki, T., Uemura, Y., Murata, K., Tanaka, T., Nakayama, T., and Shibata, H. (1993). Purification and characterization of endogenous protein



activator of human platelet proteasome. *J. Biochem. (Tokyo)* 114, 317-323.

Zaiss, D.M., Standera, S., Holzhütter, H.G., Kloetzel, P.M., and Sijts, A.J. (1999). The proteasome inhibitor PI31 competes with PA28 for binding to 20S proteasomes. *FEBS Lett.* 457, 333-338.

Zinkernagel, R.M. and Doherty, P.C. (1974). Restriction of in vitro T cell-mediated cytotoxicity in lymphocytic choriomeningitis within a syngeneic or semiallogeneic system. *Nature* 248, 701-702.

Zwickl, P., Ng, D., Woo, K.M., Klenk, H.P., and Goldberg, A.L. (1999). An archaeobacterial ATPase, homologous to ATPases in the eukaryotic 26 S proteasome, activates protein breakdown by 20 S proteasomes. *J. Biol. Chem.* 274, 26008-26014.

## 5 Thanks to...

Prof. Dr. Hans-Georg Rammensee: You find a perfect balance between giving the members of your lab a great deal of scientific freedom and still staying in close contact to their research. Your openness to any problems is encouraging, your creativity in case of experimental dead-end-roads is inspiring. Last but not least: Thank you for not making a perfect grade-point average a precondition for joining your lab in 1997.

PD Dr. Hansjörg Schild: Thank you for your guidance and sharing with me your excitement for basic research. Thank you also for the relaxed atmosphere in the group. I wish you continuous success with the recruitment of motivated students and with keeping them motivated. Hau rein!

Prof. Rammensee and Dr. Schild for creating the financial frame for great research. Your support for the participation in congress visits and summer schools provide your students with stimulating opportunities to present data, exchange ideas and even make friends.

Prof. G. Jung, for agreeing to review this dissertation.

The "Sequenzladen": Many thanks to PD Dr. Stefan Stevanović, Markus Schirle, Dr. Wieland Keilholz, Patricia Hrستیć, Melanie Kraft and former member Martin Priemer for introducing me to the world of mass spectrometry and Edman sequencing, for peptide synthesis and for analyzing myriads of cleavage products from proteasomal digests. Your know-how has contributed fundamentally to the success of my Ph.D. research.

Dr. Tobias P. Dick, for introducing me to and raising my interest in the fascinating "protein shredder" proteasome, for showing me what analytical thinking looks like at its best and for continuous, successful collaboration.

Beate Pömmerl for the construction and purification of many recombinant proteins.

All other (former and present) members of the group I have had the pleasure to work with. You will always remain a part of my fond memories of the lab. Sorry for not mentioning each one of you here by name.

All my collaborators:

Dr. René E.M. Toes, for productive teamwork on immuno-proteasomes, for bringing Dutch humor into the lab and for good discussions about a German hero: Rudi Völler.

Dr. Christina Kuttler and Prof. Dr. Karl-Peter Haderer (Biomathematics Department, University Tübingen) and Dr. Can Kesmir (Copenhagen and Utrecht) for close and fruitful teamwork on the prediction of proteasomal cleavages.

Dr. W. Heinemeyer and Prof. Dr. D.H. Wolf (Biochemistry Department, University of Stuttgart), Dr. M. Groll and Prof. Dr. R. Huber (Max-Planck-Institute for Biochemistry, Martinsried, Germany) for providing yeast proteasomes, structural data and expert advice.

The outdoor section: Lars Stoltze for wonderfully organized lab trips by bike (through the Schönbuch), by canoe (on the Neckar) and by car (to L'Alsace). Markus Schirle (Thailand 1998, Iceland 1999) and Lars Stoltze (Iceland 1999, Tatra 2000) for company during exciting vacation trips and beautiful fall hikes on the Swabian Alb. Thanks also to both of you for (up to now unsuccessfully) trying to cure my respect of heights at several climbing and hiking occasions.

The culture club: Markus Schirle for joining me for countless entertaining and relaxing nights at the clubs DEPOT and Jazzkeller ("Bag of Goodies") in Tübingen and for visits at exciting modern art exhibitions in Venice, London and all over Germany.

The French connection: Merci beaucoup à Arnaud Moris, Cécile Gouttefangeas, Steve Pascolo et Patrice Decker pour le flair international dans la laboratoire et l'humeur française, pour le Beaujolais Primeur, le fromage et les croissants.

BATIIa/2 (short for "too little to live, too much to die"): Thank you for paying my rent, vacations, music CDs and for always reminding me that I really wanted to finish my Ph.D. thesis in due time.

My parents: Thank you for your support through all the years. I am eternally grateful for having you.

Last but not least: Sventja, for your unbroken patience with me and for maintaining my emotional balance.

## 6 Abbreviations

aa	aa residue(s) = amino acid(s)
AMC	7-amino-4-methylcoumarin
APC	antigen presenting cell
BrAAP	Branched amino acid peptidase
Bz	Benzoyl
CAPROS	Casein-derived artificial proteasome substrate
ChT	chymotryptic activity
CTL	cytotoxic T lymphocyte
EBNA1	Epstein-Barr virus nuclear antigen 1
EBV	Epstein-Barr virus
EDTA	ethylenediamine-tetraacetate
ER	endoplasmic reticulum
Fmoc	9-fluorenylmethoxycarbonyl
FPLC	fast performance liquid chromatography
GARR	Gly-Ala-rich region
HEPES	N-2-hydroxyethylpiperazin-N'-2-ethansulfonic acid
HLA	human leukocyte antigen
IFN	interferon
Ig	immunoglobulin
JAK	Janus-kinase
LMP	low molecular weight polypeptide
MALDI	matrix-assisted laser desorption ionization
MHC I/II	major histocompatibility complex class I/II
MS	mass spectrometry / mass spectrometric
NP	nucleoprotein
Ntn	N-terminal nucleophile
PAGE	polyacrylamide gel electrophoresis
PAProC	Prediction Algorithm for Proteasomal Cleavages
PDI	protein disulfide isomerase
PGPH	peptidylglutamylpeptide-hydrolysing
polyQ	poly-glutamine
RP-HPLC	reversed phase high performance liquid chromatography
SDS	sodiumdodecylsulfate
SNAAP	small neutral amino acid peptidase
Suc	succinyl
T	tryptic activity
TAP	transporter associated with antigen processing
TCR	T cell receptor
T-like	Trypsin-like

TOF	time of flight
Ub	ubiquitin
VSV	vesicular stomatitis virus
wt	wild-type
WWW	world-wide web
Z	benzyloxycarbonyl
$\beta_2m$	$\beta_2$ -microglobulin

### Abbreviations for amino acid names:

1-letter code	3-letter code	Full name	Selected properties
A	Ala	Alanine	small
C	Cys	Cysteine	small, polar
D	Asp	Aspartic acid	negative charge
E	Glu	Glutamic acid	negative charge
F	Phe	Phenylalanine	bulky, hydrophobic
G	Gly	Glycine	small
H	His	Histidine	positive charge
I	Ile	Isoleucine	branched-chain, hydrophobic
K	Lys	Lysine	positive charge
L	Leu	Leucine	branched-chain, hydrophobic
M	Met	Methionine	hydrophobic
N	Asn	Asparagine	polar
P	Pro	Proline	kinked, polar
Q	Gln	Glutamine	polar
R	Arg	Arginine	positive charge
S	Ser	Serine	small, polar
T	Thr	Threonine	small, polar
V	Val	Valine	branched-chain, hydrophobic
W	Trp	Tryptophan	bulky, hydrophobic
Y	Tyr	Tyrosine	bulky, hydrophobic

## 7 My Academic Teachers

Anderer, Bayer, Bisswanger, Bock, Bohley, Breyer-Pfaff, Büsen, Decker\*, Eisele, Gauglitz, Grabmayer, Günzl, Hagenmaier, Hall\*, Hamprecht, Hanack, Harris\*, Hoffmann, Jung, Kalbacher, Lindner, Little\*, Mayer, Mecke, Metzner, Nakel, Oberhammer, Parker\*, Pelling, Peschel, Pfaff, Pfeiffer, Pommer, Probst, Rammensee, Reutter, Schwarz, Soulages\*, Staudt, Strähle, Stegmann, Voelter, Wells\*, Weser, Wiesinger.

\* During my study-abroad year at the University of Arizona, Tucson, Arizona, USA.

## 8 CV + Publications

### CURRICULUM VITAE



Date of birth:	17.02.1970	Losheim am See, Saarland, Germany
High School:	1980-1990	Hochwaldgymnasium Wadern, Saarland, Germany
	1987-1988	Posen Consolidated High School, Posen, Michigan, USA (High School Graduation), scholarship awarded by the "Parlamentarisches Patenschaftsprogramm" between the German Parliament (Bundestag) and the US Congress
	May 1990	Abitur (German equivalent of High School Graduation) at the Hochwaldgymnasium Wadern
Social Service:	1990-1991	Home for alcoholics, Scheiden, Saarland, Germany
University	1991-1997	Undergraduate and graduate studies of Biochemistry at the University of Tübingen, Germany
	1994-1995	Graduate studies of Biochemistry and Immunology at the University of Arizona, Tucson, Arizona, USA
Diploma thesis:	Apr.-Dec. 1997	Title: "Investigation of the protein degradation by yeast 20S proteasomes using the model substrate yeast enolase-1", in the lab of Prof. Dr. H.-G. Rammensee, Dept. of Immunology, University of Tübingen, Germany
Scholarship:	Nov. 1997-Oct. 2000	Member of the "Graduiertenkolleg Zellbiologie in der Medizin" at the University of Tübingen, Germany
Ph.D. thesis:	Jan. 1998-Dec. 2000	Title: "From the test tube to the World Wide Web - The cleavage specificity of the proteasome", in the lab of Prof. Dr. H.-G. Rammensee in Tübingen, Germany

## LIST OF PUBLICATIONS

T.P. Dick\*, [A.K. Nussbaum](#)\*, M. Deeg, W. Heinemeyer, M. Groll, M. Schirle, W. Keilholz, S. Stevanović, D.H. Wolf, R. Huber, H.-G. Rammensee, and H. Schild (1998). Contribution of proteasomal  $\beta$ -subunits to the cleavage of peptide substrates analyzed with yeast mutants. *J. Biol. Chem.* **273**, 25637-25646.

[A.K. Nussbaum](#)\*, T.P. Dick\*, W. Keilholz, M. Schirle, S. Stevanović, K. Dietz, W. Heinemeyer, M. Groll, D.H. Wolf, R. Huber, H.-G. Rammensee, and H. Schild (1998). Cleavage motifs of the yeast 20S proteasome  $\beta$ -subunits deduced from digests of Enolase-1. *Proc. Natl. Acad. Sci. USA* **95**, 12504-12509.

D. Seebach, S. Abele, J.V. Schreiber, B. Martinoni, [A.K. Nussbaum](#), H. Schild, H. Schulz, H. Henneke, R. Woessner, and F. Bitsch (1998). Biological and pharmacological studies with  $\beta$ -peptides. *Chimia* **52**, 734-739.

C. Kuttler\*, [A.K. Nussbaum](#)\*, T.P. Dick, H.-G. Rammensee, H. Schild and K.-P. Haderl (2000). An algorithm for the prediction of proteasomal cleavages. *J. Mol. Biol.* **298**, 417-429. [Erratum appeared in: *J. Mol. Biol.* **301**, 251]

N.P.N. Emmerich, [A.K. Nussbaum](#), S. Stevanović, M. Priemer, R.E.M. Toes, H.-G. Rammensee and H. Schild (2000). The human 26S and 20S proteasomes generate overlapping but different sets of peptide fragments from a model protein substrate. *J. Biol. Chem.* **275**, 21140-21148.

L. Stoltze, [A.K. Nussbaum](#), A. Sijts, N.P.N. Emmerich, P.-M. Kloetzel and H. Schild (2000). Meeting report: The function of the proteasome system in MHC class I antigen processing. *Immunol. Today* **21**, 317-319.

[A.K. Nussbaum](#), C. Kuttler, K.P. Haderl, H.-G. Rammensee and H. Schild. PProC: A prediction algorithm for proteasomal cleavages available on the World Wide Web. *Immunogenetics* (in press).

C. Kesmir, [A.K. Nussbaum](#), H. Schild and Soren Brunak. Prediction of proteasome cleavage motifs by neural networks. (submitted for publication)

D. Bruder, [A.K. Nussbaum](#), D.M. Gakamsky, M. Schirle, S. Stevanović, H. Singh-Jasuja, H. Schild, A. Darji, T. Chakraborty, I. Pecht, J. Wehland and S. Weiss. Loss of immunodominance of the listeriolysin MHC class I epitope by a conservative Ile-Ala replacement: involvement of multiple synergizing factors (submitted for publication)

R.E.M. Toes\*, [A.K. Nussbaum](#)\*, M. Schirle, M. Kraft, N.P.N. Emmerich, A. Zwinderman, T.P. Dick, J. Müller, B. Schoenfisch, S. Stevanovic, H.-G. Rammensee and H. Schild. Discrete cleavage motifs of constitutive and immunoproteasomes revealed by quantitative analysis of cleavage products (submitted for publication)

\* The first two authors contributed equally.