PRODUCTION AND CHARACTERIZATION OF MHC CLASS II-INVARIANT CHAIN COMPLEX

Dušana Majera

Doctoral Dissertation
Jožef Stefan International Postgraduate School
Ljubljana, Slovenia, May 2012

Evaluation Board:

Prof. DDr. Boris Turk, Jožef Stefan Institute, Ljubljana
Prof. Dr. Janko Kos, Jožef Stefan Institute, Ljubljana
Prof. Dr. Ana Plemenitaš, Institute of Biochemistry, Medical Faculty, University of Ljubljana

MEDNARODNA PODIPLOMSKA ŠOLA JOŽEFA STEFANA
JOŽEF STEFAN INTERNATIONAL POSTGRADUATE SCHOOL

Dušana Majera

PRODUCTION AND CHARACTERIZATION OF MHC CLASS II-INVARIANT CHAIN COMPLEX

Doctoral Dissertation

PROIZVODNJA IN KARAKTERIZACIJA KOMPLEXA INVARIANTE VERIGE Z MHC II MOLEKULO

Doktorska disertacija

Supervisor: Prof. Dr. Dušan Turk

Co-Supervisor: Prof. Dr. Jaques Neefjes

Ljubljana, Slovenia, May 2012

Index

Abstract

Major histocompatibility complex (MHC) class II molecules are polymorphic cell surface glycoproteins expressed only on the surface of the professional antigen presenting cells and they play a crucial role in the adaptive immune responses. MHC class II molecules undergo a very complex maturation process. They are synthesized in the endoplasmic reticulum (ER), where α and β subunits of MHC class II molecules form heterodimers and associate with the trimers of invariant chain (Ii), which targets the complex to late endosomes.

Ii trimerization may be induced by the Ii transmembrane domain. Once the complex arrives at the edosomes, Ii is proteolytically processed by any of the cysteine cathepsins B, K, L, V and S undergoing a very complex regulation control. To enable biochemical studies of the maturation and composition of the MHC class II molecules, its soluble form, complexed with the two major forms, p31 and p41 of Ii, has been expressed in the HEK293T cells at levels of 5 mg/l of the cell culture.

Here we analyzed the assembly and oligomeric state of the Ii and MHC class II complexes lacking the transmembrane region. We have shown by in-vivo and in-vitro analysis that Ii trimerizes also in the absence of the transmembrane part, prior to binding of MHC class II dimers. The biochemical analysis supports the suggestion that the MHC class II-Ii complexes are not necessarily trimers of trimer, as suggested before, but that the Ii trimer can also be occupied by one or two MHC class II dimers. We also demonstrate that p41 inhibitory fragment present on p41 form of Ii inhibited activity of cathepsins L and V, but not cathepsins B and S.

Activity of cysteine cathepsins is regulated also by endogenous inhibitors stefins. Therefore I was involved in biochemical and structural characterization of interactions between cathepsins B and V and stefin A.

Povzetek

Molekule glavnega histokompatibilnega kompleksa (MHC) tipa II so polimorfni površinski glikoproteini, primarno izraženi na specializiranih antigen predstavitvenih celicah. Igrajo bistveno vlogo v adaptivnem imunskem odgovoru. MHC molekule tipa II so podvržene zelo zapletenemu procesu zorenja. Njihova sinteza poteka v endoplazmatskem retikulumu (ER), kjer α in β podenote tvorijo dimere, ki se združijo s trimeri invarijantne verige (Ii), ki usmeri kompleks v pozne lisozome

Trimerizacija Ii je najbrž vzpodbujena s transmembransko regijo Ii. Po prohidu kompleksa v lisozome, se Ii proteolitsko procesira s cisteinskimi katepsini B, K, L, V ali S. Procesiranje Ii je verjetno regulirano precej zapleteno. Za biokemijske študije zorenja in sestave MHC II molekul , smo topno obliko le teh izrazili v kompleksu s p31 in p41 obliko Ii v HEK293T celicah. Nivo ekspresije obeh kompleksov je bil 5 mg/l celične kulture.

V okviru doktorskega dela smo analizirali kompleks MHC molekul razreda II z Ii brez transmembranske regije in njegovo oligomerno sestavo. Pokazali smo, da Ii trimerizira tudi brez transmembranske regije, pred vezavo dimerov MHC molekul razreda II. Biokemijska analiza podpira predlog, da kompleksi MHC molekul razreda II in Ii niso nujno trimer trimerov kot je bilo objavljeno prej, temveč da je trimer Ii lako zaseden tudi le z eno ali dvema dimeroma MHC molekul razreda II. Pokazali smo tudi, da inhibicijski fragment p41 vsebovan v p41 obliki Ii lahko inhibira katepsina L in V ne pa katepsinov B in S.

Aktivnost cisteinskih katepsinov je regulirana tudi z endogenimi inhibitrojikot so stefini. V okviru dokotrkega dela sem bil vključena tudi v študije biokemične in strukturne karaterizacije interakcij me katepsinoma B in V in stefinom A.

Abbreviation

AMC	=	7-amino-4-methyl ccoumarin
APC	=	Antigen presenting cell
BCR	=	B cell receptor
Cath	=	Cathepsin
CD4	=	Cluster of differentiation 4
CLIP	=	Class II associated Ii chain peptide
Co-IP	=	Co-immunoprecipitation
DCs	=	Dentritic cells razlaga
DMEM	=	Dulbecco´s modified Eagle medium
DMSO	=	Dimethyl sulfoxide)
DRA	=	MHC class II alpha chain
DRB	=	MHC class II beta chain
DTT	=	Ditiotreitol
EDTA	=	Ethylenediaminetetraacetic acid
ER	=	Enodplasmic reticulum
FBS	=	Fetal bovine serum
GILT	=	Gamma-interferon inducible lysosomal thiol reductase
HEK	=	Human embryonic kydney
HLA	=	Human leukocyte antigen
HRP	=	Horse radish peroxidase
IgG	=	Immunoglobulin G
Ii	=	Invariant chain
IMAC	=	Immobilized metal-affinity chromatography
MCS	=	Multi cloning site
MHC	=	Major histocompatibility comlex
NFκB	=	Nuclear factor κB
PAGE	=	Polyacrylamide gel electrophoresis
PAMP	=	Pathogen asocciated molecular patherns
PCR	=	Polymerase chain reaction
PBS	=	Phosphate buffered saline
PEI	=	Polyethylenimine
PENSTREP	=	Mixture of Penicillin and Streptomycin
SDS	=	Sodium dodecyl sulfate
T cells	=	Thymic cells
TBST	=	Tris-buffered saline Tween 20
TCR	=	T cell receptor
TM	=	Transmembrane domain
TRIS	=	Tris (hidroksimetil) aminometan
TRL	=	Toll-like receptor

1 Introduction

1.1 Immune system

The immune system comprises a network of cells, tissues, organs and processes within an organism that protects against diseases by identifying and inactivating or killing pathogens and tumor cells. Traditionally it has been divided into innate and adaptive components, each with a different function and role. The effector mechanisms of innate immunity, which include antimicrobial peptides, epithelial cells, granulocytes, phagocytes, and the alternative complement pathway, are activated immediately after infection and rapidly control the replication of the infecting pathogen or neutralization of the toxin. This type of immunity must have appeared early in the evolution of multi-cellular organisms. The adaptive component is organized around two classes of specialized cells, T and B lymphocytes. Each lymphocyte displays a single structurally unique receptor. When it encounters the antigen that binds to this receptor, the lymphocyte is activated and proliferates. This process is called clonal selection. Clonal expansion of lymphocytes takes three to five days and is absolutely necessary for the generation of an effective immune response.

1.1.1 Strategies of innate and adaptive immune responses

The main distinction between adaptive and innate immune systems lie in mechanism of the receptors involved in the immune recognition. In the adaptive immune system, the T cell and the B cell receptors are generated somatically, during their development. Since these receptors are not encoded in the germ line, they are not predestined to recognize any particular antigen. A diverse repertoire of receptors is generated stochastically within the population of lymphocytes, and cells bearing the receptors that recognize the antigens for which they are specific are selected for clonal expansion. Since the origin of antigenic specificity is stochastic, the population of resulting receptors contains those that can bind not only to foreign antigens but also to self antigens. The latter can lead to autoimmune diseases and allergies (Medzihitov and Janeway, 2000).

The innate immune response does not recognize every possible antigen, but rather is focused on a few, highly conserved structures present in large groups of microorganisms (Janeway, 1992). These structures are reffered to as pathogen-associated molecular paterns (PAMP). The best known examples of PAMPs are bacterial polysaccharides, peptidoglycans, bacterial DNA and glucans. Although these structures are chemically quite distinct, they are produced only by pathogens and not by the mammalian host; they are essential for the survival or pathogenicity of microorganisms and are shared by entire classes of pathogens (Medzihitov R. et al., 1997). One of the receptors that recognize PAMP are the toll-like receptors (TLRs) that induces the signal-transcription factor of the nuclear factor-κB (NFκB) family (Belvin et al., 1996). Members of this family have a key role in the induction of immune and inflammatory responses in mammals (Ghosh et l., 1998).

The innate system induces costimulators, cytokines, and chemokines, which recruit and activate antigen specific lymphocytes and initiate adaptive immune responses. Thus, the adaptive immune system responds to a pathogen only after it has been recognized by the innate immune system.

1.1.2 Adaptive immune system

The main distinction between innate and adaptive immune cells is the ability of the latter to express unique, highly antigen-specific receptors on their plasma membrane. Although T and B cells have similar mechanism for producing antigen-specific receptors, the way they recognize antigen differs. B cells recognize antigens that are bound directly to the B-cell antigen receptor (BCR), a membrabe-associated form of immunoglobulin. Consequently, B cells have no need for antigen presenting cells (APC). In contrast, T cells need an antigen to be presented, because T-cell antigen receptor (TCR) recognizes antigen fragments only when they are bound to the MHC molecules on the surface of APC, such as dentritic cells, macrophages and B cells.

1.1.3 T lymphocytes

Antigen-recognition and signal-transducing elements associate at the plasma membrane to form the T-cell receptor (TCR) complex. Like other immunoglobulin genes, the TCR genes undergo DNA rearrangement of variable (V), diversity (D), joining (J), and constant (C) region segments. Diversity-generating mechanisms allow humans to produce 1016 possible antigen-specific alpha-beta T-cell receptors. V-region gene segments are highly polymorphic, and recombinant assembly of the different V, D, and J segments allows for a wide array of possible antigen receptors. Each T-cell clone expresses only one antigen receptor, and antigen-specifi c T cells develop in non-immunized or naive individuals independently of exposure to antigen. Subsequent exposure leads to clonal selection or expansion of lymphocytes with the appropriate antigen receptors. Clonal selection improves the effectivness of the immune response and is responsible for immunological memory. Mutations in the genetic machinery for V (D)J recombination have been linked to a variety of severe combined immunodeficiency conditions in humans (Kovanen and Leonard, 2004).

When the TCR engages the MHC–antigen complex of an APC, rearrangement of the plasma membrane surrounding the TCR changes the relative distribution of protein tyrosine kinases and tyrosine phosphatases near the cell surface. This redistribution of signaling molecules is referred to as the immune synapse (Davis and van der Merwe 2006) (Figure 1). T-cells are also activated by cytokines, some of which are made by other cells, and some by T-cells themselves. Occupancy of the T-cell receptor alone does not lead to T-cell activation; occupancy of other costimulatory molecules is necessary to provide a full activation signal (Chikuma and Bluestone 2003).

1.1.3.1 T cell development

Precursor T cells originate in the bone marrow from hematopoietic stem cells and migrate to the thymus. As they mature, they move from the thymic cortex to the medulla. The most immature T cells, known as doublenegative (DN) T cells, lack the surface markers CD4 and CD8; these cells do not express a mature TCR. These cells mature into doublepositive (DP) T cells, which express CD4 and CD8. DP thymocytes mature to become single-positive (SP) thymocytes, which express either CD4 or CD8, as well as a complete TCR. Mature SP T cells migrate out from the thymic medulla to populate the peripheral lymphoid tissues.

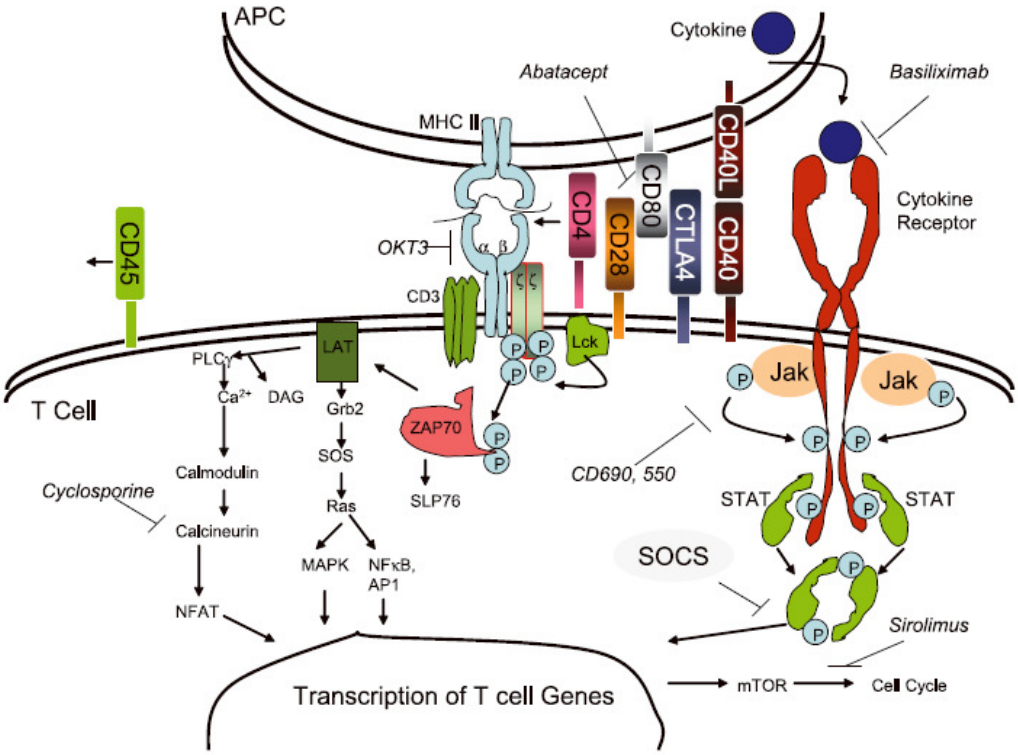


Figure 1: *T cell and APC interact at the immune synapse.* The T-cell receptor, comprised of the alpha beta subunits, CD3 invariant chains, and CD4 coreceptor, contacts the MHC-peptide complex. Costimulatory signals are transmitted through a series of molecules, including CD28 and CD40L. Engagement of peptide in the context of MHC brings CD4-linked tyrosine kinase Lck into proximity with the alpha beta subunits and CD3. This triggers a phosphorylation cascade that ultimately activates the transcription factors. Drugs that block the activation of T cells are shown in italics.(Elias et al., 2008).

The vast majority of T cell precursors generated in the thymus die there. Much of the T-cell death that occurs in the thymus is due to programmed cell death (apoptosis). To survive, potential T cells must first produce TCRs at the DP stage that can recognize self-MHC molecules, a process called positive selection. Of the SP thymocytes that do develop, some recognize self-MHC molecules and self-peptides with high avidity. These potentially autoreactive T cells are also eliminated in a process called negative selection or clonal deletion. The removal of potentially harmful T cells in the thymus is termed central tolerance.

1.1.3.2 Types of T lymphocytes

CD4+ T cells are also known as T-helper cells because of their role in promoting the function of other immune cells. Classically, they have been thought to differentiate into one of two primary effector cell types T-helper 1 (Th1) and T-helper 2 (Th2) cells. T-helper 1 cells secrete cytokines that promote cell-mediated immunity. The key Th1 cytokine is IFNγ, which enhances the ability of macrophages to kill ingested microorganisms, and up-regulates MHC class I expression on many cell types.

Another set of CD4+ cells, dubbed T regulatory (Treg) cells, is essential for limiting immune responses; as such, it is critical for peripheral tolerance (Kronenber and Rudensky 2005). Treg cells have the remarkable ability to suppress proliferation of, and cytokine production, by effector T cells. CD8+ T cells, also known as cytotoxic T lymphocytes (CTLs), recognize antigens in the context of MHC I molecules.

Because MHC I molecules present antigens synthesized by the target cell, CTLs play a prominent role in defence against intracellular pathogens, particularly viruses, protozoa, and bacteria, as well as cancer. A small number of CD8+ T cells become memory T cells. Cytotoxic T lymphocytes kill target cells directly using perforin and granzymes.

1.1.4 B lymphocytes

B lymphocytes constitute another major component of the adaptive immune response. They first develop in the fetal liver and later mature in the bone marrow. B cells produce immunoglobulins or antibodies and can also function as APCs. Immunoglobulins function as antigen receptors on B cells, but are also secreted and have many functions in host defence. Immunoglobulin molecules basically consist of four polypeptide chains: two identical light (L) chains and two identical heavy (H) chains, both of which have variable and constant regions. Structurally, the four chains assemble to form a Y-shaped molecule. The variable regions bind antigens, which, unlike in T cells, encompass many types of molecules, including proteins, lipids, carbohydrates and nucleic acids.

The BCR signal pathway is similar to that of the TCR but, unlike the TCR, it has two functions. It initiates signals that activate the B cells to proliferate but also binds and internalizes antigens. The antigens are processed, loaded onto class II molecules, and presented to CD4+ cells.

1.2 MHC molecules

APCs expose on their surface small peptides derived from antigenic proteins loaded onto the peptide binding cavity of the MHC class I or class II molecules. These antigens are recognized by T cells. Although structurally similar, MHC class I and MHC class II molecules present antigenic peptides derived from proteins that are degraded at two separate locations within the APC and they employ very different machineries.

1.2.1. MHC class I molecules

The major histocompatibility (MHC) class I molecules are heterodimers of a heavy chain, a 45 000 MW type I integral membrane glycoprotein, and β_2-microglobulin (M), a 12 000 MW soluble protein (Jones EY, 1997) and located on every nucleated cell of the body. Their function is to present peptide fragments derived from intracellular proteins. These peptides are normally derived from the cell's own ´house-keeping´ proteins but in a virally infected cell, peptides derived from viral proteins may also be presented. MHC class I molecules fold and assemble within the endoplasmic reticulum (ER) lumen and peptide binding is an integral part of the assembly process.The polypeptides that access the cytosol are degraded by the proteasome or other cytoplasmic proteases (York et al 1999). It is necessary to translocate peptides from the cytosol into the ER lumen and this function is performed by the transporter associated with antigen processing (TAP) (van Endert PM et al 2002). These peptides then bind to newly synthesized MHC class I molecules in a reaction that involves several ER-resident proteins (Purcell, 2000). Protein synthesis occurs in the cytosol and endogenous proteins or incompletely translated products can be presented via MHC class I (Reits et al., 2000).

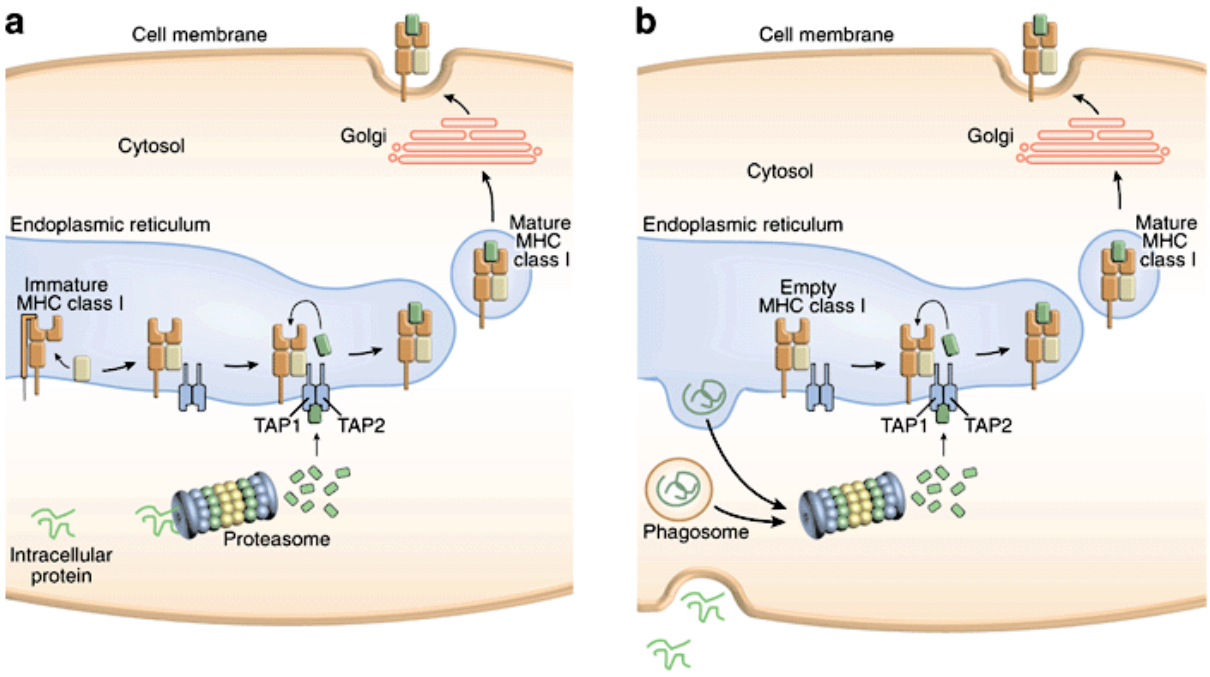


Figure 2: *MHC class I antigen processing and antigen presentation pathways.* In general, MHC class I-presented peptides are derived from intracellular proteins (**a**). These are degraded by the proteasome and transported through the transporter associated with antigen processing (TAP) in the endoplasmic reticulum. There, newly synthesized MHC class I molecules are stabilized by calnexin until β_2 microglobulin binds to the complex. The partial folded MHC class I complex binds to the TAP complex, and, after binding of peptide, the peptide/MHC complex is transported through the Golgi apparatus to the cell surface. Alternatively, exogenous proteins are phagocytosed, and endocytosed antigens may exit the endosomal pathway into the cytosol, either before or after processing, where they can enter the classical MHC class I presentation pathway (**b**). These proteins are retro-transported out of the endoplasmic reticulum and degraded by the proteasome. The degraded peptides can now enter the normal pathway through the TAP complex.(Andersen et al., 2006).

Although it is generally appropriate to refer to the MHC class I antigen pathway as ´endogenous´, in some specialized APC MHC class I present exogenous proteins by routes of cross-presentation. (Heath and Carbone, 2001). Thus, exogenous antigens can be presented by MHC class I molecules if they are endocytosed and degraded in endosomal compartments, where MHC class I molecules recycled from the plasma membrane can capture the resulting antigenic peptides or if they are translocated into the cytosol by the action of either proteins, or specialized mechanisms of transport (Yewdell et al., 1999).

MHC class I-presented peptides are of a remarkable uniform size, the vast majority being either 8 (Germain, 1994), 9 (Kloetzel, 2004), or 10 residues (Rock, 2002) depending on the particular MHC class I molecule.

1.2.2 MHC class II molecules

Major histocompatibility complex (MHC) class II molecules are polymorphic cell surface glycoproteins expressed primarily on specialized antigen presenting cells such as B lymphocytes, macrophages and dendritic cells. The peptides presented by MHC class II molecules are generated mostly by degradation of proteins that access the endocytic route, where they are processed by lysosomal proteases and other hydrolases. This antigen presentation pathway is referred as to exogenous or endosomal.

Class II molecules are heterodimers with each subunit encoded by an MHC linked gene. Both subunits (α and β) are integral membrane glycoproteins with an N-terminal extracytoplasmic region divided into two domains of 90-100 amino acids, transmembrane regions of 20-25 amino acids, and short cytoplasmic domains of 12-15 amino acids. The membrane proximal domains (α2 and β2) each contain an internal disulfide bond and are structurally similar to immunoglobulin constant region domains (Kappes and Strominger, 1988).

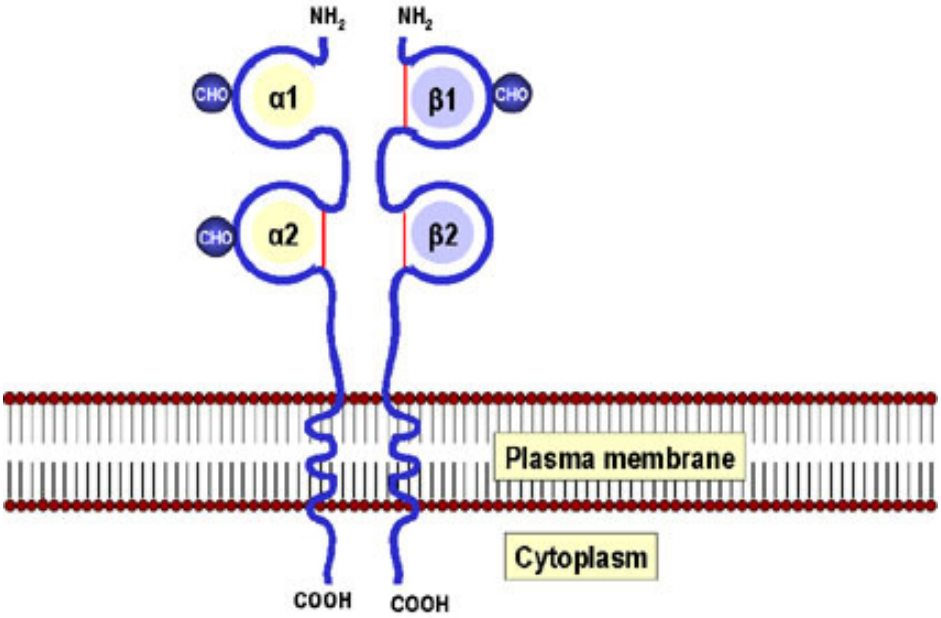


Figure 3: *MHC class II molecules*. MHC class II molecules comprise two non-identical peptides (alpha and beta) which are non-covalently associated and traverse the plasma membrane with the N terminus to the outside of the cell. The domains closest to the membrane in each chain are structurally related to immunoglobulins. With the exception of the alpha 1 domain, all domains are stabilized by disulfide bridges (red). Both the alpha and beta chains are glycosylated (Mayer and Nyland, 2010).

The membrane distal domains (α1 and β1) associate to form a structure that constitutes the peptide binding region of the molecule. X-ray crystallographic studies demonstrated that the structure of the class II peptide binding cleft consists of eight strands of anti-parallel β-sheet with two antiparallel α-helical regions overlaying them. There is a deep cleft between the α-helices which accommodates the bound peptide. Four of the β-strands and one helical region are derived from each subunit (Stern et al., 1994; Zhu et al., 2003).

Unlike class I, the class II binding groove is open at both ends. As a result, peptides binding to class II molecules tend to be of variable length, but typically between 11 and 30 residues (Rammensee et al., 1995). A hallmark of the MHC class II binding peptide groove is that there are four major pockets. These pockets accommodate side-chains of residues 1, 4, 6, and 9 of a 9-mer core region of the binding peptide. This core region interaction largely determines binding affinity and specificity (Jones et al., 2006). In addition, peptide residues immediately flanking the core region have been indicated to make contact with the MHC molecule outside of the binding groove, and to contribute to MHC-peptide interaction (Godkin et al., 2001).

MHC class II molecules are highly polymorphic, and this polymorphism largely corresponds with differences along the peptide binding groove. However, the binding motifs derived for MHC class II molecules are highly degenerate, and many promiscuous peptides have been identified that can bind multiple MHC class II molecules (Consogno et al., 2003).

1.2.3 Formation of the MHC class II-invariant chain nine-subunit complex and its transport

MHC class II molecules are synthesized on polysomes associated with the rough endoplasmic reticulum. The class α and β subunits are type I membrane glycoproteins and, when synthesized, contain classical N-terminal signal sequences responsible for their translocation into the ER (Kappes and Strominger, 1988). These sequences are cleaved co-translationally. The α and β subunits acquire their N-linked glycans during translocation and are ultimately integrated into the ER membrane by their hydrophobic transmembrane regions. The α and β subunits form heterodimers and associate with the trimers of invariant chain (Ii).

Chemical cross-linking experiments of MHC class II molecules and Ii chain from cell lysates have demonstrated that the invariant chain is trimeric, and to some extent hexameric when not associated with the class II molecules (Marks et al., 1990). The MHC class II-Ii complexes from cell lysates exist as a nine-chain structures, containing an equimolar ratio of α, β and Ii (Roche et al., 1991). But whether this occurs in solution without cross-linking support has never been shown. Ii trimerization occurs prior to the association of class α and β subunits with each other or with the invariant chain. Trimerization is completed rapidly, within a 10 minutes. Within one to two hours three αβ dimers associate successively with an individual invariant chain trimer, ultimately generating the nonamer (Lamb and Cresswell, 1992). Recently, Koch et al challenged this concept by showing that association of Ii with DRA alone precede the binding of DRB and that the Ii trimer can bind only a single MHC class II αβ dimer before leaving the ER for transport into the endocytic route(Neumann and Koch 2005; Koch, McLellan et al. 2007; Koch, Zacharias et al. 2011). Class II αβ dimers formed in the absence of invariant

chain are poorly expressed on the cell surface, but are found predominantly localized in the ER (Lamb et al., 1991).

During biosynthesis of class II molecules, Ii inhibits premature binding of endogenous peptides (Roche and Cresswell, 1992) or of unfolded polypeptides (Bush et al., 1996). It also guides folding of class II molecules to facilitate their exit from the ER (Anderson and Miller, 1992), and later serves as a targeting signal for delivery of class II molecules to endocytic compartments (Lotteau et al. 1990). The Ii does not have an N-terminal signal sequence, but the transmembrane region itself serves as a signal sequence for the translocation of its C-terminal domain into the ER (Lipp and Dobberstein, 1986).

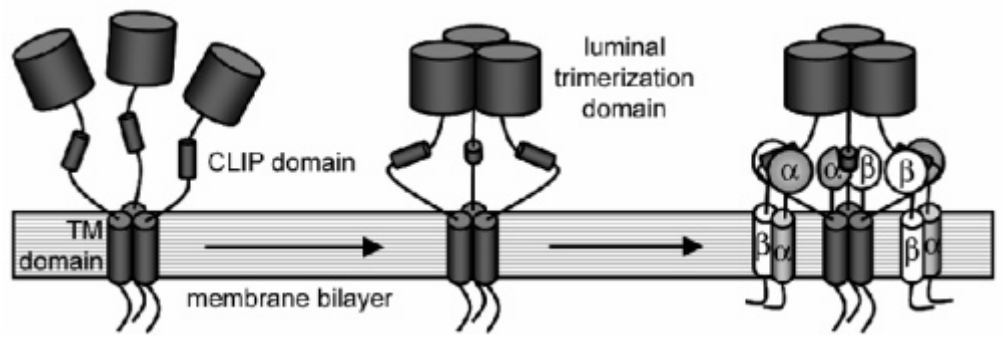


Figure 4: *Model for MHC class II-Ii complex assembly.* Trimerization of Ii is initiated by the transmembrane domain and is followed by trimerization of the lumenal trimerization domain. After Ii trimer is formed, three αβ MHC class Ii heterodimers bind to the CLIP domain of Ii to form the mature nonameric MHC class II-Ii complex (Dixon et al., 2006).

A small number of the αβ dimers synthesized in Ii free cells, either in transfected cell lines or Ii deficient mice, are expressed on the cell surface (Sekaly et al., 1986) and these constitute a small proportion of misfolded molecules which have escaped retention. Such cells have impaired ability to process and present certain protein antigens to T cells (Viville et al., 1993) and some conformation-dependent anti-class II monoclonal antibodies fail to react with these cells. (Peterson and Miller, 1990).

Class II αβ dimers have a globular conformation, while the invariant chain trimer is extended (Monacco 1992). This suggests that such an interaction could occur when the proteins are integrated together in the membrane (Chicz et al., 1993). To verify this it will be necessary to solve the structure of the MHC class II-Ii complex.

1.2.4 Invariant chain

Invariant chain is a type II transmembrane protein. (Claesson et al., 1983). In humans there aren four different forms of Ii, the predominant form being 216 amino acids long, with an N-terminal cytoplasmic domain of 30 amino acids followed by a 26 amino acid hydrophobic transmembrane region. At residues 114 and 120 in the lumenal region there are two N-linked glycans. This form is referred as p31. A second form results from an alternative initiation of translation at an upstream AUG codon (Strubin et al.,1986). It has an additional 16 amino acids at the N-terminus and is referred to as p33. A further form of

the Ii, p41, is encoded by an alternatively spliced transcript (McKnight et al., 1989). This results in the inclusion of 64 additional amino acids in the luminal region, between the two N-linked glycans of p31 and the C-terminus. The p41 specific 64 residue segment resembles a thyroglobulin type I domain, rich in cysteine residues (O'Sullivan et al., 1987). This sequence includes two N-linked glycosylation sites. In humans, p41 and p43 are derived from the same transcript by alternative initiation of translation, as in the case of p31 and p33.

APCs express two Ii isoforms, p31 and p41, in different ratios in the different types of APCs. Whereas p41 represents no more than 10 % of the total poll of Ii in splenocytes, its expression levels are considerably higher in macrophages, dendritic cells (DCs) and Langerhans cells (Pierre and Mellman, 1998). p31 and p41 isoforms of Ii have indistinguishable functions as chaperones for class II folding and intracellular trafficking (Takaesu et al., 1995,1997). Both p31 and p41 isoforms are converted into CLIP (class II associated invariant chain peptides), which is the shortest fragment of Ii remaining associated with class II molecules after proteolysis (Rudensky et al., 1991).

The capacity of Ii to form homotrimers appears to be required for an effective endosomal localization of MHC class II-Ii complexes (Arneson and Miller, 1995). Trimerization is thought to involve the C-terminal two-thirds of the luminal part of Ii (Gedde-Dahl et al., 1997). This part forms an α-helical domain and is protease-resistant (Park et al., 1995). The transmembrane domain may also contribute to the oligomerization process because the N-terminal fragment of Ii appears to retain the nonameric structure (Newcomb et al., 1996). The N-terminal third of the luminal part of Ii is susceptible to proteases and by itself does not appear to constitute a structured domain. This domain allows access of the CLIP region, which is part of it, to the class Ii binding groove (Park et al., 1995). CLIP can bind as a conventional antigenic peptide in the groove of class II molecules, in which Met91 and Met99 are the two primary anchor residues (Gosh et al., 1995).

1.2.5 Transport of the class II-invariant chain complex

The N-terminal cytoplasmic extension found on invariant chain isoforms constitutes a strong ER retention signal (Lotteau et al. 1996) and association of the class II α and β subunits with invariant chain negates the ER retention signal of the p31 form (Machamer and Cresswell, 1982).

After leaving the ER, the assembled αβIi complex traverses the Golgi apparatus. Sialic acid is added to N-linked glycans on all three chains, and also to the O-linked glycans characteristic of Ii (Machamer and Cresswell, 1984). Unlike most membrane glycoproteins, the MHC class II-Ii complex is not expressed immediately on the cell surface, with a 2-3 hours lag between transit through Golgi apparatus and expression on the cell surface (Cresswell and Blum, 1988), which suggests shuttling of the complex into the endocytic system. This system exists as a complex network of tubulo-vesicular structures (Hopkins et al., 1990) and consists of several subcompartments like early, late and dense lysosomes (Griffiths et al., 1988).

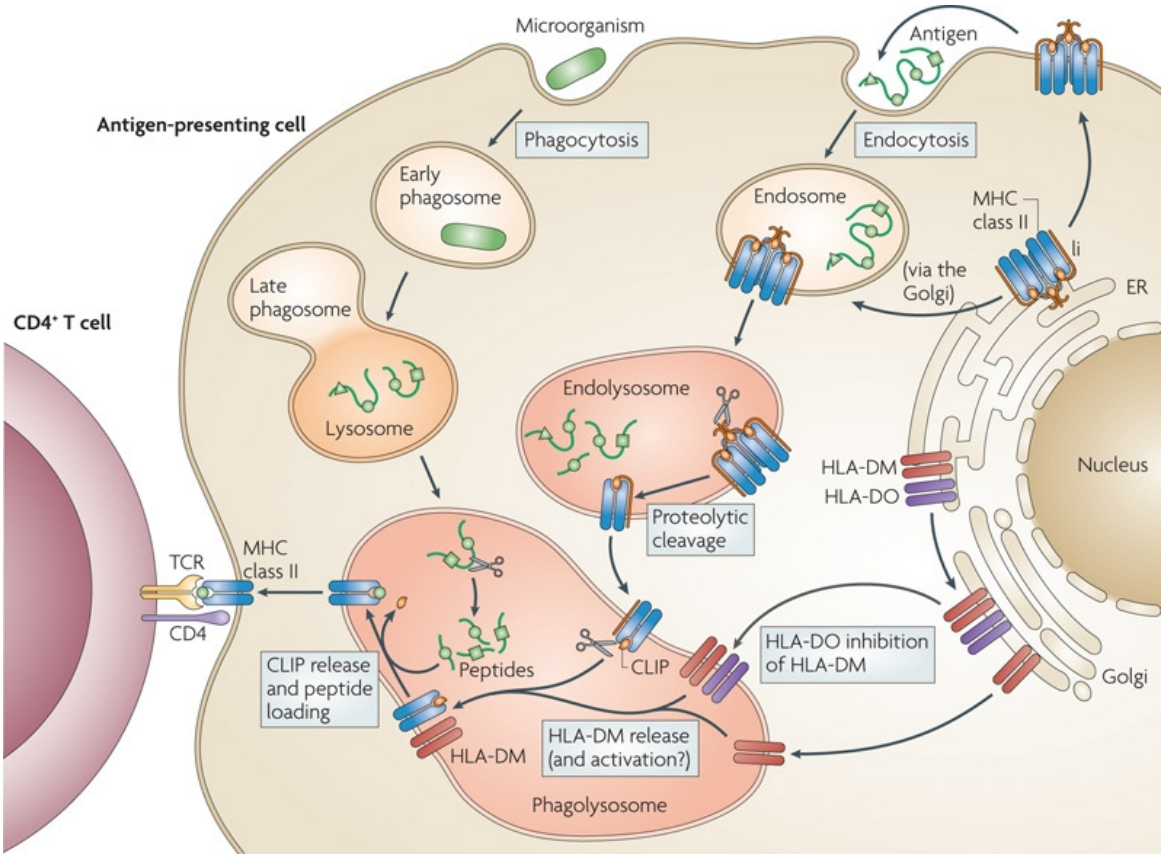


Figure 5: *MHC class II pathway*. MHC class II molecules are assembled in the endoplasmic reticulum (ER) as α-chain–β-chain heterodimers in a nonameric complex containing three heterodimers, each associated with a Ii molecule. The MHC class II-Ii complex is transported from the ER through the Golgi to reach late endocytic or lysosomal compartments where Ii is degraded by proteases, leaving an invariant chain-derived peptide, known as class II-associated Ii peptide (CLIP), that remains in the peptide-binding groove, inaccessible to proteases. Human leukocyte antigen (HLA)-DM interacts with MHC class II molecules to catalyse the dissociation of CLIP as well as the exchange of other peptides, resulting in the formation of complexes of MHC class II molecules with high-affinity peptides. MHC class II-peptide complexes are transported from vacuolar processing compartments to the plasma membrane, where they are presented to $CD4^+$ T cells.(Harding et al., 2010)

Early endosomes have a slightly acidic pH and are barely proteolytic, late endosomes are more acidic, whereas lysosomes have a low pH and are rich in hydrolytic enzymes. As noted before, Ii contains one or more sorting signals with its amino-terminal cytoplasmic domain required for localization to the endocytic system (Bakke ana Dobberstain, 1990), but it is not known precisely how complexes enter the endocytic route. One explanation is that intracellular HLA-DR invariant chain complexes are accessible to the transferrin receptor pathway, in that transferrin conjugates with neuraminidase and therefore can de-sialylate the MHC class II-Ii complex (Creswell, 1985). Transferrin receptor cycles from the plasma membrane to the early endosomes and back to the surface without entering late endosomes. Also immunoelectron microscopy showed that invariant chain and MHC class II molecules are co-localized in early endosomes (Guagliardi et al., 1990). Electron microscopy studies showed however that the large majority of intracellular HLA-DR

molecules are concentrated in compartment called the MHC class II containing compartment (Peters et al., 1991). This compartment is membrane rich and contains lysosomal enzymes. Evidence exist that class II molecules can be found throughout the endocytic system and are potentially Ii-associated and newly synthesized, or mature Ii-free class II molecules internalized from the cell surface (Harding et al., 1989). MHC class II-Ii complexes may enter the endocytic pathway at the early endosomes and then travel through the pathway to late endosomes or pre-lysosomes which contain some lysosomal markers. (Germain et al., 1993). It is possible that different cell types handle MHC class II-Ii differently.

1.2.6 Processing of Ii and transport of class II molecules to the cell surface

In the absence of α and β chains class II molecules, when cells are transfected only with p31 or p41 isoform of Ii, Ii remains in endosomes. (Marks et al., 1990). Even when all three chains are expressed, molecules also reside in endosomes for a long period. The endosomal retention signal is probably the same as the endosomal targeting signal in the cytoplasmic domain. In order that class II molecules go to the cell surface, they must dissociate from Ii. The dissociation of MHC class II molecules occurs by proteolysis of the Ii. A number of studies have been performed to explain this dissociation. One showed that when chloroquine, a lysosomotropic agent, is added to cells and then accumulated in endosomes and lysosomes, it inhibits class II-Ii dissociation, due to neutralization of the normally acidic contents of endosomes and lysosomes (Nowell and Quaranta, 1985). Many lysosomal enzymes have an acidic pH optimum for activity. The accumulation of the complex of class II molecules containing the LIP, a 21 kDa proteolytic degradation product of Ii, was found to be generated by leupeptin, an inhibitor of sulfhydryl protease (Blum and Creswell 1988). An additional 12 kDa fragment of Ii was found in complex with class II molecules in cells treated with leupeptine and antipain (Nguyen et al., 1989). These experiments showed that when Ii degradation is impaired and Ii intermediates LIP and 12kDa fragment are bound to MHC class II molecules, the MHC class II molecules are not transported to the cell surface (Neefjes and Ploegh 1992; Loss and Sant 1993). The LIP and 12 kDa fragment contains only the transmembrane region and cytoplasmic domain of Ii. This suggests that these inhibitors inhibit only partial cleavage of Ii and that other enzymes, which are not inhibited by these inhibitors, are involved. The first enzymes thought to be involved in this process are cathepsin B and cathepsin D, because they are present in endosomes (Diment and Stahl, 1985; Murphy 1985), and additionally cathepsin B is inhibited by leupeptin.

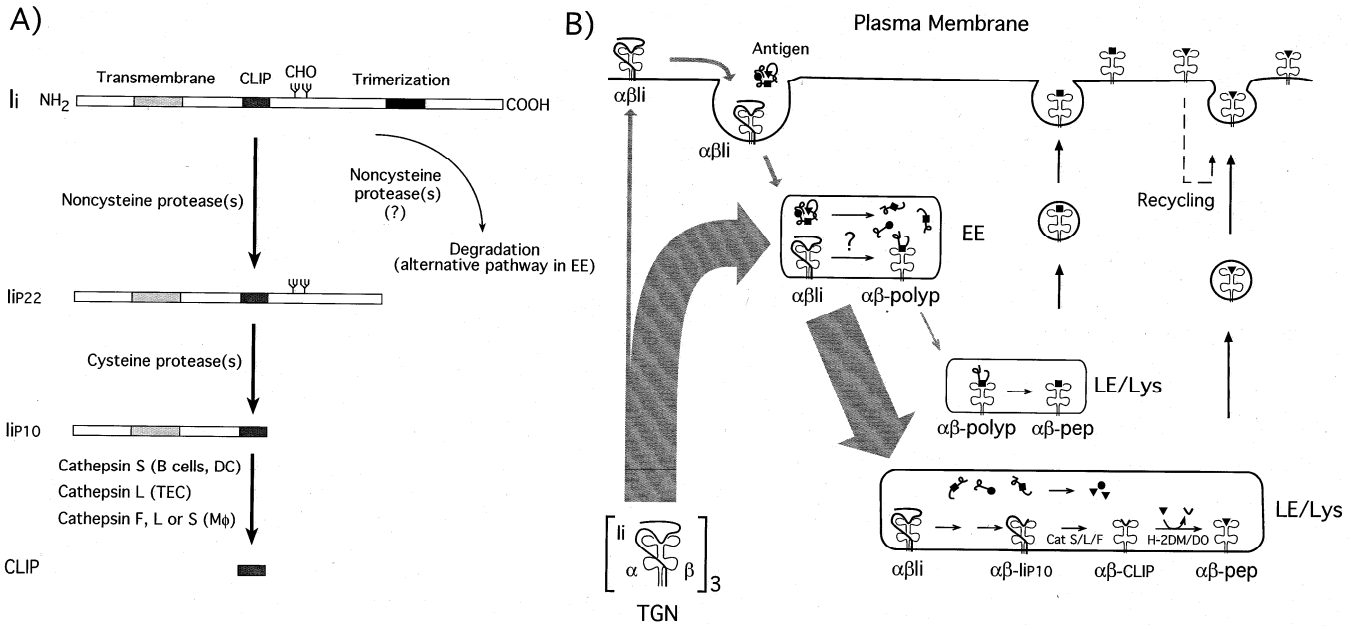


Figure 6:: A) Schematic representation of Ii intermediates generated during its degradation and B) Formation of MHC II–peptide complexes in the endocytic route. Antigens are endocytosed and the degraded along the endocytic route. The majority of the MHC II–Ii complexes reach LE/lysosomes, where Ii is degraded via the 'classical' mechanism depicted in A). (Villandagos, 2001)

In human B cell lines there is evidence that proteolysis of the MHC class II-Ii chain complex occurs in a concerted fashion, with all three Ii molecules cleaved almost simultaneously to LIP in leupeptine treated cells, and that MHC class II-LIP complex remains nonameric (Newcomb and Creswell 1993).

The MHC class II-Ii complex is exposed to a proteolytic environment possibly as a result of a vesicular fusion event. A vesicle containing active protease could be introduced into a vesicle with class II-Ii comlexes (Guagliardi et al., 1990). This could explain a prolonged period in a post-Golgi compartment with an intact MHC class II-Ii complex. Ii is highly susceptible to proteolysis (Marks et al., 1990; Marks and Cresswell 1986). Based on experiments with antibodies against C and N terminal domains of Ii, it has been shown that the latter is cleaved in the prelysosomal compartment (Romagnoli et al., 1993).

The degradation of Ii, with intermediates Iip22, Iip10 and finally CLIP, does not occur randomly, but in a stepwise fashion. These events are probably controlled by the structure of each MHC class II-Ii intermediate itself rather than by a sequential recruitment of active proteases in different compartments. The MHC class II-Ii nonamer is very compact, and only after the trimerization region of Ii is removed (MHC class II-LIP) can next step happen. The carbohydrate groups in Iip22 at the C-terminus of CLIP may block the cleavage site at the other end of CLIP (Neumann et al., 2001), so that only after Iip22 has been converted into Iip10, is this cleaved to yield MHC class II-CLIP. Only in the case of Iip10 has cleavage been attributed to particular enzymes. B cells and dendritic cells cleave Iip10 using cathepsin S (Nakagawa et al., 1999; Riese et al., 1996; Shi et al., 1999), thymic epithelial cells by cathepsin L (Nakagawa et al., 1998) and macrophages by cathepsins L and S (Shi et al, 2000). Converting a Iip10 to CLIP is highly important, since deficiency in some cathepsins impairs immune responses and causes a deficit in positive selection of CD4+ thymocytes.

After cleavage and dissociation of Ii from MHC class II molecules, αβ dimers are released from the endocytic compartment and are exposed on the cell surface. The endosomal localization signal encoded in the cytosolic portion of Ii has a retention motif that impairs delivery to the plasma membrane of MHC class II associated with full length II, Iip22 or Iip10 (Amigorena et al., 1995; Brachet et al., 1997; Neefjes and Ploegh, 1992). But when the Iip10 is cleaved to CLIP, the MHC class II-CLIP complex is free to exit to the cell surface, so the substitution of CLIP for antigenic peptide must be rapid. The peptide binding groove of the MHC class II molecule is flexible and can be opened up. When the opened conformation occurs, CLIP is released and other antigenic peptides with the right sequence are introduced. This open conformation is induced by low pH (Jensen, 1990; Mouritsen et al., 1992), but also by stabilizing the open conformation with MHC class II-like chaperones HLA-DM and HLA-DO. Stabilization of the open conformation of the peptide binding site extends the time for antigenic peptides to compete for the binding site (Alfonso et al., 1999; Kropshofer et al., 1999).

The formation of MHC class II-peptide complexes is probably very efficient, because antigenic peptides are presented as nested sets with a common core sequence varying in length (Engelhard, 1994) . They bind to the peptide binding site as long precursors and are then trimmed at their ends. If the MHC class II-precursor is exposed to the endocytic compartment too long, exopeptidase would probably trim the precursor to the very core that is protected with MHC class II molecules. So presentation as nested sets suggests that once MHC class II binds peptides ,the resulting complex is then very quickly transported to the cell surface. MHC class II-peptide complexes are shuttled to the plasma membrane in transport vehicles. They may shuttle back to the Trans Golgi compartment, then enter the constitutive transport pathway, to be transported to the cell surface (Neefjes et al., 1990). Alternatively, the MHC class II-peptide complexes enter a direct, yet uncharacterized route from a late endosome or lysosome directly to the cell surface.

Sometimes a fraction of MHC class II-Ii complex can be found on the cell surface (Saudrais et al., 1998; Wraight et al., 1990). This happens probably because the sorting mechanism for trafficking the MHC class II-Ii complex to the endocytic route is not 100 % efficient and the MHC class II/Ii complex follows the secretory pathway. However the complex does not stay there for long, because the cytoplasmic tail of Ii has internalization signals and the complexes that reach the plasma membrane are rapidly endocytosed (Bremnes et al., 1994).

1.2.7 Generation of antigens

There are different mechanisms by which antigens can access the endocytic route in APCs, such as pinocytosis, endocytosis, phagocytosis, autophagocytosis and receptor-mediated endocytosis (Watts 1997). The antigens are up taken in endocytic vesicles and they progress along early endosomes to late endosomes and lysosomes, under increasingly denaturing and proteolytic conditions. Proteolytic processing of antigens proceeds differently in each endocytic compartment (Fernandes et al., 2000; Griffin et al., 1997; Lindner and Unanue 1996; Parra-Lopez et al., 1997).

1.2.7.1 Proteolytic activities in the endocytic route

The mechanism by which step-wise degradation of antigens occurs might be a gradual

increase in the protease content between the compartments (Garin et al., 2001), because newly synthesized proteases are delivered from the trans Golgi network into different places on the endocytic route. Endosomal proteases are synthesized as zymogens, inactive precursors of proteases, and they mature into their active form by autoproteolysis or by proteolysis by other proteases (Bohley and Seglen, 1992; Turk et al., 1997).

The activity and stability of these proteases depends on pH, which differs in the various compartments along the endocytic route, so proteases become active in different locations. Besides the auto regulatory mechanism, lack of one or more proteases can affect the activity and turnover of other enzymes involved in antigen processing (Driessen et al., 2001; Honey et al., 2001; Lennon-Dumenil et al., 2001). The activation of a protease in one compartment may in turn influence the activity of other enzymes in the same location. Cathepsin S, for example, is known to be stable and active within a wide pH range in vitro, but it is incorporated only into late endosomes. Cathepsin L displays a similar pattern. Cathepsin B, though, is active in both eraly and late endosomes.

The proteolytic activity of the endocytic compartments in APC are also regulated by natural inhibitors such as members of the cystatin or serpin families (Gresser et al., 2001; Pierre and Mellman, 1998; Schick et al., 1998). Also the inhibitory thyropin-like domain of invariant chain-p41 (Bevec et al., 1996, 1997; Lenarcec and Bevec 1998) chaperones and protects the protease cathepsin L (Lennon-Dumenil et al., 2001). The gradual degradation of Ii may contribute to modulation of cathepsin activity in different endocytic compartments and consequently influence MHC class II processing (Mihelic et al, 2008; Fineshi et al, 1996).

1.2.7.2 Non-proteolytic factors involved in antigen degradation

All proteases have some degree of specificity, but many sites in an antigenic substrate are not cleaved, although with the right residue sequence, unless they are first made accessible to the protease.

Acidification is the most important factor in the process of protein unfolding (Jensen, 1993). pH decreases progressively in the lumen of the endocytic compartment and the control of acidification plays an important role in ensuring that antigen degradation occurs in stages. pH also influence the activity of proteases and protein trafficking (Forgac, 1998). It is expected that the pH of the endocytic route is tightly regulated. Control of acidification may be the mechanism in DCs and other APCs by which antigen presentation during activation is regulated (Fiebiger et al., 2001; Lutz et al., 1997). Cytokines IL-6 and IL-10 respectively increase and decrease acidity (Drakesmith et al., 1998; Fiebiger et al., 2001).

Protein unfolding sometimes requires thiol reductases that cleave disulfide bonds (Collins et al., 1991). The reducing milieu in endosomal compartments plays an important role in T cell recognition of epitopes carrying cysteine (Haque et al., 2001). The gamma-interferon inducible lysosomal thiol reductase (GILT) is, for example, constantly expressed in APCs and may have an important function in MHC class II antigen presentation. GILT is induced in non-APCs by γ-IFN (Arunachalam et al., 2000; Phan et al., 2000). Other mechanisms, such as high concentration of cysteine, that maintain a reducing environment in the endocytic route may also contribute to favour disulfide bond hydrolysis (Collins et al., 1991; Pisoni et al., 1990).

Presence of carbohydrate groups can also determine the susceptibility of an antigen to protease degradation and hinder the accessibility of the proteases to the potential cleavage sites (Surman et al., 2001). For example when the two glycosylation sites of Ii are removed, survival of Ii is decreased, suggesting that carbohydrate groups protect the

molecule against proteolytic degradation (Neumann et al., 2001).

Antigenic peptides can also be protected against proteolytic degradation by association of an antigen with the receptor that captured the antigen. Surface Ig on B lymphocytes captures and internalizes an antigen and masking certain regions in the antigen, preventing formation of one or more MHC class II-peptide complexes (Lanzavecchia, 1996).

1.2.8 Lysosomal proteases implicated in antigen presentation

Earlier *in vitro* studies revealed participation of cathepsins B, D, E, F, K, L and S in Ii degradation and antigen processing (Blum and Creswell, 1988, Lennon-Dumenil et al., 2002; Maric et al, 1994). These studies were performed with invariant chain from cell lysates and the findings still needs to be supported by *in vitro* enzyme treatment of purified Ii.

However analyse of mice with targeted mutations of these proteases revealed that some cathepsins exhibiting antigen and Ii-processing activity in cell-free assays were dispensable under physiological conditions.(Constantino et al., 2008). For example, splenocythes that lack abundant cathepsin B or cathepsin D exhibited normal Ii degradation and antigen presentation to T cells (Deussing et al., 1998).

Analyses of cathepsin S deficient and cathepsin L deficient mice revealed that these two proteases have non-redundant roles in antigen presentation. Cathepsin L is expressed in the thymus, specifically in the cortical epithelium (Benevides et al., 2001). Expression of the human homolog of cathepsin L, known as cathepsin V, is also highly abundant in the human thymus, and it was demonstrated that is involved in Ii degradation in the human thymus (Tolosa et al., 2003). In cathepsin L deficient mice, absence of cathepsin L expression in cortical thymic epithelial cells (cTECs) results in incomletely processed Ii (Nakagawa et al., 1998), accumulation of p18-p22 and p12 fragmetns tightly bound to MHC class Ii molecules and accordingly a defect in thymic positive selection of CD4+T cells. This effect is limited to CD4+T cells and is therefore MHC class II pathway-intrinsic (Honey et al., 2002). Biochemical and functional analysis of cathepsin L deficient mice point to the essential role of this enzyme in the end stages of Ii degradation in cTECs.

Cathepsin S is predominantly expressed in profesional APCs, namely DCs and B cells (Riese et al., 1998). Cathepsin S deficient mice showed an accumulation of the Ii degradation intermediates p18-p22 and p12(Nakagawa et al., 1999, Shi et al., 1999). In comparison to wild type APCs, the presentation of various exogenous protein antigens (such as Trypanosoma cruzi SA85) by B cells and DCs from cathepsin S deficient mice was markedly reduced. Deficient presentation of exogenous antigens was also observed when cathepsin S-deficient macrophages were used as APCs(Nakagawa et al., 1999). Cathepsin S deficiency also affect endogenos protein antigen presentation.

Multiple proteases can cleave full length Ii and Iip22, showing that there is protein redundancy in these two reactions. The use of inhibitors for asparaginyl endopeptidase (AEP, also known as legumain) indicated that this enzyme may be involved in full length Ii processing, but probably also this role is redundant (Manoury et al.,2003; Maehr et al., 2005).

Since Ii has other functions besides controlling the peptide loading of MHC class Ii molecules, like regulation of endosomal architecture (Lagaudriere-Gesbert et al., 2002;

Nordeng et al., 2002), modulation of DC migration(Faure-Andre et al., 2008), recognition of the pre-inflammatory cytokine macrophage migration inhibitory factor(Leng et al., 2003), and nuclear factor-κB (NF- κB) signalling (Starlets et al., 2006), use of inhibitors and knock out mice could affect the immune response indirectly through the effects of any of these activities.

Table 1: *Lysosomal proteases implicated in antigen presentation* (Honey and Rudensky, 2003).

Enzyme	Protease type	Expression	Knockout phenotype
Cathepsin B	Cysteine	B cells, DCs and macrophages	No marked immune-system phenotype reported Decreased trypsin activation and onset of pancreatitis Decreased susceptibility to TNF-induced hepatocyte apoptosis
Cathepsin D	Aspartic	B cells, DCs and macrophages	No immune-system phenotype reported Die at 21 days as a result of atrophy of the ileal mucosa
Cathepsin F	Cysteine	Ubiquitously expressed	No phenotype, artifact
Cathepsin K	Cysteine	Macrophages and osteoclasts	No immune-system phenotype reported Have osteopetrotic phenotype
Cthepsin L	Cysteine	Activity in cortical TECs,macrophages and thymociytes; protein expression by B cells and DCs	Decreased CD4+ T cell selection due to impaired Ii degradation and generation of MHC class II-bound peptides in cortical TECs Epidermal hyperplasia and hair-follicle deficiences Dilated cardiomyopathy
Cathepsin S	Cysteine	B cells,DCs, macrophages and epithelial cells	Decreased MHC class II presentation of exogenous antigens by B cells and DCs Deficient germinal-centre formation and impaired class switching Diminished susceptibility to collagen-induced arthritis
AEP	Cysteine	B cells and DCs	N.D.

Cathepsin B, a lysosomal, papain-like cysteine protease, is one of the most extensive studied human cathepsins(Vasiljeva et al.,2007). This enzyme is abundantly expressed in a variety of tissues where it takes part in protein degradation and processing. It is involved in a number of physiological and pathological processes, such as intracellular protein degradation, the immune response, prohormone processing, cancer and arthritis (Yan and Slan 2003; Turk et al.2002; Mohamed and Sloane 2006; Pozgan et al.2010; Gabrijelcic et al., 1992; Turk et al. 2004; Stoka et al., 2005; Gocheva and Joyce 2007). Its proteolytic activity is regulated by stefins and cystatins, which are endogenous inhibitors of cysteine cathepsins (Turk et al., 2002). Cathepsin B differs from other cathepsins by its dual role, exhibiting exo- as well as endopeptidase activity. Although the structures of the mature native form of cathepsin B clearly exposed the relevance of the occluding loop for the exopeptidase activity (Musil et al., 1991), they do not explain the mechanisms of endopeptidase activity, nor the inhibition of the enzyme by their endogenous protein inhibitors cystatins and stefins (Turk et al., 2008).

1.2.9 Regulation of lysosomal cysteine proteases activity

Activities of cysteine proteases need to be strictly regulated in order to prevent inappropriate proteolysis. Since they are involved in various physiological processes, including antigen presentation and Ii chain degradation, dysfunction of these proteases can be harmful or lethal to an organism. The activity of lysosomal cysteine proteases can be regulated in a number of ways, including

- the compartmentalization of cathepsins within the lysosome or other organelles,
- zymogen activation,
- the regulation of their activities by small-molecule inhibitors and various endogenous protein inhibitors,
- regulation of lysosomal proteases expression
- or a combination of all these factors for an optimal enzyme function

The cathepsins are synthesized as inactive precursors, and are activated by proteolytic removal of the N-terminal propeptide. *In vitro*, removal of the propeptide can be facilitated either by activation by other proteases such as pepsin or cathepsin D, or by autocatalytic activation at acidic pH. The latter is a bimolecular process in which one of the cathepsin molecules activates the other in a chain reaction manner (Turk et al., 2000). For example mature cathepsin L protein levels are increased in B cells in the absence of cathepsin S, although the activity of cathepsin L is not increased in absence of cathepsin S expression. This suggests that cathepsin S have impact on cathepsin L turnover (Beers et al., 2003).

Zymogen activation can be regulated also by endosomal maturation and acidification (Lankar et al., 2002; Lautwein et al., 2002; Trombetta et al., 2003). The potential for protease activation is tied to the maturation state of the phagosome and proteases are recruited to the maturing phagosome in an ordered manner. (Lennon-Dumenil et al., 2002). Cathepsins S and a L are incorporated only into late endosomes.

Limited expression of lysosomal proteases also regulates antigen presentation (Delamarre et al., 2005). This hypothesis suggests that reduced protease activity confers the ability to retain and present antigens effectivelly.

1.2.9.1 Protein inhibitors of cysteine proteases

Once activated, cathepsins have enormous disruptive potential, since their total concentration inside lysosomes can well exceed 1 mM. Their inappropriate action is controlled by cystatins, the endogenous protein inhibitors of lysosomal cysteine proteases, and not by pH (Turk e tal., 2008).

Cystatins are reversible inhibitors of papain-like cysteine proteases and are divided into four gropus: stefins, cystatins, kininogens on the basis of sequence homology and phytocystatins .(Turk et al., 2002; Rawlings et al., 2004; Turk and Bode 1991; Kordis and Turk 2009).

1.2.9.1.1 Stefins

Stefins or type 1 cystatins have a molecular weight of ~11 kDa. The two members found

in humans are stefin A and B. Besides protection of cytosolic and cytoskeleton proteins from the degradation by cysteine proteases several other functions have been suggested for stefins. Higher levels of stefin A have been determined in various tumors and cancers. These higher levels counter balance the excessive activity of cysteine cathepsins. (Jarvinen et al., 1987; Kos et al., 2000, Li et al., 2005)

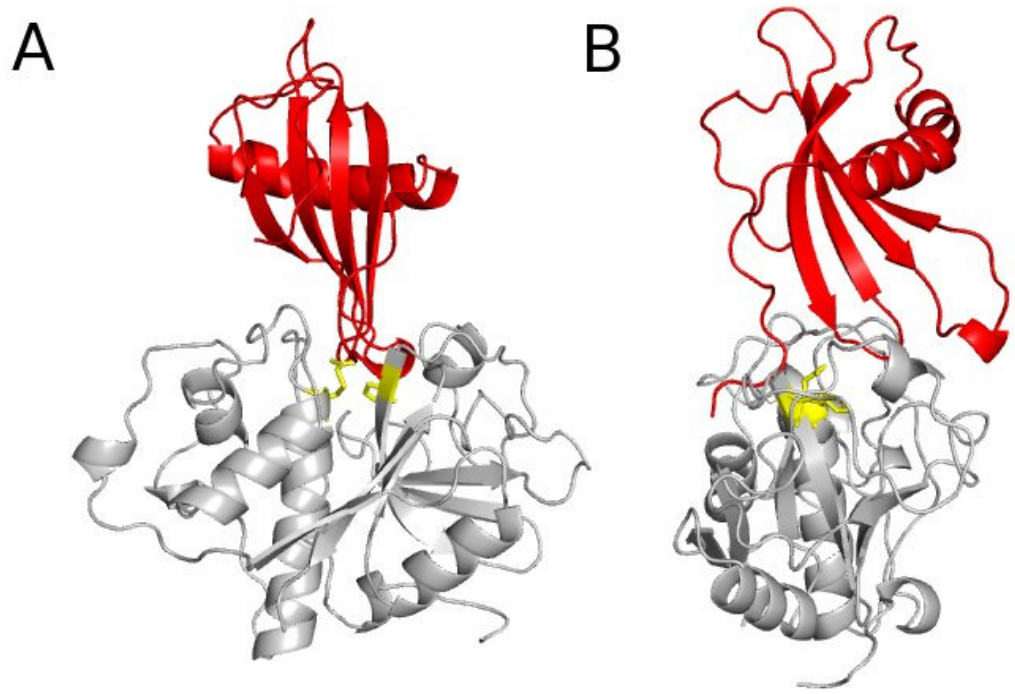


Figure 7: *A crystal structure of carboxymethylated papain with stefin B.* Papain is shown in gray, stefin B in red and active site cysteine in histidine in yellow. Stefin B fills the active site cleft in wedge like shape (A). N-terminal segment binds to the nonprimed binding sites, whereas both loops bind to primed binding sites (B).

1.2.9.1.2 Cystatins

Cystatins (or type 2 cystatins) are approximately 20-40 aminoacids longer than stefins. They are also not glycosylated, however, they have two disulfide bonds at the C-terminal part. They are mostly found in extracellular fluids (Barrett et al., 1998).

The best-researched member is cystatin C, being one of the most important extracellular inhibitors of cysteine proteases. It is used as an important and effective biomarker for kidney function (Shlipak et al., 2006). Mutations in the cystatin 3 gene are responsible for the Icelandic type of hereditary cerebral amyloid angiopathy, a condition predisposing to intracerebral haemorrhage, stroke and dementia (Levy et al., 1989). It was also discovered that certain cystatins (C, E and F) are capable of simultaneous inhibition of papain-like proteases and legumain, a cysteine protease from the C13 MEROPS family (Alvarez-Fernandez et al., 1999).

1.2.9.2 Thyropins

Thyropins (thyroglobulin type-1 domain protease inhibitors) are inhibitors with a similar structure as the thyroglobulin type-1 domain (Lenarcic and Bevec, 1998). The majority of thyropins inhibit papain-like cysteine proteases, however, a few members can inhibit aspartic protease cathepsin D (Lenarcic et al., 1999) and proteases, for whose activity cations are needed (Fowlkes et al., 1997).

One or more repeats of thyroglobulin type-1 domains are found in a number of functionally and structurally unrelated proteins (thyroglobulin, nidogen, testican, insulin-like growth factor-binding proteins, human pancreatic carcinoma marker protein (GA733-2), major histocompatibility complex class II-associated p41 invariant chain, saxifilin, cysteine protease inhibitor from salmon roe and equistatin (Lenarcic and Bevec, 1998). The p41 invariant chain, saxifilin, cysteine protease inhibitor from salmon roe and equistatin are reversible inhibitors of cysteine protease (Dubin 2005). These inhibtors belong to the I31 MEROPS family (Rawlings et al., 2004). Among these, the only known mammalian representative is the major histocompatibility (MHC) class II-associated p41 invariant chain fragment, which is a selective cathepsin L inhibitor (Bevec et al., 1996; Guncar et al., 1999; Zavasnik-Bergant et al., 2004).The regulation of antigen presentation by p41 was not confirmed *in vivo* by Rovere et al., 1998, but on the other hand was confirmed by Fineshi et al., 1996). A further proposed role for p41 fragment of was that p41 serves as chaperone molecule for cathepsin L and that it stabilizes the mature form of the protein (Fiebiger et al., 2002).

Crystallographic analyses of the p41 invariant chain bound to cathepsin L provided insight into the mechanism of interaction between thyropins and their target proteases(Guncar et al., 1999). The p41 fragment consists of two subdomains: the first one iscomposed of an α-helix and a ß-strand, the second subdomain being a three-stranded antiparallel ß-sheet. Similarly to cystatins, the molecule is wedge-shaped and interacts with the enzyme through three hairpin loops (Guncar et al,1999). When compared to cystatins, the different overall structure results in additional contacts with the surface of the protease, determining the high specificity of thyropins and distinguishing them from the relatively non-selective cystatins (Guncar et al., 1999; Mihelic and Turk, 2007).

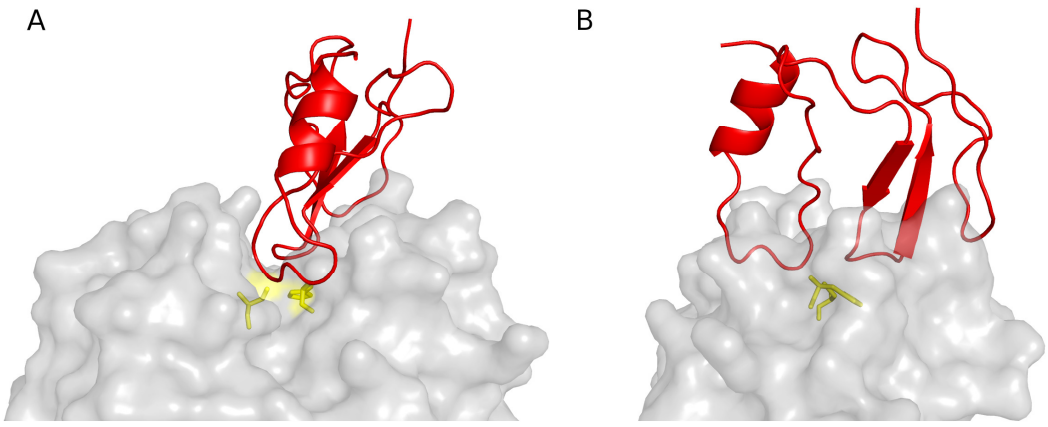


Figure 8: *Cathepsin L (gray) inhibition by p41 fragment (red).* A) View along active site and B) view perpendicular to the active site cleft. Active site residues are shown in yellow (Guncar et al., 1999)

2 Purpose of work

In order to gain insight in structure, maturation and processing of MHC class II molecules their complex with invariant chain (Ii) had to be prepared. Analysis of literature indicated that several efficient protocols for the expression of recombinant soluble MHC class II a b dimers using *E.coli* or insect cells expression systems exists, whereas for production of soluble complexes our best solution seemed the expression in a mammalian system, such as in the HEK293 cell line. Sufficient amounts of soluble material should enable us:

1. to gain insight into the composition of the structure of the complexes,
2. to reveal the importance of the transmembrane regions in the complex formation,
3. to find out the potential roles of cysteine cathepsins in the process of maturation.

The availability of the complexes with the p31 and p41forms of Ii and MHC class II molecules should enable us to establish the role and inhibitory potential of the p41 form of Ii in the maturation process. As a part of a broader study the mechanism of interactions of cysteine cathepsins with other endogenous cysteine protease inhibitors (stefins) was considered.

3 Materials and Methods

3.1 Molecular cloning

3.1.1 Theoretical background

Molecular cloning is a term in molecular biology that refers to a set of experimental methods used to assemble recombinant DNA molecules and to direct their replication within host organisms. The use of the word cloning refers to the fact that the method involves the replication of a single DNA molecule to generate a large population of identical DNA molecules. Molecular cloning generally uses DNA molecules from two different organisms, the species that is the source of the DNA to be cloned, and the species that will serve as the living host for replication of the recombinant DNA. Enzymes are used in the test tube to generate individual segments of source DNA, which are then combined with vector DNA to generate recombinant DNA molecules. The recombinant DNA is then introduced into a host organism (typically a microorganism like E. coli that is easy to grow in the laboratory) to generate a population of recombinant organisms, that is, organisms in which recombinant DNA molecules can be replicated.

Vector DNA is is an agent that can carry a DNA fragment into a host cell. If it is used for reproducing the DNA fragment, it is called a cloning vector and if it is used for expressing certain gene in the DNA fragment, it is called an expression vector. The insertion of the fragment into the cloning vector is carried out by treating the vehicle and the foreign DNA with a restriction enzyme that creates the same overhang, then ligating the fragments together. There are many types of cloning vectors. Genetically engineered plasmids and bacteriophages (such as phage λ) are perhaps most commonly used for this purpose. Other types of cloning vectors include bacterial artificial chromosomes (BACs) and yeast artificial chromosomes (YACs).

The goal of a well-designed expression vector is the production of large amounts of stable messenger RNA, and therefore proteins. Once this type of vector is inside the cell, the protein that is encoded by the gene is produced by the cellular-transcription and translation machinery ribosomal complexes. The plasmid is frequently engineered to contain regulatory sequences that act as enhancer and promoter regions and lead to efficient transcription of the gene carried on the expression vector.

3.1.2 Protocols

3.1.2.1 Gene cloning and construction of the expression vectors

cDNAs encoding human MHC class II alpha (DRA) and beta (DRB) chains without cytosolic and transmembrane domain and with native signal sequences were PCR amplified using following primers: DRA forward primer containing *Hind III* restriction site (underlined) and Kozak sequence - 5' CCCA AAGCTTGCCACC ATG GCC ATA AGT GGA GTC CCT 3', DRA reverse primer - 5' CCCA GAATTC TTA CTC TGT AGT CT containing stop codon and EcoRI restriction site (underlined), DRB forward primer containing HindIII restriction site (underlined) and Kozak sequence - 5' CCCA AAGCTT GCCACC ATG GTG TGT CTG AAG CTC CCT 3' and DRB reverse primer primer containing stop codon and EcoRI restriction site (underlined) - 5' CCCA GAATTC TTA CTT GCT CTG TGC AGA TTC AG 3'. PCR products were digested with HindIII and EcoRI restriction endonucleases and ligated into the pcDNA3 expression plasmid (Invitrogen) (Fig. 9).

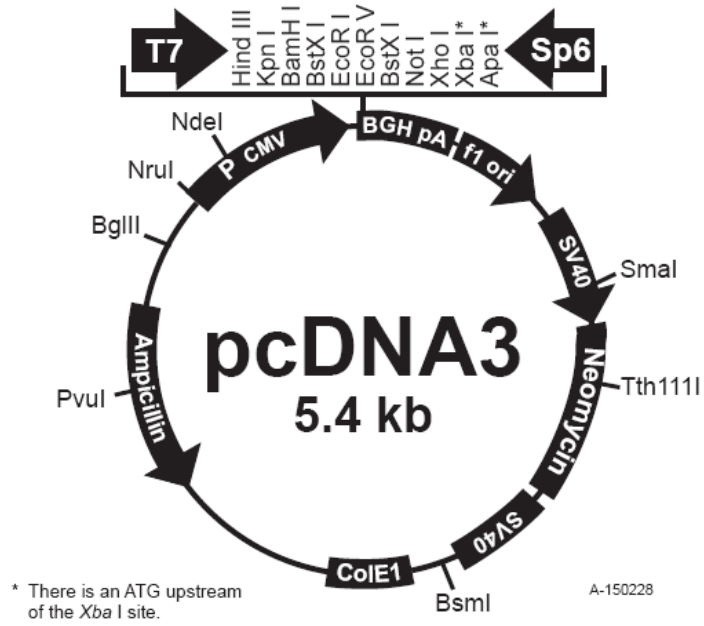
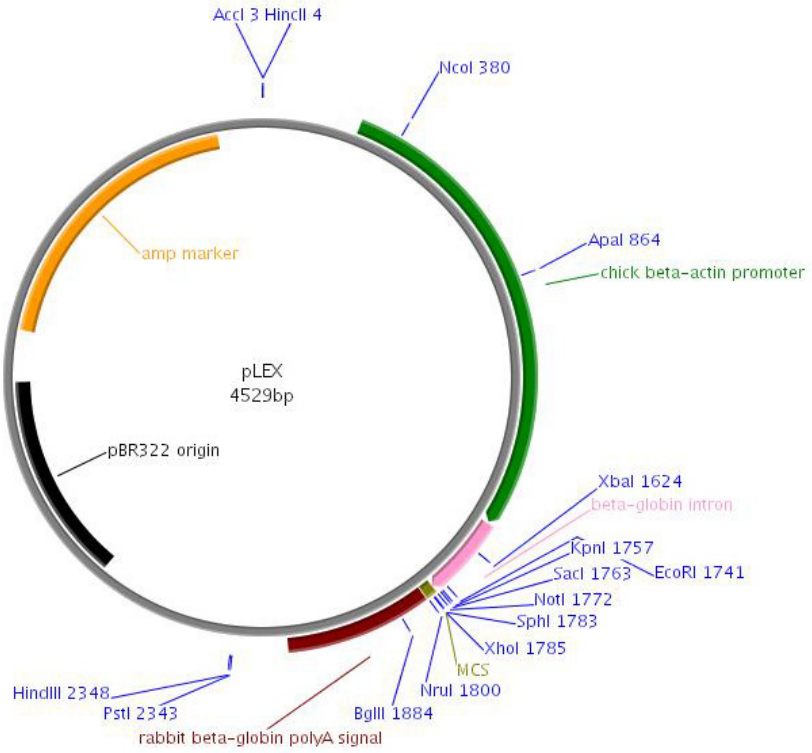


Figure 9: *Scheme of pcDNA3 expression vector.* The pcDNA vectors are designed for high-level, constitutive expression in a variety of mammalian cell lines. They offer features like cytomegalovirus (CMV) enhancer-promoter for high-level expression, large multiple cloning site, bovine growth hormone (BGH) polyadenylation signal and transcription termination sequence for enhanced mRNA stability and ampicillin resistance gene and pUC origin for selection and maintenance in *E. coli*

Human p31 and p41 isoforms of invariant chain without cytoplasmic and transmembrane regions were amplified using following primers: Ii forward primer containing AgeI restriction site (underlined) – CCCA ACCGGT CGG CTG GAC AAA CTG ACA GT and Ii reverse primer containing KpnI restriction site (underlined) - CCCA GGTACC CAT GGG GAC TGG GCC CAG ATC. PCR products were digest with AgeI and KpnI

restriction endonucleases and cloned into the pHLsec expression plasmid (Aricescu et al., 2006) (Fig. 10) in a frame with N-terminal signal sequence and C-terminal 6xHis tag.



EcoRI HindIII **Kozak** **M G I L P S P G M P A L L S**
GAATTCAAGCTT**GCCACCATGG**GGATCCTTCCCAGCCCTGGGATGCCTGCGCTGCTCTCC
L V S L L S V L L M G C V A E T G
CTCGTGAGCCTTCTCTCCGTGCTGCTGATGGGTTGCGTAGCTGAA**ACCGGT**...insert..
G T **K H H H H H H** * * XhoI
GGTACCAAGCACCACCATCACCATCACTAATGATCACTCGAG

Figure 10: *Scheme of pHLsec expression vector for secreted constructs and its MCS.* pHLsec is based on the pLEXM backbone, the MCS is however reduced because several new features are introduced like Kozak sequence, a secretion signal sequence and C-terminal His_6 tag. Constructs can be cloned into this vector, preserving and making use of the features listed above, by using the AgeI and KpnI (Aricescu et al., 2006)

Table 2: *Artificially synthesized primers*. Primers were designed for DRA, DRB, p31 and p41 isophorms of Invariant chain without transmembrane domains.

Primers	DNA sequences
DRA Forward Primer *Hind III*/Kozak sequence	CCC AAA GCT TGC CAC CAT GGC CAT AAG TGG AGT CCC T
DRA Reverse Primer *EcoRI*/stop codon	CCC AGA ATT CTT ACT CTG TAG TCT CTG GGA
DRB Forward Primer *Hind III* /Kozak sequence	CCC AAA GCT TGC CAC CAT GGT GTG TCT GAA GCT CCC T
DRB Reverse Primer *EcoRI*/stop codon	CCC AGA ATT CTT ACT TGC TCT GTG CAG ATT CAG
Invariant chains Forward Primer *AgeI*	CCC AAC CGG TCG GCT GGA CAA ACT GAC AGT
Invariant Chains Reverse Primer *KpnI*	CCC AGG TAC CCA TGG GGA CTG GGC CCA GAT C

3.1.2.2 DNA purification

DNA was purified using the Plasmid Maxi Kit (Qiagen). High-quality DNA is essential for successful transfection and only samples with an OD_{260}/OD_{280} ratio of 1.8 or higher are appropriate. The pLEXm and its derivates are very high copy number plasmids. From 250 overnight bacterial culture, the yield of DNA can be from 2-5 mg. Procedure of DNA purification starts with clearing of bacterial lysates by centrifugation. Lysates are then loaded onto the anion-exchange tip, where plasmid DNA selectively binds under appropriate low-salt and pH conditions. RNA, proteins, metabolites and other low-molecular weight impurities are removed by a medium-salt wash and ultrapure DNA is eluted in high-salt buffer. The DNA is concentrated and desalted by isopropanol precipitation and collected by centrifugation. DNA sample must be sterile, therefore DNA precipitates have to be properly washed with 70 % ethanol before dissolved in sterile 10mM Tris pH 8.0.

3.2 Protein expression

Proteins were recognized as a distinct class of biological molecules in the eighteenth century. Since that time, proteins were isolated from natural sources until a few decades ago recombinant DNA technology started to evolve. Recombinant DNA technology, followed by the increasing number of known protein sequences, allowed the production

of proteins which could not be isolated from natural sources. Basically, the DNA sequence of the protein is inserted into the appropriate vectors and then transferred into the expression cells, where the protein is expressed.

3.2.1 Theoretical backgrounds

3.2.1.1 Mammalian cell expression

Most of the proteins needed for various biological studies are produced by high-level expression in *E.coli;* however, prokaryotic expression system fails to generate correctly folded functional forms of many. Satisfactory expression system for production of stable and soluble forms of eukaryotic proteins represents major bottleneck, so invention of cost-effective and quick mammalian expression systems would be very valuable. Proteins are best expressed in their authentic cell type under physiological conditions. These conditions include control at various stages of synthesis, folding, post-translational modification and subcellular targeting of proteins.

A mammalian expression system would be ideal choice for studies focused on characterization of human cell-surface receptors and secreted proteins. Secreted mammalian proteins have been successfully produced in a large-scale format in Chinese hamster ovary cells (Cockett et al, 1990) and human embryonic kidney (HEK) cells (Meissner e tal 2001, Durocher et al., 2002, Geisse et al., 2005). Large-scale transient tranfection of mammalian cells has been demonstrated to be suitable for rapid and efficient production of proteins because of its ability to produce large amount of soluble and fully processed recombinant proteins in only a few days (Cho et al., 2003).

This rapid production is supported by rapid purification system such as affinity chromatography using synthetic tags (Braun et al., 2003). Polyhistidine (His) tag (Porath et al., 1975) is the most commonly used affinity tag for purifying recombinant proteins (Terpe, 2003). The purification of proteins bearing a His-tag is usually performed by immobilized metal-affinity chromatography (IMAC) using a Ni2+ nitrilotriacetic acid (Hochuli et al., 1987). Secreted recombinant protein form medium can by very quickly purified using IMAC, while production of secreted proteins in HEK293T cells is enhanced in the presence of fetal bovine serum (Durocher et al., 2002). However purification by IMAC is compromised by the presence of serum-derived contaminating proteins. Consequently serum-free medium can be used for secreted protein expression, that allows one-step IMAC purification of recombinant proteins to 95% homogenity (Pham et al., 2003).

3.2.1.2 Co-expression of protein complexes

Most of the functional units within the eukaryotic cell comprise of assembled proteins and nucleic acids, rather than single macromolecules (Gavin and Superti-Furga, 2003). Proper characterization of a cellular function has to be carried out on such multi-component complexes, or at least on their subsets (Romier et al., 2006). Macromolecular complexes require the purification of the different subunits in large quantities and their subsequent

assembly into functional entity. Techniques, such as *in vitro* reconstitution from separately purified components can be used to study small-size assemblies. The major disadvantage of this technique is that it is relatively slow and often requires refolding steps, because proteins that form complexes are misfolded without their cellular partners in heterologous expression system. Good compromise between endogenous purification and in vitro reconstitution from individually expressed components is co-expression of multiple proteins in the same cell.

Co-expression of proteins is an important objective for structural and biochemical studies of protein complexes because it often increases authenticity of biological activity, solubility of protein partners and yields higher amounts of the desired complex (Schwabe et al., 1997, Wang and Chong 2003). Co-expression of subunits within hosts such as *E.coli*, insect and mammalian cells have become widely used, even at the level of high-throughput projects. Scientists are trying to extend the common binary expression to the more complicated multi-expression systems.

3.2.1.3 HEK293T cells expression

HEK 293 cells are specific cell line generated in early 70s by transformation of cultures of normal human embryonic kidney cells with adenovirus 5 DNA (Graham et al., 1997). The name HEK stands for human embryonic kidney and the number 293 comes from the number of experiment. Transformation was brought by an insertion of 4.5 kb viral genomic left arm into human chromosome 19 (Louis et al., 1997).

These cells are easy to handle, have a robust growth rate, high capacity for recombinant protein expression, low-cost media requirements, shows excellent transfectability and are widely used in cell biology and biotechnology to produce therapeutic proteins and viruses for gene therapy. They are not good model for normal cells, but are good in experiments in which the behavior of the cell itself is not of interest. Typical experiment with HEK293 cells is transfection of gene (or combination of genes) of interest, and then analyzing the expressed proteins. HEK293 cells are available from all major cell banks and support repeated passages without loosing the above properties. An important variant of HEK293 cell is the 293T cell line that contains SV40 large T-antigen, that allows episomal replication of transfected plasmids containing the SV40 origin of replication.

Although HEK cell line is of epithelial origin, its biochemical machinery is capable of carrying out most of the post-translational folding and processing required to generate functional, mature protein from wide spectrum of both mammalian and non-mammalian nucleic acid. Though popular as a transient expression system, this cell type has also wide use in stably transfected form. For reconstituted proteins which require accessory proteins in order to be functional, sometimes HEK293 cells can provide it endogenously. If not, it can be co-transfected with the gene of interest, or can stably express the protein of interest (Chuang et al., 1998, Leaney et al., 2000, Lei et al., 2000). The HEK293 cells provide a robust and reliable platform in which to express receptor proteins and ion channels with high fidelity. These proteins production rates are important if all or part of a protein is to be crystallized, hybrid screened for associated molecules or used in biochemical assays. HEK cells rapidly amplify protein product and they can generate tens of milligrams of protein in weeks. All those reasons make them extensively used across many scientific disciplines and are being regarded as a first choice vehicle for the expression of many

recombinant proteins.

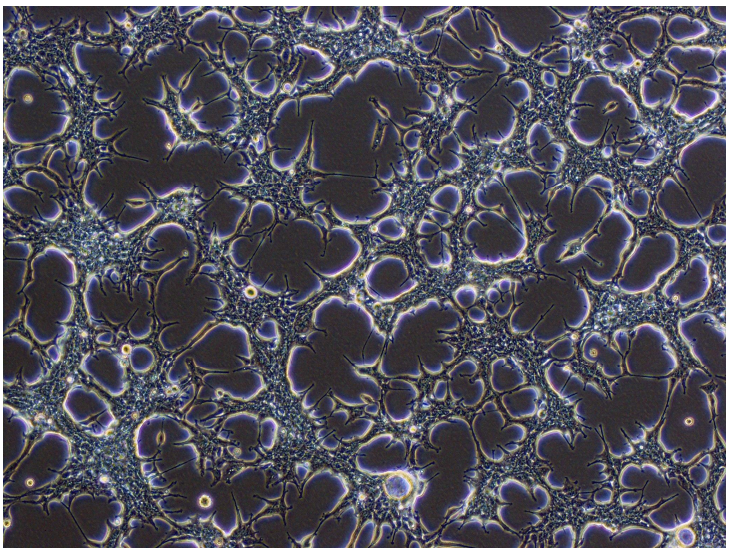


Figure 11: *HEK293T cell line under microscope*. HEK293T cells are incubated in DMEM medium for two days; they are forming monolayer and are nearly confluent.

3.2.1.4 Escherichia coli expression

The pET System is the most powerful system ever developed for the cloning and expression of recombinant proteins in *E. coli*. Target genes are cloned in pET plasmids under the control of strong bacteriophage T7 transcription; expression is induced by providing a source of T7 RNA polymerase in the host cell. T7 RNA polymerase is so selective and active that, when fully induced, almost all of the cell's resources are converted to target gene expression; the desired product can comprise more than 50% of the total cell protein a few hours after induction. Although this system is extremely powerful, it is also possible to attenuate the expression level simply by lowering the concentration of inducer. Decreasing the expression level may enhance the soluble yield of some target proteins. Another important benefit of this system is its ability to maintain target genes transcriptionally silent in the uninduced state (Baneyx 1999; Baneyx and Mujacic 2004; Pet system manual 2003).

3.2.2 Protocols

3.2.2.1 Preparing cultures of HEK293 cells

HEK293 cells growing on 10 cm culture dishes are harvested with 5 ml of trypsin-EDTA (0,25 %(w/v)(Gibco) applied for a 1 minute incubation) following removal of the culture medium (DMEM) and a PBS buffer wash. Cells are seen to lift from the dish at which point culture medium (10 ml) is added to quench trypsine activity and the cells are resuspended in culture medium. This resuspension volume is small (1ml per 10 cm) and the cells are plated directly onto 10 cm dishes. The density for stock dishes must be

sufficiently low to achieve 70 % confluence 2-3 days later. Cell cultures should be split every 2–4 days, when they reach 70–80% confluency.

3.2.2.2 Preparing Frozen Cultures of HEK 293 Cells

It is recommended to prepare frozen stock from early passage of HEK293 cells to ensure a renewable source of cells. Confluent cells from 10 cm dish are trypsinized, counted and centrifuged at 125 x g for 10 minutes. Supernatant is removed and pellet is resuspended at a density 107 cells/ml in freezing medium (10% DMSO, 40 % FCS, 50 % DMEM). 1 ml aliquots are dispensed into sterile cryovial and froze slowly 1-2 hour at -20 °C, then transferred to -70°C overnight and following day placed in liquid nitrogen for storage.

3.2.2.3 Expression of MHC class II – invariant chain complexes

HEK293T cells were maintained in DMEM (Dulbecco´s Modified Eagle Medium) medium supplemented with 4 mM L-Glutamine and 10 % fetal bovine serum in a CO2 incubator at 37C with 5% CO2.

For small-scale expression experiments, HEK293T cells were seeded in a 6-well cell culture plate containing 2 ml of medium at 80 % confluence. For each transfection experiment, 2.5 µg of plasmid DNA was diluted in 500 µl of serum free DMEM. 10 µl of Lipofectamin (Invitrogen) was added to the mixture and briefly vortexed. After 30 minutes of incubation at room temperature the transfection mixture was added to the cells, gently shacked and stored in CO2 incubator for 3 days. After 3 days the medium and cells were analyzed for protein expression.

For large-scale expression, HEK293T cells were seeded in T150 flasks and cultivated in 25 ml medium. When the cells reached 60 % confluence, co-transfection using 25 kDa linear PEI (Polysciences Inc.) was performed. 40 µg of individual plasmids containing genes for DRA, DRB, p31Ii / p41Ii were mixed in equal molar ratios in a 5 ml of serum free DMEM and 120 µg of PEI was added to the mixture. After 10 minutes incubation at room temperature the mixture was added to the cells. After 4 hours, the media was exchanged with 25 ml of DMEM without fetal bovine serum (FBS) and the cells were stored in CO2 incubator for 3 days.

3.2.2.4 Stefin A

PET3a vector containing the cDNA sequence of stefin A was transformed in *Escherichia coli* BL21(DE3) cells (Strauss et al., 1988; Martin et al., 1995). The expression was induced with 0.4 mM IPTG when OD600 reached 0.6. After 4 hours of expression, cells were centrifuged and frozen. The soluble fraction after lysis was concentrated and applied to a size-exclusion column (Superdex 75), equilibrated with the 10 mM phosphate buffer, 200 mM NaCl, pH 6.5. Eluates containing stefin A were pooled and dialysed against 10 mM phosphate buffer, pH 7.5. The dialysed sample was applied to Q-sepharose,

equilibrated with the 10 mM phosphate buffer, pH 7.5. Stefin A was eluted with 10 mM phosphate buffer, 500 mM NaCl, pH 7.5, dialysed against 10 mM acetate buffer, pH 5.5, concentrated to 1 mg/mL and frozen at -80°C until use.

3.3 Protein isolation

3.3.1 Theoretical background

Protein purification is a series of processes intended to isolate a single type of protein from a complex mixture. Protein purification is vital for the characterization of the function, structure and interactions of the protein of interest. The various steps in the purification process may free the protein from a matrix that confines it, separate the protein and non-protein parts of the mixture, and finally separate the desired protein from all other proteins. Separation of one protein from all others is typically the most laborious aspect of protein purification. Separation steps may exploit differences in protein size, physico-chemical properties, binding affinity and biological activity.

Usually a protein purification protocol contains one or more chromatographic steps. The basic procedure in chromatography is to flow the solution containing the protein through a column packed with various materials. Different proteins interact differently with the column material, and can thus be separated by the time required to pass the column, or the conditions required to elute the protein from the column. Usually proteins are detected as they are coming off the column by their absorbance at 280 nm. Many different chromatographic methods exist.

Affinity Chromatography is a separation technique based upon molecular conformation, which frequently utilizes application specific resins. These resins have ligands attached to their surfaces which are specific for the compounds to be separated. Most frequently, these ligands function in a fashion similar to that of antibody-antigen interactions. This "lock and key" fit between the ligand and its target compound makes it highly specific, frequently generating a single peak, while all else in the sample is unretained.

Metal binding is a common technique that involves engineering a sequence of 6 to 8 histidines into the N- or C-terminal of the protein. The polyhistidine binds strongly to divalent metal ions such as nickel and cobalt. The protein can be passed through a column containing immobilized nickel ions, which binds the polyhistidine tag. All untagged proteins pass through the column. The protein can be eluted with imidazole, which competes with the polyhistidine tag for binding to the column, or by a decrease in pH (typically to 4.5), which decreases the affinity of the tag for the resin. While this procedure is generally used for the purification of recombinant proteins with an engineered affinity tag (such as a 6xHis tag or Clontech's HAT tag), it can also be used for natural proteins with an inherent affinity for divalent cations

3.3.2 Protocols

3.3.2.1 Purification of MHC class II – invariant chain complexes

A Ni-Sepharose HisTrap column was used for immobilized metal affinity chromatography. The conditioned media containing His-tagged MHC class II/Ii complex was loaded onto 1-ml HisTrap column at flow rate 2 ml/min and washed with buffer A (30 mM Tris/HCl, 100 mM, 20 mM imidazole, pH 8.0) to remove proteins that were nonspecifically bound to the column. The His-tagged MHC class II-Ii complex was eluted by step elution with 50 % buffer B (30 mM Tris/HCl, 100 mM, 500 mM imidazole, pH 8.0).

Since only invariat chains have His_6 tag, the sample was further purified by ion-exchange chromatography to separate possible free invariant chains from invariant chains in complex with MHC class II molecules. The MonoQ column was equilibrated with 20 mM Tris/HCl, pH 8.0, the sample from affinity chromatography was loaded on the column and eluted with a step gradient of 0-10 M sodium chloride in 20 mM Tris/HCl, pH 8.0.The samples were buffer-exchanged on PD-10 desalting column and concentrated by centrifugation on microcons.

3.4 Protein characterization

3.4.1 SDS-PAGE, Western blotting and N-terminal sequencing

The purified MHC class II – Ii complexes were analyzed by SDS-PAGE using the 12% Bis-Tris gels. The protein bands were visualized by the Coomasie brilliant blue R-250 staining. For the Western blot analysis, the proteins were electro-transferred to the nitrocellulose membrane. The membrane was blocked with 5% (w/v) skim milk in Tris-buffer saline containing 0.05% (v/v) Tween 20 (TBST) for one hour at the room temperature. The rabbit polyclonal anti-DRA, anti-DRB and anti-Ii antibodies (Abcam) were added in dilution 1:1000 and incubated for one hour at the room temperature. The bound antibodies were detected with the horseradish peroxidase (HRP) conjugated goat anti-rabbit IgG and visualized with the peroxidase substrate.

For the N-terminal sequencing, the proteins separated on SDS-PAGE gel were electrotranfered to PVDF membrane, and visualized with the 0.1% Briliant blue G in 1% (v/v) acetic acid and 40% MeOH. The N-terminal sequencing was performed on Precise Protein Sequencing System 492 (PE Applied Biosystems).

3.4.2 Co-immunoprecipitation

3.4.2.1 Theoretical background

Immunoprecipitation is the technique of precipitating a protein antigen out of solution using an antibody that specifically binds to that particular protein. Immunoprecipitation of intact protein complexes (i.e.: antigen along with any proteins or ligands that are bound to it) is known as co-immunoprecipitation (Co-IP). Co-IP works by selecting an antibody that targets a known protein that is believed to be a member of a larger complex of proteins. By targeting this known member with an antibody it may become possible to pull the entire protein complex out of solution and thereby identify unknown members of the complex.This works when the proteins involved in the complex bind to each other tightly, making it possible to pull multiple members of the complex out of solution by latching onto one member with an antibody. This concept of pulling protein complexes out of solution is sometimes referred to as a "pull-down". Co-IP is a powerful technique that is used regularly by molecular biologists to analyze protein-protein interactions.

3.4.1.2 Protocols

The 1B5 mouse monoclonal antibody against human HLA-DR molecules were added to 1 ml of the conditioned medium and incubated for one hour on the rotary wheel at 4°C. The antibody-protein complex was pulled with Protein A-Sepharose beads (GE Healthcare) according to the manufacture's instructions and analyzed by Western blot with the IB5 antibody against MHC class II DRA chain, 2β antibody against the MHC class II DRB chain and antibody ICC against the C-terminal segment of human Ii (found in all forms of Ii).

3.4.3 Analytical size-exlusion chromatography

Superdex 200 HR 10/30 (GE Healthcare) size exclusion column was equilibrated in 30 mM HEPES, pH 7.5, 0.3 M NaCl buffer and calibrated with the gel filtration standards (Bio-Rad). The void volume (V_o) of the column was determined with Blue Dextran. The K_{av} (partition coefficient) for the individual standards was calculated from: $K_{av} = (V_r - V_o)/(V_c-V_o)$ (V_o - void volume of the column, V_r - retention volume of the protein standard, V_c - the geometric bed volume). The M_w of the complexes was determined from plots of $\log M_r$ of the standards versus K_{av} of the standards.

3.4.4 Crosslinking of truncated MHC class II-Ii complexes

10 μg of the MHC class II-Ii complex or Ii alone in 20 μl of HEPES, pH 7.5 buffer were mixed with 2.5 μl of 2.3% solution of gluteraldehyde and incubated at room temperature for 15 minutes. The crosslinking reaction was terminated by addition of 5 μl of 1 mM Tris, pH 7.5. The crosslinked complexes were separated on 7 % SDS-PAGE gels and separately detected with anti-Ii and anti-DRB antibodies as previously described.

3.4.5 Inhibitory activity of MHC class II – invariant chain complexes

Inhibitory experiments were performed using the following assay buffers: 0.1 M potassium phosphate, 1 mM EDTA, 5 mM DTT, pH 6.0 for cathepsins S and B; and 0.1 M sodium acetate, 1 mM EDTA, 5 mM DTT, pH 5.5 for cathepsins L and V. All cathepsins were mixed with the p31 or p41 Ii forms of the complexes in 1:1, 1:5 and 1:10 molar in a total volume of 200 μl. The final concentration of cathepsins in the assay mixture was 20 nM for cathepsins L and V, 50 nM for cathepsin S and 80 nM for cathepsin B. After 15 minutes incubation at the room temperature the Z-Phe-Arg-MCA substrate was added to achieve a 20 μM final concentration. The time dependent increase of fluorescence was measured on the microplate reader (Tecan) at the excitation wavelength 370 nm and the emission wavelength 460 nm.

3.4.6 Deglycosylation with *Endo H* endoglycosidase

The samples were treated with endoglycosidase EndoH (New England Biolabs, UK) according to the manufacturer's instructions. The samples were separated by SDS-PAGE and analyzed by Western blotting using anti-DRB antibodies.

3.4.6 Kinetic measurements

3.4.6.1 Theoretical background

An enzyme inhibitor is a molecule that binds to enzymes and decreases their activity. The binding of an inhibitor can stop a substrate from entering the enzyme's active site and/or hinder the enzyme from catalysing its reaction. Inhibitor binding is either reversible or irreversible. Irreversible inhibitors usually react with the enzyme and change it chemically (covalent bond formation). These inhibitors modify key amino acid residues

needed for enzymatic activity. In contrast, reversible inhibitors bind non-covalently and different types of inhibition are produced depending on whether these inhibitors bind the enzyme, the enzyme-substrate complex, or both.

Reversible inhibitors bind to enzymes with non-covalent interactions such as hydrogen bonds, hydrophobic interactions and ionic bonds. Multiple weak bonds between the inhibitor and the active site combine to produce strong and specific binding. In contrast to substrates and irreversible inhibitors, reversible inhibitors generally do not undergo chemical reactions when bound to the enzyme and can be easily removed by dilution or dialysis.

In competitive inhibition, the substrate and inhibitor cannot bind to the enzyme at the same time. This usually results from the inhibitor having an affinity for the active site of an enzyme where the substrate also binds; the substrate and inhibitor compete for access to the enzyme's active site. This type of inhibition can be overcome by sufficiently high concentrations of substrat, by out-competing the inhibitor. Competitive inhibitors are often similar in structure to the real substrate.

Reversible competitive inhibitors are competing with substrates to bind to the enzyme active site, as shown in

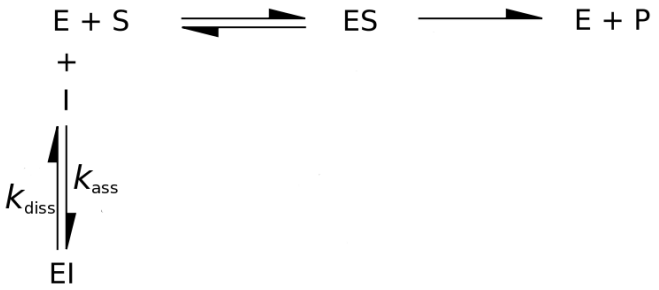


Figure 12: *Scheme of enzyme inhibition by reversible competitive inhibitors* (E – enzyme, S – substrate, P – product).

k_{diss} and k_{ass} are rate constants for inhibitor association and dissociation, respectively.

The equilibrium constant K_i is a measure of the strength of the interaction between the enzyme and inhibitor that come together to form the complex and can be calculated as:

$$K_i = \frac{k_{diss}}{k_{ass}} \tag{1}$$

When determining the equilibrium constant between enzymes and the slow tight binding inhibitors without the preequilibration step, the product formation is described as:

$$[P] = \frac{v_s \cdot t + (v_0 - v_s) \cdot (1 - e^{-kt})}{k} \tag{2}$$

where [P] is product concentration, vs is rate of product formation in stationary state, v_0 initial rate of product formation and k is pseudo-first-order rate constant describing the presteady state of the reaction.

Pseudo-first-order rate constant k has to be determined for a number of inhibitor concentrations and k_{ass} and k_{diss} are calculated with linear regression form equation:

$$k = \frac{k_{ass} \cdot [I]}{1 + [S] / K_m} + k_{diss} \qquad (3)$$

3.4.6.2 Determination of the inhibition kinetics

Inhibition kinetic of cathepsin V and stefin A was determined in continuous kinetics assay at 25°C. In kinetic experiment, cathepins V was assayed using 100 mM phosphate buffer, pH 5.5, containing 3 mM EDTA and 3 mM DTT. The active concentration of the enzyme was determined by the active site titration using synthetic inhibitor E-64. Inhibition experiments were performed under the pseudo first-order conditions with inhibitor concentration at least 40-fold higher that of the enzyme concentration. Less than 10% of the substrate was hydrolyzed during these experiments. Stefin A, was mixed at various concentrations with the substrate solution in the phosphate buffer in the fluorometric cuvette. The reaction was initiated by addition of the enzyme in a negligible volume. The progress curves were monitored at excitation and emission wavelengths of 370 and 460 nm, respectively, using a C-61 fluorimeter (Photon Technology International). Typical sigmoidal curves were observed and were analyzed by the method described previously (Morrison, 1982).

4 Results

In order to gain structural and biochemical insight to MHC class II-Ii interaction is necessary to produce soluble and secreted MHC class II-Ii complex.

First step in protein production is molecular cloning. Genes of interest are modified and amplified by PCR method and subsequently cloned into appropriate expression vectors. In this case we choose mammalian protein expression system. After co-transfection of desired genes, the expression of proteins is validated by Western blott analyzis or SDS-PAGE and if necessary, the expression is optimized. This step of validating the expression is small-scale expression. In this point we conclude if the choice of expression system was successful. If so, proteins are produced in large-scale expression in order to get reasonable amount of desired protein for purification and further experiments. Usually there are several purification steps required for pure proteins. Protein purification is vital for the characterization of the function, structure and interactions of the protein of interest. A protein purification protocol contains one or more chromatographic steps. We use affinity chormatogrpahy and ion-exchane chromatography. Once the proteins satisfy the criteria of purity, they are ready for characterization experiments. In case of protein complex production, formation of complex is verified by co-immunoprecipitation and molecular mass is assessed by size-exclusion chromatography and crosslinking experiments.. MHC class II-p41 complex has inhibiotion domain, which allows testing for inhibitory activity.

4.1 Construct design

To express the complexes in a soluble and secreted form only the luminal part of the Ii lacking the N-terminal cytosolic and transmembrane regions was cloned into the pHLSec plasmid in a frame with the N-terminal signal sequence present on the multiple cloning site of the plasmids (Fig. 13b). Similarly, the MHC class II α and β chains with the native signal sequences, but lacking the C-terminal transmembrane region, were cloned into the pcDNA3 mammalian expression plasmid (Fig. 13a). To make the affinity purification of the complexes possible, we have decided to exploit the 6xHis affinity tag. It was introduced at the C-terminus of the Ii. Selection of recombinant expression plasmids was performed by restriction analysis and DNA sequencing (Fig. 14).

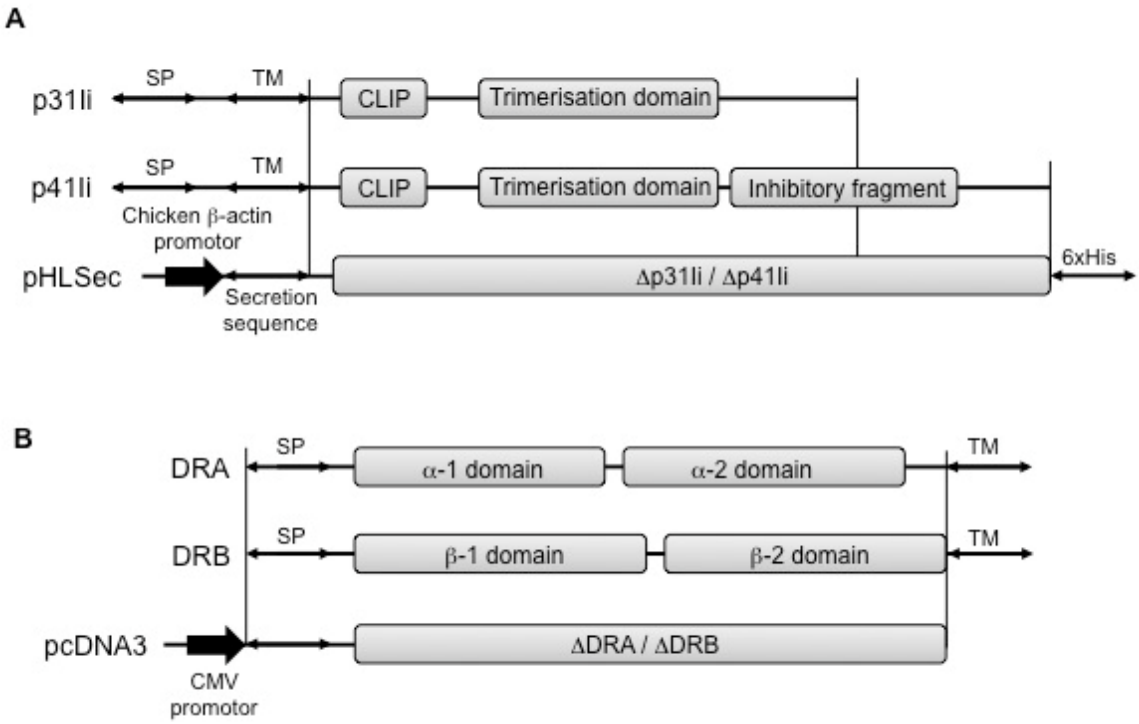


Figure 13: *Shematic representation of expression plasmids*. for (A) MHC class II associated p31 and p41 form of the invariant chain (Ii) and (B) MHC class II α (DRA) and β (DRB) chains. The regions od DRA, DRB and Ii that were cloned into the expression plasmids are annotated with Δ. SP–signal peptide, TM–transmembrane region, 6xHis–histidine affinity tag.

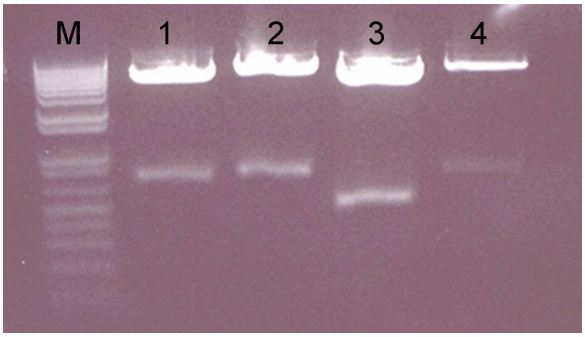


Figure 14: SDS-PAGE r*estriction analysis of expression vectors for MHC class II and Invariant chains*.M, 1Kb Plus ladder; lane 1, pCDNA3/DRA (HindIII/EcoRI) ; lane 2, pCDNA3/DRB (HindIII/EcoRI) ; lane 3, pHLsec/p31 (AgeI/KpnI) ; lane 4, pHLsec/p41 (AgeI/KpnI). DNA sizes are: pcDNA3-5.4 kb; pHLsec-4.5 kb; DRA-648 bp; DRB-681 bp; p31-474 bp; p41-663 bp.

4.2 Small-scale expression

A small scale expression screen was carried out on a 6-well cell culture plate using lipofectaminas a transfection reagent. To validate the expression system the individual constructs were transfected alone or mixed in different combinations. Three days post transfection the expression of genes was analyzed by Western blot using rabbit polyclonal antibodies against individual chains. When the HEK293T cells were transfected with the truncated forms of DRA or DRB or both together, only a weak accumulation of DRA and DRB protein products were detected inside the cells (Fig. 15a). From this we concluded that the intracellular accumulation of DRA and DRB does not result in the formation of fully active and competent MHC class II dimers. This is consistent with the previous finding that, in the Ii negative cells, MHC class II dimers could be still formed, however, they were misfolded and retained in the pre-Golgi compartments (Anderson and Miller, 1992).

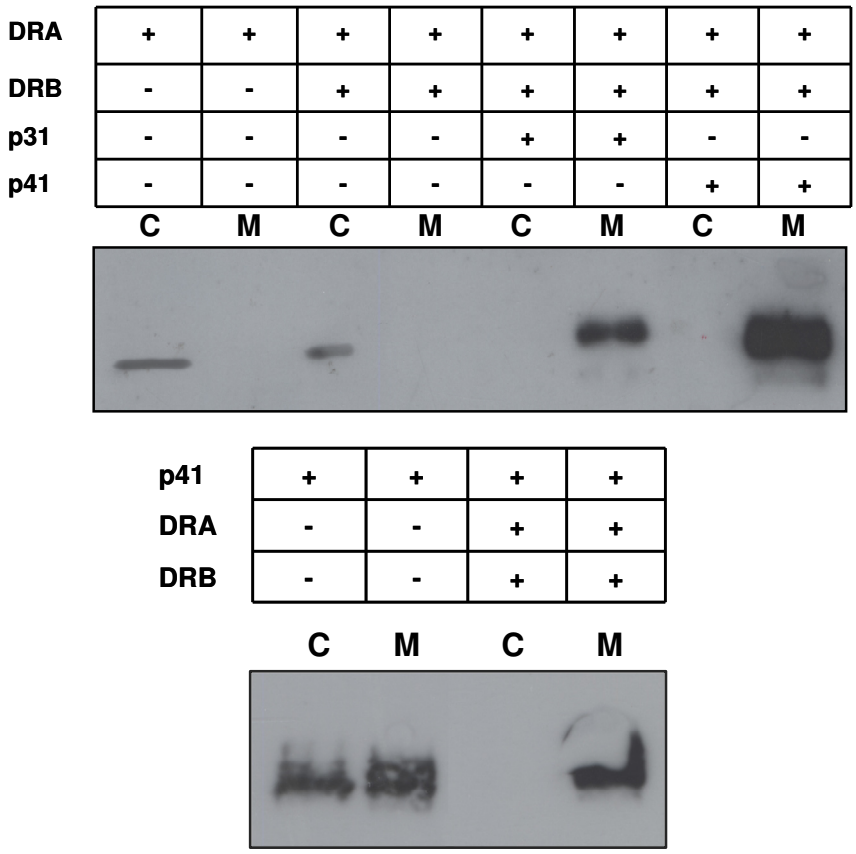


Figure 15: Western blot analysis of (A) DRA and (B) p41 Ii intra- and extra-cellular expression using different combinations of DRA, DRB and Ii plasmids. The proteins were detected using anti-DRA and anti-Ii rabbit polyclonal antibodies. Plus (+) and minus (-) indicates which combination of plasmids were used in transfection experiments. C indicates cell lysates and M medium.

When the DRA and DRB chains were cotransfected together with the Ii both chains were detected in the medium, suggesting that DRA, DRB and Ii together form a complex which is efficiently secreted from the cells. These results are consistent with

previous observations, that invariant chain is a chaperone protein essential for correct folding and assembly of MHC class II complexes.In the reverse experiment the HEK293T cells were transfected with the p41 form of Ii alone or together with DRA and DRB (Fig. 15b). The truncated form of the p41 form of Ii lacking the transmembrane region was in both cases efficiently secreted outside the cells, however the secretion was more efficient in the presence of DRA and DRB chains.

To additionally confirm that the secreted Ii, DRA and DRB chains really form the complex, we imunoprecipitated DRA from the medium of HEK293T cells cotransfected with DRA, DRB and Ii using IB5 monoclonal antibody. By the Western blot analysis of protein A-Sepharose pulled immunoprecipitated fractions we were indeed able to identify all three chains forming the complex (Fig 16).

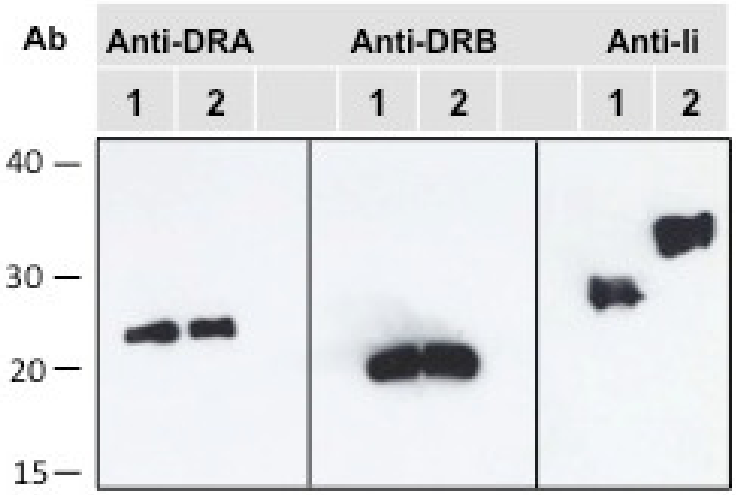


Figure 16: *Western blot analysis of co-imunoprecipitated MHC class II-Ii complexes***.** The complexes were isolated from the transfection medium by co-imunoprecepitation using IB5 monoclonal antibody against DRA and analyzed by Western blot using the monoclonal antibodies against DRA, DRB and Ii. Lane 1: MHC class Ii-p31 complex incubated with IB5 antibody against α chain, lane 2: MHC class Ii-p41 complex incubated with IB5 antibody against α chain, lane 3: MHC class Ii-p31 Ii complex incubated with 2β antibody against β chain, lane 4: MHC class Ii-p41 Ii complex incubated with 2β antibody against β chain, lane 5: MHC class Ii-p31 Ii complex incubated with ICC antibody against invariant chain, lane 6: MHC class Ii-p41 Ii complex incubated with ICC antibody against invariant chain.

4.3 Large-scale expression and purification of MHC class II – Ii invariant chain complexes

The large-scale expression was routinely performed in 20 T150 flasks yielding 500 ml of the growth medium as a source for isolation of the complexes. Plasmids DNA coding for DRA, DRB and the p31 or p41 forms of the Ii were mixed in equal ratios and co-transfected into HEK293T cells using PEI. The initial expression and purification trials showed that even a trace amount of FBS present in the growth medium interferes with the protein purification on Ni-chelating Sepharose. Therefore the transfected HEK293T cells were maintained in the DMEM medium without FBS. The absence of FBS slightly decreased the cell growth and viability, but did not influence the protein production yield. Four days post transfection, the medium was collected, concentrated and applied to the Ni-chelating column. After extensive washing with the binding buffer, the bound complexes were eluted with the binding buffer containing 300 mM imidazol. To separate the MHC class II Ii complexes and the empty Ii complexes that might co-purify on the Ni-affinity column, the samples were additionally applied on the Mono Q ion-exchange column. The calculated isoeletric points for the (αβIi)3 complex and the Ii trimer are 5.71 and 8.05, respectively. Therefore the ion-exchange chromatography performed at pH 7.5 should enable their separation. At this pH the nonameric complex is negatively charged and should strongly absorb to the positively charged ion-exchange carrier, whereas the more acidic empty Ii trimer should theoretically pass through the carrier unbound. Elution with the linear gradient of increasing NaCl concentration showed that the complexes with both forms of Ii were eluted at 0.25 – 0.3 M NaCl in a single protein peak and that only a minor amount of proteins did not bind to the ion-exchange column (Fig 17A and B). Three collected protein fractions corresponding to major eluted peak from the Mono-Q column were analyzed by SDS-PAGE (Fig 17A and B). The positions of DRA, DRB and Ii chains on the PAGE-gel were identified by comparing the results of Westeren blot analysis. Mw for DRA and DRB deduced from SDS-PAGE analysis are in good agreement with those calculated from the amino acid sequence. The positions of the p31 and p41 forms of Ii on the PAGE gel appear at higher molecular weights than expected, presumably due to their extensive N-glycosylation. The identity of the individual chains and purity of the complexes was additionally verified with the N-terminal amino acid sequencing, which revealed only amino acid sequences corresponding to the chains of DRA, DRB and the Ii. The total amount of the purified MHC class II molecules in complex with Ii is approximately 2.5 mg from 500 ml of media, hence 5 mg per liter.

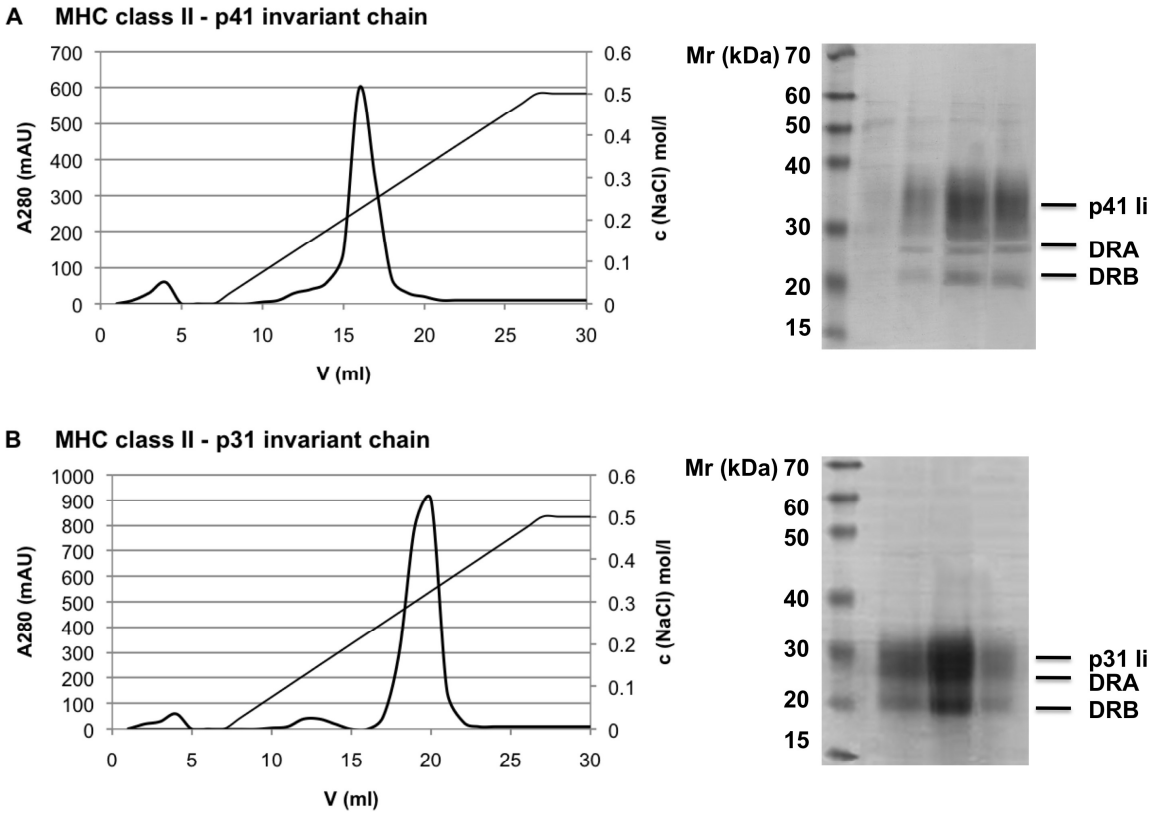


Figure 17: *Purification of MHC class II–Ii chain complexes.* Mono Q ion-exchange chromatography and SDS-PAGE analysis of the protein fractions corresponding to the major eluted peaks for (A) MHC class II – p41 Ii and (B) MHC class II – p31 Ii chain complexes.

4.4 Oligomeric state of MHC class II – Ii complexes

To analyze the oligomeric stechiometry of the complexes lacking the transmembrane regions the molecular mass of the purified products was assessed by analytical size exclusion chromatography and crosslinking experiments. The purified proteins eluted from the Superdex 200 HR size exclusion column in a single symmetric peak with elution volume around 10 ml (Fig 18A). The apparent molecular weight of these complexes was determined from the calibration curve obtained by plotting the elution volumes of the protein standards versus logarithm of their molecular weight (Fig 18B). The elution volumes of the ΔDRADRBp41Ii) complexes purified from HEK293T and HEK293S GnTl-cell lines correspond to the apparent molecular weights of 290 and 273 kDa. Since only a single peak was eluted from the size-exclusion column these results suggest that the expressed ΔDRA, ΔDRB and the ΔIi associate in high-molecular weight complex. (It should be noted that the estimation of molecular weight of proteins by size exclusion chromatography is reliable only for globular proteins as it depends on the protein radius

and its hydrodynamic properties and should therefore be considered only as an approximation of the actual weight.).

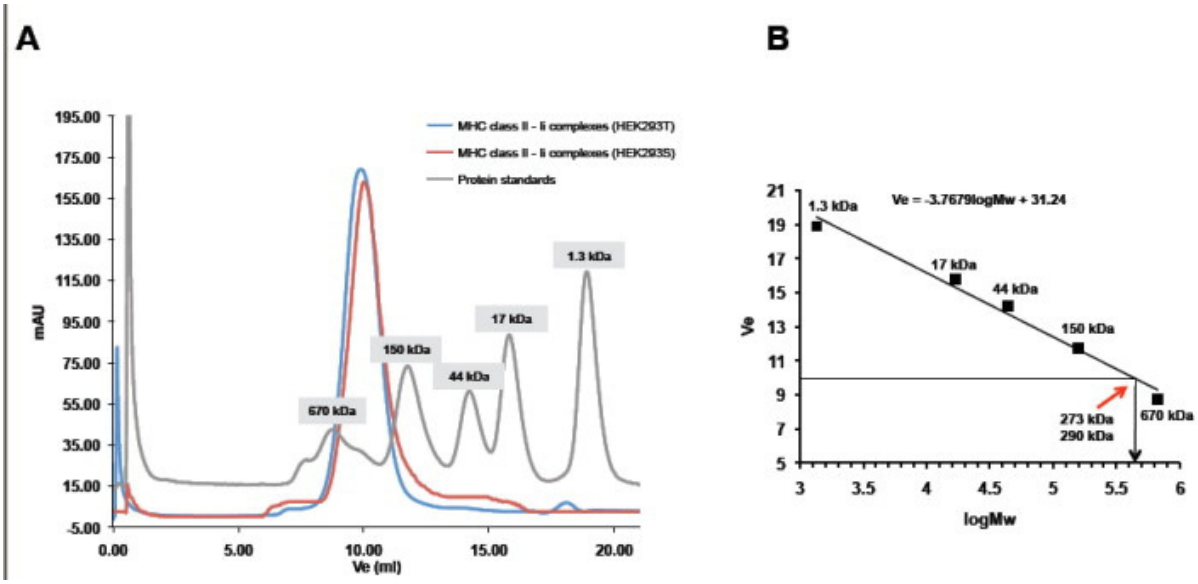


Figure 18: *Assessment of molecular weight of MHC class II–41Ii complexes using size exclusion chromatography.* (A) Superdex 200 HR elution profiles of MHC class II–p41Ii complexes expressed in HEK293T (blue) and HEK293S GnTl-cells (red). The elution profile of protein standards (thyroglobulin 667 kDa, gamma-globulin 158 kDa, ovalbumin 44 kDa, myoglobin 17 kDa and vitamin B12 1.35 kDa) is shown in gray. (B) Calibration plot obtained from plotting the elution volume of standards versus their molecular weight. The elution volume and calculated molecular weight for MHC class II–p41Ii complexes are marked.

To independently confirm the oligomeric structure of the soluble MHC class II–p41Ii complex, the purified complex was crosslinked with gluteraldehyde and DSS. Both crosslinkers gave the same results so only data with gluteraldehyde are shown (Fig 19). To reduce the unspecific crosslinking between complexes, the concentration of the protein used in these experiments was reduced to below 0.5 mg/ml. The molecular weight of the cross-linked soluble p41Ii is 100 and 75 kDa for glycosylated and non-glycosylated forms of the protein. This weight is in agreement with the expected molecular weight of the Ii trimers. Since the trimeric structure was formed in the absence of the transmembrane and cytosolic parts of the Ii, it is evident that the transmembrane region and cytosolic tail of Ii are not critical for the formation of the nonameric complex. The luminal segment of Ii should also contain information that induces trimerization. The crosslinking of the Δ(DRADRBp41Ii) complexes showed two major bands of molecular mass of 130 and 180 kDa for glycosylated form of the complexes, and two bands of 110–140 kDa for underglycosylated forms of the complexes. However, a minor bands of 210 and 170 kDa were also observed. A protein complex of 209 kDa would correspond to the calculated molecular weight of the expected $(DRA–DRB–Ii)_3$ nonameric form. The results from the cross-linking experiments suggest considerable heterogeneity in the binding of HLA–DRAB complexes on the Ii trimer with a preference for one and two

MHC class II complexes corresponding to molecular weights of 120 and 160 kDa, respectively (Table 3).

Table 3: *Calculated molecular weights and the numbers of potential N-glycosylation sites.* This is refering to the expressed constructs of the truncated MHC class II DRA, DRB chains and the p31 and p41 forms of I and their complexes.

Proteins	Mw (Da)	N- glycosylation sites
DRA	24.365	2
DRB	23.071	1
p31	18.397	2
p41	25.345	4
DRA+DRB+p31	65.833	5
DRA+DRB+p41	72.781	7
$(DRA+DRB+p31)_3$	197.499	15
$(DRA+DRB+p41)_3$	218.343	21

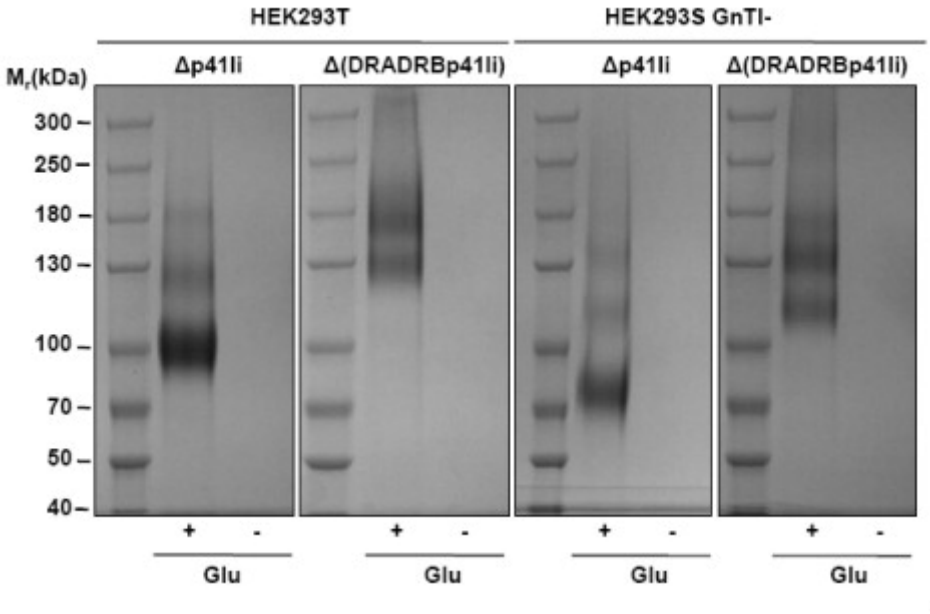


Figure 19: *Crosslinking of the soluble MHC class–p41Ii and p41Ii complexes*. The lanes with the gluteraldehyde (Glu) crosslinked material are marked on the top with "+" and the lanes without crosslinking with "–". The proteins were separated on 3–8% Tris–acetate gels and stained with Coomassie Blue.

Comparison of DRA, DRB chains and the Ii inside the cells and secreted forms shows obvious differences in molecular weight. Since all three chains contain a number of Nglycans (Table 3), the deviation in the molecular weight marks conversion of high-mannose to complex Nglycans that is the result of transport through the secretory pathway. This also indicates that DRA and DRB when transfected alone or only in pairs are retained in the ER. To provide further support that the differences in the molecular weight arise from the localization and N-glycosylation pattern, we have analyzed DRB

transfected into the HEK293S GnTl- cell line which is deficient in N-acetylglucosaminyltransferase I activity and therefore lacks complex type N-glycans. When DRB was cotransfected with DRA and the Ii in HEK293S GnTl- cells, the apparent molecular weight of the secreted DRB is identical to the molecular weight of the intracellularly accumulated DRB (Fig. 20A). The weight of intracellular DRB was further decreased by the EndoH treatment, which removes the basic mannose glycan attached to the proteins in the ER during early stages of their synthesis, wheras secreted ΔDRB (when cotransfected with ΔDRA and ΔIi) is EndoH resistant. The ER localization of DRB, when transfected alone, was additionally confirmed by staining fixed cells with anti-DRB antibodies before analyses by confocal microscopy (Fig. 20B). Also DRA was retained in the ER according to biochemical and microscopy criteria.

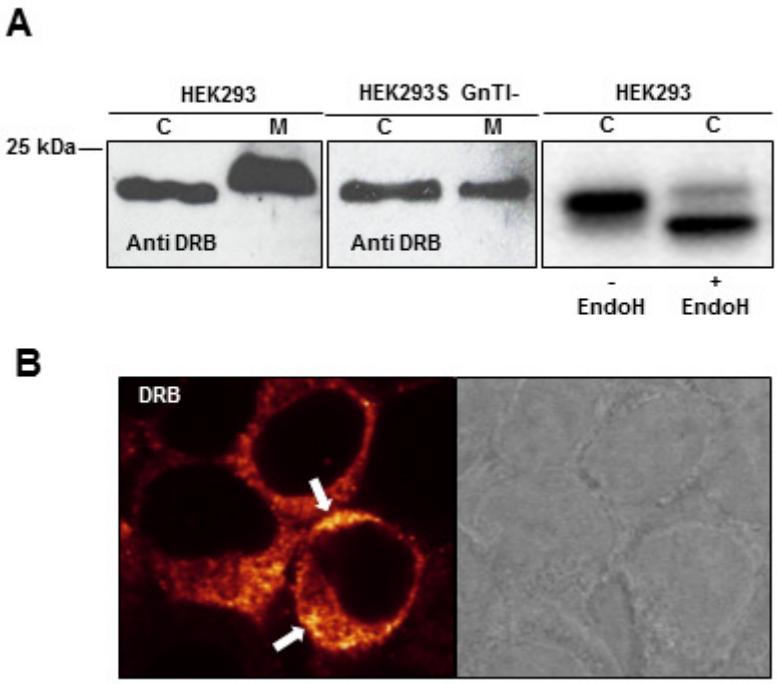


Figure 20: *Analysis of the molecular size, cellular localization and glycosylation pattern of ΔDRB*. (A) ΔDRB was transfected together with ΔDRA and the ΔIi in HEK293T and HEK293S GnTl- cells and analyzed by anti-DRB immunobloting. C and M annotates intracellular and extracellular expression. Intracellular and secreted ΔDRB was additionally treated with Endo H. (B) Immunofluorescence staining of ΔDRB transfected alone in HEK293T cells and imaged by confocal microscopy. The arrow indicates the accumulation of the ΔDRB inside the ER

4.4 Inhibitory activity of the MHC class II – Ii complexes

Cathesin L, V and S were expressed as described in Mihelic et al.2008 wheres cathepsin B was expressed as described in Kuhelj et al., . MHC class II complex with p31 and p41 isoform were expressed in HEK293T cells.

The two major Ii processing cathepsins L and V are inhibited by the inhibitory fragment present in the p41 form of the Ii in the low picomolar range with K_i=0.005nM and K_i =0.007nM, respectively, whereas the interaction between cathepsin S and the p41 inhibitory fragment is in high micromolar range (K_i =208nM) and cathepsin B is not inhibited at all. The presence of the inhibitory fragment in the p41 form of Ii enabled us to test its inhibitory activity As expected, the measurement of enzymatic activity of cathepsins B, L, S, and V by the fluorogenic substrate under the conditions tested showed that the presence of the p31 form of Ii in the complexes exhibited no effect, whereas the presence of the p41 Ii form in the samples inhibited cathepsins L and V while leaving the activity of cathepsins B and S unaffected (Fig. 21-24). This test of inhibitory activity has provided a strong evidence for the porper folding of the complexes. Furthermore, this test demonstrates that the two splice variants of the Ii-MHC class II complex indeed differ by their effect on the activity of endosomal proteases in the areas in MIIC.

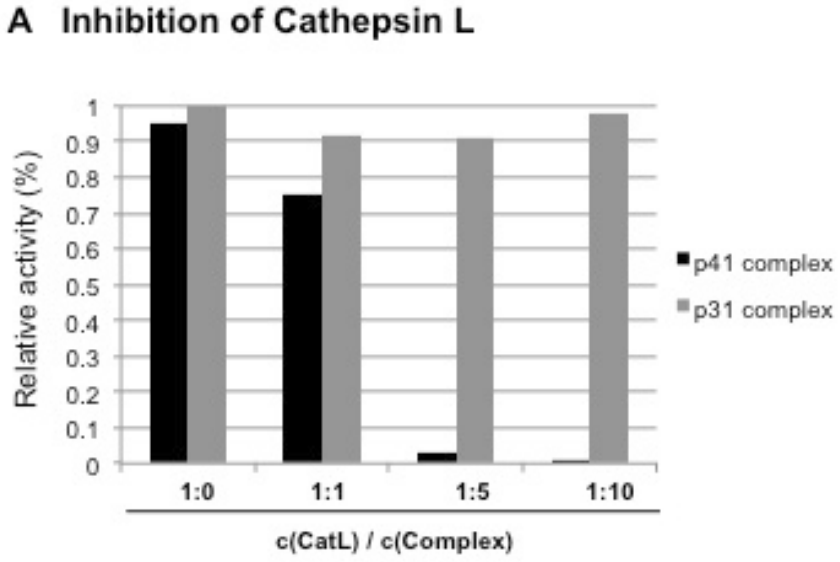


Figure 21: *Inhibition of cathepsin L by MHC class II-p41 and MHC class II-p31 Ii complexes.* Cathepsins L and MHC class II-Ii complexes were mixed in the different molar ratios. The activity of cathepsins L was monitored using the substrate Z-Phe-Arg-AMC.

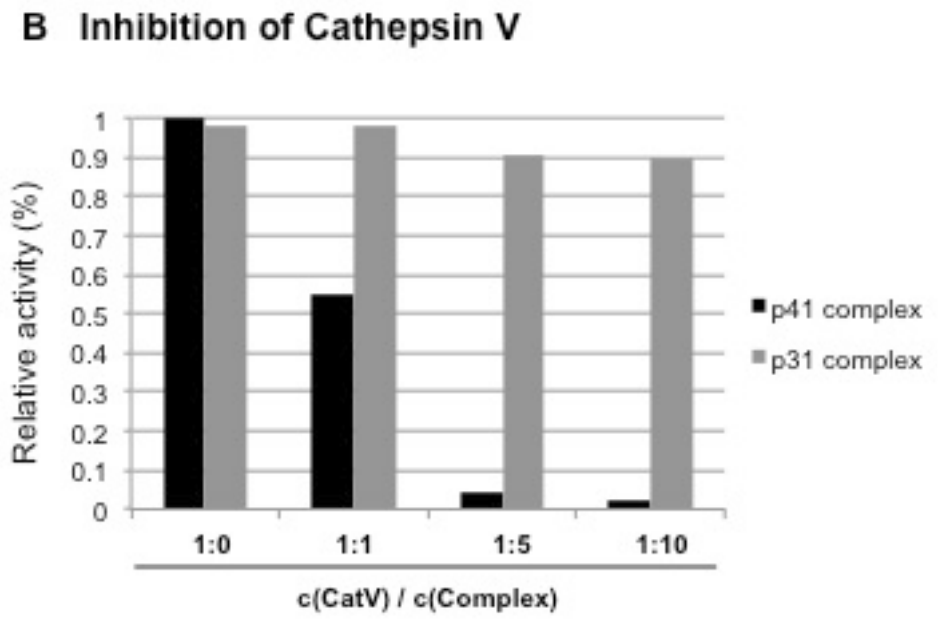


Figure 22: *Inhibition of cathepsin V by MHC class II-p41 and MHC class II-p31 Ii complexes*. Cathepsins V and MHC class II-Ii complexes were mixed in the different molar ratios. The activity of cathepsins V was monitored using the substrate Z-Phe-Arg-AMC.

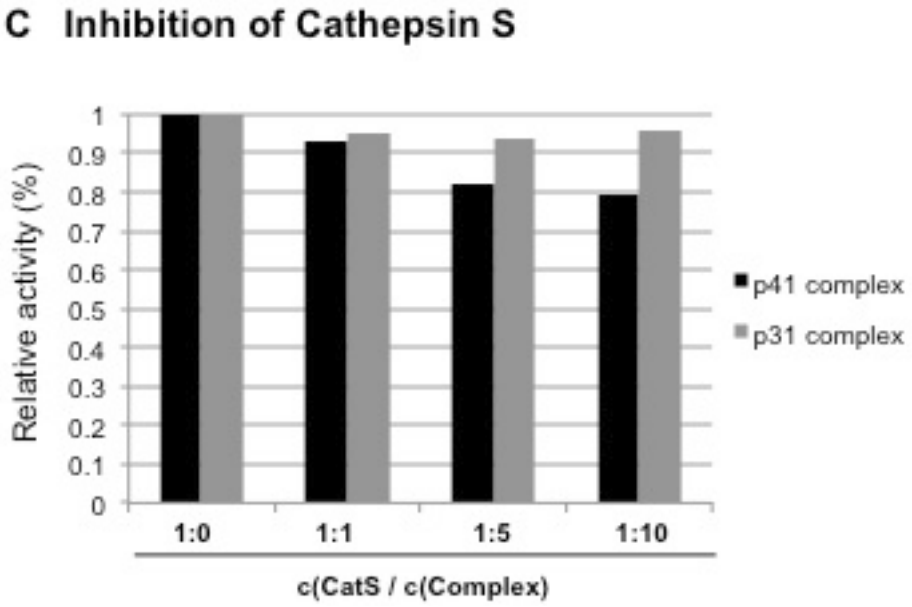


Figure 23: *Inhibition of cathepsin S by MHC class II-p41 and MHC class II-p31 Ii complexes*. Cathepsins S and MHC class II-Ii complexes were mixed in the different molar ratios. The activity of cathepsins S was monitored using the substrate Z-Phe-Arg-AMC.

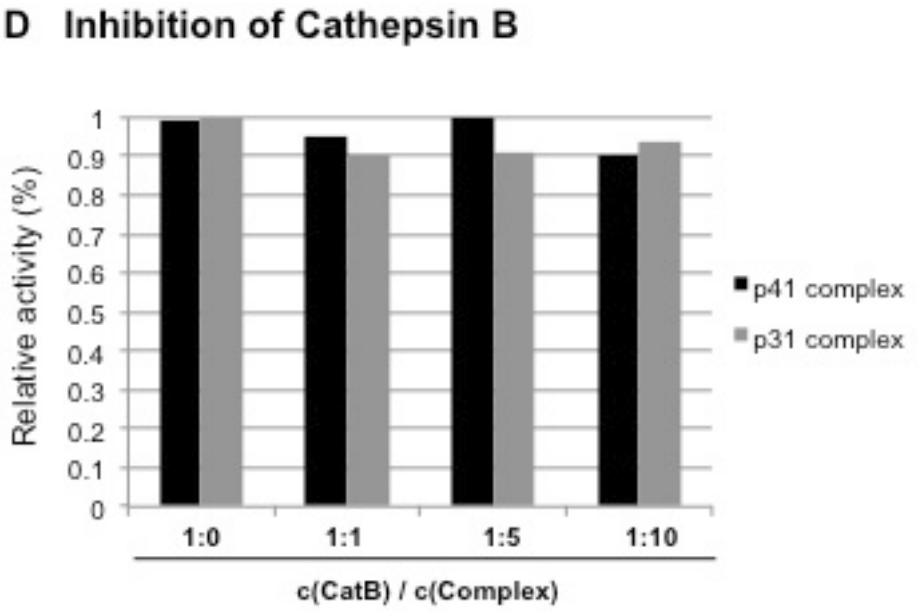
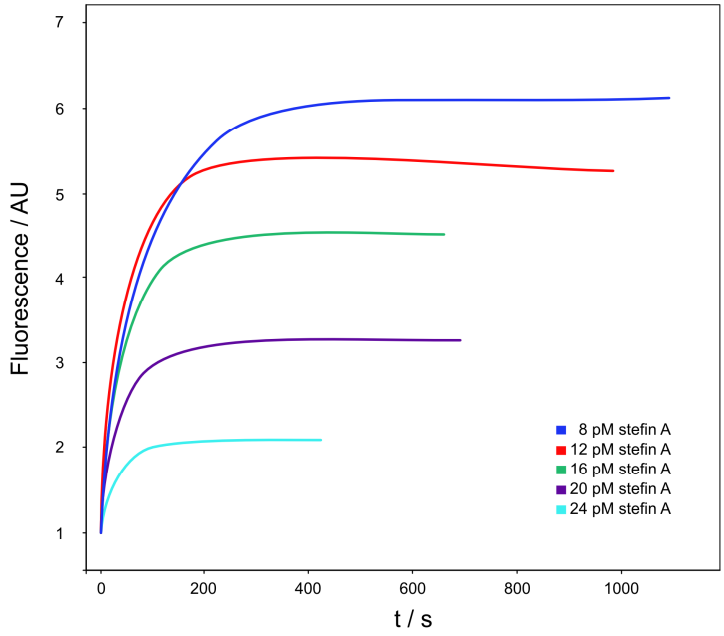


Figure 24: *Inhibition of cathepsin B by MHC class II-p41 and MHC class II-p31 Ii complexes*. Cathepsins B and MHC class II-Ii complexes were mixed in the different molar ratios. The activity of cathepsins B was monitored using the substrate Z-Phe-Arg-AMC.

4.5 Kinetics of stefin A inhibition of cathepsin V

The active concentration of cathepsin V was determined by the active site titration using the synthetic inhibitor E-64. Cathepsin V was typically 75-85% active and stefin A 50-70% active. The percentage varied from batch to batch and was strongly influenced by repeating freeze – thaw cycles. To study the inhibition of cathepsin V and papain, we assumed a simple, competitive mechanism of inhibition without a preequilibrium step. Where k_{ass}, and k_{diss} are the rate constantsf or complex formation and dissociation, respectively. Inhibition kinetic of cathepsin V and stefin A was determined in continuous kinetics assays at 25°C (Morrison, 1982). Enzyme was added to a mixture of the inhibitor and fluorogenic substrate and the concentration of product (P) was continuously monitored by fluorescence. Figure 25 shows typical biphasic progress curves recorded following the reaction of cathepsin V with stefin A.



.

Figure 25: *Progress curves for the inhibition of cathepsin V by stefin A*

The progress curves were then fitted by nonlinear regression with program Grafiti.

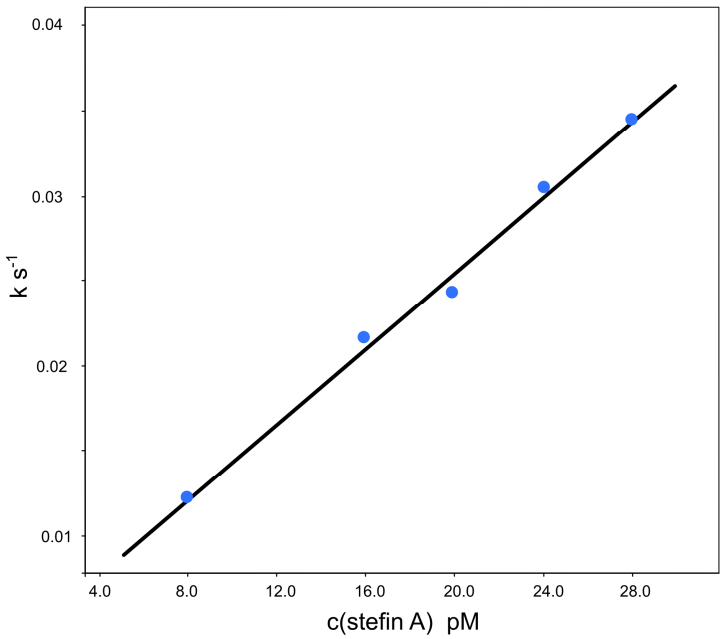


Figure 26: *Dependence of k from concentration of stafin A inhibitor*

From theoretical curves from figure 24 and line from figure 25 we were able to calculate following kinetic constants of cathepsin V inhibition by stefin A:

k_{ass} (10^6 $M^{-1}s^{-1}$) =2.4±0.2;
k_{diss} (10^{-4} s^{-1})=2.9±0.7;
K_i (nM)=1.24±0.31

The inhibition constants for cathepsin V inhibition by stefin A is similar to that already published (Cheng et al., 2006). There are some differences (2.0 nM vs. 0.11 nM for cathepsin V), but these can be explained by systemic errors and differences between measurements in different labs.

Expressed stefin A was used to determine crystal structure of cathepsin B-stefin A complex to clarify the structural properties of the occluding loop upon the binding of stefins. The crystal structure of the complex between wild-type human stefin A and wild-type human cathepsin B was determined at 2.6 Å resolution. The papain-like part of cathepsin B structure remains unmodified, whereas the occluding loop residues are displaced. The part enclosed by the disulfide bridge containing histidines 110 and 111 (i.e. the 'lasso' part) is rotated by ~ 45° away from its original position. (Renko et al., 2010).

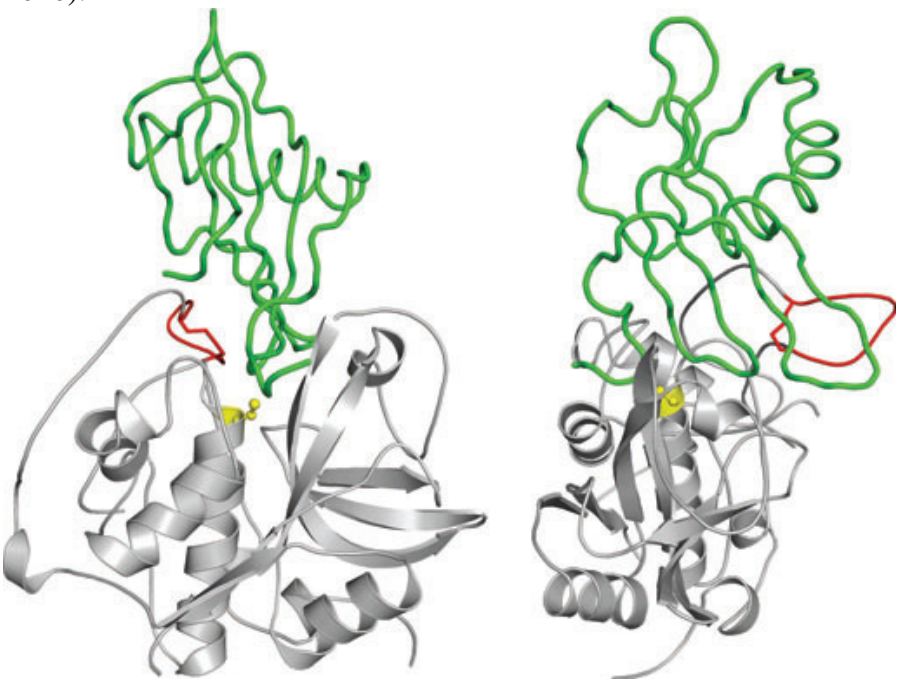


Figure 27: *Structure of the cathepsin B–stefin A complex.* (A) View along the active site cleft. (B) View perpendicular to the active site cleft. Cathepsin B is shown in gray and stefin A in green. The catalytic cysteine is shown in yellow. The wedge-shaped structure of stefin A fills the active site cleft along the whole length and displaces the occluding loop (the 'lasso' is shown in red).

5 Discussion

Major histocompatibility class (MHC) II – invariant chain (Ii) complexes are transmembrane glycoproteins that play a central role in adaptive immunity (Creswell et al., 1987; Rocha and Neefjes, 2008). They are synthesized in the lumen of endoplasmic reticulum (ER). The molecular mechanism of MHCII α (DRA) and β (DRB) chain and Ii association has been extensively studied for the past decades, however the exact mechanism of their assembly still remains unclear. The classic model of the MHCII–Ii assembly suggested that the Ii first forms a trimer via its trimerisation domain. This trimers then act as a chaperone-like scaffold for the assembly of three α and β chain dimers that concurrently bind to the Ii trimmer forming a nonameric complex (Romagnoli et al., 1994, Anderson and Miller 1992). It should be noted that these results were based on cross-linking studies and not confirmed by analyses of MHC class II-Ii complexes in solutions. Once fully assembled, the MHC class II-Ii complexes are transported to late endosomal compartments called MIIC of professional antigen presenting cells (APCs) (Neefjes et al., 1990; Peters et al., 1992). Here, Ii is degraded in a step-wise manner by lysosomal cysteine proteases (Pieters et al., 1991, Hsing and Rudensky, 2005)

5.1 Choice of expression system and construct design

The MHC class II molecules exist in a number of allelic variants (Marsh et al., 2010). They all are transmembrane complexes. The most frequently studied alleles are the HLA DR molecules. For our model system we have chosen to work with the HLA DR1 variants. All three chains, alpha (DRA), beta (DRB) and Ii, contain an N-terminal signal sequence that targets the synthesis of nascent polypeptide chains into the lumen of the ER. In the Ii chain the signal sequence is followed by a transmembrane region (G47 to Y72) and a luminal domain, whereas in the MHC class II α and β chains the transmembrane regions are located at their C-terminal ends of the molecules. They include residues from N217 to I239 and from M228 to F250 of the α and β chains respectively.

So far there have been no reports of the successful expression of MHC class II-Ii complexes in quantities sufficient for their structural and biochemical characterization. There have been attempts to produce MHC class II dimers and there are available protocols using bacterial, insect and mammalian cells (Frayser et al, 1999; Kalandadze et al., 1996). The most of these protocols rely on the in vitro refolding of the bacterial expressed and denaturated MHC class II chains in the presence or absence of a peptide with the high affinity for the MHC class II dimers. The expression of soluble MHC class II α and β chains lacking the transmembrane region in the eukaryotic systems has been difficult due to the tendency of MHC class II chains to aggregate even in the presence of peptide ligands.

Production of MHC class II/Ii chain complexes in E.coli does not appear to be suitable. Expression of the p41 inhibitory fragment failed when expression was induced in the

bacterial cytosol. It could not be produced in soluble form and also the refolding protocol was also not efficient. But when expression of p41 inhibitory fragment was targeted to the periplasmic space of bacterial cells, results were positive (Mihelic et al., 2008). However, expression of MHC class II /Ii complex in periplasm seems impossible due to the size and complexity of this human complex.

For the expression of MHC class II–invariant chain complexes, human embryonic kidney cell line HEK293T was used. In contrast to the standard HEK293 cell line, HEK293T cells are transformed with SV40 Large T-antigen and are capable of episomal replication of transfected plasmids carrying the SV40 origin of replication. This results in an increased number of plasmid copies inside the transfected cells, thereby increasing the level of protein production.

HEK293 cells are highly suitable for large scale production of secreted soluble proteins and receptors (Ariescu et al., 2006). When compared to insect cells and yeast expression systems, the expression in HEK293 cells is more rapid and proteins are modified post translationally, including glycosylation which is identical to that of the authentic human proteins.

The exact positions of the transmembrane regions inside the protein sequences were identified using TMpred server (Hofmann & Stoffel, 1993). To express these complexes in a soluble form, cDNA sequences of DRA and DRB with authentic N-terminal sequences, but without the C-terminal transmembrane portion, were cloned into the pcDNA3 expression plasmid (Fig. 1A). In the case of Ii, the transmembrane region is inserted between the N-terminal signal sequence and the rest of the protein, making it impossible to use the authentic signal sequence to target the synthesis of Ii without a transmembrane region into the lumen of ER. To bypass this problem, cDNA sequence of the Ii chain lacking N-terminal signal sequence and transmembrane region was cloned in frame with the N-terminal secretion sequence located on the pHLsec expression plasmid and fused with C-terminal 6xHis tag (Fig. 1B).

To purify the complexes, we have exploited the 6xHis affinity tag. It was introduced at the C-terminus of the Ii. On Ii, the major interaction site with MHC class II dimers on the Ii is the CLIP region located at the N-terminal part of the luminal domain, whereas the C-terminal part of the Ii most probably does not interact with other parts of the complex and is exposed to the solvent. On the other hand, the sequence of the 6 histidines at the C-termini of the α and β chains might interfere with dimer formation or binding to the Ii.

5.2 Role of Ii chain transmembrane domain

Several studies evaluated the importance of the transmembrane (TM) region in the trimerisation of Ii. It was shown that the TM regions of the Ii strongly self-associate as a trimer (Dixon et al., 2006). Destabilization of this TM interaction prevented trimerisation of Ii and assembly of the complexes (Ashman and Miller 1999),), leading to disruption of antigen presentation (Frauwirth and Shastri, 2001). These studies were performed by mutations of highly conserved amino acid residues within the Ii TM domain and do not fully support the role of the TM domain in the assembly of the complexes.To unambiguously address the role of the TM domain of the Ii in the assembly of the MHC class II–Ii complexes, we produced soluble complexes lacking the TM region.

To independently confirm the oligomeric structure of the soluble p41Ii protein, the purified protein was crosslinked with gluteraldehyde and DSS. The molecular weight of the cross-linked soluble p41Ii is 100 and 75 kDa for glycosylated and non-glycosylated forms of the protein. This weight is in agreement with the expected molecular weight of

the Ii trimers. Since the trimeric structure was formed in the absence of the transmembrane and cytosolic parts of the Ii, it is evident that the transmembrane region and cytosolic tail of Ii are not critical for the formation of the MHC class II-Ii complex. The luminal segment of Ii should also contain information that induces trimerization.
Our results do not exclude the importance of the transmembrane domain for Ii trimerization; however, they do show that trimers can be formed in the absence of the transmembrane region. This further indicates that, in the MHC class II–Ii complexes, more than one region contributes to the formation of the HDA, HDB, Ii trimers.

5.3 Oligomeric state of MHC class II/Ii complexes

Previously it was shown that the first step in the assembly of MHC class II–Ii complexes is the trimerization of the Ii and that MHC class II dimers can associate only with a trimeric form of the Ii. The formation of invariant chain is rapid and is the first step in assembly of a nine-chain αβIi complex. Three higher molecular weight complexes are then progressively formed, that correspond to an invariant chain trimer with one αβ dimer, two αβ dimers and three αβ dimers, respectively. (Lamb and Cresswell, 1992). Chemical cross-linking experiments have demonstrated that the invariant chain is trimeric when not associated with class Ii molecules (Marks et al., 1990) and that αβIi complexes from cell lysates exist as a nine-chain structure, containing an equimolar ratio of α, β and invariant chains (Roche et al., 1991).

We have analyzed the oligomeric stoichiometry of the complexes lacking the transmembrane regions. The molecular mass of the purified products was assessed by analytical size exclusion chromatography. The 250 kDa form of the p31 complex predominates, whereas the predominant form of the p41 form of the complex is the 310 kDa form. This is in approximate agreement with the expected sizes of nonameric complexes 197.5 kDa for the p31 form of the complex and 218.3 kDa for the p41 form of the complex.

The differences between the calculated Mws and those determined from the analytical size exclusion chromatography are probably a consequence of glycosylation. The p31 form of $(\alpha\beta Ii)_3$ complex namely contains 15 glycosylation sites, whereas the p41 form of the complex contains 21 glycosylation sites (Table 1). Hence, the differences in the number of glycosylation sites and the heterogeneous pattern of glycosylation are probably responsible for the deviation of the molecular size peaks of the p31 and p41 forms of the complexes from the theoretical values.
With two independent experiment approaches, namely size-exclusion chromatography and crosslinking, we have estimated the composition of secreted MHC class II–Ii protein complexes. The size exclusion experiments suggest that the complex resembles a trimer of trimers with three MHC class II complexes associated to three Ii chains that form the core of the complex. Chemical cross-linking and more precise molecular weight determination by SDS–PAGE reveals a more complex picture with an Ii trimer associated with mainly one or two MHC class II complexes and only a small proportion in the non-americ state. This would correspond to the recent observations Koch and coworkers (Koch et al., 2011) who suggested that the stoichiometry of the MHC class II–Ii chain composition may not be a trimer of trimers (a DRA–DRA–Ii trimer) (Roche et al., 1991) but a pentamer composed of a trimer of Ii with a one pair of ΔDRADRB) bound. As Ii is present in the ER in considerably higher amounts than DRA and DRB chains, two options exist. MHC class II-Ii complexes are only considered correctly folded to leave the ER when in a nanomeric complex (the trimer of trimers) or the Ii trimer can leave the ER when already one or two MHC class II complexes are associated. Our results suggest that

the latter option is most likely. The ratio between fully and partially MHC class II Ii-loaded complexes may depend on the relative amounts of Ii and MHC class II DRA and DRB expressed, which may vary between cells and likely depending on the DRA and DRB sequences. The existence of intermediate MHC class II–Ii complexes of various stoichiometries was also observed before. Jasanoff et al. showed that under in vitro conditions MHC class II dimers binds to the Ii trimmers independently and that nonameric complexes are formed only in the excess of MHC class II dimmers (Jasanoff et al., 1999).

5.4 Regulation of activity of cysteine cathepsins by p41 inhibitory fragment

Invariant chain degradation is a critical step in MHC class II maturation for antigen presentation. It has been shown that different APCs use different cysteine proteases for this process in the thymus and in the periphery. cathepin L is necessary for Ii degradation in mouse cortical epithelial cells (Nakagawa et al, 1998), whereas cat S is essential for antigen presentation in peripheral APCs (Riese 1996). Cathepsin V as a novel member of the papain family (Bromme et al., 1999). Although this protease shares a protein sequence identity as high as 80 % with human cathepsin L, its tissue distribution is strikingly different from that of cathepsin L. Although human cathepsin L was found in most tissues, the expression of cathepsin V was restricted to thymus and testis. Furthermore, the expression of cathepsin V was restricted to TECs, indicating that in this compartment of the human thymus, cathepsin V, but not cathepsin L, is the predominant cysteine protease. Cortical TECs are mainly responsible for positive selection of T cells. cathepsin V has overtaken the role of cathepsin L in the positive selection of thymocytes.

Degradation of the Ii is believed to be a consequence of proteolytic activity of three cysteine cathepsins L, V and S in a cell-type specific manner (Hsing and Rudensky, 2005). Degradation of the Ii ultimately results in mature MHC class II molecules arriving at the cell surface (Neefjes and Ploegh, EMBO J 1992). But the process is more complex. Ii exists in two alternatively spliced forms, the p31 and p41 form common to mice and humans and their variants, p33 and p43, specific for humans (ref). The two major Ii forms differ by the insertion of a 64 amino acid residues long fragment in the p41 form of Ii (Strubin et al., 1986). This fragment, also termed the p41 inhibitory fragment, exhibits sequence homology to the thyropins, thyroglobulin type-I domain inhibitors of cysteine proteinases (Lenarcic and Bevec, 1998), and acts as a strong and reversible inhibitor of the majority of lysosomal cysteine endopeptidases (Bevec et al., 1996; Mihelic et. al., 2008). The two major Ii processing cathepsins L and V are inhibited by the p41 inhibitory fragment in low picomolar range with K_i=0.005nM and K_i =0.007nM, respectively, whereas the interaction between cathepsin S and the p41 inhibitory fragment is in high micromolar range (K_i =208nM) (Mihelic et al., 2008).

The presence of the inhibitory fragment in the p41 form of Ii suggests that the MHC class II-Ii complexes may self-regulate their own proteolytic processing, but the exact mechanism remains elusive. It was shown before that the extent of intracellular cysteine protease activity is dramatically reduced in the cells expressing the p41 form of Ii and that this p41 form enhances antigen presentation by reducing the proteolytic activity of lysosomal compartments of antigen presenting cells, likely due to the protection over-excessive degradation of antigenic epitopes (Fineschi et.al. 1996). Even more relevant is the notion that the degradation pattern of the p41 and p31 forms of Ii differ and that the p41 form posses the ability to impart the antigen presentation pattern to the p31 form

(Fineschi et. al., 1995).

The presence of the inhibitory fragment in the p41 form of Ii allows testing for its inhibitory activity. Therefore the soluble MHC class II-Ii complexes were tested for the in vitro inhibition of cathepsins L, V, S and B. As expected the measurement of enzymatic activity of cathepsins by flurogenic substrate showed that the presence of the p31 form of Ii in the complexes had no effect on the activity of the tested cathepsins. In contrast to p31 form of the complex, the p41Ii form inhibited cathepsins L and V while the activity of cathepsins B and S remained unaffected.

5.5 Inhibition mechanism of cathepsins by stefins

The stefin proteins, a subgroup (Type 1) of the cystatin superfamily, act as cytoplasmic inhibitors of cysteine proteases such as cathepsins (Zavasnik-Bergant 2007). Given their inhibitory role toward cathepsins, stefin proteins could regulate antigen presentation processes involved in immune response and autoimmune diseases (Turk et al., 2008, Roper et al., 2008). A classical biochemical determination of the inhibition constant using hydrolysis of the fluorogenic substrate was employed. The inhibition constant for cathepsins V inhibition by stefin A is similar to that already published.

The crystal structures of complexes between stefin A and cathepsin B reveal the geometric mode of binding of inhibitors into the active sites of proteases. In the case of endopeptidases the structure of the enzymes remained unaltered compared to the structures of the free enzymes (Somoza et al., 2000; Fujishima et al., 1997). The occluding loop is specific feature of cathepsin B, responsible for its exopeptidase activity. In order to form the complex with cathepsin B, inhibitors must displace the occluding loop residues. The crystal structure has shown that the occluding loop can adopt different positions in the complex between cathespin B and stefin A. It is highly likely that the occluding loop is fully flexible and can move further away in solution, but its movement in the crystals of cathepsin B – stefin A complexes is limited because of the extremely tight crystal packing.

5.6 Further studies

With the purified MHC class II/Ii chain complexes further research can be initiated:

Studies will show which proteases can perform invariant chain degradation. By *in vitro* cleavage it will be possible to detect invariant chain intermediates and show the order of specific cleavages.

We believe that based on the expression of MHC class II/ Ii complex described in the thesis, it will be possible to determine the crystal structure.

The follow-up study will focus on construction of an *in vitro* assay for antigen presentation. The goal is to produce all components of antigen presentation in endosomes and analyze the pattern of their interaction. These components are MHC class II molecules, antigens, various proteases and the HLA-DM molecule, that catalyses exchange of CLIP for antigenic peptide. This study could bring new insight into endosomal processes and clarify the roles of the molecules involved.

6 Conclusions

HEK293T cells expression of constructs containing DRA, DRB and the p31 and p41 forms of Ii results in folded proteins secreted to media

The large scale production in flasks can provide sufficient amounts of material for further structural and biochemical studies

We have shown by in-vivo and in-vitro analysis that the Ii chain lacking the transmembrane and cytoplasmic regions is efficiently secreted as a trimer.

Moreover, we have shown that soluble Ii can assemble and be secreted with soluble DRA and DRB in complexes that may be more heterogeneous than assumed previously.

7 Acknowledgements

Construct design and initial expression and purification of MHC class II / Ii chain complexes were done during my stay at the Department of tumor biology, Amsterdam, Netherlands in the frame of Marie Curie Training network MRTN-CT-512385.

Further expressions, characterization and inhibition experiments were performed at the Department of biochemistry and molecular and structural biology at Jozef Stefan Institute.

A thanks goes to:

My supervisor Prof. Dr. Dušan Turk and
My co-supervisor Prof. Dr. Jaques Neefjes
The Examination Board (Prof. DDr. Boris Turk, Prof. Dr. Ana Plemenitaš, Prof. Dr. Janko Kos)
My coworkers (Dr. Marko Mihelič, Dr. Miha Renko, Andreja Doberšek, and others coworkers from our department)
Aleksandar Pechkov, Roger Pain, Katja Galeša
My coworkers at NKI-AVL (Lennert Jansen, Dr. Nuno Rocha, Nickolas Karantzelis and others)
My friends and all who helped me in any way

Ďakujem mojej mame a rodine za podporu a porozumenie

References

Alfonso, C.; Liljedahl, M.; Winqvist, O.; Surh, C. D.; Peterson, P. A.; Fung-Leung, W.P. et al. The role of H2-O and HLA-DO in major histocompatibility complex class II-restricted antigen processing and presentation. *Immunol. Rev*. **172**, 255–266 (1999).

Alvarez-Fernandez, M.; Barrett, A. J.; Gerhartz, B.; Dando, P. M.; Ni, J.; Abrahamson, M. Inhibition of mammalian legumain by some cystatins is due to a novel second reactive site. *J. Biol. Chem.* **274**, 19195–203(1999).

Amigorena, S.; Webster, P.; Drake, J.; Newcomb, J.; Cresswell, P.; Mellman, I. Invariant chain cleavage and peptide loading in major histocompatibility complex class II vesicles. *J. Exp. Med.* **181**, 1729–1741 (1995).

Anderson, M. S. and Miller, J. Invariant chain can function as a chaperone protein for class II MHC molecules, *Proc. Natl. Acad. Sci. USA* **89**, 2282–2286 (1992).

Andersen M. H.; Schrama D.; Straten P. T. and Jurgen C. Cytotoxin T cell. *J. Invest. Derm.* **126**, 32–41(2006).

Antoniou, A. N.; Blackwood, S. L.; Mazzeo, D. and Watts, C. Control of antigen presentation by a single protease cleavage site. *Immunity* **12**, 391–398 (2000).

Aricescu, A. R.; Lu, W. and Jones E. Y. A time- and cost-efficient system for high-level protein production in mammalian cells. *Acta Cryst. Sec. D*, **62**, 1243–1250 (2006).

Arunachalam, B.; Phan, U. T.; Geuze, H. J.; Cresswell, P. Enzymatic reduction of disulfide bonds in lysosomes: characterization of a gamma-interferon-inducible lysosomal thiol reductase (GILT). *Proc. Natl. Acad. Sci. USA* **9**7, 745–750 (2000).

Beers, C.; Honey, K.; Fink, S.; Forbush K.; Rudensky A. Differential regulation of cathepsin S and cathepsin L in interferone-gamma-treated macrophages. *J. Exp. Med.* **197**, 169–179 (2003).

Bakke, O. and Ploegh, H. L. Uncoating ATPase Hsc70 is recruited by invariant chain and controls the size of endocytic compartments. *Proc. Natl. Acad. Sci. USA* **99**, 1515–1520 (2002).

Bakke, O. and Dobberstein, B. MHC class II-associated invariant chain contains a sorting signal for endosomal compartments, *Cell* **63,** 707−716 (1990).

Baneyx, F. Recombinant protein expression in Escherichia coli. *Curr. Opin. Biotechnol.* **10**, 411–21 (1999).

Baneyx, F.; Mujacic, M. Recombinant protein folding and misfolding in Escherichia coli. *Nat. Biotechnol.* **22**, 1399–408 (2004).

Barrett, A. J.; Rawlings, N. D.; Woessner, J. F. Handbook of proteplyc enzymes (London, UK, academic Press, **1867**, 1998).

Benavides, F. et al. Impaired hair follicle morphogenesis and cycling with abnormal epidermal differentiation in nackt mice, a cathepsin L-deficient mutation. *Am. J. Pathol*. **161**, 693–703 (2002).

Benavides, F. et al. The CD4 T cell-deficient mouse mutation nackt (nkt) involves a deletion in the cathepsin L (CtsI) gene. *Immunogenetics* **53**, 233–242 (2001).

Bennett, K. et al. Antigen processing for presentation by class II major histocompatibility complex requires cleavage by cathepsin E. *Eur. J. Immunol*. **22**, 1519–1524 (1992).

Bevec, T.; Stoka, V.; Pungercic, G.; Dolenc, I. and Turk, V. Major histocompatibility complex class II-associated p41 invariant chain fragment is a strong inhibitor of lysosomal cathepsin L. *J. Exp. Med*. **183**, 1331–1338 (1996).

Bevec, T.; Stoka, V.; Pungercic, G.; Cazzulo, J. J.; Turk, V. A fragment of the major histocompatibility complex class II-associated p41 invariant chain inhibits cruzipain, the major cysteine proteinase from *Trypanosoma cruzi*. *FEBS Lett.* **401**, 259–261 (1997).

Bevec, T.; Stoka, V.; Pungercic, G.; Dolenc, I.; Turk, V. Major histocompatibility complex class II-associated p41 invariant chain fragment is a strong inhibitor of lysosomal cathepsin L. *J. Exp. Med.* **183**, 1331–1338 (1996).

Blum J. S.; Cresswell P. Role for intracellular proteases in the processing and transport of class II HLA antigens. *Proc. Natl. Acad. Sci. USA* **85**, 3975–79 (1988).

Bohley, P.; Seglen, P. O. Proteases and proteolysis in the lysosome. *Experientia* **48**, 151–157 (1992).

Brachet, V.; Raposo, G.; Amigorena, S.; Mellman, I. Ii chain controls the transport of major histocompatibility complex class II molecules to and from lysosomes. *J. Cell. Biol.* **137**, 51–65 (1997).

Braun, B.; LaBaer, J. High throughput protein production for functional proteomics, *Trends Biotechnol*. **21**, 383–388 (2003).

Bremnes, B.; Madsen, T.; Gedde-Dahl, M.; Bakke, O. An LI and ML motif in the cytoplasmic tail of the MHC-associated invariant chain mediate rapid internalization. *J. Cell. Sci.* **107**, 2021–2032 (1994).

Bromme, D.; Li, Z.; Barnes, M.; Mehler, E. Human cathepsin V functional expression, tissue distribution, electrostatic surface potential, enzymatic characterization, and chromosomal localization. *Biochemistry* **38**, 2377–2385 (1999).

Cheng, T.; Hitomi, K.; van Vlijmen-Willems, I. M.; de Jongh, G. J.; Yamamoto, K.; Nishi, K.; Watts, C.; Reinheckel, T.; Schalkwijk, J.; Zeeuwen, P. L. Cystatin M/E is a high affinity inhibitor of cathepsin V and cathepsin L by a reactive site that is distinct from the legumain-binding site. A novel clue for the role of cystatin M/E in epidermal cornification. *J. Biol. Chem.* **281**, 15893–9 (2006).

Chicz, R. M., Urban, R. G., Gorga J. C., Vignali, D. A. A, Lane, W. S. and Strominger, J. C. Specificity and promiscuity among naturally precessed peptides bound to HLA-DR alleles. *J. Exp. Med.* **178**, 27–47 (1993).

Chicz, R. M., Urban, R. G., Lane, W. S., Gorga, J. C., Stern, L., Vignali, D. A. A. and Strominger, J. L. Predominant naturally processed peptides hound to HLA-DRI are denved from MHC-related molecules and heterogeneoub in size. *Nature* **358**, 764 (1992).

Cho, M. S., Yee, H., Brown, C., Mei, B., Mirenda, C., Chan, S. Versatile expression system for rapid and stable production of recombinant proteins. *Biotechnol. Prog*. **19**, 229–232 (2003).

Chuang, H. H.; Yu, M.; Jan, Y. N.; Jan, L. Y. Evidence that the nucleotide exchange and hydrolysis cycle of G proteins causes acute desensitization of G-protein gated inward rectifier K+ channels. *Proceedings of the National Academy of Sciences* **29**, 11727–11732 (1998).

Cockett, M. I.; Bebbington, C. R.; Yarranton, G. T. The use of engineered E1A genes to

transactivate the hCMV-MIE promoter in permanent CHO cell lines. *Biotechnology* **8**, 662–667 (1990).

Collins, D. S.; Unanue, E. R.; Harding, C. V. Reduction of disulfide bonds within lysosomes is a key step in antigen processing. *J. Immunol.* **147**, 4054–4059 (1991).

Consogno, G; Manici, S.; Facchinetti, V; Bachi, A; Hammer, J. et al. Identification of immunodominant regions among promiscuous HLA-DR-restricted CD4+ T-cell epitopes on the tumor antigen MAGE-3. *Blood* **101**, 1038–1044 (2003).

Costantino, C. M.; Hang, H. C.; Kent, S. C.; Hafler, D. A.; Ploegh, H. L. Lysosomal cysteine

and aspartic proteases are heterogeneously expressed and act redundantly

to initiate human invariant chain degradation. *J. Immunol.* **180**, 2876

85 (2008).

Cresswell, P. Assembly, Transport, and Function of MHC Class II Molecules. Annual *Review of Immunology* **12**, 259–291 (1994).

Cresswell, P.; Blum, J. S.; Kelner, D. N.; Marks, M. S. Biosynthesis and processing of class II histocompatibility antigens, *Crit. Rev. Immun.* **7**, 31–53 (1987).

Delamarre, L.; Pack, M.; Chang, H.; Mellman, I.; Trombetta, E. S. Differential lysosomal proteolysis in antigen-presenting cells determines antigen fate. *Science* **307**, 1630–1634 (2005).

Deussing, J. et al. Cathepsins B and D are dispensable for major histocompatibility complex class II-mediated antigen presentation. *Proc. Natl. Acad. Sci. USA* **95**, 4516–4521 (1998).

Dixon, A. M.; Stanley, B. J.; Matthews, E. E.; Dawson, J. P.; Engelman, D. M. Invariant chain transmembrane domain trimerisation: a step in Mhc class II assembly. *Biochemistry* **45**, 5228–5234(2006).

Drakesmith, H.; O'Neil, D.; Schneider, S. C.; Binks, M.; Medd, P.; Sercarz, E. et al. In vivo priming of T cells against cryptic determinants by dendritic cells exposed to interleukin 6 and native antigen. *Proc. Natl. Acad. Sci. USA* **95**, 14903–14908 (1998).

Driessen, C. et al. Cathepsin S controls the trafficking and maturation of MHC class II molecules in dendritic cells. *J. Cell Biol*. **147**, 775–790 (1999).

Driessen, C.; Lennon-Dumenil, A. M.; Ploegh, H. L. Individual cathepsins degrade immune complexes internalized by antigen-presenting cells via Fcγ receptors. *Eur. J. Immunol*. **31**, 1592–1601 (2001).

Driessen, C.; Lennon-Dumenil, A. M.; Ploegh, H. L. Individual cathepsins degrade immune complexes internalized by antigenpresenting cells via Fcg-receptors. *J. Eur. Immunol*. **31**, 1592–1601 (2001).

Dubin, G. Proteinaceous cysteine protease inhibitors. *Cell Mol. Life Sci.* **62**, 653–69 (2005).

Durocher, Y.; Perret, S.; Kamen, A. High-level and high-throughput recombinant protein production by transient transfection of suspension-growing human 293-EBNA1 cells. *Nucleic Acids Res*. **30** (2002).

Elias K.; Siegel R.; O'Shea J. J. Primer on the Rheumatic Diseases, 94–107, DOI: 10.1007/978-0-387-68566-3_4 (2008).

Engelhard, V. H. Structure of peptides associated with class I and class II MHC molecules. *Annu. Rev. Immunol*. **12**, 181–207 (1994).

Faure-Andre, G. et al. Regulation of dendritic cell migration by CD74, the MHC class II-

associated invariant chain. *Science* **322**, 1705–1710 (2008).

Fernandes, D. M.; Vidard, L.; Rock, K. L. Characterization of MHC class II-presented peptides generated from an antigen targeted to different endocytic compartments. *Eur. J. Immunol*. **30**, 2333–2343 (2000).

Fiebiger E, et al. Invariant chain controls the activity of extracellular cathepsin L. *J Exp Med* **196**, 1263–1269 (2002).

Fiebiger, E.; Meraner, P.; Weber, E.; Fang, I. F.; Stingl, G.; Ploegh, H. et al. Cytokines regulate proteolysis in major histocompatibility complex class ii-dependent antigen presentation by dendritic cells. *J. Exp. Med*. **193**, 881–892 (2001).

Fineschi, B.; Arneson, L. S.; Naujokas, M. F.; Miller, J. Proteolysis of major histocompatibility complex class II-associated invariant chain is regulated by the alternatively spliced gene product p41. *Proc. Natl. Acad. Sci. USA* **92**, 10257–10261 (1995).

Fineschi, B.; Sakaguchi, K.; Appella, E.; Miller, J. The proteolytic environment involved in MHC class II-restricted antigen presentation can be modulated by the p41 form of invariant chain. *J. Immunol*. **157**, 3211–3215 (1996).

Finesshi B.; Sakaguchi, K.; Appella, E.; Miler, J. The proteolytic environment involved in MHC class II-restricted antigen presentation can be modulated by the p41 form of invariant chain. *J. Immunol.* **157**, 3211–3215 (1996).

Finley, E. M.; Kornfeld, S. Subcellular localization and targeting of cathepsin E. *J. Biol. Chem*. **269**, 31259–31266 (1994).

Forgac, M. Structure, function and regulation of the vacuolar (H+)-ATPases. *FEBS Lett.* **440**, 258–263 (1998).

Fowlkes, J. L.; Thrailkill, K. M.; Serra, D. M.; Nagase, H. Insulin-like growth factor binding protein (IGFBP) substrate zymography. A new tool to identify and characterize IGFBP-degrading proteinases. *Endocrine* **7**, 33–6 (1997).

Frayser, M.; Sato, A. K.; Xu, L.; Stern, L. J. Empty and peptide-loaded class II major histocompatibility complex proteins produced by expression in Escherichia coli and folding in vitro, *Protein Expr. Purif.* **15**, 105–114 (1999).

Gabrijelcic, D.; Svetic, B.; Spaic, D.; Skrk, J.; Budihna, M.; Dolenc, I.; Popovic, T.; Cotic, V.; Turk, V. Cathepsins B, H and L in human breast carcinoma. *Eur. J. Clin. Chem. Clin. Biochem.* **30**, 69–74 (1992).

Garin, J.; Diez, R.; Kieffer, S.; Dermine, J. F.; Duclos, S.; Gagnon, E. et al. The phagosome proteome: insight into phagosome functions. *J. Cell Biol*. **152**, 165–180 (2001).

Gavin, A. C.; Superti-Furga, G. Protein complexes and proteome organization from yeast to man. *Curr. Opin. Chem. Biol*. **7**, 21–27 (2003).

Gedde-Dahl, M.; Freisewinkel, I.; Staschewski, M.; Schenck, K.; Koch N.; Bakke, O. Exon 6 is essential for invariant chain trimerization and induction of large endosomal structure, *J. Biol. Chem*. **272,** 8281–8287 (1997).

Geisse, S.; Henke, M. Large-scale transient transfection of mammalian cells: a newly emerging attractive option for recombinant protein production. *J. Struct. Funct. Genomics* **6**, 165–170 (2005).

Germain, R. N.; Castellino, F.; Han, R.; Reis e Sousa, C.; Romangoli, P.; Sadegh-Nasseri, S.; Zhong, G. M. Processing and presentation of endocytically acquired protein antigens by MHC class II and class I molecules. *Immunol. Rev*. **151**, 5–30 (1996).

Germain, R. N. MHC-dependent antigen processing and peptide presentation: providing ligands for T lymphocyte activation. *Cell* **76**, 287–299 (1994).

Gingras, R.; Richard, C.; El-Alfy, M.; Morales, C. R.; Potier, M.; Pshezhetsky, A. V. Purification, cDNA cloning, and expression of a new human blood plasma glutamate carboxypeptidase homologous to *N*-acetyl-aspartyl-alpha-glutamate carboxypeptidase/ prostate-specific membrane antigen. *J. Biol. Chem.* **274**, 11742–11750 (1999).

Godkin, A. J.; Smith, K. J.; Willis, A.; Tejada-Simon, M. V.; Zhang, J. et al. Naturally processed HLA class II peptides reveal highly conserved immunogenic flanking region sequence preferences that reflect antigen processing rather than peptide-MHC interactions. *J. Immunol.* **166**, 6720–6727 (2001).

Gocheva, V.; Joyce, J. A. Cysteine cathepsins and the cutting edge of cancer invasion. *Cell Cycle* **6**, 60–64 (2007).

Gowen, M. et al. Cathepsin K knockout mice develop osteopetrosis due to a deficit in matrix degradation but not demineralization. *J. Bone Miner. Res*. **14**, 1654–1663 (1999).

Graham, F. L.; Smiley, J.; Russell, W. C.; Nairn R. Characteristics of a human cell line transformed by DNA from human adenovirus type 5. *J. Gen. Virol*. **36**, 59–74 (1977).

Gresser, O.; Weber, E.; Hellwig, A.; Riese, S.; Regnier-Vigouroux, A. Immunocompetent astrocytes and microglia display major differences in the processing of the invariant chain and in the expression of active cathepsin L and cathepsin S. *Eur. J. Immunol.* **31**, 1813–1824 (2001).

Griffin, J. P.; Chu, R.; Harding, C. V. Early endosomes and a late endocytic compartment generate different peptide-class II MHC complexes via distinct processing mechanisms. *J. Immunol*. **158**, 1523–1532 (1997).

Guicciardi, M. E. et al. Cathepsin B contributes to TNF-α-mediated hepatocyte apoptosis by promoting mitochondrial release of cytochrome c. *J. Clin. Invest*. **106**, 1127–1137 (2000).

Guncar, G.; Pungercic, G.; Klemencic, I.; Turk V.; Turk D. Crystal structure of MHC class II-associated p41 Ii fragment bound to cathepsin L reveals the structural basis for differentiation between cathepsins L and S. *Embo. J*. **18**, 793–803 (1999).

Halangk, W. et al. Role of cathepsin B in intracellular trypsinogen activation and the onset of acute pancreatitis. *J. Clin. Invest*. **106**, 773–781 (2000).

Harding, C. V.; Boom, W. H. Regulation of antigen presentation by Mycobacterium tuberculosis: a role for Toll-like receptors. *Nat. Rev. Micro.* **8**, 296–307 (2010).

Haque, M. A.; Hawes, J. W.; Blum, J. S. Cysteinylation of MHC class II ligands: peptide endocytosis and reduction within APC influences T cell recognition. *J. Immunol.* **166**, 4543–4551 (2001).

Hochuli E.; Dobeli H.; Schacer A. New metal chelate absorbent selective for proteins and peptide containing neighbouring histidine residues. *J. Chromatogr.* **411,** 177–184 (1987).

Honey K.; Rudensky A. Y. Lysosomal cysteine proteases regulate antigen presentation. *Nat. Immunol. Rev.* **3**, 472–482 (2003).

Honey, K. et al. Thymocyte expression of cathepsin L is essential for NKT cell development. *Nature Immunol*. **3**, 1069–1074 (2002).

Honey, K.; Duff, M.; Beers, C.; Brissette, W. H.; Elliott, E. A.; Peters, C. et al. Cathepsin S regulates the expression of cathepsin L and the turnover of gamma-interferon-inducible lysosomal thiol reductase in B lymphocytes. *J. Biol. Chem*. **276**, 22573–

22578 (2001).

Honey, K.; Nakagawa, T.; Peters, C.; Rudensky, A. Cathepsin L regulates CD4+ T cell selection independently of its effect on invariant chain: a role in the generation of positively selecting peptide ligands. *J. Exp. Med*. **195**, 1349–1358 (2002).

Hsieh, C. S.; deRoos, P.; Honey, K.; Beers, C.; Rudensky, A. Y. A role for cathepsin L and cathepsin S in peptide generation for MHC class II presentation. *J. Immunol*. **168**, 2618–2625 (2002).

Hsing, L. C.; Rudensky, A. Y. The lysosomal cysteine proteases in MHC class II antigen presentation. *Immunol. Rev*. **207**, 229–241 (2005).

Jarvinen, M.; Rinne, A.; Hopsu-Havu, V. K. Human cystatins in normal and diseased tissues-a review. *Acta Hisochem.* **82**, 5–18 (1987).

Jensen, P. E. Regulation of antigen presentation by acidic pH. *J. Exp. Med.* **171**, 1779–1784 (1990).

Jensen, P. E. Acidification and disulfide reduction can be sufficient to allow intact proteins to bind class II MHC. *J. Immunol.* **150**, 3347–3356 (1993).

Jones, E. Y.; Fugger, L.; Strominger, J. L.; Siebold C. MHC class II proteins and disease: a structural perspective. *Nat. Rev. Immunol.* **6**, 271–282 (2006).

Kalandadze, A.; Galleno, M.; Foncerrada, L.; Strominger, J. L.; Wucherpfennig, K. W. Expression of recombinant HLA-DR2 molecules. Replacement of the hydrophobic transmembrane region by a leucine zipper dimerization motif allows the assembly and secretion of soluble DR alpha beta heterodimers. *J. Biol. Chem.* **271**, 20156–20162 (1996).

Kloetzel, P. M. The proteasome and MHC class I antigen processing. *Biochim. Biophys. Acta* **1695**, 225–233 (2004).

Koch, N.; McLellan, A. D. et al. A revised model for invariant chain-mediated assembly of MHC class II peptide receptors. *Trends in biochemical sciences* **32**, 532–537 (2007).

Koch, N.; Zacharias, M. et al. Stoichiometry of HLA class II-invariant chain oligomers. *PloS one* **6**, 17257 (2011).

Koike, M. et al. Cathepsin D deficiency induces lysosomal storage with ceroid lipofuscin in mouse CNS neurons. *J. Neurosci*. **20**, 6898–6906 (2000).

Kordis, D.; Turk, V. Phylogenomic analysis of the cystatin superfamily in eukaryotes and prokaryotes. *BMC Evol. Biol.* **9**, 266 (2009).

Kos, J.; Krasovec, M.; Cimerman, N.; Nielsen, H. J.; Christensen, I. J.; Brunner, N. Cysteine proteinase inhibitor stefin A, stefin B and cystatin C in sera from patients with colorectal cancer: relation to prognosis. *Clin. cancer Res.* **6**, 505–11 (2000).

Kropshofer, H.; Hammerling, G. J.; Vogt, A. B. The impact of the non-classical MHC proteins HLA-DM and HLA-DO on loading of MHC class II molecules. *Immunol. Rev.* **172**, 267–278 (1999).

Kuhelj, R.; Dolinar, M.; Pungercar, J.; Turk, V. The preparation of catalytically active human cathepsin B from its precursor expressed in Escherichia coli in the form of inclusion bodies. *Eur. J. Biochem.* **229**, 533–9 (1995).

Kvist, S.; Wiman, K.; Claesson, L.; Peterson, P. A.; Dobberstein, B. Membrane insertion and oligomeric assembly of HLA-DR histocompatibility antigens. *Cell* **29**, 61–69 (1982).

Lamb, C. A.; Cresswell, P. Assembly and transport properties of invariant chain trimers

and HLA-DR-invariant chain complexes. *J. Immuno*. **148**, 3478–82 (1992).

Lankar, D.; Vincent-Schneider, H.; Briken, V.; Yokozeki, T.; Raposo, G.; Bonnerot, C. Dynamics of major histocompatibility complex class II compartments during B cell receptor-mediated cell activation. *J. Exp. Med.* **195**, 461–472 (2002).

Lautwein A. et al. Inflammatory stimuli recruit cathepsin activity to late endosomal compartments in human dendritic cells. *Eur. J. Immunol.* **32**, 3348–3357 (2002).

Leaney, J. L.; Milligan, G.; Tinker, A. The G protein alpha subunit has a key role in determining the specificity of coupling to, but not the activation of, G protein-gated inwardly rectifying K+ channels. *Journal of Biological Chemistry* **275**, 921–929 (2000).

Lei, Q.; Jones, M. B.; Talley, E. M.; Schrier, A. D.; McIntire, W. E.; Garrison, J. C. et al. Activation and inhibition of G proteincoupled inwardly rectifying potassium (Kir3) channels by G protein beta gamma subunits. *Proceedings of the National Academy of Sciences* **97**, 9771–9776 (2000).

Lenarcic, B.; Bevec, T. Thyropins – new structurally related proteinase inhibitors. *Biol. Chem*. **379**, 105–111 (1998).

Lenarcic, B.; Turk, V. Thyroglobulin type-1 domains in equistatin inhibit both papain-like cysteine proteinases and cathepsin D. *J. Biol. Chem.* **274**, 563–6 (1999).

Leng, L. et al. MIF signal transduction initiated by binding to CD74. *J. Exp. Med.* **197**, 1467–1476 (2003).

Lennon-Dumenil, A. M.; Bryant, R. A. R.; Valentijn, K.; Driessen, C.; Overkleeft, H. S.; Erickson, A. et al. The p41 isoform of Invariant Chain is a chaperone for Cathepsin L. *Embo. J*. (2001).

Lennon-Dumenil, A. M.; Bakker, A. H.; Wolf-Bryant, P.; Ploegh, H. L. Lagaudriere-Gesbert C. A closer look at proteolysis and MHC-class-II-restricted antigen presentation. *Curr. Opin. Immunol.* **14**, 15–21 (2002).

Lennon-Dumenil, A. M. et al. Analysis of protease activity in live antigen-presenting cells shows regulation of the phagosomal proteolytic contents during dendritic cell activation. *J. Exp. Med.* **196**, 529–540 (2002).

Levy, E.; Lopez-Otin, C.; Ghiso, J.; Geltner, D.; Frangione, B. Stroke in Icelandic patients with hereditary amyloid angiopathy is related to a mutation in the cystatin C gene, an inhibitor of cysteine proteases. *J. Exp. Med.* **169**, 1771–8 (1989).

Li, C.; Schwabe, J. W.; Banayo, E.; Evans, R. M. Coexpression of nuclear receptor partners increases their solubility and biological activities. *Proc. Natl. Acad. Sci. U. S. A* **94**, 2278– 2283 (1997).

Li, W.; Ding, F.; Zhang, L.; Liu, Z.; Wu, Y.; Luo, A.; Wu, M.; Wang, M.; Zhan, Q.; Liu, Z. Overexpression of stefin A in human esophageal squamous cell carcinoma cells inhibits tumor cell growth, angiogenesis, invasion, and metastasis. *Clin. Cancer Res.* **11**, 8753–62 (2005).

Lindner, R.; Unanue, E. R. Distinct antigen MHC class II complexes generated by separate processing pathways. *Embo. J.* **15**, 6910–6920 (1996).

Lipp, J.; Dobberstein, B. Signal recognition particle-dependent membrane insertion of Mouse invariant chain: a membrane-spanning protein with a cytoplasmically exposed amino terminus. *J. Cell Biol.* **102**, 2169–75 (1986).

Louis, N.; Evelegh, C.; Graham, F. L. Cloning and sequencing of the cellular-viral junctions from the human adenovirus type 5 transformed 293 cell line. *Virology* **233**, 423–9 (1997).

Lutz, M. B.; Rovere, P.; Kleijmeer, M. J.; Rescigno, M.; Assmann, C. U.; Oorschot, V. M. et al. Intracellular routes and selective retention of antigens in mildly acidic cathepsin D/lysosome associated membrane protein-1/MHC class II- positive vesicles in immature dendritic cells. *J. Immunol.* **159**, 3707–3716 (1997).

MacLennan, I. C. et al. Human major histocompatibility complex class II invariant chain is expressed on the cell surface. *J. Biol. Chem.* **265**, 5787–5792 (1990).

Maehr, R. et al. Asparagine endopeptidase is not essential for class II MHC antigen presentation but is required for processing of cathepsin L in mice. *J. Immunol.* **174**, 7066–7074 (2005).

Manoury, B. et al. Asparagine endopeptidase can initiate the removal of the MHC class II invariant chain chaperone. *Immunity* **18**, 489–498 (2003).

Manoury, B. et al. Destructive processing by asparagine endopeptidase limits presentation of a dominant T cell epitope in MBP. *Nature Immunol*. **3**, 169–174 (2002).

Maric, M. A.; Taylor, M. D.; Blum, J. S. Endosomal aspartic proteinases are required for invariant- chain processing. *Proc. Natl. Acad. Sci. USA* **91**, 2171–2175 (1994).

Marks, M. S.; Blum, J. S.; Cresswell, P. Invariant chain trimers are sequestered in the rough endoplasmic reticulum in the absence of association with HLA class II antigens. *J. Cell Biol.* **111**, 8391990–55 (1990).

Marsh, S. G.; Albert, E. D.; Bodmer, W. F.; Bontrop, R. E.; Dupont, B.; Erlich, H. A.; Fernández-Viña, M.; Geraghty, D. E.; Holdsworth, R.; Hurley, C. K.; Lau, M.; Lee, K. W.; Mach, B.; Maiers, M.; Mayr, W. R.; Müller, C. R.; Parham, P.; Petersdorf, E. W.; Sasazuki, T.; Strominger, J. L.; Svejgaard, A.; Terasaki, P. I.; Tiercy, J. M.; Trowsdale, J. Nomenclature for factors of the HLA system. *Tissue Antigens* **75**, 291–455 (2010).

Martin, J. R.; Craven, C. J.; Jerala, R.; Kroon-Zitko, L.; Zerovnik, E.; Turk, V.; Waltho, J. P. The three-dimensional solution structure of human stefin A. *J. Mol. Biol.* **246**, 331–43 (1995).

Meissner, P.; Pick, H.; Kulangara, A.; Chatellard, P.; Friedrich, K.; Wurm, F. M. Transient Gene Expression in Suspension Hek293 Cells: Application to Large-Scale Protein Production. *Biotechnol. Bioeng.* **75**, 197–203 (2001).

Mihelic, M.; Dobersek, A.; Guncar, G.; Turk, D. Inhibitory fragment from the p41 form of invariant chain can regulate activity of cysteine cathepsins in antigen presentation. *J. Biol. Chem.* **283**, 14453–14460 (2008).

Mihelic, M.; Turk, D. Two decades of thyroglobulintype-1 domain research. *Biol. Chem.* **388**, 1123–30 (2007).

Mohamed, M. M.; Sloane, B. F. Cysteine cathepsins: multifunctional enzymes in cancer. *Nat. Rev. Cancer* **6**, 764–775 (2006).

Monaco, J. J. Pathways for the processing and presentation of antigens to T cells. *J. Leukocyte Biol.* **57**, 543–547, pmid:7722412 (1995).

Mouritsen, S.; Hansen, A. S.; Petersen, B. L.; Buus, S. pH dependence of the interaction between immunogenic peptides and MHC class II molecules. Evidence for an acidic intracellular compartment being the organelle of interaction. *J. Immunol*. **148**, 1438–1444 (1992).

Musil, D.; Zucic, D.; Turk, D.; Engh, R. A.; Mayr, I.; Huber, R.; Popovic, T.; Turk, V.; Towatari, T.; Katunuma, N. et al. The refined 2.15 A X-ray crystal structure of human liver cathepsin B: the structural basis for its specificity. *Embo. J*. **10**, 2321–2330 (1991).

Nakagawa, T. Y. et al. Impaired invariant chain degradation and antigen presentation and diminished collagen-induced arthritis in cathepsin S null mice. *Immunity* **10**, 207–217 (1999).

Nakagawa, T.; Roth, W.; Wong, P.; Nelson, A.; Farr, A.; Deussing, J. et al. Cathepsin L: critical role in Ii degradation and CD4 T cell selection in the thymus. *Science* **280**, 450–453 (1998).

Nakagawa, T. Y.; Brissette, W. H.; Lira, P. D.; Griffiths, R. J.; Petrushova, N.; Stock, J. et al. Impaired invariant chain degradation and antigen presentation and diminished collagen-induced arthritis in cathepsin S null mice. *Immunity* **10**, 207–217 (1999).

Neefjes, J. J.; Ploegh, H. L. Inhibition of endosomal proteolytic activity by leupeptin blocks surface expression of MHC class II molecules and their conversion to SDS resistance alpha beta heterodimers in endosomes. *Embo. J.* **11**, 411–416 (1992).

Neumann, J.; Schach, N.; Koch, N. Glycosylation signals that separate the trimerization from the mhc class II-binding domain control intracellular degradation of invariant chain. *J. Biol. Chem.* **276**, 13469–13475 (2001).

Neumann, J.; Koch, N. Assembly of major histocompatibility complex class II subunits with invariant chain. *FEBS letters* **579**, 6055–6059 (2005).

Nguyen, Q. V.; Knapp, W.; Humphreys, R. E. Inhibition by leupeptin and antipain of the intracelullar proteolysis of Ii. *Human Immunol*. **24**, 153–63 (1989).

Nordeng, T. W. et al. The cytoplasmic tail of invariant chain regulates endosome fusion and morphology. *Mol. Biol. Cell* **13**, 1846–1856 (2002).

O´Sulivan, D. M.; Noonan, D.; Quaranta, V. Four Ia invariant chain forms derive from a single gene by alternate slicing and alternate initiation of transcription/translation. *J. Exp. Med*. **166**, 444–460 (1987).

Pham, P. L.; Perret, S.; Doan, H. C.; Cass, B.; St-Laurent, G.; Kamen, A.; Durocher, Y. Large-scale transient transfection of serumfree suspension-growing HEK293 EBNA1 cells: peptone additives improve cell growth and transfection efficiency. *Biotechnol. Bioeng*. **84**, 332–342 (2003).

Park, S. J.; Sadegh-Nasseri, S.; Wiley, D. C. Invariant chain made in Escherichia coli has an exposed N-terminal segment that blocks antigen binding to HLA-DR1 and trimeric C-terminal segment that binds empty HLA-DR1. *Proc. Natl. Acad. Sci. USA* **92**, 11289–11293 (1995).

PET System Manual. http//www.novagen.com (2003).

Peterson K.; Hakansson, M.; Nilsson, H.; Forsberg, G.; Svensson, L. A.; Liljas, A.; Walse, B. Crystal structure of a superantigen bound to MHC class Ii displays zinc and peptide dependence. *Embo. J*., **20**, 3306–3312 (2001).

Phan, U. T.; Arunachalam, B.; Cresswell, P. Gamma-interferon-inducible lysosomal thiol reductase (GILT). Maturation, activity, and mechanism of action. *J. Biol. Chem*. **275**, 25907–25914 (2000).

Pierre, P., Mellman, I. Developmental regulation of invariant chain proteolysis controls MHC class II trafficking in mouse dendritic cells. *Cell* **93**, 1135–1145 (1998).

Pieters, J.; Horstmann, H.; Bakke, O.; Griffiths, G.; Lipp, J. Intracellular transport and localization of major histocompatibility complex class II molecules and associated invariant chain. *J. Cell Biol.* **115**, 1213–1223 (1991).

Pisoni, R. L.; Acker, T. L.; Lisowski, K. M.; Lemons, R. M.; Thoene, J. G. A cysteine-specific lysosomal transport system provides a major route for the delivery of thiol to human fibroblast lysosomes: possible role in supporting lysosomal proteolysis. *J. Cell.*

Biol. **110**, 327–335 (1990).

Porath, J.; Carlsson, J.; Olsson, I.; Belfrage, G. Metal chelate afinity chromatography, a new approach to protein fractionation. *Nature* **258**, 598–599 (1975).

Pozgan, U.; Caglic, D.; Rozman, B.; Nagase, H.; Turk, V.; Turk, B. Expression and activity profiling of selected cysteine cathepsins and matrix metalloproteinases in synovial fluids from patients with rheumatoid arthritis and osteoarthritis. *Biol. Chem.* **391**, 571–579 (2010).

Rammensee, H.; Friede, T.; Stevanović, S. MHC ligands and peptide motifs: first listing. *Immunogenetics* **41**, 178–228 (1995).

Rawlings, N. D.; Tolle, D. P.; Barrett, A. J.: Evolutionary families of peptidase inhibitors. *Biochem. J.* **378**, 705–16 (2004).

Reinheckel, T.; Deussing, J.; Roth, W.; Peters, C. Towards specific functions of lysosomal cysteine peptidases: phenotypes of mice deficient for cathepsin B or cathepsin L. *Biol. Chem*. **382**, 735–741 (2001).

Riberdy, J. M.; Newcomb J. R.; Surman, M. J.; Barbosa, J. A.; Creswell, P. HLA-DR molecules from an antigen-processing mutant cell line are associated with invariant chain peptides. *Nature* **360**, 474–476 (1992).

Riese, R. J. et al. Cathepsin S activity regulates antigen presentation and immunity. *J. Clin. Invest.* **101**, 2351–2363 (1998).

Riese, R. J. et al. Regulation of CD1 function and NK1.1+ T cell selection and maturation by cathepsin S. *Immunity* **15**, 909–919 (2001).

Riese, R. J.; Wolf, P. R.; Bromme, D.; Natkin, L. R.; Villadangos, J. A.; Ploegh, H. L. et al. Essential role for cathepsin S in MHC class II-associated invariant chain processing and peptide loading. *Immunity* **4**, 357–366 (1996).

Roche, P. A.; Marks, M. S.; Cresswell, P. Formation of a nine-subunit complex by HLA class II glycoproteins and the invariant chain. *Nature* **354**, 392–394 (1991).

Rock, K. L.; York, I. A.; Saric, T.; Goldberg, A. L. Protein degradation and the generation of MHC class I-presented peptides. *Adv. Immunol.* **80**, 1–70 (2002).

Rodriguez, G. M.; Diment, S. Role of cathepsin D in antigen presentation of ovalbumin. *J. Immunol.* **149**, 2894–2898 (1992).

Roper, R. J.; McAllister, R. D.; Biggins, J. E.; Michael, S. D.; Min, S. H. et al. Aod1 controlling day 3 thymectomy-induced autoimmune ovarian dysgenesis in mice encompasses two linked quantitative trait loci with opposing allelic effects on disease susceptibility. *J. Immunol.* **170**, 5886–5891 (2003).

Rovere, P. et al. Dendritic cell maturation and antigen presentation in the absence of invariant chain. *Proc. Natl. Acad. Sci. USA* **95**, 1067–1072 (1998).

Rudensky, A. Y.; Preston-Hurlburt, P.; Hong, S. C.; Barlow, A.; Janeway, C. A. Sequence analysis of peptides bound to MHC class II molecules. *Nature* **353**, 622–627 (1991).

Saftig, P. et al. Impaired osteoclastic bone resorption leads to osteopetrosis in cathepsin-K-deficient mice. *Proc. Natl Acad. Sci. USA* **95**, 13453–13458 (1998).

Saftig, P. et al. Mice deficient for the lysosomal proteinase cathepsin D exhibit progressive atrophy of the intestinal mucosa and profound destruction of lymphoid cells. *Embo. J*. **14**, 3599–3608 (1995).

Santamaria, I. et al. Cathepsin L2, a novel human cysteine proteinase produced by breast and colorectal carcinomas. *Cancer Res.* **58**, 1624–1630 (1998).

Saudrais, C.; Spehner, D.; de la Salle, H.; Bohbot, A.; Cazenave, J. P.; Goud, B. et al. Intracellular pathway for the generation of functional MHC class II peptide complexes in immature human dendritic cells. *J. Immunol.* **160**, 2597–2607 (1998).

Schick, C.; Pemberton, P. A.; Shi, G. P.; Kamachi, Y.; Cataltepe, S.; Bartuski, A. J. et al. Cross-class inhibition of the cysteine proteinases cathepsins K, L, and S by the serpin squamous cell carcinoma antigen 1: a kinetic analysis. *Biochemistry* **37**, 5258–5266. (1998).

Sette, A.; Southwood, S.; Miller J.; Appella, E. Binding of major histocompatibility complex class II to the invariant chain-derived peptide, CLIP, is regulated by allelic polymorphism in class II. *J. Exp. Med*. **181**, 677–683 (1995).

Shi, G. P. et al. Cathepsin S required for normal MHC class II peptide loading and germinal center development. *Immunity* **10**, 197–206 (1999).

Shi, G. P.; Bryant, R. A.; Riese, R.; Verhelst, S.; Driessen, C.; Li, Z. et al. Role for cathepsin F in invariant chain processing and major histocompatibility complex class II peptide loading by macrophages. *J. Exp. Med.* **191**, 1177–1186 (2000).

Shi, G. P.; Villadangos, J. A.; Dranoff, G.; Small, C.; Gu, L.; Haley, K. J. et al. Cathepsin S required for normal MHC class II peptide loading and germinal center development. *Immunity* **10**, 197–206 (1999).

Shlipak, M. G.; Katz, R.; Sarnak, M. J.; Fried, L. F.; Newman, A. B.; Stehman-Breen, C.; Seliger, S. L.; Kestenbaum, B.; Psaty, B.; Tracy, R. P.; Siscovick, D. S. Cystatin C and prognosis for cardiovascular and kidney outcomes in elderly persons without chronic kidney disease. *Ann. Intern. Med.* **145**, 237–46 (2006).

Starlets, D. et al. Cell-surface CD74 initiates a signaling cascade leading to cell proliferation and survival. *Blood* **107**, 4807–4816 (2006).

Stern, L. J.; Brown, J. H.; Jardetzky, T. S.; Gorga, J. C.; Urban, R. G. et al. Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. *Nature* **368**, 215–221 (1994).

Stoka, V.; Turk, B.; Turk, V. Lysosomal cysteine proteases: structural features and their role in apoptosis. *IUBMB Life* **57**, 347–353 (2005).

Strauss, M.; Stollwerk, J.; Lenarcic, B.; Turk, V.; Jany, K. D.; Gassen, H. G. Chemical synthesis of a gene for human stefin A and its expression in E. coli. *Biol. Chem. Hoppe Seyler* **369**, 1019–30 (1988).

Strubin, M.; Berte, C.; Mach B. Alternative splicing and alternative initiation of translation explain the four forms of the Ia antigen-associated invariant chain. *Embo. J.* **5**, 3483–3488 (1986).

Strumptner, P.; Benaroch, P. Interaction of MHC class II molecules with the invariant chain (81-90) region. *Embo. J.* **16**, 5807–5818 (1997).

Stypmann, J. et al. Dilated cardiomyopathy in mice deficient for the lysosomal cysteine peptidase cathepsin L. *Proc. Natl. Acad. Sci. USA* **99**, 6234–6239 (2002).

Surman, S.; Lockey, T. D.; Slobod, K. S.; Jones, B.; Riberdy, J. M.; White, S. W. et al. Localization of CD4+ T cell epitope hotspots to exposed strands of HIV envelope glycoprotein suggests structural influences on antigen processing. *Proc. Natl. Acad. Sci. USA* **98**, 4587–4592 (2001).

Tao, K.; Stearns, N. A.; Dong, J.; Wu, Q. L.; Sahagian, G. G. The proregion of cathepsin L is required for proper folding, stability, and ER exit. *Arch. Biochem. Biophys.* **311**, 19–27 (1994).

Terpe, K. Overview of tag protein fusions: from molecular and biochemical fundamentals

to commercial systems. *Appl. Microbiol. Biotechnol*. **60**, 523–533 (2003).

Teyton, L.; O'Sullivan, D.; Dickson, P. W.; Lotteau, V.; Sette, A.; Fink, P.; Peterson, P. A. Invariant chain distinguishes between the exogenous and endogenous antigen presentation pathways. *Nature* **348**, 39–44 (1990).

Tolosa, E., et al. Cathepsin V is involved in the degradation of invariant chain in human thymus and is overexpressed in myasthenia gravis. *J Clin Invest* **112**, 517–526 (2003).

Trombetta, E. S.; Ebersold, M.; Garrett, W.; Pypaert, M.; Mellman, I. Activation of lysosomal function during dendritic cell maturation. *Science* **299**, 1400–1403 (2003).

Turk, V.; Bode, W. The cystatins: protein inhibitors of cysteine proteases. *FEBS Lett.* **285**, 213–9 (1991).

Turk, V.; Kos, J.; Turk, B. Cysteine cathepsins (proteases) – on the main stage of cancer? *Cancer Cell* **5**, 409–410 (2004).

Turk, V.; Stoka, V.; Turk, D. Cystatins: biochemical and structural properties, and medical relevance. *Front. Biosci.* **13**, 5406–5420 (2008).

Turk, B.; Turk, D.; Salvesen, G. S. Regulating cysteine protease activity: essential role of protease inhibitorsas guardians and regulators. *Curr. Pharm. Des.* **8**, 1623–37 (2002).

Turk, V.; Turk, B.; Guncar, G.; Turk, D.; Kos, J. Lysosomal cathepsins: structure, role in antigen processing and presentation, and cancer. *Adv. Enzyme Regul.* **42**, 285–303 (2002).

Unanue, E. R. A. Antigen-Presenting Function of the Macrophage, *Rev. Immun.* **2**, 395–428 (1984).

Vasiljeva, O.; Reinheckel, T.; Peters, C.; Turk, D.; Turk, V.; Turk, B. Emerging roles of cysteine cathepsins in disease and their potential as drug targets. *Curr. Pharm. Des.* **13**, 387–403 (2007).

Van Noort, J. M.; Jacobs, M. J. Cathepsin D, but not cathepsin B, releases T cell stimulatory fragments from lysozyme that are functional in the context of multiple murine class II MHC molecules. *Eur. J. Immunol*. **24**, 2175–2180 (1994).

Villadangos, J. A.; Riese, R. J.; Peters, C.; Chapman, H. A.; Ploegh, H. L. Degradation of mouse invariant chain: roles of cathepsins S and D and the influence of major histocompatibility complex polymorphism. *J. Exp. Med*. **186**, 549–560 (1997).

Villadangos, J. A.; Bryant, R. A. R.; Deussing, J.; Driessen, C.; Lennon-Dumenil, A. M.; Riese, R. J. et al. Proteases involved in MHC class II antigen presentation. *Immunol. Rev.* **171**, 109–120 (1999).

Villadangos, J. A. Presentation of antigens by MHC class II molecules: getting the most of them. *Mol. Immunol.* **38**, 329–345 (2001).

Wang, B. et al. Human cathepsin F. Molecular cloning,functional expression, tissue localization, and enzymatic characterization. *J. Biol. Chem*. **273**, 32000–32008 (1998).

Wang, H.; Chong, S. Visualization of coupled protein folding and binding in bacteria and purification of the heterodimeric complex. *Proc. Natl. Acad. Sci. U. S. A*. **100**, 478–483 (2003).

Watts, C. Capture and processing of exogenous antigens for presentation on MHC molecules. *Annu. Rev. Immunol*. **15**, 821–850 (1997).

Wolf, P. R.; Ploegh, H. L. How MHC class II molecules acquire peptide cargo: biosynthesis and trafficking through the endocytic pathway. *Annu. Rev. Cell Dev. Biol.* **11**, 267–306 (1995).

Wraight, C. J.; van Endert, P.; Moller, P.; Lipp, J.; Ling, N. R.; MacLennan, I. C. et al. Human major histocompatibility complex class II invariant chain is expressed on the

cell surface. *J. Biol. Chem.* **265**, 5787–5792 (1990).

Yan, S.; Sloane, B. F. Molecular regulation of human cathepsin B: implication in pathologies. *Biol. Chem.* **384**, 845–854 (2003).

Zavasnik - Bergant, T. Cystatin protease inhibitors and immune functions. *Front. Biosci.* **13**, 4625–4637 (2008).

Zavasnik - Bergant, V.; Schweiger, A.; Bevec, T.; Golouh, R.; Turk, V.; Kos, J. Inhibitory p41 isoform of invariant chain and its potential target enzymes cathepsins L and H in distinct populations of macrophages in human lymph nodes. *Immunology* **112**, 378–385 (2004).

Zhu, Y.; Rudensky, A. Y.; Corper, A. L.; Teyton, L.; Wilson, I. A. Crystal structure of MHC class II IAb in complex with a human CLIP peptide: prediction of an I-Ab peptide-binding motif. *J, Mol, Biol.* **236**, 1157–74 (2003).

Index of Figures

Index of Tables

Appendix

Published articles:

Renko, M.; Požgan, U.; Majera, D.; Turk, D. Stefin A displaces the occluding loop of cathepsin B only by as much as required to bind to the active site cleft. *FEBS J.* **277**, 4338–45 (2010).

Majera, D.; Crnigoj Kristan, K.; Neefjes, J.; Turk, D.; Mihelic, M. Expression, purification and assembly of soluble multimeric MHC class II-invariant chain complexes. DOI 10.1016/j.febslet.2012.03.038. (2012).

the FEBS Journal

Stefin A displaces the occluding loop of cathepsin B only by as much as required to bind to the active site cleft

Miha Renko, Urška Požgan, Dušana Majera and Dušan Turk

Department of Biochemistry and Molecular and Structural Biology, Jozef Stefan Institute, Ljubljana, Slovenia

Keywords
cathepsin B; complex; conformational flexibility; crystal structure; occluding loop; stefin A

Correspondence
D. Turk, Department of Biochemistry and Molecular and Structural Biology, Jozef Stefan Institute, Jamova 39, SI-1000 Ljubljana, Slovenia
Fax: +386 1 477 3984
Tel: +386 1 477 3215
E-mail: dusan.turk@ijs.si

Database
The coordinates and structure factors are available in the Protein Data Bank database under accession number 3K9M

(Received 14 June 2010, revised 11 August 2010, accepted 16 August 2010)

doi:10.1111/j.1742-4658.2010.07824.x

Cathepsin B (EC 3.4.22.1) is one of the most versatile human cysteine cathepsins. It is important for intracellular protein degradation under normal conditions and is involved in a number of pathological processes. The occluding loop makes cathepsin B unique among cysteine cathepsins. This $\sim$ 20 residue long insertion imbedded into the papain-like protease scaffold restricts access to the active site cleft and endows cathepsin B with its carboxydipeptidase activity. Nevertheless, the enzyme also exhibits endopeptidase activity and is inhibited by stefins and cystatins. To clarify the structural properties of the occluding loop upon the binding of stefins, we determined the crystal structure of the complex between wild-type human stefin A and wild-type human cathepsin B at 2.6 Å resolution. The papain-like part of cathepsin B structure remains unmodified, whereas the occluding loop residues are displaced. The part enclosed by the disulfide bridge containing histidines 110 and 111 (i.e. the 'lasso' part) is rotated by $\sim$ 45° away from its original position. A comparison of the structure of the unliganded cathepsin B with the structure of the proenzyme, its complexes with chagasin and stefin A shows that the magnitude of the shift of the occluding loop is related to the size of the binding region. It is smallest in the procathepsin structures and increases in the series of complexes with stefin A and chagasin, although it has no impact on the binding constant. Hence, cathepsin B can dock inhibitors and certain substrates regardless of the size of the binding region.

Structured digital abstract
- MINT-7990451: *Stefin-A* (uniprotkb:P01040) and *Cathepsin B* (uniprotkb:P07858) *bind* (MI:0407) by *x-ray crystallography* (MI:0114)

Introduction

Cathepsin B (EC 3.4.22.1), a lysosomal, papain-like cysteine protease, is one of the most extensive studied human cathepsins [1]. This enzyme is abundantly expressed in a variety of tissues where it takes part in protein degradation and processing. It is involved in a number of physiological and pathological processes, such as intracellular protein degradation, the immune response, prohormone processing, cancer and arthritis [2–9]. Its proteolytic activity is regulated by stefins and cystatins, which are endogenous inhibitors of cysteine cathepsins [10]. Cathepsin B differs from other cathepsins by its dual role, exhibiting exo- as well as endopeptidase activity. The crystal structure of this human enzyme [11] has revealed that an $\sim$ 20 residues long insertion, termed the 'occluding loop', occupies the part of the active site cleft on the primed side and blocks access to the active site cleft beyond the S2′ substrate binding site [11,12]. The occluding loop is

Abbreviation
PDB, Protein Data Bank.

held together by the disulfide bond between C108 and C119. Its attachment to the body of the enzyme is stabilized by two salt bridges, between H110 and D22, and between R116 and D224. The crystal structure suggested that two histidines, H110 and H111, positioned within the active site cleft, are responsible for the docking of the C-terminal carboxylic group of peptidyl substrates. This observation was confirmed by the crystal structure of the complex of a substrate-mimicking inhibitor, CA030, interacting through its C-terminal carboxylic group with the two histidine residues [13]. The concept of utilizing additional structural features to block part of the active site cleft aiming to restrict the binding of peptidyl substrates and facilitating binding of the substrate termini is not unique to cathepsin B [14]. Dipeptidyl peptidase I (DDPI), also known as cathepsin C, contains a large segment of the proregion [15,16], termed the exclusion domain [17], which is associated with the mature enzyme and blocks the active site cleft beyond the S2 site, as shown in crystal structures of DPPI alone and in complex with the inhibitor Gly-Phe-CHN2 [18]. The amino peptidase cathepsin H has a covalently attached stretch of eight residues originating from the propeptide, termed the mini chain, which blocks the unprimed binding site [19]. The mini loop in carboxypeptidase cathepsin X blocks the primed side of the active site, restricting access to only one residue [20].

Although the structures of the mature native form of cathepsin B clearly exposed the relevance of the occluding loop for the exopeptidase activity [11], they do not explain the mechanisms of endopeptidase activity, nor the inhibition of the enzyme by their endogenous protein inhibitors cystatins and stefins [21]. A further step in understanding of these mechanisms was provided by the crystal structures of human [22] and rat procathepsins B [23]. They revealed that, in the zymogen form, the propeptide rather than the occluding loop fills the active site cleft. It was shown that the single and double mutations D22A, H110A, R116A and D224A disrupted the salt bridges between the occluding loop and the body of the enzyme, resulting in enhanced endopeptidase activity [24]. Furthermore, the deletion mutant lacking 12 central residues of the 'lasso' region between the disulfide C109–C118 confirmed that their absence yields an enzyme with pure endopeptidase activity, completely lacking exopeptidase activity, and with a 40-fold increase of affinity for cystatins [12]. These results indicated that loop flexibility must be responsible for the endopeptidase activity of cathepsin B, as well as that endopeptidase activity should be associated with the occluding loop displacement from the active site cleft. Recently, the crystal structure of the complex between chagasin, a cysteine protease inhibitor from *Trypanosoma cruzi*, and human cathepsin B, a multiple mutant with destabilized affinity of the occluding loop residues towards the active site cleft, has shown that, on binding to cathepsin B, chagasin displaces the occluding loop from the active cleft [25]. In the present study, we report the crystal structure of the complex between two human proteins: wild-type stefin A and wild-type human cathepsin B. A structural comparison suggests that the structure of the occluding loop residues adapts to each binding ligand in its own way and swings out only as much as is mandatory.

Results and Discussion

Crystals of the complex of stefin A and cathepsin B contain complete wild-type protein sequences. The positioning of the main chains of nearly all residues is clearly revealed by the electron density maps, with the exception of E95, a stretch of four occluding loop residues from V112 to S115 in the first molecule of cathepsin B; G75 and Q76 in the molecule A of stefin A; and M1 and E78 in the molecule B of stefin A. Additionally, eleven side chains lack adequate electron density. The r.m.s.d. between all pairs of superimposed CA atoms of cathepsin B molecules, excluding residues 105–125 of the occluding loop, is 0.34 Å, whereas the r.m.s.d. between all pairs of superimposed CA atoms of stefin A molecules exhibits a somewhat larger value of 0.88 Å. The r.m.s.d. between the equivalent CA atoms from the occluding loop region (I105–D124) and the second binding loop of stefin A (F70–V81) are 1.4 and 1.2 Å, respectively. This comparison shows that the differences between the two molecules of cathepsin B are confined to the occluding loop region, whereas the differences between the two stefin A molecules are spread out through the entire structure, with slightly increased variability in the S72–D79 region that forms the second binding loop.

Cathepsin B structure exhibits a two-domain, papain-like fold [11]. The N-terminal domain includes the central helix that contains, on its N-terminus, the active site C29. The C-terminal domain is based on a four-stranded β-barrel fold, contributing H199, the other active site residue. The active site cleft is formed at the interface between the two domains, which are also named L- and R- (left and right), in accordance with the standard view used to present the papain-like folds.

The structure of stefin A exhibits the cystatin-like fold composed of a five-stranded β-sheet embracing an α-helix (Fig. 1). This arrangement creates a wedge-shaped

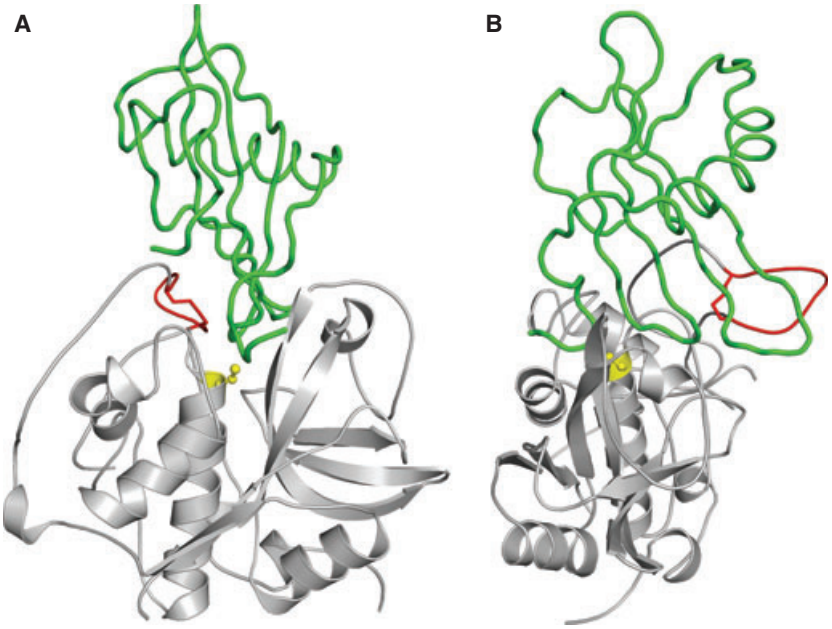


Fig. 1. Structure of the cathepsin B–stefin A complex. (A) View along the active site cleft. (B) View perpendicular to the active site cleft. Cathepsin B is shown in gray and stefin A in green. The catalytic cysteine is shown in yellow. The wedge-shaped structure of stefin A fills the active site cleft along the whole length and displaces the occluding loop (the 'lasso' is shown in red).

structure with the N-terminal trunk and two hairpin loops at its narrow edge [26]. This narrow edge docks into the active site cleft of cathepsin B (Fig. 1). The binding mode is equivalent to those from the related complexes of stefin B-papain [27] and stefin A-cathepsin H [28]. A superimposition of complexes of cathepsin B and H with stefins showed that stefin A binds to cathepsin B as deeply as stefin B does to cathepsin H. To illustrate this, we calculated the average distances between CA atoms of the active site cysteine and histidine residues in cathepsins B and H and the center of CA atoms of stefins in the structures of both complexes. The average distance is 23.4 Å, which is the same for both enzymes (Table 1). The comparison shows that the final positions of stefin A molecules in the complexes are not affected by the additional features of the exopeptidases, occluding loop and mini chain, which occupy parts of the active site cleft (Fig. 2). These additional features hinder binding along the whole interdomain interface, although they both are displaced upon binding of the ligand.

The N-terminal trunk and the first binding loop occlude the active site C29, blocking enzymatic activity. The N-terminal trunk binds into the nonprimed

Table 1. Average distances between CA atoms of the stefins and catalytic residues of cysteine proteases.

Distance calculated	*d* (Å)
Papain–stefin B	23.93
Cathepsin H–stefin A	23.36 ± 0.23
Cathepsin B–stefin A	23.34 ± 0.15

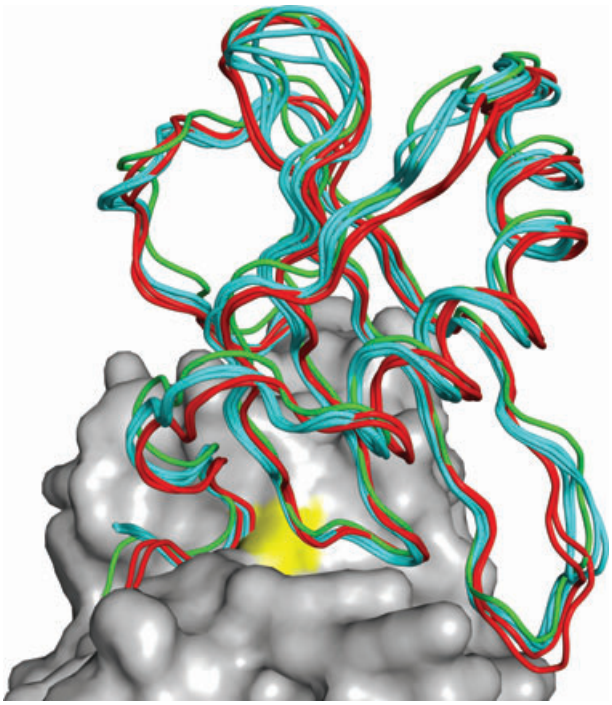


Fig. 2. Flexibility of stefin structures. Papain surface (PDB code: 1STF) [27] is shown in gray with the part of the reactive cysteine residue shown in yellow. Four structures of stefin A from the complex with cathepsin H are shown in cyan (PDB code: 1NB3) [28]. The two structures of stefin A from the complex with cathepsin B are shown in red. The stefin B structure from the complex with papain is shown in green. Six stefin A molecules were moved onto the scaffold of papain using transformation parameters obtained from the superimpositions of their enzymatic partners on the papain structure.

substrate binding sites, whereas the two loops bind into the primed sites. They occlude the catalytic C29 (Fig. 2, surface colored in yellow) in the middle and thereby prevent the approach of substrate molecules. The same approach is utilized by the p41 fragment, a representative of thyropins [29], chagasin [30,31] and mycocypins [32].

The N-terminal trunk comes down the S1 binding area of cathepsin B, occupies the S2 binding site with proline residue P3, and continues through the S2 binding site upwards (away from the cathepsin B surface). Two hydrogen bonds between the stefin A amide hydrogen (G4) and carbonyl (P3) with cathepsin B carbonyl atom (G198) and amide hydrogen (G74) attach the first loop to the active site cleft.

The first binding loop of stefin A (V47–Q51) fills the S1′ site with V48. In addition to this hydrophobic interaction, the loop is fastened to the cathepsin B surface by the hydrogen bond between the stefin A A49 amide and cathepsin B G24 carbonyl. The binding of this loop is further stabilized by a hydrogen bond between the stefin A N52 side chain amide and the cathepsin B S25 carbonyl group.

The second binding loop (L73–D79) comes down to the area beyond the S2′ site and displaces the occluding loop residues of cathepsin B. It is firmly anchored by the β-sheet hydrogen-bonding pattern formed between the three loops in stefin A and an additional hydrogen bond formed between the amide hydrogen of L73 and the side chain carbonyl of E109. A layer of solvent molecules mediates the contacts between the C-terminal part of the second binding loop and cathepsin B.

The occluding loop differs from the native structure [Protein Data Bank (PDB) code: 1HUC] [11] in the region from S104 and D124 (Figs 3 and 4). The lasso structure between the C108–C119 disulfide is rotated by $\sim 45°$ and pushed aside. This movement dramatically changes the position of the two occluding loop histidines, H110 and H111. Instead of a parallel positioning within the active site cleft, these two side chains now point in different, almost opposing directions. The side chain of H110 points away from the active site cleft to the back of the molecule, whereas the side chain of H111 points upwards and away from the surface. In the complex, two stefin A residues, A49 from the tip of the first binding loop and L73 from the second binding loop, fill the places that the two histidines occupy in the native structure. Besides the lasso, the inhibitor also pushes away the chain from C119 to D124. The position of CA atom of E122 is changed by almost 7 Å from the position that it occupies in the native cathepsin B structure. In this respect, stefin interactions with exopeptidases are not unique. The N-terminal trunk of stefin A can displace the

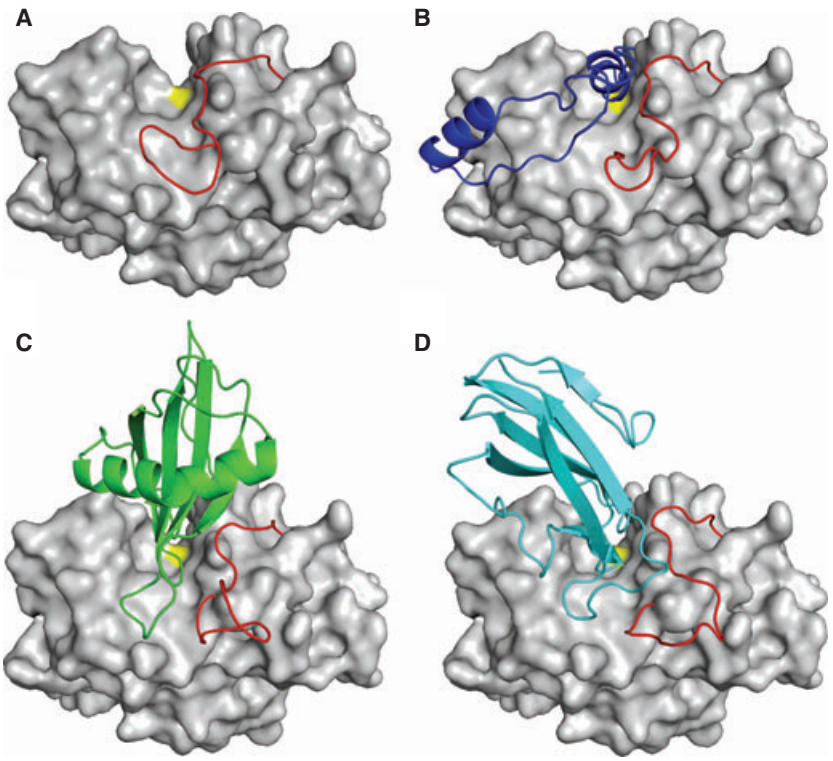


Fig. 3. The extent of the occluding loop displacement in the unliganded and liganded structures. The occluding loop (red) is shown in on the surface of the papain-like part of the structure (gray). (A) Unliganded cathepsin B (PDB code: 1HUC) [11]. (B) Propeptide in dark blue (PDB code: 3PBH) [22]. (C) Complex with stefin A, with stefin A in green. (D) A complex with chagasin (shown in cyan) (PDB code: 3CBJ) [25].

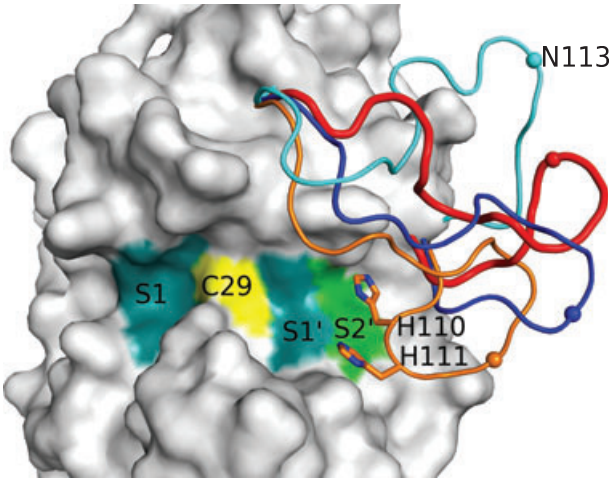


Fig. 4. The extent of the occluding loop displacement (superimposed). The papain-like part of cathepsin B is shown as a gray surface with the catalytic cysteine part shown in yellow, whereas the S1, S1′ and S2′ binding sites are shown in green and cyan. The occluding loops from various cathepsin B structures (proenzyme, complex with stefin A, complex with chagasin) are shown in dark blue, red and cyan, respectively. The occluding loop residues, H110 and H111, from the naked cathepsin B, are shown in orange. Spheres represent the position of CA atom of N113, to indicate the extent of movement of the occluding loop.

mini chain which blocks part of the binding cleft in cathepsin H [28].

Two salt bridges, H110–D22 and R116–D224, which additionally stabilize the attachment of the loop to the body of the enzyme, are disrupted in the complex. R116 and D224, however, compensate for the loss of the salt bridge interaction by finding electrostatically favorable partners in K184 of cathepsin B and E78 of stefin A, respectively. The structure presented here shows that a weakening of the embedded occluding loop in the active site cleft is not mandatory for the formation of the crystals of the complex, even though it is associated in a drop of K_i from 0.93 to 0.35 nM, as shown by the chagasin–cathepsin B study. The stefin A–cathepsin B complex contains the wild-type sequences and physiologically occurring interactions, as opposed to the crystal structure of chagasin, a parasite inhibitor from *T. cruzi*, and cathepsin B complex [25] (PDB code: 3CBJ). In that complex, the first salt bridge interaction has been disrupted by the H110A mutant and the reactive site of the enzyme is turned off by the C29A mutant. (it is assumed that the cathepsin B mutations do not affect the geometry of binding of chagasin). The wild-type sequences have also been preserved in the related structural studies of procathepsin B [22].

These three structures, as well as the structure of the native cathepsin B (Figs 3 and 4), demonstrate that the occluding loop can adopt a variety of positions, with the moving part consisting of residues between E109 and D124. The extent of the occluding loop shifts from their position in the native enzyme (PDB code: 1HUC), as demonstrated by the displacement of the CA atom of N113, are 7 Å in the proenzyme structure (PDB code: 3PBH); 14.7 and 15.3 Å in both molecules of the complex with stefin A reported in the present study; 14.5 Å in the monoclinic crystal form of the complex with chagasin (PDB code: 3CBJ); and 22.5 Å in the tetragonal crystal form of the complex with chagasin (PDB code: 3CBK) (Figs 3 and 4). The molecular weight of the stefin A and chagasin are similar (11 kDa versus 12 kDa); however, the structure of L6 loop in chagasin is different from the structure of the second binding loop in stefin A. Stefin A forms a V-shaped structure that fills the active site cleft, whereas the S97–S100 region in L6 loop of chagasin (shown in orange in Fig. 3D) expands the interactions region and, additionally, pushes the occluding loop away. Compared with the second binding loop of stefins, the larger and broader L6 loop of chagasin requires an additional shift of residues R116 and P117. The CA atoms of R116 residues from the two cathepsin B structures are almost 10 Å apart. It is concluded is that the occluding loop is rather flexible and will adapt to structural features of the inhibitors as well as to the packing constraints of the environment. The larger and wider the features of the ligands that compete with the occluding loop for binding to the active site, the farther away the occluding loop residues are shifted. As seen in the tetragonal form of the cathepsin B chagasin complex (3CBK), the depth of the binding of inhibitor as well as the shift of the occluding loop can be additionally extended by the crystal packing constraints. Hence, these structures demonstrate that the occluding loop residues can adopt a variety of conformations, whereas the rest of the structure of cathepsin B appears to be rigid.

A comparison of the interaction constants of the binding of chagasin (K_i = 0.93 nM [25]) and stefins (1.7 and 2 nM [33,34], 0.91 nM [35]) to cathepsin B indicates that the extent of the shift does not affect the inhibition constants. This observation suggests that the energy cost of ligand binding associated with occluding loop removal is not related to the magnitude of the occluding loop shift from the active site cleft. Cathepsin B can bind certain ligands along the whole interdomain interface. During docking, size alone most likely plays no role. Cathepsin B will accept inhibitors or substrates, whatever is available.

Materials and methods

Cathepsin B and stefin A were expressed as described previously [36,37], mixed in a molar ratio 1 : 1.1, and concentrated to 30 mg·mL^{-1} in 10 mM sodium acetate (pH 5.5). Crystals were grown in 0.2 M sodium sulfate, 24% PEG3000. The initial crystals grown by the sitting drop method were highly mosaic, and thereby of no use for structural determination. Accordingly, the hanging drop method was used in combination with the controlled evaporation approach [38], which greatly improved crystal quality. The crystals, which grew in the form of thin plates, were soaked in mother liquor supplemented with 20–30% glycerol and frozen in liquid nitrogen before data collection.

Diffraction data were collected at the XRD1 workstation at Synchrotron Elletra (Trieste, Italy) and processed using HKL2000 software [39]. Determination of the space group was nontrivial. The data were first processed in the $P2_1$ space group as a result of the higher symmetry, with an acceptable R_{merge} of 0.132 and data completeness of 96.7%. The structure was determined by molecular replacement using AMORE [40] with cathepsin B [13] and stefin A [28] as search models. The crystals are extremely dense, having only 28% of solvent, resulting in Matthews coefficient (V_M) of 1.70 [41]. It was unexpected that such tightly packed crystals only diffracted to 2.6 Å. The protein database analysis took into account 10 471 crystal forms of proteins, deposited in the PDB in 2002 [42]. It showed that more tightly packed crystals (i.e. lower V_M) tend to diffract to higher resolutions.

Initially, we processed the data and attempted to refine the structure in the $P2_1$ space group. The refinement presented difficulties and the crystal packing in the occluding loop region suggested that it might be advisable to determine the structure in a lower symmetry space group, namely diffraction data in the lower symmetry space group, P1. These data had a lower R_{merge} of 0.084 and slightly lower completeness (92.4%). The lower completeness of the P1 data set is a consequence of highly anisotropic diffraction, which forced us to discard part of the collected data to maintain reasonable merging statistics. The anisotropy was a consequence of the shape of the crystals, which were thin plates diffracting poorly in one orientation. The P1 space group data resulted in an improved electron density map for the occluding loop residues and were used for further refinement and model building. Superimposition of the two cathepsin B molecules reveals an almost perfect two-fold rotational symmetry (r.m.s.d of 0.36 Å for CA atoms with the occluding loop residues excluded; rotational polar angle 179.9°) and a screw component of 15.62 Å essentially equal to half of the b cell axis (31.08 Å). However, the two inhibitor structures are further apart. The two-fold rotational symmetry is almost preserved (r.m.s.d. of 0.58 Å for CA atoms with the third loop residues from 71 to 80 excluded; rotational polar angle 179.6°), whereas the screw component of 15.38 Å indicates a deviation from the ideal screw shift. When the cathepsin B molecules superimposition parameters were applied on stefin A molecules, their superimposition shows deviation in the position of the two molecules from those observed in the crystal structure. The largest separations between equivalent atoms are visible at the parts furthest apart from active site cleft (e.g. slightly over 0.8 Å for CA atoms of the residue D88). Hence, the lower space group symmetry is not only justified by the improved resolution of the occluding loop residues, but also by the difference in the position of the two stefin A molecules. The structure was refined using REFMAC [43] and MAIN [44].

Table 2. Data collection and refinement statistics for the complex of cathepsin B with stefin A. Numbers in parentheses are for the highest resolution shell.

Data collection	
PDB ID	3K9M
Space group	P1
Cell dimensions	
a, b, c (Å)	62.0, 31.0, 70.9
α, β, γ (°)	90.0, 104.5, 90.0
Resolution (Å)	68.6–2.5
R_{merge} (%)	8.4 (20.6)
$I/\sigma I$	9.5 (2.6)
Completeness (%)	92.1 (66.7)
Redundancy	2.6 (2.2)
Refinement	
Resolution	40.5 – 2.61
Number of reflections (work/free)	24360/713
R_{work}/R_{free}	19.8/25.0
B factor (Å 2)	42.0
Number of atoms	
Protein	5454
Water	127
r.m.s.d.	
Bond lenghts (Å)	0.013
Bond angles (°)	1.71

Data collection and refinement statitistics are summarized in Table 2. The coordinates and structure factors were deposited in the PDB (ID 3K9M). Distance *d* (Table 1) between stefin A and the different enzymes is the average distance between the centre of mass of CA atoms of the stefin molecule and the centre of mass of the CA atoms of the reactive site cysteine and histidine residues.

Acknowledgements

This work was supported by Slovenian Research Agency Grant Nos. P1-0048 and P1-0140; a Marie Curie Fellowship of the European Community programme Drugs for Therapy (MRTN-CT-2004-512385)

to D.M.; and a Young Researcher fellowship to M.R. and U.P.

References

1 Vasiljeva O, Reinheckel T, Peters C, Turk D, Turk V & Turk B (2007) Emerging roles of cysteine cathepsins in disease and their potential as drug targets. *Curr Pharm Des* **13**, 387–403.

2 Yan S & Sloane BF (2003) Molecular regulation of human cathepsin B: implication in pathologies. *Biol Chem* **384**, 845–854.

3 Turk V, Turk B, Guncar G, Turk D & Kos J (2002) Lysosomal cathepsins: structure, role in antigen processing and presentation, and cancer. *Adv Enzyme Regul* **42**, 285–303.

4 Mohamed MM & Sloane BF (2006) Cysteine cathepsins: multifunctional enzymes in cancer. *Nat Rev Cancer* **6**, 764–775.

5 Pozgan U, Caglic D, Rozman B, Nagase H, Turk V & Turk B (2010) Expression and activity profiling of selected cysteine cathepsins and matrix metalloproteinases in synovial fluids from patients with rheumatoid arthritis and osteoarthritis. *Biol Chem* **391**, 571–579.

6 Gabrijelcic D, Svetic B, Spaic D, Skrk J, Budihna M, Dolenc I, Popovic T, Cotic V & Turk V (1992) Cathepsins B, H and L in human breast carcinoma. *Eur J Clin Chem Clin Biochem* **30**, 69–74.

7 Turk V, Kos J & Turk B (2004) Cysteine cathepsins (proteases) – on the main stage of cancer? *Cancer Cell* **5**, 409–410.

8 Stoka V, Turk B & Turk V (2005) Lysosomal cysteine proteases: structural features and their role in apoptosis. *IUBMB Life* **57**, 347–353.

9 Gocheva V & Joyce JA (2007) Cysteine cathepsins and the cutting edge of cancer invasion. *Cell Cycle* **6**, 60–64.

10 Turk B, Turk D & Salvesen GS (2002) Regulating cysteine protease activity: essential role of protease inhibitors as guardians and regulators. *Curr Pharm Des* **8**, 1623–1637.

11 Musil D, Zucic D, Turk D, Engh RA, Mayr I, Huber R, Popovic T, Turk V, Towatari T, Katunuma N *et al.* (1991) The refined 2.15 A X-ray crystal structure of human liver cathepsin B: the structural basis for its specificity. *EMBO J* **10**, 2321–2330.

12 Illy C, Quraishi O, Wang J, Purisima E, Vernet T & Mort JS (1997) Role of the occluding loop in cathepsin B activity. *J Biol Chem* **272**, 1197–1202.

13 Turk D, Podobnik M, Popovic T, Katunuma N, Bode W, Huber R & Turk V (1995) Crystal structure of cathepsin B inhibited with CA030 at 2.0-A resolution: A basis for the design of specific epoxysuccinyl inhibitors. *Biochemistry* **34**, 4791–4797.

14 Turk D & Guncar G (2003) Lysosomal cysteine proteases (cathepsins): promising drug targets. *Acta Crystallogr D Biol Crystallogr* **59**, 203–213.

15 Paris A, Strukelj B, Pungercar J, Renko M, Dolenc I & Turk V (1995) Molecular cloning and sequence analysis of human preprocathepsin C. *FEBS Lett* **369**, 326–330.

16 Dolenc I, Turk B, Pungercic G, Ritonja A & Turk V (1995) Oligomeric structure and substrate induced inhibition of human cathepsin C. *J Biol Chem* **270**, 21626–21631.

17 Turk D, Janjic V, Stern I, Podobnik M, Lamba D, Dahl SW, Lauritzen C, Pedersen J, Turk V & Turk B (2001) Structure of human dipeptidyl peptidase I (cathepsin C): exclusion domain added to an endopeptidase framework creates the machine for activation of granular serine proteases. *EMBO J* **20**, 6570–6582.

18 Molgaard A, Arnau J, Lauritzen C, Larsen S, Petersen G & Pedersen J (2007) The crystal structure of human dipeptidyl peptidase I (cathepsin C) in complex with the inhibitor Gly-Phe-CHN2. *Biochem J* **401**, 645–650.

19 Guncar G, Podobnik M, Pungercar J, Strukelj B, Turk V & Turk D (1998) Crystal structure of porcine cathepsin H determined at 2.1 A resolution: location of the mini-chain C-terminal carboxyl group defines cathepsin H aminopeptidase function. *Structure* **6**, 51–61.

20 Guncar G, Klemencic I, Turk B, Turk V, Karaoglanovic-Carmona A, Juliano L & Turk D (2000) Crystal structure of cathepsin X: a flip-flop of the ring of His23 allows carboxy-monopeptidase and carboxy-dipeptidase activity of the protease. *Structure* **8**, 305–313.

21 Turk V, Stoka V & Turk D (2008) Cystatins: biochemical and structural properties, and medical relevance. *Front Biosci* **13**, 5406–5420.

22 Podobnik M, Kuhelj R, Turk V & Turk D (1997) Crystal structure of the wild-type human procathepsin B at 2.5 A resolution reveals the native active site of a papain-like cysteine protease zymogen. *J Mol Biol* **271**, 774–788.

23 Cygler M, Sivaraman J, Grochulski P, Coulombe R, Storer AC & Mort JS (1996) Structure of rat procathepsin B: model for inhibition of cysteine protease activity by the proregion. *Structure* **4**, 405–416.

24 Nagler DK, Storer AC, Portaro FC, Carmona E, Juliano L & Menard R (1997) Major increase in endopeptidase activity of human cathepsin B upon removal of occluding loop contacts. *Biochemistry* **36**, 12608–12615.

25 Redzynia I, Ljunggren A, Abrahamson M, Mort JS, Krupa JC, Jaskolski M & Bujacz G (2008) Displacement of the occluding loop by the parasite protein, chagasin, results in efficient inhibition of human cathepsin B. *J Biol Chem* **283**, 22815–22825.

26 Bode W, Engh R, Musil D, Thiele U, Huber R, Karshikov A, Brzin J, Kos J & Turk V (1988) The 2.0 A X-ray crystal structure of chicken egg white cystatin and its possible mode of interaction with cysteine proteinases. *EMBO J* **7**, 2593–2599.

27 Stubbs MT, Laber B, Bode W, Huber R, Jerala R, Lenarcic B & Turk V (1990) The refined 2.4 A X-ray

crystal structure of recombinant human stefin B in complex with the cysteine proteinase papain: a novel type of proteinase inhibitor interaction. *EMBO J* **9**, 1939–1947.
28 Jenko S, Dolenc I, Guncar G, Dobersek A, Podobnik M & Turk D (2003) Crystal structure of Stefin A in complex with cathepsin H: N-terminal residues of inhibitors can adapt to the active sites of endo- and exopeptidases. *J Mol Biol* **326**, 875–885.
29 Guncar G, Pungercic G, Klemencic I, Turk V & Turk D (1999) Crystal structure of MHC class II-associated p41 Ii fragment bound to cathepsin L reveals the structural basis for differentiation between cathepsins L and S. *EMBO J* **18**, 793–803.
30 Redzynia I, Ljunggren A, Bujacz A, Abrahamson M, Jaskolski M & Bujacz G (2009) Crystal structure of the parasite inhibitor chagasin in complex with papain allows identification of structural requirements for broad reactivity and specificity determinants for target proteases. *FEBS J* **276**, 793–806.
31 Ljunggren A, Redzynia I, Alvarez-Fernandez M, Abrahamson M, Mort JS, Krupa JC, Jaskolski M & Bujacz G (2007) Crystal structure of the parasite protease inhibitor chagasin in complex with a host target cysteine protease. *J Mol Biol* **371**, 137–153.
32 Renko M, Sabotic J, Mihelic M, Brzin J, Kos J & Turk D (2010) Versatile loops in mycocypins inhibit three protease families. *J Biol Chem* **285**, 308–316.
33 Lenarcic B, Krizaj I, Zunec P & Turk V (1996) Differences in specificity for the interactions of stefins A, B and D with cysteine proteinases. *FEBS Lett* **395**, 113–118.
34 Turk B, Ritonja A, Bjork I, Stoka V, Dolenc I & Turk V (1995) Identification of bovine stefin A, a novel protein inhibitor of cysteine proteinases. *FEBS Lett* **360**, 101–105.
35 Estrada S, Pavlova A & Bjork I (1999) The contribution of N-terminal region residues of cystatin A (stefin A) to the affinity and kinetics of inhibition of papain, cathepsin B, and cathepsin L. *Biochemistry* **38**, 7339–7345.
36 Kuhelj R, Dolinar M, Pungercar J & Turk V (1995) The preparation of catalytically active human cathepsin B from its precursor expressed in Escherichia coli in the form of inclusion bodies. *Eur J Biochem* **229**, 533–539.
37 Jerala R, Kroon-Zitko L & Turk V (1994) Improved expression and evaluation of polyethyleneimine precipitation in isolation of recombinant cysteine proteinase inhibitor stefin B. *Protein Expr Purif* **5**, 65–69.
38 Govada L & Chayen E (2009) Crystallization by controlled evaporation leading to high resolution crystals of the C1 domain of cardiac myosin binding protein-C (cMyBP-C). *Cryst Growth Des* **2009**, 3.
39 Otwinowski Z & Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* **276**, 21.
40 Navaza J & Saludjian P (1997) AMoRe: an automated molecular replacement program package. *Methods Enzymol* **276**, 581–594.
41 Matthews BW (1968) Solvent content of protein crystals. *J Mol Biol* **33**, 491–497.
42 Kantardjieff KA & Rupp B (2003) Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Sci* **12**, 1865–1871.
43 Murshudov GN, Vagin AA & Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**, 240–255.
44 Turk D (1992) Weiterentwicklung eines Programms fuer Molekuelgraphik und Elektrondichte-Manipulation and Seine Anwendung auf Verschiedene Protein-Strukturaufklerungen. PhD thesis, Technische Universitaet Muenchen, Germany.

FEBS Letters xxx (2012) xxx–xxx

journal homepage: www.FEBSLetters.org

Expression, purification and assembly of soluble multimeric MHC class II–invariant chain complexes

Dušana Majera [a], Katarina Črnigoj Kristan [a], Jacques Neefjes [c], Dušan Turk [a,b,*], Marko Mihelič [a,b]

[a] Department of Biochemistry, Molecular and Structural Biology and Centre for Protein and Structure Production, Jozef Stefan Institute, Jamova 39, SI-1000 Ljubljana, Slovenia
[b] Centre of Excellence CIPKEBIP, Jamova 39, SI-1000 Ljubljana, Slovenia
[c] Division of Cell Biology and Centre for Biomedical Genetics, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 7 March 2012
Accepted 14 March 2012
Available online xxxx

Edited by Gianni Cesareni

Keywords:
Immune response
MHC class II molecule
Invariant chain
Transmembrane domain
Invariant chain trimerization

ABSTRACT

Major histocompatibility class (MHC) II molecules are essential for running adaptive immune response. They are produced in the ER and targeted to late endosomes with the help of invariant chain (Ii) trimers. Ii trimerization may be induced by the Ii TM domain. To enable mechanistic and structural studies of MHC class II–Ii assembly, soluble forms of the complexes were expressed. We show that Ii trimerizes in the absence of the transmembrane part, prior to binding of α/β chains. The biochemical analysis supports the suggestion that the MHC class II–Ii complexes are not necessarily trimers of trimers, but that the Ii trimer can also be occupied by one or two MHC class II complexes.

1. Introduction

The major histocompatibility class (MHC) II–invariant chain (Ii) complexes are transmembrane glycoproteins that play a central role in adaptive immunity [1,2]. They are synthesized in the endoplasmic reticulum (ER) where the nascent MHC class II α and β chains associate with Ii trimers [3,4]. Once fully assembled, the MHC class II–Ii complexes are transported to the late endosomal compartments called MIIC of professional antigen presenting cells (APCs) [5,6]. Here, Ii is degraded in a step-wise manner by the lysosomal cysteine proteases [7,8] resulting in the release of the active MHC class II dimers. Ii consists of several distinct segments that play specific roles in the assembly and cellular localization of MHC class II complexes, including the N-terminal cytosolic tail which contains signals for the delivery of MHC class II into the endocytic route [9], a single transmembrane helix, a trimerisation domain located at the luminal site of the protein and a short segment termed CLIP which associate with the MHC class II peptide binding groove and prevents premature binding of antigenic peptides [10]. Ii exists in two alternatively spliced forms, the p31 and p41 form, which differ by the insertion of a 64 amino acid long domain in the p41 form of Ii [11]. This domain is a strong inhibitor of cysteine endopeptidases suggesting that the invariant chain might regulate its own proteolytic processing [12,13].

The molecular mechanism of MHC class II α (DRA) and β (DRB) chain and Ii association has been extensively studied for the past decades, however the exact mechanism of their assembly still remains unclear. The classic model of the MHC class II–Ii assembly suggests that the Ii first forms a trimer via its trimerisation domain. This trimers then act as a chaperone-like scaffold for the assembly of three α and β chain dimers that concurrently bind to the Ii trimers forming a nonameric complex [14,15]. Recently, Koch et al. challenged this concept by showing that association of Ii with DRA alone precede the binding of DRB and that the Ii trimer can bind only a single MHC class II αβ dimer before leaving the ER for transport into the endocytic route [16–18].

Several studies evaluated the importance of the transmembrane (TM) region in the trimerisation of Ii. It was shown that the TM regions of the Ii strongly self-associate as a trimer [19]. Destabilization of this TM interaction prevented trimerisation of Ii and assembly of the complexes [20]. These studies were performed by mutations of highly conserved amino acid residues within the Ii TM domain and do not fully support the role of the TM domain in the assembly of the complexes.

To unambiguously address the role of the TM domain of the Ii in the assembly of the MHC class II–Ii complexes, we produced soluble complexes lacking the TM region. The analysis of the purified MHC class II–Ii complexes lacking the TM regions showed that soluble Ii still formed a trimer that associate with MHC class II α and β

* Corresponding author at: Department of Biochemistry, Molecular and Structural Biology and Centre for Protein and Structure Production, Jozef Stefan Institute, Jamova 39, SI-1000 Ljubljana, Slovenia. Fax: +386 1 477 3984.
E-mail address: dusan.turk@ijs.si (D. Turk).

0014-5793/$36.00
http://dx.doi.org/10.1016/j.febslet.2012.03.038

chains. These complexes are efficiently secreted by cells allowing rapid purification in quantities sufficient for further biochemical studies.

2. Materials and methods

2.1. Gene cloning and construction of the expression vectors

The truncated MHC class II α and β chains (ΔDRA and ΔDRB) containing the N-terminal signal sequence but without the C-terminal TM and cytosolic domains were amplified from the full length DRA and DRB cDNAs, using the following set of primers: 5′CCCA AAGCTTGCC**ACC*ATGG***CCATAAGTGGAGTCCCT3′ and 5′CCCAGAATTCT TACTCTGTAGTCT3′ as a forward and reverse primers for ΔDRA and 5′CCCAAAGCTTGCC**ACC*ATGG***TGTGTCTGAAGCTCCCT3′ and 5′CCCA GAATTCTTACTTGCTCTGTG CAGATTCAG3′ as a forward and reverse primers for ΔDRB. The underlined sequences correspond to the HindIII (forward primers) and EcoRI (reverse primers) endonuclease recognition site. The region presented in bold is a Kozak consensus sequence. The PCR products were cloned into the pcDNA3 expression plasmid (Invitrogen). The human p31 and p41 forms of Ii (Δp31Ii and Δp41Ii) without the cytoplasmic and TM regions were PCR amplified using 5′CCCAACCGGTCGGCTGGACAAACTGACAGT3′ as a forward and 5′CCCAGG TACCCATGGGGACTGGGCCCAGATC3′ as a reverse primers. The underlined regions represent AgeI and KpnI restrictions sites. The PCR products were cloned into the pHLsec expression plasmid [21] in frame with the N-terminal H2-κ^b signal sequence and the C-terminal His-tag.

2.2. Expression of truncated MHC class II – invariant chain complexes

HEK293 cells were maintained in DMEM medium supplemented with 4 mM L-glutamine and 10% fetal bovine serum in a CO_2 incubator at 37 °C with 5% CO_2.

For the small-scale transfection, the HEK293T and HEK293S GnTI- cells were seeded in a 6-well cell culture plate and transfected with Lipofectamin (Invitrogen) according to the manufacturer's instructions. For large-scale expression the cells were seeded in T150 flasks and cultured in 25 ml medium. When the cells reached 60% confluency, co-transfection using 25 kDa linear PEI (Polysciences Inc.) was performed. Forty micrograms of individual plasmids containing cDNAs for ΔDRA, ΔDRB, or p41 form of ΔIi were mixed in equimolar ratios in a 5 ml of serum-free DMEM. One hundred and twenty micrograms of PEI was added to the mixture. After 10 min of incubation at the room temperature the mixture was added to the cells. After 4 h, the media was exchanged with 25 ml of DMEM without FBS and the cells were cultured for 3 days.

2.3. Purification of MHC class II – invariant chain complexes

The conditioned medium from the HEK293T cells seeded in 20 T150 flasks (approximately 500 ml) was collected and concentrated on the Amicon ultra filtration device and dialyzed against the 30 mM Tris, pH 7.5, 100 mM NaCl, 20 mM imidazol buffer (binding buffer). The sample was applied to the 1 ml HiTrap IMAC FF column (GE Healthcare) charged with Ni^{2+} at a flow rate 1 ml/min. The unbound proteins were washed with the 20 column volumes of the binding buffer. The bound proteins were eluted with the binding buffer containing 300 mM imidazol. The eluted complexes were dialyzed against the 20 mM Tris, pH 7.5 buffer and additionally purified on the Mono Q column (GE Healthcare) equilibrated in the same buffer. The recombinant complexes were eluted with a linear gradient of NaCl from 0 to 0.5 M in the 20 mM Tris buffer, pH 7.5. The eluted samples were desalted by a PD-10 desalting column, concentrated and stored at −70 °C.

2.4. SDS–PAGE, Western blotting and N-terminal sequencing

The purified complexes were analyzed by SDS–PAGE using the 15% Bis–Tris gels and visualized by the Coomassie Brilliant Blue staining. For Western blot analysis, the proteins were electrotransferred to nitrocellulose membrane. The membrane was blocked with 5% (w/v) skim milk in Tris-buffer saline containing 0.05% (v/v) Tween 20 (TBST) for one hour at room temperature. The rabbit polyclonal anti-DRA, anti-DRB [5] and anti-Ii antibodies (a gift from P. Cresswell, Yale University) were added in dilution 1:1000 and incubated for one hour at room temperature. The bound antibodies were detected with the horseradish peroxidase (HRP) conjugated goat anti-rabbit IgG and visualized with peroxidase substrate.

For N-terminal sequencing, the proteins separated on SDS–PAGE gel were electrotransfered to the PVDF membrane, and visualized with 0.1% Brilliant Blue G in 1% (v/v) acetic acid and 40% (v/v) methanol. The N-terminal sequencing was performed on Precise Protein Sequencing System 492 (PE Applied Biosystems).

2.5. Co-Immunoprecipitation of the complexes

The IB5 mouse monoclonal antibody against the human DRA or Tu36 mouse monoclonal antibody against DR dimers were added to 1 ml of the conditioned medium and incubated for one hour on the rotary wheel at 4 °C. The antibody–protein complex was isolated by protein A Sepharose beads (GE Healthcare) according to the manufacturer's instructions and analyzed by Western blot using anti-DRA, anti-DRB and anti-Ii rabbit polyclonal antibodies.

2.6. Analytical size-exclusion chromatography

Superdex 200 HR 10/30 (GE Healthcare) size exclusion column was equilibrated in 30 mM HEPES, pH 7.5, 0.3 M NaCl buffer and calibrated with the gel filtration standards (Bio-Rad). The M_r of the complexes was determined from the plots of elution volumes of standards versus log of their M_r.

2.7. Crosslinking of truncated MHC class II – invariant chain complexes

Ten micrograms of the Δ(DRA/DRB/p41Ii) complex or the Δp41Ii alone in 20 μl of HEPES, pH 7.5 buffer were mixed with 2.5 μl of 2.3% solution of gluteraldehyde or 5 mg/ml solution of DSS and incubated at the room temperature for 15 min. The crosslinking reaction was terminated by addition of 5 μl of 1 M Tris, pH 7.5. The crosslinked complexes were separated on the 3–8% Tris–Acetate PAGE gels and stained with Coomassie Blue.

2.8. Deglycosylation with Endo H

The samples were treated with endoglycosidase Endo H (New England Biolabs, UK) according to the manufacturer's instructions. The samples were separated by SDS–PAGE and analyzed by Western blotting using anti-DRB antibodies.

3. Results and discussion

3.1. Design of the expression constructs

To promote the expression of MHC class II–Ii chain complexes in the soluble form, sequences of the individual chains have been modified. In the case of Ii the N-terminal cytosolic tail and the transmembrane region (encompassing amino acids from G47 to Y72) followed by three consecutive Gln residues were removed.

Please cite this article in press as: Majera, D., et al. Expression, purification and assembly of soluble multimeric MHC class II–invariant chain complexes. FEBS Lett. (2012), http://dx.doi.org/10.1016/j.febslet.2012.03.038

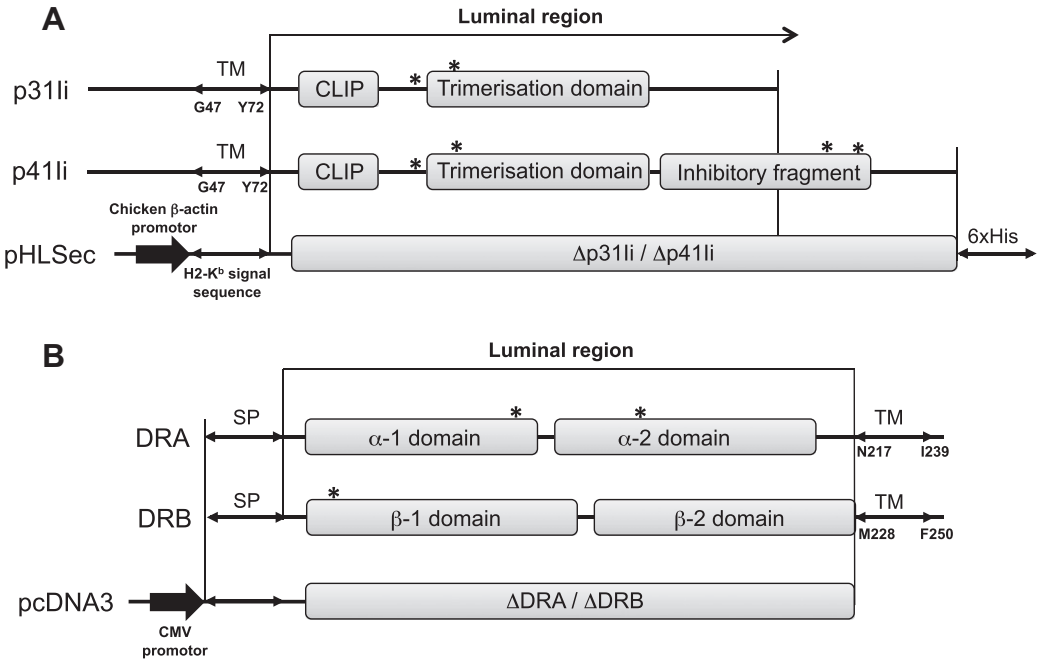


Fig. 1. Schematic representation of expression plasmids for (A) MHC class II associated p31 and p41 forms of the invariant chain (Ii) and (B) MHC class II α (DRA) and β (DRB) chains. The regions of DRA, DRB and Ii that were cloned into the expression plasmids are annotated with Δ. SP marks the signal peptide, TM the transmembrane region, * the sites of N-linked glycosylation and 6xHis the histidine affinity tag.

The remaining luminal part of the Ii starting from G76 was cloned into the pHLSec plasmid in frame with the N-terminal H2-Kb signal sequence (Fig. 1A). The truncated Ii lacks the N-terminal type II signal sequence essential for the delivery of the complexes into the endocytic route and we therefore introduced a new type I signal sequence for insertion in the ER and transport for secretion. In contrast to Ii, the transmembrane regions of MHC class II α and β chains are located at the C-termini. They encompass residues from N217 to I239 and from M228 to F250 of the α (DRA) and β (DRB) chains respectively. To facilitate expression of the soluble DRA and DRB chains both sequences were removed by terminating the chains at amino acid E216 (DRA) and K227 (DRB), respectively. Hence, only the luminal parts of DRA and DRB chains with the native N-terminal signal sequence were cloned into the pcDNA3 plasmid (Fig. 1B). All constructs contain the N-terminal signal sequences that will target the synthesized polypeptide chains into the lumen of ER, where the assembly of the complexes takes place.

3.2. Association of soluble MHC class II αβ and Ii in vivo

Individual constructs of the truncated ΔDRA, ΔDRB and ΔIi chains were transfected in HEK293 cells either alone or mixed in different combinations. Expression of individual chains was analyzed by immunoblotting (Fig. 2). When the p41 and p31 isoforms of the ΔIi were transfected alone or in combination with ΔDRA and ΔDRB chains the majority of the invariant chain was secreted (Fig. 2, Anti Ii blots). The absence of the cytosolic targeting signals thus resulted in targeting of the Ii into the secretory pathway [22]. In contrast to the ΔIi, a different distribution pattern was observed when expression of ΔDRA and ΔDRB chains was analyzed. When ΔDRA chain was transfected alone or in combination of only one of the other subunits (ΔDRB chain or the ΔIi), it was only detected inside the cells, whereas expression of all three subunits (the ΔDRA, ΔDRB and ΔIi chains) resulted in secretion as detected by anti-DRA antibodies (Fig. 2, Anti DR blot). Similar results were obtained for ΔDRB. The secreted DRB was detected only when ΔDRB was co-transfected together with ΔDRA and the ΔIi (Fig. 2, Anti DRB blot). When transfected alone or only with the ΔIi or ΔDRA, the ΔDRB chain was retained and detected only intracellularly. Due to the weak affinity and high cross reactivity of DRB antibodies, the intracellular expression of ΔDRB could be detected only by increasing the exposure time for blot development which resulted in the increased background and non-specific staining of other proteins. However, from Fig. 2, (Anti DRB* blot) it can be clearly seen that ΔDRB, when expressed alone or in combination with ΔDRA or ΔIi, is retained intracellularly. We have also observed increased intracellular expression of ΔDRA as well as ΔDRB when transfected with ΔIi alone (Fig. 2, blot Anti DRA and blot Anti DRB), suggesting that ΔIi somehow stabilize individual DRA and DRB. Comparison of ΔDRA, ΔDRB chains and the ΔIi inside the cells and secreted forms shows obvious differences in molecular weight. Since all three chains contain a number of *N*-glycans (Table 1), the deviation in the molecular weight marks conversion of high-mannose to complex *N*-glycans that is the result of transport through the secretory pathway. This also indicates that ΔDRA and ΔDRB when transfected alone or only in pairs are retained in the ER. To provide further support that the differences in the molecular weight arise from the localization and N-glycosylation pattern, we have analyzed ΔDRB transfected into the HEK293S GnTI- cell line which is deficient in *N*-acetylglucosaminyltransferase I activity and therefore lacks complex type *N*-glycans. When ΔDRB was cotransfected with ΔDRA and the ΔIi in HEK293S GnTI- cells, the apparent molecular weight of the secreted ΔDRB is identical to the molecular weight of the intracellularly accumulated ΔDRB (Fig. 3A). The weight of intracellular ΔDRB was further decreased by the Endo H treatment, which removes the basic mannose glycan attached to the proteins in the ER during early stages of their synthesis, whereas secreted ΔDRB (when cotransfected with ΔDRA and ΔIi) is Endo H resistant. The ER localization of ΔDRB, when transfected alone, was additionally confirmed by staining fixed cells with anti-DRB antibodies before analyses by confocal microscopy (Fig. 3B). Also ΔDRA was retained in the ER according to biochemical and microscopy criteria. These observations are consistent with the previous reports that MHC class II dimers can

Please cite this article in press as: Majera, D., et al. Expression, purification and assembly of soluble multimeric MHC class II–invariant chain complexes. FEBS Lett. (2012), http://dx.doi.org/10.1016/j.febslet.2012.03.038

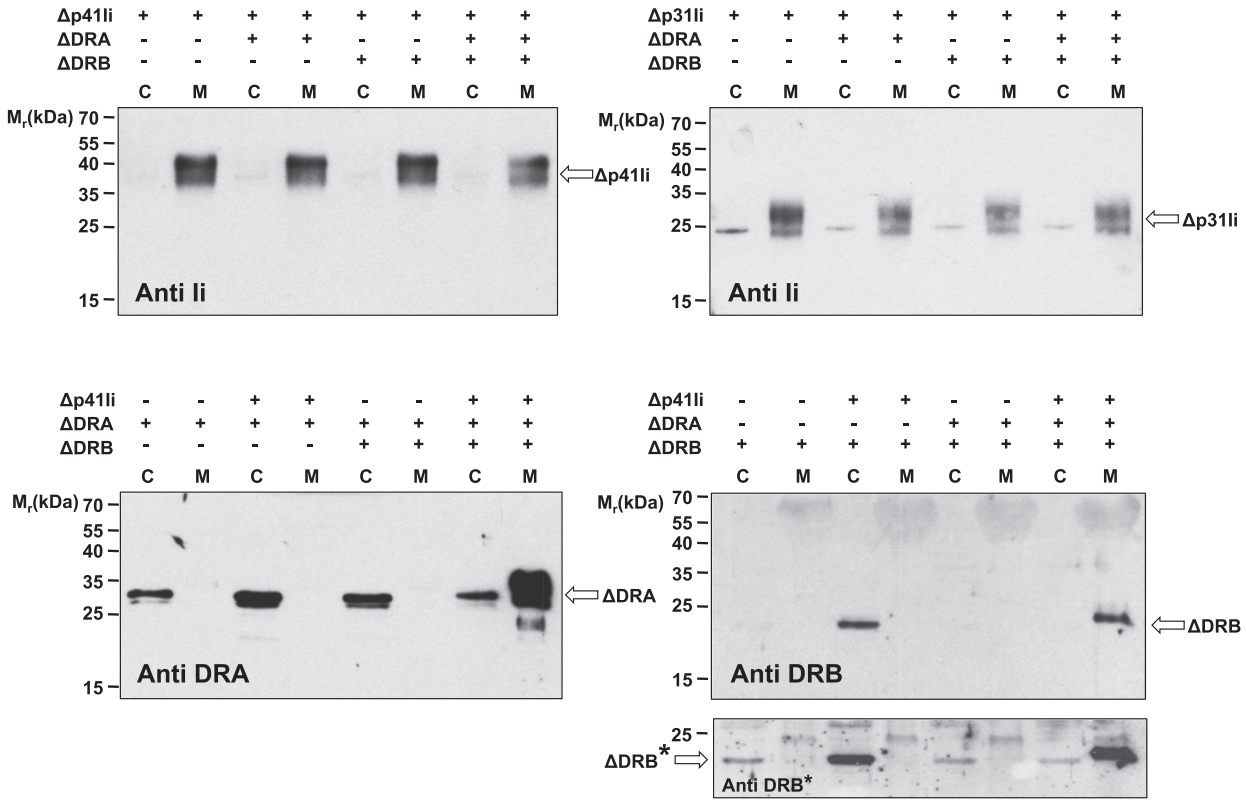


Fig. 2. Expression analysis of soluble MHC class II and Ii proteins. Western blot analysis of intra- (C) and extracellular (M) expression of Δp41Ii, Δp31Ii, ΔDRA and ΔDRB after transfection of HEK293T cells with different combinations of plasmids. The proteins were detected using the anti-Ii, anti-DRA and anti-DRB antibodies. Plus (+) and minus (−) signs indicate which combination of plasmids was used in the transfection experiments. The anti-DRB blot developed with extended exposure time is marked with ∗.

Table 1
Calculated molecular weights and number of potential N-glycosylation sites of truncated forms of MHC class II DRA and DRB and p41 invariant chain.

	Mr (kDa)	N-glycosylation sites
ΔDRA	22	2
ΔDRB	23	1
Δp41Ii	25	4
$\Delta(p41Ii)_3$	75	12
$\Delta(p41Ii)_3\Delta(DRADRB)$	120	15
$\Delta(p41Ii)_3\Delta(DRADRB)_2$	165	18
$\Delta(p41Ii)_3\Delta(DRADRB)_3$	210	21

fold and leave – though inefficiently – the ER whereas the free DRA and DRB chains are retained in the ER [15,23].

The existence of the complex between secreted ΔDRA, ΔDRB and the ΔIi was verified by immunoprecipitation using Tu36 mouse monoclonal antibody that recognizes only properly folded and loaded MHC class II DR dimers and the 1B5 mouse monoclonal antibody recognizing DRA [5]. Immunoblots of immuno-isolated fractions using anti-DRA, DRB and Ii antibodies showed that all three chains associated in a complex (Fig. 4).

3.3. Large-scale expression and purification of soluble MHC class II – Ii invariant chain complexes

The MHC class II–Ii complexes secreted from HEK293T and glycosylation defective HEK293S GnTI- cell lines were purified on the Ni-chelating column. We have attached a His-tag at the C-terminus of Ii for isolation purposes. Co-purified ΔDRA and ΔDRB should then be in a complex with ΔIi. To separate the soluble MHC class II–Ii chain complexes from the soluble Ii alone that might co-purify on the Ni-affinity column, the samples were subjected to further purification. The calculated isoelectric points for the truncated MHC class II–Ii complex and the Ii are 5.71 and 8.05, respectively which suggest that they can be separated by ion-exchange chromatography (Mono-Q) performed at pH 7.5. At this pH, the complex is negatively charged and should bind to the positively charged ion-exchange carrier, whereas the more acidic empty Ii trimer should theoretically flow through. The fractions corresponding to the major eluted peak from the Mono-Q column were collected and analyzed by SDS–PAGE (Fig. 5A). The positions of ΔDRA, ΔDRB and ΔIi on the PAGE-gel were identified by combining the results of the Western blot analysis and the N-terminal amino acid sequencing. The N-terminal sequences confirmed the presence of all three chains in the sample (Fig. 5B).

The molecular weights of ΔDRB – as deduced from SDS–PAGE analysis – corresponded to the theoretical masses (including the carbohydrates) calculated from the amino acid sequences (Table 1). The estimated molecular weight for ΔDRA is higher than calculated, however expression in HEK293S GnTI- cells results in the formation of a sharp band at 30 kDa (estimated Mw). The apparent weight of the glycosylated p41 form of ΔIi is between 35 and 40 kDa, which is substantially higher than the theoretical value. Expression of the complexes in N-glycosylation deficient cell line resulted in the decrease of the molecular weight of the Ii, however, the Ii is still present in several bands with higher weight than expected. The single O-glycosylation site in the Ii that contributes a considerable number of negative charges may cause the faster migration by SDS–PAGE. Immunoblot analysis of the purified complexes using anti-His antibodies, which recognize the affinity tag attached to the C-terminus of the Ii, showed identical staining as Coomassie Blue. By the N-terminal sequencing, we were able to identify only a single amino acid sequence that starts at the last three amino acids of the H2-κ^b signal sequence followed by the

Please cite this article in press as: Majera, D., et al. Expression, purification and assembly of soluble multimeric MHC class II–invariant chain complexes. FEBS Lett. (2012), http://dx.doi.org/10.1016/j.febslet.2012.03.038

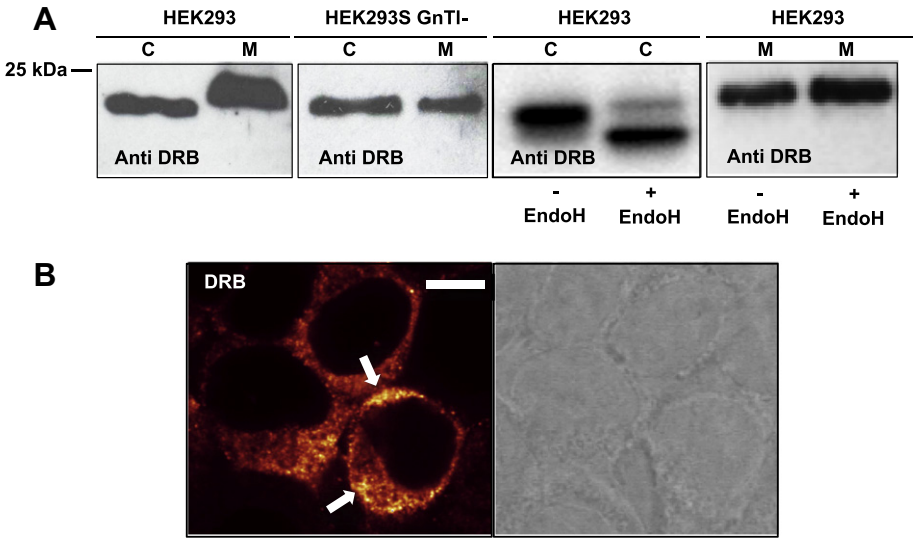


Fig. 3. Analysis of the molecular size, cellular localization and glycosylation pattern of ΔDRB. (A) ΔDRB was transfected together with ΔDRA and the Δli in HEK293T and HEK293S GnTI- cells and analyzed by anti-DRB immunobloting. C and M annotates intracellular and extracellular expression. Intracellular and secreted ΔDRB was additionally treated with Endo H. (B) Immunofluorescence staining of ΔDRB transfected alone in HEK293T cells and imaged by confocal microscopy. The arrow indicates the accumulation of the ΔDRB inside the ER. Bar: 100 μm.

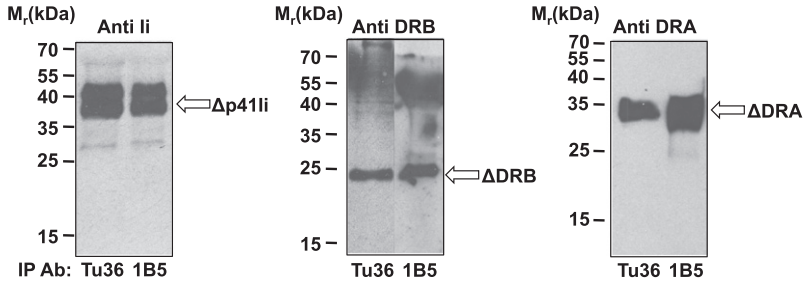


Fig. 4. Secreted MHC class II molecules lacking the transmembrane region associate with the soluble Ii. Western blot analysis of immunoprecipitated soluble MHC class II–Ii complexes is shown. The complexes were immunoprecipitated with Tu36 and 1B5 mouse monoclonal antibodies. The individual chains of the complex were detected with the rabbit polyclonal antibodies specific to individual chain.

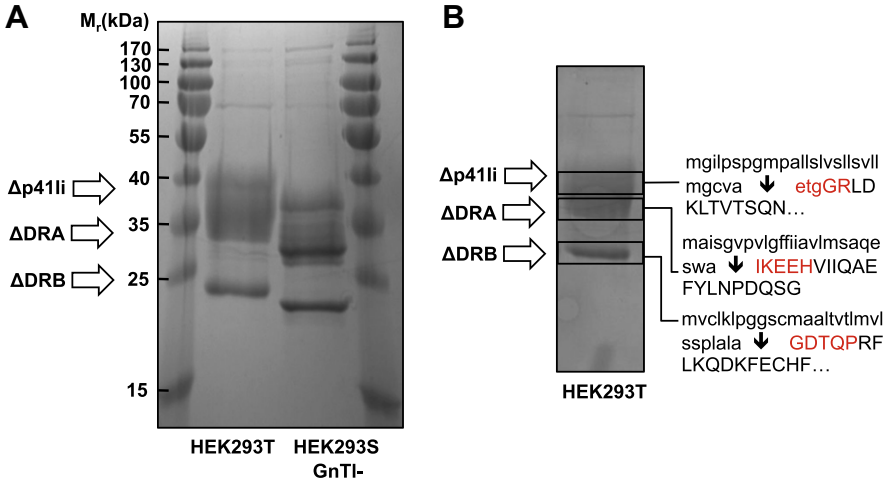


Fig. 5. SDS–PAGE analysis and N-terminal amino acid sequencing of MHC class II–Ii chain complexes. (A) The purity of the complexes expressed in HEK293T (lane 1) and HEK293S GnTI- cell line (lane 2) was analyzed by SDS–PAGE. The positions of ΔDRA, ΔDRB and the Δp41Ii are marked. (B) The complex purified from cotransfected HEK293T cells was transferred to PVDF membrane. The bands that were cut out and sequenced are shown in frames. The determined N-terminal amino acid sequences are shown in red. The predicted cleavage sites for signal peptidase are marked with arrows.

Ii sequence. This further indicates that multiple and diffused bands corresponding to the p41 Ii form of the invariant chain are the consequence of posttranslational modifications and not of proteolytic degradation. The N-terminal sequence was also determined for

Please cite this article in press as: Majera, D., et al. Expression, purification and assembly of soluble multimeric MHC class II–invariant chain complexes. FEBS Lett. (2012), http://dx.doi.org/10.1016/j.febslet.2012.03.038

ΔDRA and ΔDRB. In both cases the sequence matched the sequences of DRA and DRB lacking the signal peptide (Fig. 5b).

3.4. Oligomeric state of soluble MHC class II–Ii complexes

To analyze the stoichiometry of the expressed proteins, we assessed the molecular weight of purified proteins by analytical size exclusion chromatography and crosslinking experiments. The purified proteins eluted from the Superdex 200 HR size exclusion column in a single symmetric peak with elution volume around 10 ml (Fig. 6A). The apparent molecular weight of these complexes was determined from the calibration curve obtained by plotting the elution volumes of the protein standards versus logarithm of their molecular weight (Fig. 6B). The elution volumes of the ΔDRA-DRBp41Ii complexes purified from HEK293T and HEK293S GnTI- cell lines correspond to the apparent molecular weights of 290 and 273 kDa. Since only a single peak was eluted from the size-exclusion column these results suggest that the expressed ΔDRA, ΔDRB and the ΔIi associate in high-molecular weight complex. (It should be noted that the estimation of molecular weight of proteins by size exclusion chromatography is reliable only for globular proteins as it depends on the protein radius and its hydrodynamic properties and should therefore be considered only as an approximation of the actual weight.)

To independently confirm the oligomeric structure of the soluble MHC class II–p41Ii complex, the purified complex was crosslinked with gluteraldehyde and DSS. Both crosslinkers gave the same results so only data with gluteraldehyde are shown (Fig. 7). To reduce the unspecific crosslinking between complexes, the concentration of the protein used in these experiments was reduced below 0.5 mg/ml. The molecular weight of the cross-linked soluble p41Ii is 100 and 75 kDa for glycosylated and non-glycosylated forms of the protein. This weight is in agreement with the expected molecular weight of the Ii trimers. Since the trimeric structure was formed in the absence of the transmembrane and cytosolic parts of the Ii, it is evident that the transmembrane region and cytosolic tail of Ii are not critical for the formation of the nonameric complex. The luminal segment of Ii should also contain information that induces trimerization. The crosslinking of the Δ(DRADRBp41Ii) complexes showed two major bands of molecular mass of 130 and 180 kDa for glycosylated form of the complexes, and two bands of 110–140 kDa for underglycosylated forms of the complexes. However, a minor bands of 210 and 170 kDa were also observed. A protein complex of 209 kDa would correspond to the calculated molecular weight of the expected $(DRA–DRB–Ii)_3$ nonameric form. The results from the cross-linking experiments suggest considerable heterogeneity in the binding of HLA–DRAB complexes on the Ii trimer with a preference for one and two MHC class II complexes corresponding to molecular weights of 120 and 160 kDa, respectively (Table 1).

With two independent experiment approaches, namely size-exclusion chromatography and crosslinking, we have estimated the composition of secreted MHC class II–Ii protein complexes. The size exclusion experiments suggest that the complex resembles a trimer of trimers with three MHC class II complexes associated to three Ii chains that form the core of the complex. Chemical cross-linking and more precise molecular weight determination by SDS–PAGE reveals a more complex picture with an Ii trimer associated with mainly one or two MHC class II complexes and only a small proportion in the nonameric state. This would correspond to the recent observations Koch and coworkers [18], who suggested that the stoichiometry of the MHC class II–Ii chain composition may not be a trimer of trimers (a DRA–DRB–Ii trimer) [3] but a pentamer composed of a trimer of Ii with a one pair of Δ (DRADRB) bound. As Ii is present in the ER in considerably higher amounts than DRA and DRB chains, two options exist. MHC class II-Ii complexes are only considered correctly folded to leave the ER when in a nonameric complex (the trimer of trimers) or the Ii trimer can leave the ER when already one or two MHC class II complexes are associated. Our results suggest that the latter option is most likely. The ratio between fully and partially MHC class II–Ii loaded complexes may depend on the relative amounts of Ii and MHC class II DRA and DRB expressed, which may vary between cells and likely depending on the DRA and DRB sequences. The existence of intermediate MHC class II–Ii complexes of various stoichiometries was also observed before. Jasanoff et al. showed that under in vitro conditions MHC class II dimers bind to the Ii trimmers independently and that nonameric complexes are formed only in the excess of MHC class II dimmers [24].

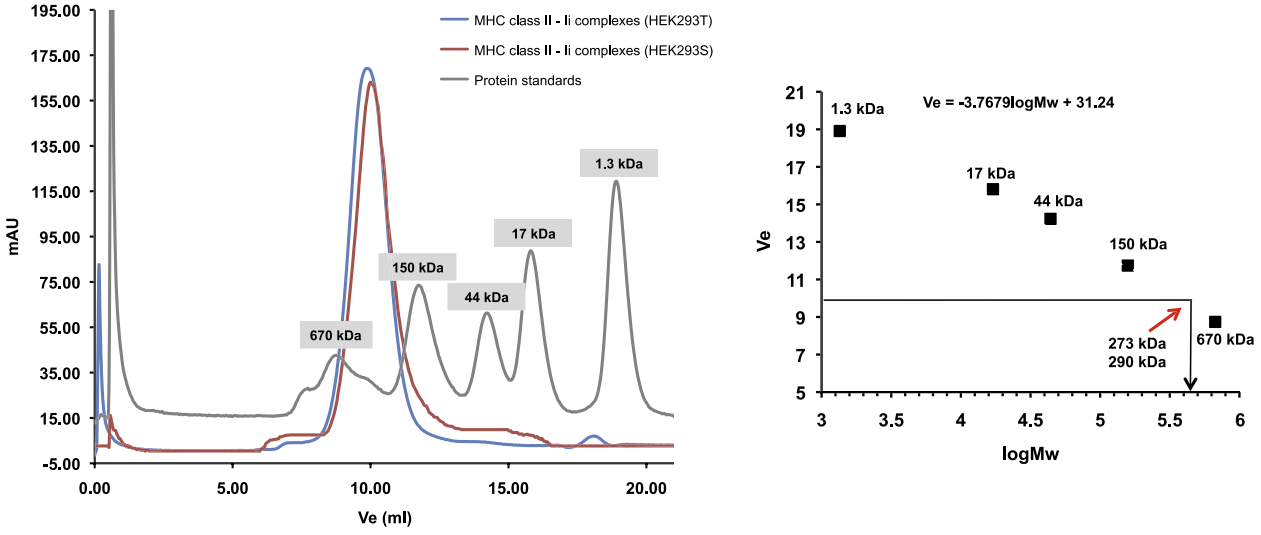


Fig. 6. Assessment of molecular weight of MHC class II–41Ii complexes using size exclusion chromatography. (A) Superdex 200 HR elution profiles of MHC class II–p41Ii complexes expressed in HEK293T (blue) and HEK293S GnTI- cells (red). The elution profile of protein standards (thyroglobulin 667 kDa, gamma-globulin 158 kDa, ovalbumin 44 kDa, myoglobin 17 kDa and vitamin B12 1.35 kDa) is shown in gray. (B) Calibration plot obtained from plotting the elution volume of standards versus their molecular weight. The elution volume and calculated molecular weight for MHC class II–p41Ii complexes are marked.

Please cite this article in press as: Majera, D., et al. Expression, purification and assembly of soluble multimeric MHC class II–invariant chain complexes. FEBS Lett. (2012), http://dx.doi.org/10.1016/j.febslet.2012.03.038

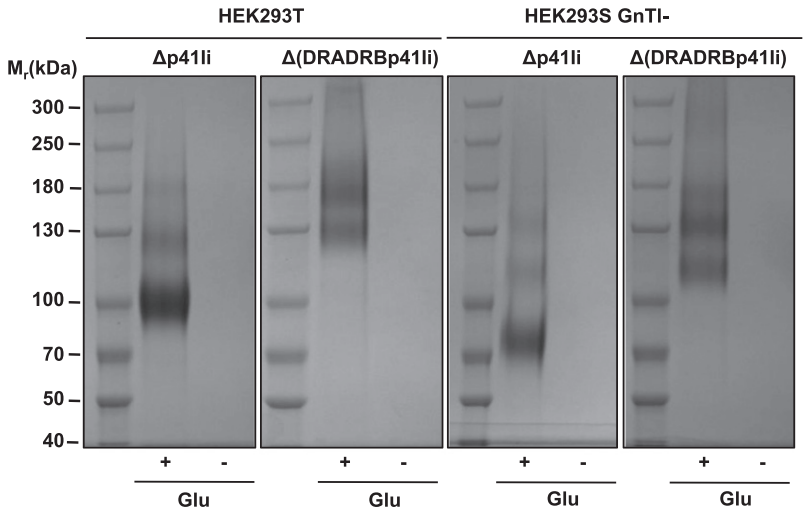


Fig. 7. Crosslinking of the soluble MHC class–p41Ii and p41Ii complexes. The lanes with the gluteraldehyde (Glu) crosslinked material are marked on the top with "+" and the lanes without crosslinking with "−". The proteins were separated on 3–8% Tris–Acetate gels and stained with Coomassie Blue.

4. Conclusions

Several studies that were addressing the role of Ii TM region in the trimerisation of Ii and assembly of MHC class II–Ii complexes suggested that this domain plays an essential role in these processes [19]. These studies utilized point mutations in the highly conserved patch of polar amino acids in the TM domain of the Ii. These mutations possibly disrupted the interaction between the transmembrane helices of Ii, which might disturb the orientation of Ii in the luminal domain to prevent trimerisation and binding of MHC class II dimers. Here we analyzed the assembly and oligomeric state of the Ii and MHC class II complexes lacking the transmembrane region. We have shown by in vivo and in vitro analysis that the Ii chain lacking the transmembrane and cytoplasmic regions is efficiently secreted as a trimer. Moreover, we have shown that soluble Ii can assemble and be secreted with soluble DRA and DRB in complexes that may be more heterogeneous than assumed previously.

Acknowledgements

This work was financially supported by the Marie Curie Fellowship MRTN-CT-512385 (to D.M.) and Slovenian Research Agency.

References

[1] Rocha, N. and Neefjes, J. (2008) MHC class II molecules on the move for successful antigen presentation. EMBO J. 27, 1–5.
[2] Cresswell, P., Blum, J.S., Kelner, D.N. and Marks, M.S. (1987) Biosynthesis and processing of class II histocompatibility antigens. Crit. Rev. Immunol. 7, 31–53.
[3] Roche, P.A., Marks, M.S. and Cresswell, P. (1991) Formation of a nine-subunit complex by HLA class II glycoproteins and the invariant chain. Nature 354, 392–394.
[4] Lamb, C.A. and Cresswell, P. (1992) Assembly and transport properties of invariant chain trimers and HLA-DR-invariant chain complexes. J. Immunol. 148, 3478–3482.
[5] Neefjes, J.J., Stollorz, V., Peters, P.J., Geuze, H.J. and Ploegh, H.L. (1990) The biosynthetic pathway of MHC class II but not class I molecules intersects the endocytic route. Cell 61, 171–183.
[6] Pieters, J., Horstmann, H., Bakke, O., Griffiths, G. and Lipp, J. (1991) Intracellular transport and localization of major histocompatibility complex class II molecules and associated invariant chain. J. Cell Biol. 115, 1213–1223.
[7] Manoury, B., Mazzeo, D., Li, D.N., Billson, J., Loak, K., Benaroch, P. and Watts, C. (2003) Asparagine endopeptidase can initiate the removal of the MHC class II invariant chain chaperone. Immunity 18, 489–498.
[8] Hsing, L.C. and Rudensky, A.Y. (2005) The lysosomal cysteine proteases in MHC class II antigen presentation. Immunol. Rev. 207, 229–241.
[9] Bakke, O. and Dobberstein, B. (1990) MHC class II-associated invariant chain contains a sorting signal for endosomal compartments. Cell 63, 707–716.
[10] Stumptner-Cuvelette, P. and Benaroch, P. (2002) Multiple roles of the invariant chain in MHC class II function. Biochim. Biophys. Acta 1542, 1–13.
[11] Strubin, M., Berte, C. and Mach, B. (1986) Alternative splicing and alternative initiation of translation explain the four forms of the Ia antigen-associated invariant chain. EMBO J. 5, 3483–3488.
[12] Bevec, T., Stoka, V., Pungercic, G., Dolenc, I. and Turk, V. (1996) Major histocompatibility complex class II-associated p41 invariant chain fragment is a strong inhibitor of lysosomal cathepsin L. J. Exp. Med. 183, 1331–1338.
[13] Mihelic, M., Dobersek, A., Guncar, G. and Turk, D. (2008) Inhibitory fragment from the p41 form of invariant chain can regulate activity of cysteine cathepsins in antigen presentation. J. Biol. Chem. 283, 14453–14460.
[14] Romagnoli, P. and Germain, R.N. (1994) The CLIP region of invariant chain plays a critical role in regulating major histocompatibility complex class II folding, transport, and peptide occupancy. J. Exp. Med. 180, 1107–1113.
[15] Anderson, M.S. and Miller, J. (1992) Invariant chain can function as a chaperone protein for class II major histocompatibility complex molecules. Proc. Natl. Acad. Sci. USA 89, 2282–2286.
[16] Neumann, J. and Koch, N. (2005) Assembly of major histocompatibility complex class II subunits with invariant chain. FEBS Lett. 579, 6055–6059.
[17] Koch, N., McLellan, A.D. and Neumann, J. (2007) A revised model for invariant chain-mediated assembly of MHC class II peptide receptors. Trends Biochem. Sci. 32, 532–537.
[18] Koch, N., Zacharias, M., Konig, A., Temme, S., Neumann, J. and Springer, S. (2011) Stoichiometry of HLA class II-invariant chain oligomers. PLoS ONE 6, e17257.
[19] Dixon, A.M., Stanley, B.J., Matthews, E.E., Dawson, J.P. and Engelman, D.M. (2006) Invariant chain transmembrane domain trimerization: a step in MHC class II assembly. Biochemistry 45, 5228–5234.
[20] Ashman, J.B. and Miller, J. (1999) A role for the transmembrane domain in the trimerization of the MHC class II-associated invariant chain. J. Immunol. 163, 2704–2712.
[21] Aricescu, A.R., Lu, W. and Jones, E.Y. (2006) A time- and cost-efficient system for high-level protein production in mammalian cells. Acta crystallogr. D: Biol. Crystallogr. 62, 1243–1250.
[22] Pieters, J., Bakke, O. and Dobberstein, B. (1993) The MHC class II-associated invariant chain contains two endosomal targeting signals within its cytoplasmic tail. J. Cell Sci. 106 (Pt 3), 831–846.
[23] Viville, S., Neefjes, J., Lotteau, V., Dierich, A., Lemeur, M., Ploegh, H., Benoist, C. and Mathis, D. (1993) Mice lacking the MHC class II-associated invariant chain. Cell 72, 635–648.
[24] Jasanoff, A., Song, S., Dinner, A.R., Wagner, G. and Wiley, D.C. (1999) One of two unstructured domains of Ii becomes ordered in complexes with MHC class II molecules. Immunity 10, 761–768.

Please cite this article in press as: Majera, D., et al. Expression, purification and assembly of soluble multimeric MHC class II–invariant chain complexes. FEBS Lett. (2012), http://dx.doi.org/10.1016/j.febslet.2012.03.038