

Structural and functional characterization
of the protein-protein interaction between the
HCMV immunoevasin UL16 and the
NKG2D ligand MICB

*Strukturelle und funktionelle Charakterisierung
der Protein-Protein Interaktion zwischen
dem HCMV Immunevasin UL16 und dem NKG2D
Liganden MICB*

DISSERTATION

der Fakultät für Chemie und Pharmazie
der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines Doktors
der Naturwissenschaften

2010

vorgelegt von
Steffen Müller

Tag der mündlichen Prüfung:

16.07.2010

Dekan:

1. Berichterstatter:

2. Berichterstatter:

Prof. Dr. Lars Wesemann

Prof. Dr. Thilo Stehle

Prof. Dr. Alexander Steinle

Preface

A number of paragraphs, figures and tables of this thesis have been published before in the following article and are indicated and referenced accordingly (see also 8. *Publications*).

1. Muller S, Zocher G, Steinle A, Stehle T (2010) Structure of the HCMV UL16-MICB complex elucidates select binding of a viral immunoevasin to diverse NKG2D ligands. PLoS Pathog 6: e1000723.

The contributions of the individual authors to this publication were as follows: SM, AS and TS conceived and designed the experiments; SM performed the experiments and generated the figures and tables; SM, GZ, AS and TS analyzed the data; SM, AS and TS wrote the paper.

Data, figures and tables that were not generated by the author of this thesis are indicated and referenced accordingly.

Contents

1. INTRODUCTION	1
1.1 The human immune system	1
1.1.1 Immunoreceptors employed by the innate immune system	2
1.1.2 Immunoreceptors employed by the adaptive immune system	3
1.2 Natural killer cells	3
1.2.1 NK cell receptors	5
1.2.1.1 Inhibitory NK cell receptors	8
1.2.1.2 Activating NK cell receptors	8
1.2.2 The activating NK cell receptor NKG2D	8
1.2.3 Ligands of the activating NK cell receptor NKG2D	9
1.3 Herpes viruses	12
1.3.1 Human cytomegalovirus	12
1.3.2 HCMV immune evasion strategies	13
1.3.2.1 Interference with MHC class I peptide presentation	13
1.3.2.2 Molecular decoys of inhibitory NK cell receptor ligands	15
1.3.2.3 Downmodulation of activating NK cell receptor ligands	16
1.3.2.4 The UL16 protein	17
1.4 Aims	18
2. MATERIALS AND METHODS	19
2.1 Materials	19
2.1.1 Hardware	19
2.1.1.1 General purpose equipment	19
2.1.1.2 Molecular biology	19
2.1.1.3 X-ray crystallography	19
2.1.1.4 Surface Plasmon Resonance	20
2.1.1.5 Protein production: Eukaryotic	20
2.1.1.6 Protein production: Prokaryotic	20
2.1.1.7 Chromatography	20
2.1.1.8 Protein Analytics	20
2.1.2 Software	21
2.1.2.1 X-ray data processing and structural analysis	21
2.1.2.2 Surface Plasmon Resonance	21
2.1.2.3 Data analysis	21
2.1.2.4 Chromatography	21
2.1.2.5 Reference management	21
2.1.2.6 Online tools and databanks	21
2.1.3 Consumables	22
2.1.3.1 Chemicals	22
2.1.3.2 Eukaryotic cell culture	23
2.1.3.3 Prokaryotic (bacterial) cell culture	23
2.1.3.4 Protein purification and analysis	23
2.1.3.5 Molecular biology	24
2.1.3.6 Crystallization Screens (commercial)	24
2.1.3.7 Surface Plasmon Resonance	24

2.1.3.8	Enzymes	24
2.1.3.9	Proteins	25
2.1.4	Plasmids and primers	25
2.1.4.1	Plasmids (eukaryotic expression)	25
2.1.4.2	Plasmids (prokaryotic expression)	26
2.1.4.3	Primers	26
2.1.5	Growth media, buffers and solutions	26
2.1.5.1	Cultivation of bacterial cells	26
2.1.5.2	Cultivation of eukaryotic cells	27
2.1.5.3	Protein expression, refolding and purification	27
2.1.5.4	Native protein deglycosylation	29
2.1.5.5	SDS-PAGE	29
2.1.5.6	Agarose gel electrophoresis	30
2.1.5.7	Surface Plasmon Resonance	30
2.1.5.8	Crystallization	30
2.1.6	Bacterial strains and eukaryotic cell lines	31
2.1.6.1	Bacterial strains	31
2.1.6.2	Eukaryotic cell lines	31
2.2	Methods	32
2.2.1	Methods in molecular biology	32
2.2.1.1	Competent cells	32
2.2.1.2	Polymerase chain reaction (PCR)	33
2.2.1.3	Purification of PCR fragments	34
2.2.1.4	Agarose gel electrophoresis	34
2.2.1.5	Agarose gel extraction	35
2.2.1.6	Restriction digestion of plasmids and PCR products	35
2.2.1.7	Ligation	36
2.2.1.8	Agar plates	37
2.2.1.9	Transformation	37
2.2.1.10	Colony PCR	38
2.2.1.11	Bacterial overnight cultures	39
2.2.1.12	Glycerol stocks of <i>E.coli</i> cells	39
2.2.1.13	Isolation of plasmid DNA from <i>E.coli</i>	39
2.2.1.14	Photometric determination of DNA concentration and purity	40
2.2.2	Methods in cell biology	40
2.2.2.1	Cultivation of eukaryotic cells	40
2.2.2.2	Roller bottle setup with CHO Lec 3.2.8.1 cells	41
2.2.2.3	Cryoconservation of eukaryotic cells	41
2.2.2.4	Counting of living cells	42
2.2.2.5	Kill curve	42
2.2.2.6	Transfection	43
2.2.2.7	Subcloning	44
2.2.3	Methods in protein biochemistry	45
2.2.3.1	Protein concentration determination	45
2.2.3.2	Discontinuous SDS-PAGE	47
2.2.3.3	General aspects of protein purifications performed with FPLC systems	48
2.2.3.4	Constructs, expression, refolding and purification of MICB _{short} and MICB _{long}	49
2.2.3.5	Constructs, expression and purification of UL16	54
2.2.3.6	Purification of the glycosylated UL16 _{mon} -MICB _{short} complex	59
2.2.3.7	Purification of deglycosylated UL16 _{mon} -MICB _{short} complex	60
2.2.3.8	Deglycosylation of CV-1 cell produced UL16	61
2.2.4	Surface Plasmon Resonance	61
2.2.5	X-ray crystallography	63
2.2.5.1	X-rays and data collection	63
2.2.5.2	Protein crystal symmetry	64
2.2.5.3	Protein crystal growth	65

2.2.5.4	Protein crystal freezing	68
2.2.5.5	X-ray diffraction	69
2.2.5.6	Data collection	71
2.2.5.7	Data processing	72
2.2.5.8	Phasing	77
2.2.5.9	Structure refinement	81
2.2.5.10	Glycan modeling.	84
3.	RESULTS	85
3.1	Expression and purification of MICB_{long} and MICB_{short}	85
3.1.1	Bacterial expression of MICB _{long} and MICB _{short}	85
3.1.2	IMAC purification of MICB _{long} and MICB _{short}	87
3.1.3	SEC purification of MICB _{long} and MICB _{short}	89
3.2	Expression and purification of UL16 and the UL16-MICB complex	90
3.2.1	Expression of UL16 in CV-1 cells	90
3.2.2	Deglycosylation of UL16 produced in CV-1 cells	91
3.2.3	Expression of UL16 in CHO Lec 3.2.8.1 cells	92
3.2.4	Subcloning of UL16-Fc expressing CHO Lec 3.2.8.1 cells	94
3.2.5	UL16-Fc purification by Protein A affinity chromatography	95
3.2.6	UL16 purification by Protein A and benzamidine affinity chromatography	96
3.2.7	Monomeric and dimeric forms of UL16	97
3.2.8	The monomer represents the active form of UL16	97
3.2.9	Standard methods fail to separate the UL16 monomer from the dimer	99
3.2.10	Ni-NTA purification of the UL16 _{mon} -MICB _{long} complex	101
3.2.11	SEC purification of the glycosylated UL16 _{mon} -MICB _{long} complex	102
3.2.12	Ni-NTA purification of the UL16 _{mon} -MICB _{short} complex	104
3.2.13	HIC purification of the UL16 monomer	105
3.2.14	SEC purification of glycosylated UL16 _{mon} -MICB _{short} complex	106
3.2.15	SEC purification of deglycosylated UL16 _{mon} -MICB _{short} complex	107
3.3	Crystallization of the UL16-MICB complex	108
3.3.1	Glycosylated UL16 _{mon} -MICB _{long} complex	108
3.3.2	Glycosylated UL16 _{mon} -MICB _{short} complex	110
3.3.3	Deglycosylated UL16 _{mon} -MICB _{short} complex	111
3.4	SPR measurements of UL16-NKG2D ligand interactions	112
3.5	Structure determination of the UL16-MICB complex	114
3.5.1	Data collection	114
3.5.2	Data processing	115
3.5.3	Matthews coefficient and solvent content	118
3.5.4	Structure determination	118
3.5.5	Structural refinement	122
3.6	The crystal structure of UL16	123
3.7	The crystal structure of MICB_{short}	124
3.8	The UL16-MICB_{short} interface	125
4.	DISCUSSION	128
4.1	The UL16-MICB complex - purification and crystallization obstacles	128

4.1.1	CV-1 cell produced UL16 is not suitable for crystallization	128
4.1.2	The UL16 dimer is likely formed by locally misfolded UL16 monomers	130
4.1.3	MICB binding sites are inaccessible in the UL16 dimer	132
4.1.4	Purification and crystallization of the UL16 _{mon} -MICB _{long} complex	133
4.1.5	Separation of UL16 _{dim} from UL16 _{mon} -MICB _{short} complex by Ni-NTA AC	134
4.1.6	Purification of the UL16 monomer	136
4.1.7	Surface plasmon resonance measurements	137
4.1.8	Purification and crystallization of the UL16 _{mon} -MICB _{short} complex	138
4.1.9	Purification and crystallization of the deglycosylated UL16 _{mon} -MICB _{short} complex	138
4.2	The UL16-MICB complex - structural insights into NKG2D ligand binding by UL16 and NKG2D	139
4.2.1	Comparison between the UL16-MICB and the NKG2D-MICA complex	139
4.2.2	The promiscuous binding mode of UL16	141
4.2.3	Key determinants of NKG2D-ligand binding to UL16	142
4.2.4	UL16 and the evolutionary diversification of NKG2D ligands	145
4.2.5	Accession numbers	146
5.	CONCLUSIONS	147
6.	ZUSAMMENFASSUNG	150
7.	REFERENCES	154
8.	PUBLICATIONS	161
9.	ABBREVIATIONS	162
10.	ACKNOWLEDGEMENTS	165
11.	ACADEMIC TEACHERS	166
12.	CURRICULUM VITAE	167
13.	LEBENS LAUF	168

1. Introduction

1.1 The human immune system

The human immune system (Figure 1-1) consists of specialized molecules, cells and organs that form a complex network of cellular and humoral defense mechanisms that protects the body against diseases caused by pathogens (bacteria, viruses, fungi and parasites) and harmful substances (toxins).

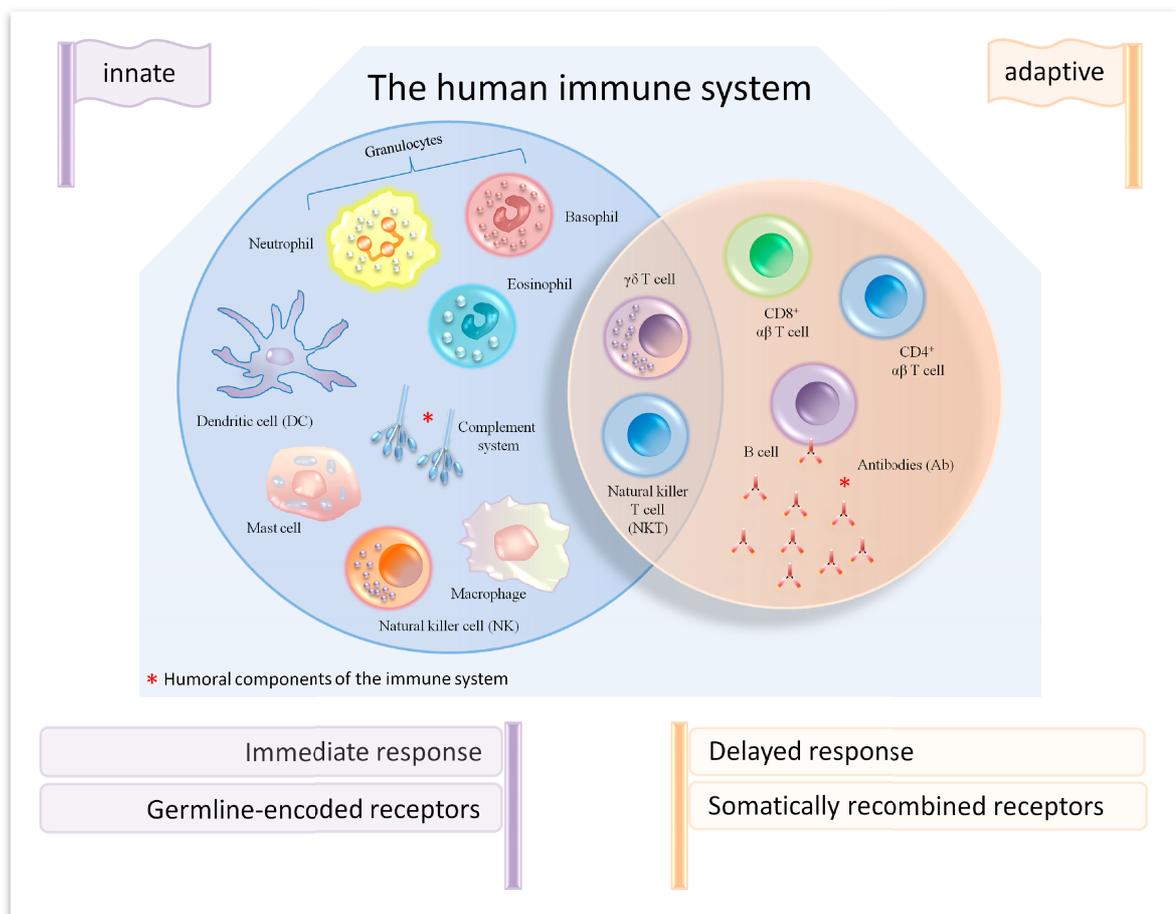


Figure 1-1. The human immune system. Shown are the cellular and humoral components (marked with a red asterisk) of both the innate (orange background) and adaptive (blue background) human immune system. Key features of each subsystem are indicated at the bottom of the panel. Figure modified from Dranoff, 2004¹.

Furthermore, based on its ability to distinguish between healthy and malignant host cells, the immune system also plays a significant role in the defense against cancer cells. In order to perform its protective functions, the immune system is equipped with a wide array of mechanisms that normally enable it to recognize, neutralize, and eventually eliminate immunological threats. The

recognition of pathogens or cancer cells is mediated through highly specific so-called immunoreceptors that engage pathogen- or cancer cell-specific antigens, which, by definition, are self-molecules (or parts thereof) that can stimulate an immune response. The human immune system is divided into an innate and an adaptive immune system (Figure 1-1). The innate immune system is characterized by (i) an immediate response after an immunological challenge, (ii) a constant set of germline-encoded immunoreceptors that recognize a limited number of specific antigens, and (iii) the absence of an immunological memory. The opposite is true for the adaptive immune system, which features (i') a delayed immune response against an immunological threat, (ii') a vast immunoreceptor repertoire, obtained by somatic recombination, that allows the recognition of an almost indefinite number of antigens, and (iii') the presence of immunological memory. Although both systems have their own characteristics and distinct arms of cellular and humoral defense mechanisms (Figure 1-1), it is the complex yet dynamic and powerful interplay of both systems that makes the immune system so effective.

1.1.1 Immunoreceptors employed by the innate immune system

The recognition of pathogens by components of the innate immune system is mediated by invariable, germline-encoded and constitutively expressed pattern recognition receptors (PRRs). PRRs detect evolutionary conserved pathogen-associated molecular patterns (PAMPs) that are common to many pathogens. This explains why a relatively small number of invariant PRRs can detect a large number of diverse pathogens. Furthermore, the constitutive expression of PRRs by a large population of innate immune cells explains why the innate immune system can essentially respond instantaneously to an immunological threat. An example for a PRR are the Toll-like receptors (TLR), which can for instance recognize lipopolysaccharides present on the outer membrane of Gram-negative bacteria or double-stranded (ds) RNA, which is typical for some virus families with ds RNA genomes. A different type of PRR is represented by the mannose-binding lectin (MBL), which recognizes a broad array of pathogens based on their carbohydrate surface patterns.

The NK cell receptors, which in addition to the PRRs also constitute a large group of innate immunoreceptors, are discussed in Section 1.2.1.

1.1.2 Immunoreceptors employed by the adaptive immune system

The adaptive immune system employs receptors that have been selected based on their high specificity for a particular antigen. The membrane-bound B cell receptors (BCRs) or their soluble derivatives, the antibodies, recognize antigens that are exposed on the pathogen's surface, while the membrane-bound T cell receptors (TCRs) recognize pathogen-derived peptides that are presented by MHC class I and/or II molecules on the surface of antigen-presenting cells (APCs). At any given time, a healthy human individual features billions of naive B and T cells, each of which is equipped with a unique receptor on its cellular surface that entails a unique specificity for a particular antigen^{2, 3}. Furthermore, each day millions of new B and T cells with new receptor specificities are generated, while at the same time old B and T cells are eliminated. In this way, the receptor repertoire is continuously turned over and can therefore easily adapt to new immunological challenges. This diversity in the B and T cell repertoire is achieved by a genetic mechanism termed somatic recombination during which V, (D) and J gene segments are randomly selected from corresponding gene pools, followed by their ordered assembly to produce a unique, functional immunoreceptor phenotype. During the subsequent process of clonal selection, only those naive B and T cell clones start to proliferate (clonal expansion) that were selected from the human body's pool of B and T cells based on their ability to bind to one of the pathogen's antigens via their highly specific BCRs and TCRs, respectively. This in turn explains the high antigen specificity that is so typical for B and T cell receptors^{4, 5}. The clonal expansion process takes time and is therefore the reason for the significantly longer time delay between immunologic challenge and immune response, when compared to the innate immune system.

1.2 Natural killer cells

Natural killer (NK) cells were first described in 1975 and named according to their ability to 'naturally' kill tumor cells without prior sensitization and in a major histocompatibility complex (MHC) molecule independent fashion⁶⁻⁸. Nowadays, NK cells are recognized as a population of large granular cytotoxic lymphocytes that (i) belong to the innate immune system, (ii) are characterized by the absence of the (T cell) surface marker CD3, and (iii) can be subdivided into two subpopulations depending on the abundance and presence of the surface markers CD56 and CD16 (the Fc γ III receptor), respectively⁹⁻¹¹. NK cells comprise 15 % of the overall lymphocytes and are of pivotal importance in the cellular immunosurveillance of viral pathogens and cancer cells¹²⁻¹⁶. They recognize virus infected or malignant host cells either indirectly via their Fc γ III receptors that bind to target cell-bound IgG antibodies and trigger a process known as antibody

dependent cellular cytotoxicity (ADCC), or directly via sets of activating and inhibitory NK cell receptors^{17, 18}. Upon activation, NK cells stimulate other immune cells through the release of cytokines or destroy their target cells through the release of cytotoxic granules filled with proteins called perforins and granzymes⁹ (Figure 1-2). The perforins create pores in the target cell membrane, while the simultaneously released granzymes diffuse through these pores inside the target cell where they trigger apoptotic cascades that lead to the self destruction of the virus infected or degenerated host cell. Whether an individual NK cell exhibits cytotoxic or immune stimulatory effects depends on its phenotype. CD56^{bright} CD16⁻ NK cells produce high amounts of cytokines but no perforins and consequently exhibit immunomodulatory functions, while the opposite is true for CD56^{dim} CD16⁺ NK cells, which are predominantly involved in NK cell mediated cytotoxicity^{9, 11}.

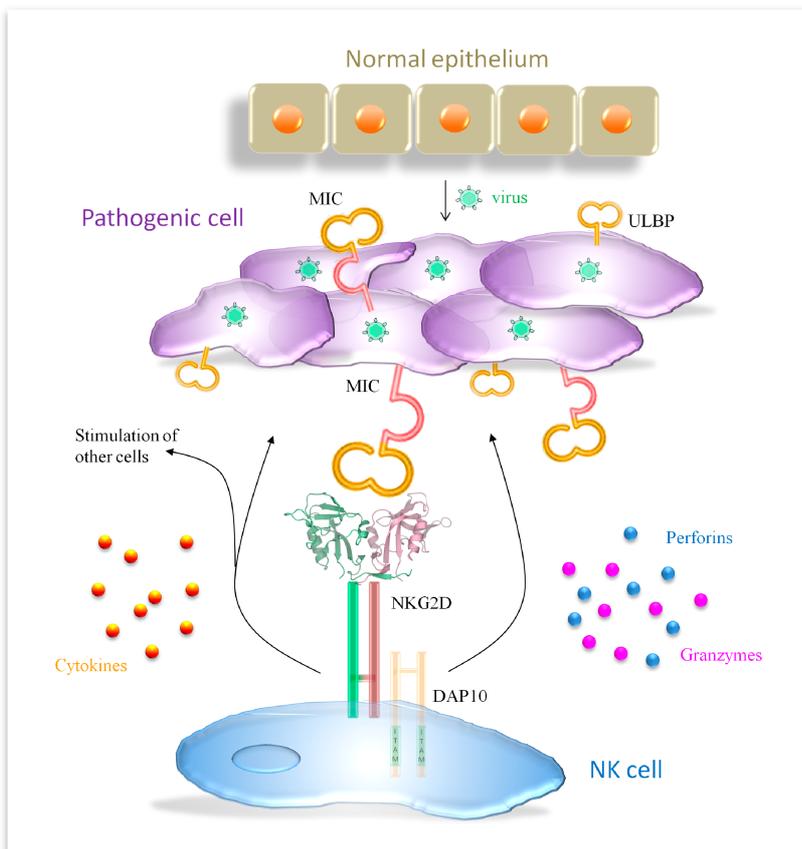


Figure 1-2. NK cell recognition. Shown is one possible way by which NK cells detect viral pathogens. Once infected with the virus, pathogenic host cells upregulate stress-inducible MHC class I-like self-ligands such as MIC and ULBP molecules on their cellular surface. These self-ligands can activate patrolling NK cells by engaging the the activating NKG2D immunoreceptor. Upon activation the NK cells release cytokines that stimulate other immune functions and/or kill the virus infected cells through the release of perforins and granzymes. The schematic representations of the MIC and ULBP molecules were modified from Eagle & Trowsdale, 2007¹⁹. The figure was modified from Cerwenka & Lanier, 2001¹². The displayed NKG2D structure is based on pdb code 1hyr²⁰.

1.2.1 NK cell receptors

NK cells as well as B and T cells all constitute subpopulations of human lymphocytes. However, in contrast to B and T cells, NK cells lack the ability of somatic recombination and therefore the ability to generate highly antigen-specific receptors. Instead NK cells mainly comprise large sets of germline-encoded activating and inhibitory receptors that, based on their molecular structures, either belong to the immunoglobulin superfamily (IgSF) or the superfamily of proteins containing a C-type lectin-like domain (CTLD), respectively^{17, 18, 21-24}. An additional feature by which the NK cell receptors differ from BCRs and TCRs (as well as from PRRs) is that most of them do not detect foreign antigens but instead monitor the abundance of particular activating and inhibitory self ligands on the surface of host cells. An increase in activating or a decrease in inhibitory self ligands upon viral infection or transformation of a host cell will result in the activation of NK cells and the subsequent NK cell-mediated killing of the abnormal cell.

A) C-type lectin-like proteins

CTLD containing proteins (CTLDcps) were first discovered as carbohydrate recognition domains (CRDs) in proteins that bind carbohydrates (lectins) in a Ca^{2+} (C-type) dependent fashion²⁵⁻²⁹. Subsequently, many CTLDcps that were deficient in both their carbohydrate and Ca^{2+} binding capabilities were identified. The corresponding domains instead had evolved to recognize other ligands such as proteins, lipids and even inorganic compounds²⁷. A prominent group of non-carbohydrate binding CTLDcps is represented by the protein-binding, homo- and heterodimeric C-type lectin-like NK cell receptors (CTLRs) which, based on their overall domain architecture, constitute a distinct subfamily of vertebrate CTLDcps^{26, 27, 30}. The CTLR protomers are type II transmembrane proteins and consist of four characteristic regions: an N-terminal cytoplasmic tail, a transmembrane region, a short stalk region and a C-terminal CTLD (Figure 1-4).

B) Immunoglobulins

The IgSF is comprised of proteins that feature a common Ig-like fold. Among them are the classical immunoglobulins, or Igs, (antibodies, and their membrane-bound forms, the B-cell receptors) that gave the fold and the superfamily their names. Over time, many proteins were identified that were not classical Igs but nonetheless were partly or entirely composed of Ig-like domains³¹. Nowadays the heterogenic group of IgSF proteins includes enzymes, chaperones, cytokine receptors, adhesion molecules, muscle proteins, viral coat proteins, proteins relevant for the nervous system, proteins relevant for the immune system (classical Igs, T-cell receptors, MHC

molecules a.s.o.) and many others³¹⁻³³. The term 'Ig-fold' is often used instead of 'Ig-like fold' when one refers to the fold of classical immunoglobulins.

Ig-like domains feature seven to nine antiparallel β -strands that assemble to form two distinct β -sheets. The two sheets pack against each other, forming a β -sandwich. Although Ig-like domains all have the same topology, their domain size varies considerably, owing to the varying lengths of their loops and additional β -strands present in some variants^{31, 34-36}. Nevertheless, typical characteristics for all Ig-like domains can be defined that include (i) their extreme sequence heterogeneity that makes it difficult to define a sequence signature of this fold and therefore often hampers the prediction of Ig-like folds based on a protein's primary sequence^{33, 35}, (ii) their common structural core comprised of β -strands b/e (sheet 1) and c/f (sheet 2)³⁵ and (iii) an additional 3 - 5 β -strands that form the positional and conformationally flexible strands at the edges of the sheets³⁵. Based on the number and location of the edge strands (i. e. differences in the molecular shape) as well as sequence criteria, at least 5 topological subtypes of Ig-like domains, the C1, C2, V, H and I types, can be distinguished from one another³¹⁻³³. A selection of these sybtypes is presented schematically in Figure 1-3³⁵. Classical Ig domains are either of the C1 (constant) or V (variable) type, which in addition to the abovementioned criteria can also be distinguished by the number of residues that lie between two b and f-strand cysteines that in many Ig-like proteins form a disulfide bridge. A residue number of 65 - 75 indicates a V-type, while 55 - 65 residues indicate the presence of a C-type domain³¹. The mentioned disulfide bridge between strands b and f was originally considered as one of the cornerstones of the Ig-like fold, and was thought to be essential for its stability³⁵. Over time, however, many more structures of proteins containing Ig-like domains were solved, and these revealed that the number and the locations of the disulfide bridges are quite variable, with some Ig-like domains having no disulfide bridge at all³⁵.

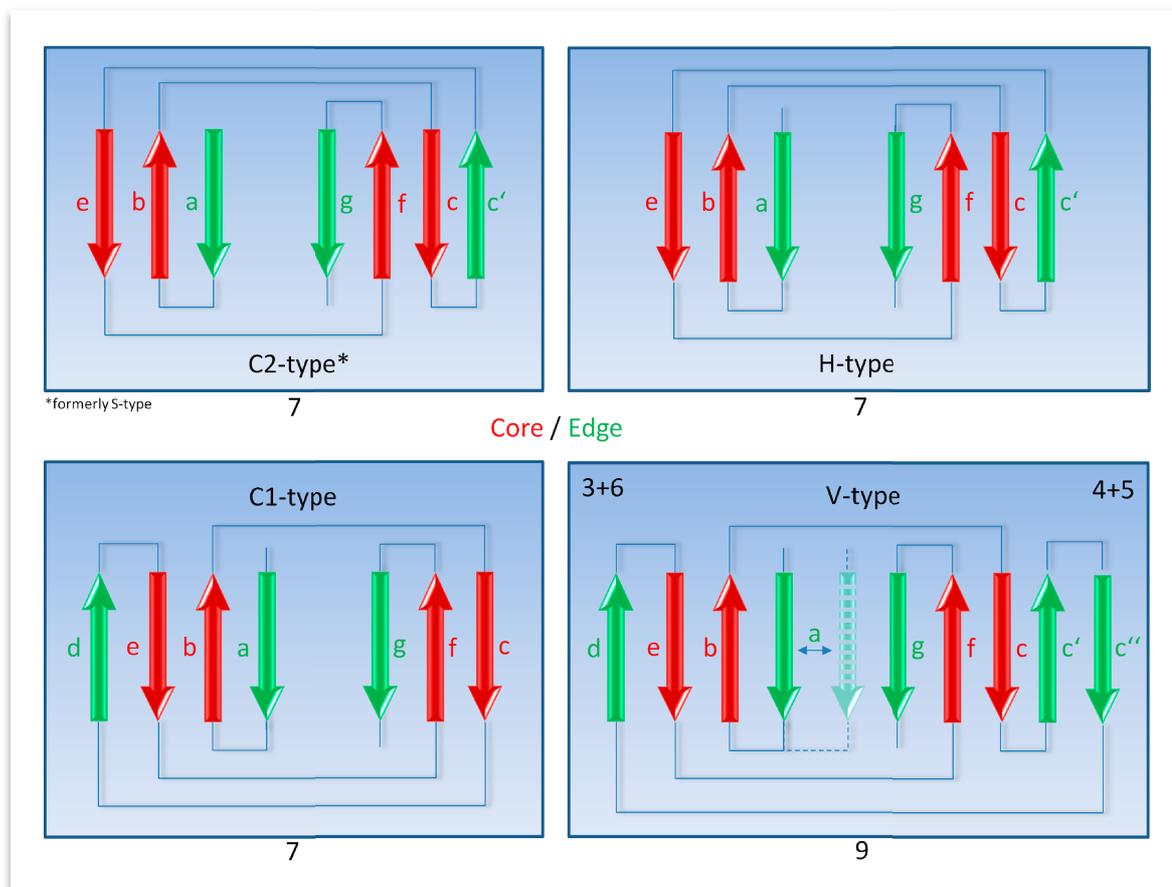


Figure 1-3. Ig-like domain subtypes. The four panels show the 2D β -strand topologies as observed in the C1 (C – constant), C2, H (H – hybrid) and V (V – variable) Ig-like subtypes. The core and edge β -strands of the Ig-like fold are represented by red and green arrows, respectively, and are labeled on the left (sheet 1) or the right (sheet 2) handside. The number below each panel indicates the number of β -strands that is typical for each particular subtype. Depending on the position of β -strand *a* in the V-type immunoglobulins, two subspecies that are either composed of 3+6 or 4+5 β -strands can be distinguished from each other. The figure was drawn after Bork et al., 1994³⁵ and modified based on Halaby *et al.*, 1999³³.

C) Gene loci of NK cell receptors

In humans, the genes encoding NK cell receptors of the IgSF or CTLD are genetically clustered³⁷. The CTLRs are encoded in the natural killer gene complex (NKC)^{17, 21} on chromosome 12, while Ig-like receptors such as the killer Ig-like receptors (KIR), the Ig-like transcripts (ILT) that are also called leukocyte Ig-like receptors (LIR) and the natural cytotoxicity receptors (NCR) are encoded in the leukocyte receptor complex (LRC) on chromosome 19^{21, 24}.

1.2.1.1 Inhibitory NK cell receptors

An immune response by cytotoxic T lymphocytes (CTL) is critically dependent on the presentation of viral or tumor cell specific antigens by MHC class I molecules on the surface of virally infected or malignant host cells, respectively. Therefore, in order to subvert an antiviral or anti-tumor immune response viruses and tumor cells have devised strategies to downregulate the peptide presenting MHC class I self proteins^{38, 39}. However, according to the ‘missing-self hypothesis’ that was first formulated by Ljunggren and Kärre, a hampered MHC class I surface expression results in a decreased engagement of MHC class I-specific inhibitory NK cell surface receptors, and consequently in a loss of inhibitory signals that, in the receptor’s unengaged state, are usually provided by tyrosine phosphorylated immunoreceptor tyrosine-based inhibitory motifs (ITIMs) located in the receptor’s cytoplasmic tail⁴⁰⁻⁴³. The loss of inhibitory signals may in turn lead to the activation of the engaging NK cell, and therefore the elimination of the abnormal cell. However, NK cell activity is not solely dependent on receptors that inhibit NK cell activation, but rather tightly regulated by the integration of signals from a multitude of both inhibitory and activating NK cell receptors^{44, 45}.

1.2.1.2 Activating NK cell receptors

In contrast to inhibitory NK cell receptors, activating NK cell receptors do not monitor the downregulation of self-ligands but instead their upregulation. According to the ‘induced-self hypothesis’, an increased host cell surface expression of stress-inducible self-ligands results in an increased engagement of activating NK cell surface receptors. Different from the inhibitory receptors the signalling of activating receptors is often mediated by adaptor molecules. These normally pass on the signal via tyrosine phosphorylated immunoreceptor tyrosine-based activation motifs (ITAMs) that are, similar to the ITIMs of inhibitory receptors, located in the proteins cytoplasmic tail^{41-44, 46}.

1.2.2 The activating NK cell receptor NKG2D

The immunoreceptor NKG2D (natural-killer group 2, member D) is a powerful activating, homodimeric CTLR that mediates NK surveillance of stressed cells such as virus infected or degenerated host cells (Figure 1-4)^{19, 43, 47, 48}. However, in addition to its original classification as a NK-cell receptor, NKG2D is also expressed by distinct subsets of T cells such as CD8⁺ αβ and γδ T cells as well as natural killer T cells, NKT^{47, 49}. *Ex vivo* studies with human NK cells showed that

the surface expression of NKG2D is modulated upon exposure to certain types of cytokines. While interleukins (IL) 2, 10, 12, 15 and IFN- α stimulate the upregulation of NKG2D^{42, 50-52}, the opposite was observed for the cytokines TGF- β and IL21^{53, 54}. In humans, NKG2D transduces signals into the interior of the cell by associating with DAP10 adaptors (DAP12 in mice⁴⁷) that pass on the signal via a phosphatidylinositol-3-kinase (PI-3K) dependent pathway (Figure 1-4)^{41, 42, 47, 55}. However, the quality of the signal is different. While NKG2D transmits activation signals in case of NK cells, the signal transmitted in case of the T cell subsets has a costimulatory character^{56, 57}.

As is typical for human CTLRs, NKG2D is encoded in the NKC. Together with its distant homologs NKG2A, C and E, NKG2D comprises the NKG2 receptor family of CTLRs. Although the NKG2 receptor family members are genetically clustered in the NKC and structurally related due to their C-type lectin-like fold, NKG2D stands clearly out among them as a distinct receptor for the following reasons^{47, 58}: (i) it is only distantly related to the other NKG2 receptors, (ii) it is a homodimeric protein whereas the other members are CD94-NKG2 heterodimers, (iii) it specifically binds to MIC (MHC class I chain related molecule) and ULBP (UL16 binding protein) ligands (Figure 1-4) instead of human leukocyte antigen E (HLA-E), and (iv) in contrast to many other immunoreceptors, NKG2D features a fairly promiscuous ligand binding mode, allowing it to recognize an array of at least eight diverse MHC class I-like ligands that are upregulated in a cell stress-associated manner and in this way allow for the immunorecognition of infected or transformed cells (for details see Section 1.2.3)¹⁹. Two rather different explanations of the mechanism behind this promiscuous ligand binding mode exist^{47, 59-61}. One group claims that NKG2D recognizes its diverse array of ligands by an induced fit mechanism that requires NKG2D to undergo certain conformational adjustments, while another group argues for a rigid adaptation mechanism⁵⁹⁻⁶¹. The latter mechanism is based on a rigid NKG2D binding site that consists of a set of predominantly hydrophobic residues, which can adjust their spatial positions in order to accommodate the largely different interface amino acid residues provided by the set of diverse NKG2D ligands^{59, 60}. The question arises what the biologic meaning of this remarkable redundancy of NKG2D ligands might be? A common hypothesis addressing this question proposes that the NKG2D ligand diversity may be explained as a host response to the selective pressure exerted by viral immunoevasins (see Section 1.2.3)^{14, 19}. However, it was also argued that cancer cells may be an additional driving force for the observed NKG2D ligand diversification¹⁹.

1.2.3 Ligands of the activating NK cell receptor NKG2D

As mentioned above one of the most intriguing characteristics of the human NKG2D immunoreceptor is its promiscuous ligand binding mode allowing it to engage eight MHC class I-

like ligands (Figure 1-4)^{14, 19, 43, 47}. These ligands belong to either of two diverse protein families, the MIC and ULBP families, that are upregulated upon cellular stress, as caused for example by viral infections, heat shock, DNA damage or ultraviolet (UV) radiation (Figures 1-2 and 1-4)^{16, 62}.

The MIC family includes MICA and MICB, both of which comprise an ectodomain, a transmembrane domain and a cytoplasmic tail^{63, 64}. The ectodomain is composed of a membrane distal MHC class I-like $\alpha 1\alpha 2$ platform domain that is connected to a membrane proximal MHC class I-like $\alpha 3$ Ig-like domain^{63, 64}. The $\alpha 1\alpha 2$ and $\alpha 3$ domains are linked together via a flexible hinge region that leads to an elevated interdomain flexibility (Figure 1-4)^{63, 64}. In contrast to classical MHC class I molecules, MICs do not associate with β_2 -microglobulin ($\beta 2m$) or bind peptides^{20, 61, 63, 64}.

The ULBP family, as the name implies, was originally discovered based on the ability of some family members to engage the human cytomegalovirus (HCMV) protein UL16⁶⁵. It currently consists of ULBP members 1 to 6 that, based on their relationship to the retinoic acid early inducible 1 (Rae1) proteins in mice, are also sometimes referred to as retinoic acid early transcript (RAET) proteins (Figure 1-4)^{19, 43, 66, 67}. The ULBP ectodomains, like those of the MICs, are comprised of an MHC class I-like $\alpha 1\alpha 2$ platform domain^{47, 61}. In contrast to MICA and MICB, however, the ULBPs lack $\alpha 3$ Ig-like domains (Figure 1-4). Other aspects in which ULBPs show similarities to or differ from the MICs concern the way in which particular ULBP ectodomains are linked to the cellular membrane, and their allelic diversity (i. e. their genetic polymorphism)^{19, 47, 57, 68}. Like the MICs, ULBPs 4 and 5 represent type I transmembrane proteins and are inserted into the membrane by their particular transmembrane domains (Figure 1-4). ULBPs 1-3 and ULBP6, on the other hand, are anchored into the membrane via fatty acids that are part of the glycolipid glycosylphosphatidylinositol (GPI)⁶⁹. These so-called GPI-anchors represent a covalent posttranslational modification at the C-terminus of a protein that allows the protein to become membrane-associated even in the absence of a transmembrane domain. In terms of polymorphism, MICA and MICB can be considered highly polymorphic as they give rise to 70 and 22 allelic variants (www.ncbi.nlm.nih.gov/gv/mhc), respectively, whereas only very few allelic variants have so far been described for the ULBPs⁷⁰.

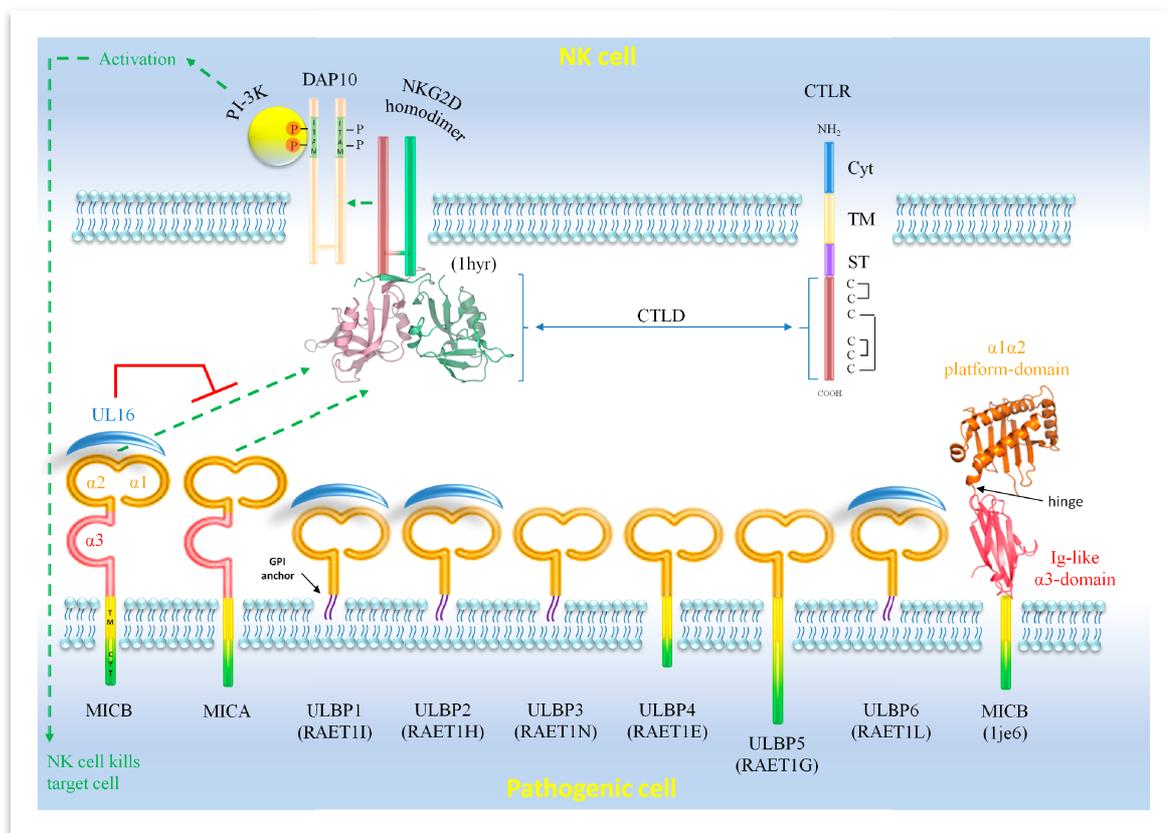


Figure 1-4. Ligands of the activating NK cell immunoreceptor NKG2D. Pathogenic cell membrane (bottom): Shown are the schematic representations of the eight, currently known MHC class I-like NKG2D ligands. The three-dimensional structure of MICB on the right hand side illustrates the MHC class I-like fold. Also indicated is the selective engagement of the NKG2D ligands by the HCMV immunoevasin UL16, which is drawn on top of its NKG2D ligand targets as a blue half-moon. NK cell membrane (top): Presented on the left is a schematic of the activating NK cell immunoreceptor NKG2D and its associated adaptor molecule DAP10. The C-type lectin-like domains (CTLDs) of the NKG2D protomers (purple and green) are displayed as three-dimensional structures. The schematic on the right shows the four regions that typically constitute the C-type lectin-like receptors (CTLR) encoded in the natural killer gene complex (NKC)¹⁷: Cyt – cytoplasmic tail, TM – transmembrane domain, ST – stalk region and CTLD. The CTLD of the CTLRs is further characterized by six invariant cysteines that form intra-chain disulfide bonds. ‘C’, cysteine. The NKG2D structure is based on pdb code 1hyr²⁰, while the MICB structure is based on pdb code 1je6⁶⁴. The figure was modified and updated from Eagle & Trowsdale, 2007¹⁹.

Many studies have shown that even proteins with little similarity in their primary amino acid sequences may nonetheless give rise to protein families with the same structural fold (or parts thereof) and as a consequence of this structural similarity may eventually exhibit very similar functions. A prominent example of this type can be found in the IgSF. However, this is also true for the ULBP and MIC families. Both families exhibit surprisingly similar structures even though the MICs share as little as 21 - 29 % identical residues in their $\alpha 1\alpha 2$ platform domain with the ULBPs. The sequence homology within each group is significantly higher: the MICs share 83 % identical residues in their $\alpha 1\alpha 2$ platform domain, while the ULBPs share 38 – 59 % amino acid sequence

identity in the same region. Based on these data one may wonder how tightly the MICs and ULBPs interact with NKG2D. A parameter that indicates the binding strength or complex stability, is the dissociation constant (K_D) of the interaction that is also commonly referred to as the affinity of the complex. Although all MICs and ULBPs were shown in functional studies to bind NKG2D, affinity data are unfortunately only available for MICA (0.9 μ M), MICB (0.8 μ M) and ULBP1 (1.1 μ M)^{57, 60}. Nevertheless, this data suggests that NKG2D is able to bind both NKG2D ligand families with similar affinity.

1.3 Herpes viruses

The order *Herpesvirales* is comprised of the virus families *Alloherpesviridae* (fish and frog viruses) *Malacoherpesviridae* (a bivalve virus) and *Herpesviridae* (mammal, bird and reptile viruses)⁷¹. The family of *Herpesviridae* consists again of three subfamilies, the α , β and γ *Herpesvirinae*, each of which is in turn composed of four genera. Herpesvirus species belonging to eight out of these twelve genera have been shown to cause illness in humans, and are named human herpes virus (HHV) 1-8^{71, 72}.

Herpes viruses are 200 - 250 nm in diameter and contain a linear dsDNA genome of 125 - 245 kbp⁷³. Based on these features, herpes viruses are among the largest and most complex viruses currently known. Additional characteristics of herpes viruses are (i) an icosahedral nucleocapsid that is 125 nm in diameter and contains the dsDNA, (ii) the tegument, which consists of amorphous virus-coded proteins and encloses the capsid, and (iii) a lipid bilayer envelope on the outside that contains several viral glycoproteins^{73, 74}.

All herpes viruses replicate in the nucleus of the host cell⁷⁵. Upon infection the herpesvirus undergoes a complete replication cycle that leads to the production of functional viral particles and entails the lysis of host cells and then establishes a latent life-long infection, the latter being characterized by the presence of replication-competent viral genomes in the absence of infectious viral particles^{74, 76}. However, latency phases alternate periodically with rather short acute phases during which the virus reenters its lytic replication cycle⁷⁶.

1.3.1 Human cytomegalovirus

The ubiquitous human cytomegalovirus (subfamily *Betaherpesvirinae*, genus *Cytomegalovirus*) or human herpesvirus 5 (HHV5) is the largest virus in the family of *herpesviridae*^{71, 72, 75}. In case of the strain AD169 the HCMV genome is comprised of 230 kbp and encodes 208 functional open reading frames (ORF)⁷⁷. Two adjacent unique long (UL) and unique short (US) ORF sequence

clusters, which are separated by other ORFs referred to as internal repeat sequences (ILS and IRS), form the center of the linear HCMV genome. This center is flanked on both sides by ORFs named terminal repeat sequences (TLS and TRS)⁷⁷. Depending on the time of expression these HCMV genes can also be classified in early and late genes⁷⁵. The HCMV life cycle is illustrated in Ref.⁷⁵. After infection, HCMV usually causes lifelong asymptomatic (latent) infections in healthy individuals but can also lead to severe illness in immunocompromised individuals and very young children⁷⁴. HCMV is usually transmitted via saliva, body excretions or sexual contact but can also infect embryos or fetuses by transmission through the mother's placenta (intrauterine transmission)⁷⁴. The latter type of infection typically causes the cytomegalic inclusion disease (CID) that is often characterized by acute manifestations such as pneumonia or hepatitis and (in case of survival) long-term effects such as mental retardation or blindness⁷⁴. The mortality rates range from 5 % in case of congenital and 10 % in case of transplantation-associated HCMV infections⁷⁴.

1.3.2 HCMV immune evasion strategies

A feature that distinguishes HCMV from other viruses is its unique multitude of strategies to evade both innate and adaptive mechanisms of human immunosurveillance. Accordingly, HCMV is also referred to as a paradigm of viral immune evasion⁷⁸. A small selection of these immune-modulatory strategies is presented in the following paragraphs and in Figure 1-5^{14, 39, 78-80}.

1.3.2.1 Interference with MHC class I peptide presentation

The detection and killing of a virus infected host cell by cytotoxic T cells (CTL) is critically dependent on the presentation of viral peptides by MHC class I molecules on the surface of the infected cell. HCMV efficiently subverts CTL responses by means of several US and UL immunoevasins, which interfere with the intracellular processes that lead to the generation of functional MHC class I-peptide complexes^{14, 78}. The immunoevasin UL83 interferes with the entry of viral proteins in the proteasome, thereby hampering the production of viral peptides (Figure 1-5). An additional immunoevasin, US6, blocks the 'transporter associated with antigen processing' (TAP) that is required to transport peptides generated by the proteasome from the cytosol into the ER, the cellular compartment where the obtained viral peptides are mounted onto the MHC class I molecules (Figure 1-5). Finally, a number of Ig-like glycoproteins named US2, US3, and US11 can directly interact with the MHC class I molecules and lead to their retention in the ER (US3) or facilitate the degradation of MHC class I (US2, US11) molecules by mediating their dislocation to

the cytoplasm (Figure 1-5)⁷⁸. Although this paragraph has been about evasive maneuvers of HCMV affecting MHC class I molecules it needs to be pointed out that also the peptide presentation by MHC class II molecules, which is pivotal in eliciting immune responses by T helper cells, is compromised⁸¹.

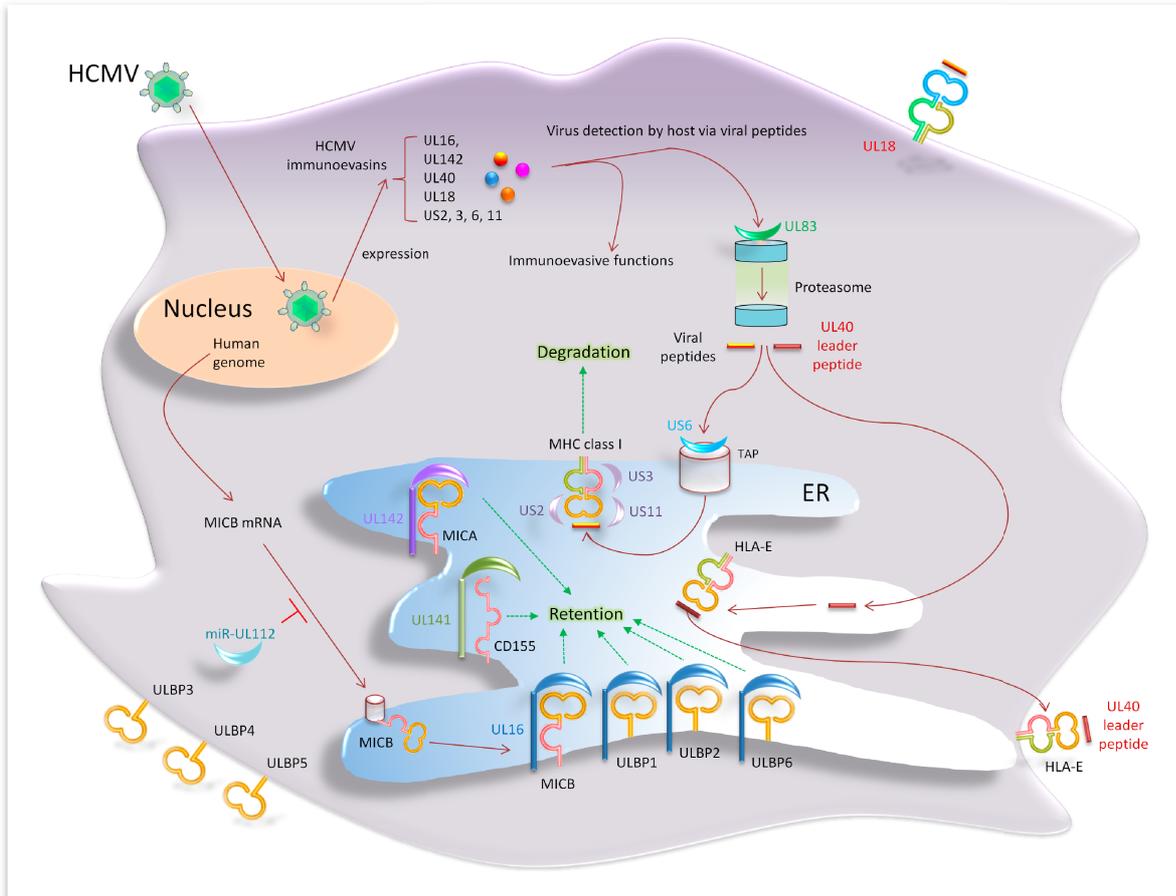


Figure 1-5. Mechanisms of immune evasion by HCMV. Presented is a selection of the strategies employed by HCMV to subvert antiviral immune responses mediated by T cells and NK cells. T cell responses are hampered by HCMV immunoevasins US2, US3, US6, US11 and UL83 that interfere with the presentation of viral peptides by MHC class I molecules. NK cell responses on the other hand are impeded by immunoevasins UL18, UL40, UL16, UL142 and UL141. UL18 and the HLA-E-UL40 leader peptide complex both serve as ligands for inhibitory NK cell receptors, while immunoevasins UL16, UL142, UL141 and the miRNA of UL112 downmodulate the surface expression of activating NK cell receptor ligands by retaining those ligands in the ER (UL16, UL112 and UL141) or by sequestering their mRNA (miR-UL112), respectively. UL16 selectively targets the NKG2D ligands MICB, ULBP1, ULBP2 and ULBP6, while other NKG2D ligands such as ULBP3, ULBP4 and ULBP5 can escape UL16 engagement. The NKG2D ligand MICA, which can also escape UL16, is retained by yet another immunoevasin, UL142. The schematic representations of the MIC and ULBP molecules were modified from Eagle & Trowsdale, 2007¹⁹. The figure was modified from Lanier, 2008¹⁴.

1.3.2.2 Molecular decoys of inhibitory NK cell receptor ligands

As outlined above, the downmodulation of surface MHC class I molecules will result in a decreased engagement of NK cell inhibitory receptors and, as a consequence, to the activation of NK cells. To avoid this, HCMV employs the immunoevasin UL18, a structural homolog of classical MHC class I proteins that, like the majority of other structurally defined virulence factors that are homologous to human proteins, was most likely acquired through hijacking of MHC class I encoding host genes. As is typical for MHC class I molecules, UL18 also has the ability to bind peptides and associate with β 2m. In HCMV immune evasion UL18 is believed to serve as a molecular decoy for the inhibitory NK cell receptor ILT-2 (also known as LIR-1) that specifically recognizes an epitope in the α 3 domain of most MHC class I molecules (Figure 1-5)¹⁴. Thus, UL18 serves as a means to conceal the HCMV caused downregulation of MHC class I molecules from NK cells. This function of UL18 as an ILT-2 decoy is supported by the crystal structure of the UL18- β 2m-ILT-2 complex and affinity measurements that place the affinity between UL18 and ILT-2 in the nanomolar range, which is significantly higher than observed for the interactions between ILT-2 and its native ligands^{82, 83}.

In addition to the monitoring of the MHC class I protein surface expression through direct interactions of the MHC class I molecules with inhibitory NK cell receptors, another, more indirect, monitoring strategy involves the non-classical MHC class I protein HLA-E, which has the special ability to bind a set of peptides that are derived from signal sequences of MHC class I molecules (Figure 1-5)^{39, 78}. The formation of HLA-E-peptide complexes is a prerequisite for the transport of HLA-E to and its expression on the cellular surface, where it suppresses the activation of NK cells by engaging the inhibitory C-type lectin-like receptor CD94-NKG2A^{39, 78}. The MHC class I derived peptides will only be mounted onto HLA-E after their transport into the ER by TAP. Unfortunately, upon infection of a cell with HCMV, the latter is blocked by US6³⁹. It follows that the inability to mount HLA-E with peptides will result in a reduced surface presentation of HLA-E leading to a decreased engagement of CD94-NKG2D and therefore the activation of NK cells⁷⁹. To avert such an NK cell response, HCMV encodes the immunoevasin UL40, which has a signal peptide (VMAPRTLIL) that is identical to the signal peptides typical for classical HLA-C proteins^{78, 79}. Moreover, the UL40 peptide can be mounted onto HLA-E after its transport into the ER in a TAP-independent fashion (Figure 1-5)⁷⁸. In this manner, HCMV is able to maintain high surface expression levels of HLA-E when MHC class I derived peptides are missing in the ER due to US6 activity.

1.3.2.3 Downmodulation of activating NK cell receptor ligands

Several diverse and independent immune evasive strategies employed by HCMV aim to avert an NKG2D-mediated immunorecognition, which emphasizes the importance of this particular surveillance system. As described above, the activating immunoreceptor NKG2D stimulates anti-viral and anti-tumor effector responses of human cytotoxic lymphocytes such as NK cells when engaged by one of its eight stress-inducible MHC class I-like ligands. One of the most well-known and extensively studied strategies of HCMV to subvert NKG2D-mediated immune responses is mediated by the HCMV glycoprotein UL16, which counteracts NKG2D-mediated antiviral immune responses by intracellular retention of the distantly related NKG2D ligands MICB, ULBP1, ULBP2 and ULBP6 (RAET1L) in the ER and Golgi network compartments (Figures 1-4 and 1-5⁸⁴⁻⁸⁷). However, other NKG2D ligands such as the MICB-homolog MICA, the ULBP2-homolog ULBP5 (RAET1G) and other ULBPs such as ULBP3 and ULBP4 are not targeted by UL16^{65, 88-90}. Currently, a common hypothesis addressing the remarkable redundancy of NKG2D ligands proposes that NKG2D ligand diversification may be explained as a host response to the selective pressure exerted by viral immunoevasins^{14, 19}. The importance of the NKG2D-mediated immunosurveillance of HCMV infections was further emphasized when other immunoevasins such as UL142 and the microRNA (miRNA) of the UL112 gene (hcmv-miR-UL112) were discovered (Figure 1-5)^{91, 92}. UL142 retains full-length MICA alleles in the cis-Golgi apparatus and in this way complements the function of UL16 to prevent the surface expression of NKG2D ligands^{91, 93}. Interestingly, UL142 is unable to downregulate MICA*008, a common MICA allele that is characterized by a truncated transmembrane region and the absence of a cytoplasmic tail^{91, 93}. This result suggests that MICA*008 represents an evolutionarily selected MICA escape variant that can bypass UL142 retention and therefore gives additional support to the hypothesis that the diversification of NKG2D ligands has been driven by viral immunoevasins. MicroRNAs are small noncoding RNAs that allow the host to posttranscriptionally regulate the expression of particular genes⁹². In order to perform this function, specific miRNAs bind to their corresponding mRNAs and in this way label them for degradation. Recent reports indicated that viruses can also express miRNAs⁹⁴⁻⁹⁶. In this respect the HCMV encoded UL112 miRNA was shown to specifically target MICB mRNAs during viral infection therefore leading to a decreased surface expression of MICB and subsequently to a decreased engagement of NKG2D and activation of NK cells⁹².

Another activating NK cell receptor ligand that is targeted by HCMV is CD155. CD155, also known as the poliovirus receptor (PVR), is a cellular surface protein involved in intercellular interactions⁷⁹. Under normal conditions these interactions prevent the exposure of CD155 to the environment. It is assumed that these intercellular interactions are at least partially disrupted upon

viral infection, which entails the exposure of CD155 and therefore its accessibility by patrolling NK cells⁷⁹. These recognize CD155 via their activating receptors DNAM-1 (CD226) and CD96 (TACTILE) and in this way become activated^{78,79}. The immunoevasin UL141 impedes this kind of detection as it accumulates in the ER and hinders CD155 to reach the cellular surface (Figure 1-5)⁷⁸.

1.3.2.4 The UL16 protein

UL16 was first described in 1992 as a HCMV encoded protein that is dispensable for the growth of human fibroblasts⁹⁷. UL16's immunomodulatory function (see Section 1.3.2.3) was reported nine years later in context with a new family of human NKG2D ligands, the ULBPs⁶⁵. UL16 is a type 1 transmembrane glycoprotein that was shown to be expressed by HCMV infected fibroblasts but to be nonessential for HCMV replication⁹⁷. The UL16 protein consist of 230 amino acids (aa) that correspond to a molecular weight of 26.1 kDa and can be divided into four distinct sections (Figure 1-6): a signal peptide (26 aa), a large luminal domain (164 aa) with eight potential N-glycosylation sites that indicates heavy glycosylation, a transmembrane domain (17 aa) and a cytoplasmic tail (23 aa) with a YQRL motif that could explain why UL16 preferentially accumulates in the ER and Golgi network⁸⁴⁻⁸⁶. Additional reports suggested that UL16 can also traffic through the plasma membrane and the nuclear envelope^{98, 99}. The structure of the UL16 luminal domain has so far remained elusive as no structural homologs could be predicted from UL16's primary amino acid sequence and because a crystal structure of UL16 or a UL16-NKG2D ligand complex has yet to be determined.

1	MERRRGTVPL GWVFFVLCLS ASSSCAVDLG SKSSNETCRL NVTELASIHP	50
51	GETWTLHGMC ISICYE NVT EDEIIGVAFT WQHNSVVDL WLYQNDTVIR	100
101	NPSDITTNIL QDGLKMRTVP VTKLYTSRMV TRLVGRYDC LRCENGTTKI	150
151	IERLYVRLGS LYPRPPGSL AKHPSVSADE ELSATLARDI VLVSAILLFF	200
201	FLALRIPQR LQRLRIRLP HRYQRLRTED	230

Figure 1-6. UL16 full length sequence. Shown is the full length amino acid sequence of UL16 as encoded by HCMV strain AD169 prior to signal peptide cleavage in the ER. Indicated by the four different background colors are the distinct regions of the UL16 protein (see also Figure 2-4A): ER signal peptide (SP) – bright orange, ER/Golgi-luminal domain – blue; transmembrane domain (TM) – yellow; cytoplasmic tail (Cyt). Cysteine residues are colored in pink. Potential N-glycosylation sites are colored in yellow. ER, Endoplasmic reticulum.

1.4 Aims

The primary objectives of this work were (i) to elucidate the molecular basis for the promiscuous but highly selective binding of the HCMV immunoevasin UL16 to diverse NKG2D immunoreceptor ligands, (ii) to obtain insights into the molecular adjustments of NKG2D ligands that escape UL16 engagement, and (iii) to address the question whether selective pressure exerted by UL16 may have contributed to the diversification of NKG2D ligands. In order to achieve these goals, structural and functional analyses of the protein-protein interaction between UL16 and the NKG2D ligand MICB were performed using X-ray crystallography and surface plasmon resonance (SPR).

In detail, the experiments presented in this thesis were designed to answer the following questions:

- ⇒ **What is the three-dimensional structure of UL16?** Why can the UL16 structure not be predicted from its primary amino acid sequence? Is this because UL16 exhibits a new type of structural fold or because there is too little amino acid sequence identity between UL16 and its structural homologs?
- ⇒ **What are the structural features allowing UL16 to (partially) recapitulate the promiscuous engagement of highly diverse ligands by NKG2D?** Do UL16 and NKG2D (despite the lack of any sequence similarity between them) share certain structural elements to engage similar epitopes on NKG2D ligands, or do they exhibit completely different binding modes?
- ⇒ **What are the binding parameters (kinetic rate constants and affinities) of the individual UL16-NKG2D ligand interactions?**
- ⇒ **What is the structural basis for the binding of UL16 to some diverse NKG2D ligands (MICB, ULBP1, ULBP2) and for the inability to bind other NKG2D ligands (MICA, ULBP3, RAET1G) that are highly related to some of the former ligands?** How are the latter ligands able to escape from UL16 binding without compromising NKG2D binding?
- ⇒ **Can structural data provide support for the hypothesis that the diversification of NKG2D ligands resulted from selective pressure exerted by viral immunoevasins?**

2. Materials and Methods

2.1 Materials

2.1.1 Hardware

2.1.1.1 General purpose equipment

Pipettes Pipetman (2 µl - 1 ml)	Gilson, Middleton, USA
Ultrapure water Biocell	Millipore, Schwabach, SUI
Autoclave VX-95	Systec, Wettenberg, GER
Scale 323S-OCE	Sartorius, Göttingen, GER
Scale 4202S-OCE	Sartorius, Göttingen, GER
Scale Genius	Sartorius, Göttingen, GER
Centrifuge 5414 D	Eppendorf, Wesseling-Bertzdorf, GER
Centrifuge Multifuge 1L-R	Kendro, Langenselbold, GER
Icemaker AF 80	Scotsman, Milan, ITA
Microwave NN-E203 WB	Panasonic, Hamburg, GER
pH meter PB-11	Sartorius, Göttingen, GER
Transilluminator Universal Hood II	Bio-Rad, Munich, GER
Transsonic T 460 (ultrasonic water bath)	Elma, Singen, GER

2.1.1.2 Molecular biology

Gel Doc XR imaging system	Bio-Rad, Munich, GER
PCR iCycler	Bio-Rad, Munich, GER
Photometer NanoDrop ND-1000	Thermo Scientific, Waltham, USA
Photometer SmartSpec Plus	Bio-Rad, Munich, GER
Quartz cuvette (75 µl)	Helma, Jena, GER
Mini Sub-Cell GT	Bio-Rad, Munich, GER

2.1.1.3 X-ray crystallography

24 well crystallization plates	Hampton research, Aliso Viejo, USA
22 mm siliconized cover slides	Hampton research, Aliso Viejo, USA
96 well Intelli-Plates Art	Robbins, Sunnyvale, USA
Sealing for 96 well plates	HJ-Bioanalytik, Mönchengladbach, GER
Capillaries (Ø 0.3 and 0.5 mm)	Hampton Research, Aliso Viejo, USA
Crystallization robot Freedom Evo 150	Tecan, Männedorf, SUI
Depression wells	Hampton research, Aliso Viejo, USA
Fiber loops (0.1 - 0.7 mm)	Hampton research, Aliso Viejo, USA
Magnetic bases for fiber loops	Hampton research, Aliso Viejo, USA
Microscope MZ16	Leica, Wetzlar, GER
Microscope S6E	Leica, Wetzlar, GER
Cryosystem X-Stream	Rigaku/MSC, The Woodlands, USA
Detector Mar345dtb image plate	Mar research, Hamburg, GER
Detector PILATUS at SLS	Paul Scherrer Institute (PSI), Villigen, SUI
VariMax HF optics	Rigaku/MSC, The Woodlands, USA
Rotating Anode Micromax-007 HF	Rigaku/MSC, The Woodlands, USA
Vacuum grease Baysilone	GE Bayer Silicones, Leverkusen, GER

2.1.1.4 Surface Plasmon Resonance

Biacore 2000

GE Healthcare, Uppsala, SWE

2.1.1.5 Protein production: Eukaryotic

Hera Safe Clean bench

ThermoFisher Sci., Waltham, USA

Hera Cell culture flask incubators

ThermoFisher Sci., Waltham, USA

Hera Cell roller bottle incubators

ThermoFisher Sci., Waltham, USA

Steritop Express Plus filter (0.22 µM)

Millipore, Schwabach, SUI

Neubauer improved counting chamber

Marienfeld, Lauda-Königshofen, GER

Beckmann Centrifuge J2-21

BeckmanCoulter GmbH, Krefeld, GER

Water bath SWB20

PD-Industriegesellsch. mbH, Dresden, GER

Microscope DM IL

Leica, Wetzlar, GER

Mr. Frosty freezing container

ThermoFisher Sci., Waltham, USA

2.1.1.6 Protein production: Procaryotic

High pressure homogenizer

EmulsiFlex Avestin, Ottawa, CAN

Sonicator Digital Sonifier 250

Branson, Danbury, USA

Incubation Shaker Unitron

InforsHT, Bottmingen-Basel, SWI

Shaker DOS-10L

NeoLab, Heidelberg, GER

Shaker Intelli-Mixer

NeoLab, Heidelberg, GER

Centrifuge Sorvall RC-6

Kendro, Langenselbold, GER

Sorvall rotor SS34

Kendro, Langenselbold, GER

Sorvall rotor SLC4000

Kendro, Langenselbold, GER

2.1.1.7 Chromatography

Analytical FPLC SMART

GE Healthcare, Uppsala, SWE

Analytical FPLC Ettan

GE Healthcare, Uppsala, SWE

Preparative FPLC Äkta prime plus

GE Healthcare, Uppsala, SWE

Preparative FPLC BioLogic Duo Flow

Bio-Rad, Munich, GER

Peristaltic EconoPump

Bio-Rad, Munich, GER

HiTrap Protein A HP 5 ml affinity column

GE Healthcare, Uppsala, SWE

HiTrap Benzamidine FF 1 ml affinity column

GE Healthcare, Uppsala, SWE

HiTrap HisTrap HP 1 ml affinity column

GE Healthcare, Uppsala, SWE

HiTrap Phenyl HP 5 ml affinity column

GE Healthcare, Uppsala, SWE

Superdex 75 PC 3.2/30 column

GE Healthcare, Uppsala, SWE

Superdex 200 PC 3.2/30 column

GE Healthcare, Uppsala, SWE

Superdex 75 10/300 GL column

GE Healthcare, Uppsala, SWE

2.1.1.8 Protein Analytics

Mini-PROTEAN 3 Cell

Bio-Rad, Munich, GER

Mini-PROTEAN 3 Multi-Casting Chamber

Bio-Rad, Munich, GER

PowerPac Basic

Bio-Rad, Munich, GER

2.1.2 Software

2.1.2.1 X-ray data processing and structural analysis

B-factor analysis	CCP4 package ¹⁰⁰ : Bavarge
Mar345dtb	Marresearch GmbH, Hamburg, GER
Indexing, Integrating, Scaling	XDS ¹⁰¹
Model building	Coot ¹⁰²
Molecular Replacement	CCP4 package ¹⁰⁰ : PHASER
Refinement	PHENIX ¹⁰¹
Structural figures	Pymol ¹⁰³
Sugar modeling	GlyProt ¹⁰⁴ : www.glycosciences.de
TLS group assignment	TLSMD: www.skuld.bmsc.washington.edu
TLS refinement	PHENIX ¹⁰¹
Matthews Probability Calculator	www.ruppweb.org/Mattprob/
Structure validation	WHAT_IF: www.swift.cmbi.ru.nl
Comparison of 3D protein structures	CCP4 package ¹⁰⁰ : PROCHECK
Exploration of molecular interfaces	SSM: www.ebi.ac.uk
	DALI: www.ehkhidna.biocenter.helsinki.fi
	PISA: www.ebi.ac.uk

2.1.2.2 Surface Plasmon Resonance

BIAsimulation 2.1	GE Healthcare, Uppsala, SWE
BIACORE 2000 control 3.2.1	GE Healthcare, Uppsala, SWE
BIAevaluation 4.1	GE Healthcare, Uppsala, SWE

2.1.2.3 Data analysis

Origin 8 Pro	OriginLab, Northampton, USA
--------------	-----------------------------

2.1.2.4 Chromatography

PrimeView	GE Healthcare, Uppsala, SWE
-----------	-----------------------------

2.1.2.5 Reference management

EndNote X2	Thomson Reuters, Philadelphia, USA
------------	------------------------------------

2.1.2.6 Online tools and databanks

Protein properties	ProtParam: www.expasy.org
Protein profile search	Blastp: www.ncbi.nlm.nih.gov/blast
3D protein structures	SMART: www.smart.embl-heidelberg.de
N-glycosylation site prediction	PDB: www.pdb.org
Literature search	NetNGlyc 1.0 Server: www.cbs.dtu.dk
Codon usage optimization	PubMed: www.ncbi.nlm.nih.gov
	GCUA: www.gcu.schoedl.de

Signal peptide prediction	SignalIP: www.cbs.dtu.dk
Transmembrane region prediction	TMHMM: www.cbs.dtu.dk
Align two sequences (protein)	SIM-Prot: www.expasy.org
	Blastp: www.ncbi.nlm.nih.gov
Align two sequences (nucleotide)	Blastn: www.ncbi.nlm.nih.gov
Protease cleavage sites	Peptide cutter: www.expasy.org
DNA to protein translation	www.expasy.org/tools/dna.html
Primer design	DNACalc: www.sigma-genosys.com
Restriction enzyme specificity	NEB cutter: www.tools.neb.com
Invert / complement DNA sequences	Manipulate DNA: www.vivo.colostate.edu
Protein to DNA translation	Backtranslation tool: www.entelechon.com
Protease cleavage sites in proteins	PeptideCutter: www.expasy.org

2.1.3 Consumables

2.1.3.1 Chemicals

Unless noted otherwise, all chemicals were analytical reagent (AR) grade. Additional chemicals and solutions that were of importance to a specific application are listed in the corresponding sections.

Ammoniumperoxodisulphate (APS)	Roth, Karlsruhe, GER
Ammonium sulphate	Sigma Aldrich, Munich, GER
Arginine	Roth, Karlsruhe, GER
Bromophenol blue	Roth, Karlsruhe, GER
Disodium hydrogen phosphate (Na_2HPO_4)	Sigma Aldrich, Munich, GER
EDTA, disodium salt	Roth, Karlsruhe, GER
Ethanol, abs.	J. T. Baker, Deventer, NET
Ethylene glycol	Sigma Aldrich, Munich, GER
Glacial acetic acid	Roth, Karlsruhe, GER
Glutathione, reduced	Sigma Aldrich, Munich, GER
Glutathione, oxidized	Sigma Aldrich, Munich, GER
Glycerol	Roth, Karlsruhe, GER
Glycine	Roth, Karlsruhe, GER
HEPES	Roth, Karlsruhe, GER
Imidazole	Sigma Aldrich, Munich, GER
L-Arginine	Roth, Karlsruhe, GER
Methanol, p.a.	Sigma Aldrich, Munich, GER
MgCl_2	Sigma Aldrich, Munich, GER
Nickel chloride	Roth, Karlsruhe, GER
Polyethylene glycols (various)	Hampton research, Aliso Viejo, USA
Rotiphorese Gel 30 37.5:1	Roth, Karlsruhe, GER
SDS	Roth, Karlsruhe, GER
Sodium cacodylate trihydrate	Sigma Aldrich, Munich, GER
Sodium chloride (NaCl)	Roth, Karlsruhe, GER
Sodium dihydrogen phosphate (NaH_2PO_4)	Sigma Aldrich, Munich, GER
Sodium hydroxide	Roth, Karlsruhe, GER
TEMED	Roth, Karlsruhe, GER
TRIS base	Sigma Aldrich, Munich, GER
Triton X-100	Sigma Aldrich, Munich, GER
Urea	Roth, Karlsruhe, GER
β -mercaptoethanol (β -ME)	Sigma Aldrich, Munich, GER

2.1.3.2 Eukaryotic cell culture

Geneticin (G-418)	Invitrogen, Karlsruhe, GER
PSN antibiotic mixture (100x)	Invitrogen, Karlsruhe, GER
Penicillin/Streptomycin (100x)	Invitrogen, Karlsruhe, GER
Ultra-Low IgG FCS	Invitrogen, Karlsruhe, GER
Opti-MEM-I	Invitrogen, Karlsruhe, GER
Dulbecco's PBS (without Ca ²⁺ /Mg ²⁺)	PAA, Cölbe, GER
Trypan blue stain (0,4 %)	Invitrogen, Karlsruhe, GER
Lipofectamine 2000	Invitrogen, Karlsruhe, GER
Trypsin /EDTA (1x) 0,05 % in DPBS	Invitrogen, Karlsruhe, GER
L-glutamine 200 mM (100x), liquid	Invitrogen, Karlsruhe, GER
Sodium pyruvate 100 mM (100x), liquid	Invitrogen, Karlsruhe, GER
CELLMASTER roller bottles	Greiner, Frickenhausen, GER
CELLSTAR cell culture flasks	Greiner, Frickenhausen, GER
α-MEM	Invitrogen, Karlsruhe, GER
DMSO	Sigma Aldrich, Munich, GER
CELLSTAR cell culture dishes	Greiner, Frickenhausen, GER
Protein A/G UltraLink resin	Thermo Scientific (Pierce), Waltham, USA

2.1.3.3 Prokaryotic (bacterial) cell culture

Isopropyl-β-D-thiogalactopyranoside (IPTG)	PeqLab, Erlangen, GER
Agar-agar	Roth, Karlsruhe, GER
Ampicillin, sodium salt	Sigma, Munich, GER
Kanamycin sulphate	Sigma Aldrich, Munich, GER
Chloramphenicol	Sigma Aldrich, Munich, GER
LB medium, complete	Sigma Aldrich, Munich, GER
Peptone	Roth, Karlsruhe, GER
Yeast extract	Roth, Karlsruhe, GER
Complete protease inhibitor tablets (EDTA)	Roche, Mannheim, GER

2.1.3.4 Protein purification and analysis

Spin-X centrifuge tube filters (0.22 μm)	Corning, Corning, USA
Microcon YM3	Millipore, Schwabach, GER
Amicon Ultra 4 and Ultra 15	Millipore, Schwabach, GER
Membrane filters (0.45 and 5 μm)	Millipore, Schwabach, GER
Syringe top filters (0.22 μm)	VWR International, Vienna, AUT
Steritop Express Plus filter (0.22 μM)	Millipore, Schwabach, SUI
Acrylamide-Bisacrylamide Rotiphorese Gel 30	Roth, Karlsruhe, GER
Bradford Protein Assay	Bio-Rad, Munich, GER
Coomassie Brilliant Blue R-250	Bio-Rad, Munich, GER
PageRuler unstained protein ladder	Fermentas, St. Leon-Rot, GER
Rotiphorese SDS-PAGE buffer (10x)	Roth, Karlsruhe, GER
Bradford reagent: Bio-Rad Protein Assay	Bio-Rad, Munich, GER
Spectra/Por Dialysis Membrane 6 - 8 kDa MWCO	Spectrum, DG Breda, NED

2.1.3.5 Molecular biology

StarPure Agarose, research grade	StarLab, Ahrensburg, GER
6x loading dye for agarose gels	Fermentas, St. Leon-Rot, GER
QIAquick PCR Purification Kit	Qiagen, Hilden, GER
QIAquick Gel Extraction Kit	Qiagen, Hilden, GER
dNTPs	Promega, Mannheim, GER
Minipreps Wizard Plus SV	Promega, Mannheim, GER
O ⁺ GeneRuler 1 kb DNA Ladder	Fermentas, St. Leon-Rot, GER
O ⁺ GeneRuler 100bp DNA Ladder	Fermentas, St. Leon-Rot, GER
6x Orange Loading Dye solution	Fermentas, St. Leon-Rot, GER
10x Pfu Ultra HF reaction buffer	Stratagene, La Jolla, USA
Buffers for restriction enzymes from NEB	NEB, Frankfurt/Main, GER
BSA solution for restriction enzymes from NEB	NEB, Frankfurt/Main, GER
Quick Ligation Kit	NEB, Frankfurt/Main, GER
Rapid DNA Ligation Kit	Fermentas, St. Leon-Rot, GER
Plasmid Maxi Kit	Qiagen, Hilden, GER

2.1.3.6 Crystallization Screens (commercial)

Crystal Screens 1, 2 and light	Hampton research, Aliso Viejo, USA
Grid Screen Ammonium Sulfate	Hampton research, Aliso Viejo, USA
Wizard I and II	Emerald BioSystems, Bainb. Island, USA
PEG/Ion 1	Hampton research, Aliso Viejo, USA

2.1.3.7 Surface Plasmon Resonance

Sensor Chip CM5 (research grade)	GE Healthcare, Uppsala, SWE
Surfactant P20	GE Healthcare, Uppsala, SWE
Protein A/G (soluble)	BioVision, Mountain View, USA
Amine Coupling Kit	GE Healthcare, Uppsala, SWE
Biacore Maintenance Kit	GE Healthcare, Uppsala, SWE
EDTA 0.5 M	Sigma Aldrich, Munich, GER

2.1.3.8 Enzymes

DNAse I (10 mg/ml)	Roche, Mannheim, GER
EndoH (500 ,000 U/μl)	NEB, Frankfurt/Main, GER
PfuUltra HF DNA polymerase (2.5 U/μl)	Stratagene, La Jolla, USA
Taq LC polymerase (1U/μl)	Fermentas, St. Leon-Rot, GER
T4 DNA ligase (1 U/μl)	Fermentas, St. Leon-Rot, GER
Thrombin (1U/μl)	GE Healthcare, Uppsala, SUI
PreScission protease	GE Healthcare, Uppsala, SUI
Restriction enzymes XhoI and NdeI (Table 2-1)	NEB, Frankfurt/Main, GER

Table 2-1. Restriction enzymes

Name	Recognition sequence	Digestion type	BSA	Heat inactivation	NEB buffer
Nde I	5'...CATATG...3' 3'...GTATAC...5'	Sticky ends	No	65 °C; 20 min	4
Xho I	5'...CTCGAG...3' 3'...GAGCTC...5'	Sticky ends	Yes	65 °C; 20 min	2 ¹⁾

¹⁾In case of a double digestion together with NdeI NEB buffer 4 was used as recommended by NEB.

2.1.3.9 Proteins

ULBP1/Fc chimera (50 µg)

ULBP2/Fc chimera (50 µg)

ULBP4/Fc chimera

ULBP5/Fc chimera

R&D Systems, Minneapolis, USA

R&D Systems, Minneapolis, USA

provided by Dr. M. Wittenbrink

(Working group of Prof. Dr. A. Steinle)

provided by Dr. M. Wittenbrink

(Working group of Prof. Dr. A. Steinle)

2.1.4 Plasmids and primers

2.1.4.1 Plasmids (eukaryotic expression)

pcDNA3.1(-)

Invitrogen, Karlsruhe, GER

Features:

- ⇒ Human cytomegalovirus (CMV) promoter permits efficient, high-level expression of recombinant proteins in eukaryotic cells through high level transcription by RNA polymerase II
- ⇒ SV40 origin allows episomal replication of plasmid in eukaryotic cells expressing the SV40 large T antigen (e. g. HEK293T cells)
- ⇒ Neomycin resistance gene allows for selection of the vector in eukaryotic cells with geneticin (G418)
- ⇒ pUC origin allows for high-copy number of plasmid in *E.coli*
- ⇒ Ampicillin resistance gene allows for selection of the vector in *E.coli* with ampicillin

2.1.4.2 Plasmids (prokaryotic expression)

pET-21a(+)

Merck (Novagen), Darmstadt, GER

Features:

- ⇒ T7 promoter permits efficient, high-level expression of recombinant proteins in *E.coli* through high level transcription by the T7 RNA polymerase
- ⇒ Ampicillin resistance gene allows for selection of the vector in *E.coli* with Ampicillin.
- ⇒ C-terminal hexahistidine (His₆-tag) for affinity purification of the recombinant protein with Ni-NTA columns

2.1.4.3 Primers

All primers (Table 2-2 and Table 2-3) were purchased from biomers.net (Ulm, GER).

Table 2-2. Cloning primers

Source DNA	Sequence (5' → 3')		T _m [°C]
MICB _{α1-α3} in pET-21a(+)	Forward	TCAGCGTACATATGGAGCCCCACAGTCTTCGTTA	71
	Reverse-1 ¹⁾	CAGAAGAACCACGCGGCACCAGACCACCTGTTCTCC TGATGGCCACCCCGGATTTC	90
	Reverse-2 ¹⁾	TCATTCTACTCGAGTCAATGATGATGATGATGATGAT GATGGCCAGAAGAACCACGCGGCACCAGACCA	89

¹⁾ Reverse overhanging primers were used to introduce a thrombin cleavage site and an His₈-tag

Table 2-3. Sequencing primers

Plasmid	Sequence (5' → 3')		T _m [°C]
pcDNA3.1(-)	Forward	TAATACGACTCACTATAGGG	39
	Reverse	TAGAAGGCACAGTCGAGG	45
pET-21a(+)	Forward	TAATACGACTCACTATAGG	34
	Reverse	GCTAGTTATTGCTCAGCGG	48

2.1.5 Growth media, buffers and solutions

If not stated otherwise, buffer solutions were adjusted to the desired pH at 4 °C since the pH values of many of the commonly used buffer solutions (e. g. Tris or HEPES buffers) are significantly temperature dependent. Furthermore, buffers were passed through a 0.45 µm filter and degassed (using either an ultrasonic water bath or a water driven vacuum pump) before usage in FPLC applications.

2.1.5.1 Cultivation of bacterial cells

Table 2-4. LB medium

Composition	Preparation (pro 1 ddH ₂ O)
LB medium, complete	25g

Table 2-5. LB agar plates

Composition	Preparation (pro 1 ddH ₂ O)
LB medium, complete	see Table 2-4
Agar-agar	15g

Table 2-6. SOB medium

Composition	Preparation (1 l)
LB medium, complete	see Table 2-4
KCl	10 mM
MgSO ₄	20 mM

2.1.5.2 Cultivation of eukaryotic cells

Table 2-7. CHO Lec 3.2.8.1 cell media

Composition	Selection medium Concentration	Non-selection medium Concentrations
Medium: α -MEM	-	-
Penicillin G	100 U/ml	100 U/ml
Streptomycin	100 μ g/ml	100 μ g/ml
L-glutamine	2 mM	2 mM
Sodium pyruvate	1 mM	1 mM
Ultra-Low IgG FCS	7.5 %	10 %
Geneticin (G418)	1.5 mg/ml ¹⁾	

¹⁾ To determine the concentration of G418 a kill curve was generated (see Section 2.2.2.5)

Table 2-8. CV-1 cell media

Composition	Concentrations
Medium: DMEM	-
Penicillin G	100 U/ml
Streptomycin	100 μ g/ml
L-glutamine	2 mM
Sodium pyruvate	1 mM
Ultra-Low IgG FCS	7.5 %
Geneticin (G418)	2.0 mg/ml

Table 2-9. Cryo solution

Composition	Concentrations
Ultra-low IgG FCS	90 % (v/v)
DMSO	10 % (v/v)

2.1.5.3 Protein expression, refolding and purification

Table 2-10. TBS buffer

Composition	Concentrations
Tris base ¹⁾ at 4 °C	20 mM
NaCl	130 mM
KCl	2 mM

¹⁾ The pH is indicated in the respective Sections

Table 2-11. Lysis buffer

Composition	Concentrations
Tris base pH 8 at 4 °C	20 mM
NaCl	130 mM
KCl	2 mM
DNaseI grade II	10 µg/ml
MgCl ₂	10 mM
PMSF	0.1 mM
Glycerol	5 %
Complete Protease Inhibitor tablet (EDTA free)	1 x

Table 2-12. Triton X-100 and Triton-free wash buffers

Composition	Triton X-100 wash buffer Concentrations	Triton X-100 free wash buffer Concentrations
Tris base pH 8 at 4 °C	50 mM	50 mM
NaCl	50 mM	50 mM
Triton X-100	0.5 % (w/v)	-

Table 2-13. Urea buffer

Composition	Concentrations
Tris base pH 8 at 4 °C	100 mM
Glycine	50 mM
Urea	8 M

Table 2-14. Refolding buffer

Composition	Concentrations	Dialysis step
Tris base pH 8 at 4 °C	100 mM	-
L-Arginine	0.4→0.1 M	Step 4
PMSF	0.1 mM (1x)	-
EDTA	1 mM	-
Urea	4→2→1→0 M	Step 1-3

Table 2-15. IMAC (Ni-NTA) chromatography buffers

Composition	Ni-NTA binding buffer Concentrations	Ni-NTA elution buffer Concentrations
NaH ₂ PO ₄ pH 7.4 at 4 °C	20 mM	20 mM
NaCl	0.5 M	0.5 M
Imidazole	30 mM	500 mM

Table 2-16. Protein A chromatography buffers

Composition	Protein A binding buffer Concentration	Protein A elution solution Concentration
Glycine pH 9	173 mM	-
Arginine pH 4.1 ¹⁾	-	0.5 M
NaCl	0.5 M	10 mM

¹⁾ adjust pH carefully as no buffer is present

Table 2-17. HEPES reservoir buffer

Composition	Concentration
HEPES pH 9	0.5 M

Table 2-18. HIC chromatography buffers

Composition	HIC binding buffer Concentration	HIC elution buffer Concentration
Na ₂ HPO ₄ pH 7.4	50 mM	50 mM
NaCl	1 M	-

2.1.5.4 Native protein deglycosylation

Table 2-19. Endo H buffer

Composition	Concentration
NaAc pH 5.2 at 25 °C	0.1 M

2.1.5.5 SDS-PAGE

Table 2-20. Transformation buffer solutions

Composition	Concentrations Tfb I	Concentrations Tfb II
RbCl	100 mM	10 mM
CaCl ₂	20 mM	75 mM
MnCl ₂	50 mM	-
KAc	30 mM	-
Glycerol	15 mM	15 mM
MOPS	-	10 mM
pH	5.8 at RT with KOH	6.8 at RT with KOH

Table 2-21. SDS sample buffer compositions

Composition	Reducing SDS sample buffer Volume	Non-reducing SDS sample buffer Volume
SDS 10 %	1.6 ml	1.6 ml
Glycerol	2.8 ml	2.8 ml
Tris 1.5 M pH 6.8	0.5 ml	0.5 ml
β-mercaptoethanol, β-ME	0.4 ml	-
ddH ₂ O	2.5 ml	2.9 ml
Bromophenol blue	Add spatula tip	Add spatula tip
Total volume	7.8 ml	7.8 ml

Table 2-22. SDS gel compositions

Composition	4 % stacking gel (4 gels)	15 % separation gel (4 gels)
ddH ₂ O	6.1 ml	3.5 ml
Tris 1.5 M pH 6.8	2.5 ml	-
Tris 1.5 M pH 8.8	-	3.7 ml
SDS 10 % (w/v)	100 µl	100 µl
Acrylamide-bisacrylamide	1.3 ml	7.5 ml
TEMED	10 µl	7.5 µl
APS 10 % (w/v)	100 µl	150 µl
Total volume	~10.1 ml	~15.0 ml

Table 2-23. Coomassie staining and destaining solutions

Composition	Staining Volume / Mass	Destaining Volume
Coomassie brilliant blue	0.5 g	-
Methanol	250 ml	25 ml
Glacial acetic acid	30 ml	37 ml
ddH ₂ O	220 ml	438 ml
Total volume	500 ml	500 ml

2.1.5.6 Agarose gel electrophoresis

Table 2-24. TAE buffer

Composition	Concentration
Tris base ¹⁾	2 M
EDTA pH = 8	50 mM
Glacial acetic acid	57 % (w/v)

¹⁾ The pH of this buffer is not adjusted and should be close to 8.5

2.1.5.7 Surface Plasmon Resonance

Table 2-25. HBS-EP buffer

Composition	Concentrations
HEPES pH 7.4 at 25 °C	10 mM
NaCl	150 mM
EDTA	3 mM
P20	0.005 % (v/v)

2.1.5.8 Crystallization

Table 2-26. Mother liquor of crystallization condition

Composition	Final concentrations
Sodium cacodylate trihydrate, pH 6.5	10 mM
Ammonium sulphate	20 mM
PEG8000	25 % (v/v)

2.1.6 Bacterial strains and eukaryotic cell lines

2.1.6.1 Bacterial strains

For a detailed description of the features of the used *E.coli* strains please refer to the Novagen Competent Cells brochure.

Rosetta 2 (DE3)

Merck (Novagen), Darmstadt, GER

Features:

The *E.coli* Rosetta 2(DE3) cells are IPTG inducible and supply seven rare tRNAs on two separate plasmids with chloramphenicol (Cam) resistance. The rare tRNAs will allow for the efficient translation of mRNAs containing eukaryotic codons rarely used in *E.coli*.

NovaBlue

Merck (Novagen), Darmstadt, GER

Features:

The *E.coli* NovaBlue cells are used for cloning purposes and are characterized by a high transformation efficiency and a tetracycline (Tet) resistance.

2.1.6.2 Eukaryotic cell lines

CHO Lec 3.2.8.1 cells

Personal stock of Prof. Dr. T. Stehle

A number of distinct CHO Lec cell lines were developed by Pamela Stanley in 1989¹⁰⁵. Each cell line is deficient in a number of Golgi glycosyltransferases and/or Golgi nucleotide-sugar transporters and therefore produces a specific glycoprotein phenotype. The phenotypic composition of each cell line was determined by testing the cells ability to bind individual members from a set of defined lectins (Lec). The CHO Lec 3.2.8.1 cell line, which exhibits four individual Lec phenotypes, has been shown to be useful for X-ray crystallographic applications¹⁰⁶. Its inability to produce hybrid and complex glycans usually leads to a significantly reduced and much more homogeneous glycan moiety of a glycoprotein, which in turn can facilitate the glycoprotein's crystallization. For more details see Section 3.2.3.

CV-1 cells stably transfected with UL16-Fc

Personal stock of Prof. Dr. A. Steinle

The CV-1 cells, stably transfected with the UL16-Fc construct (described in Section 2.2.3.5.1), were provided by Dr. Jessica Spreu. At the beginning of the UL16-MICB project Dr. Jessica Spreu

also provided purified UL16-Fc protein that had been produced in CV-1 cells in the laboratory of Prof. Dr. Alexander Steinle. During the later stages of the project the author used the UL16-Fc expressing CV-1 cells (provided by Dr. Jessica Spreu, see above) for the production of UL16-Fc in the laboratory of Prof. Dr. Thilo Stehle.

2.2 Methods

2.2.1 Methods in molecular biology

2.2.1.1 Competent cells

Bacterial cells that are harvested during their logarithmic growth phase and that up to that point have grown in media supplemented with monovalent and divalent cations (e. g. Ca^{2+} , Mn^{2+} and Rb^{+}), have the ability to take up foreign DNA in a process called transformation.

Protocol: A volume of 100 ml SOB medium was inoculated with 1 ml of bacterial o/n (overnight) culture and grown up to an OD_{600} of 0.4 - 0.6. Bacteria were cooled on ice for 10 min and harvested by a 10 min centrifugation step at 1,700x g at 4 °C. The supernatant was discarded and the bacterial pellet resuspended in 20 ml ice-cold TfbI (Table 2-20). After an incubation time of 10 min the bacteria were again harvested by a 10 min centrifugation step at 1700x g at 4 °C, followed by their resuspension in 4 ml ice-cold TfbII (Table 2-20). Finally 100 μl aliquots of this bacterial suspension were frozen at -80 °C and used for transformation processes (see Section 2.2.1.9).

Note: *E.coli* strains that carry no antibiotic resistance genes whatsoever (e. g. NovaBlue) have to be treated under sterile conditions (i. e. by using a clean bench with laminar flow capabilities) to avoid contamination with other bacteria. However, some *E.coli* strains as for example Rosetta 2(DE3) cells feature special characteristics such as the ability to produce rare tRNAs and to express the T7 lysozyme for a more controlled expression of proteins. These features are based on additional plasmids that usually carry additional resistance genes. These plasmids should be selected for at all times and allow for a convenient way to avoid contamination during the preparation of competent cells.

2.2.1.2 Polymerase chain reaction (PCR)

The polymerase chain reaction (PCR) allows for the exponential amplification of a DNA sequence *in vitro*. It is based on a simple three step cycle, which is usually repeated 30-40 times in a so-called thermocycler: (Step 1) During ‘denaturation’ the temperature of the DNA containing PCR solution is increased to 95 °C, which leads to the efficient thermal denaturation of the DNA double helix and therefore eventually results in two single DNA strands. The PCR solution also contains two short synthetic oligonucleotides (primers). One of the two primers is hereby complementary to the sense DNA strand while the other is complementary to the antisense DNA strand and both primer together flank the region of the DNA that is supposed to be amplified. (Step 2) During ‘annealing’ the temperature is lowered to 50 - 70 °C so that both synthetic primers can specifically anneal to their corresponding single DNA template strand. (Step 3) Finally, during ‘elongation’ the temperature is increased again to 72 °C, which allows a thermostable DNA polymerase to create the complementary DNA strand of each of the two single template DNA strands by efficiently elongating each of the two primers in 5’→3’ direction.

The DNA polymerase classically used in PCR reactions is Taq from *Thermus aquaticus* (processivity: 2800 n/min). However, a clear disadvantage of Taq is its missing 3’→5’ proofreading activity, which makes the elongation of the newly synthesized DNA strand more error prone. This is the reason why Taq is nowadays used less frequently in standard cloning procedures. Instead, Pfu from *Pyrococcus furiosus* (processivity: 4500 n/min), which features a 3’→5’ proofreading activity, is used for these applications.

Protocol: PCR samples were prepared according to the pipetting scheme shown in Table 2-27 and then transferred to a thermocycler. The subsequent standard PCR program is described in Table 2-28.

Table 2-27. PCR pipetting scheme

Components	Stock concentration	Final concentration	Volume per PCR vial [μl]
DNA (plasmid)	100 ng/μl	100 ng	1
Primer 1	100 μM ²⁾	2 μM	1
Primer 2	100 μM ²⁾	2 μM	1
dNTPs ¹⁾	10 mM	200 μM	1
Pfu Ultra HF buffer	10 x	1 x	5
Pfu Ultra HF	2.5 u/μl	2.5 u	1
ddH ₂ O	-	-	40
Total volume			50

¹⁾ total concentration: 10 μM dNTP corresponds to 2.5 μM of each of the four deoxynucleotides

²⁾ 100 μM primer equals 100 pmol/μl

Table 2-28. Standard PCR program

Temp. [°C]	Time [s]	Step	Number of cycles
95	60	Initial denaturation	1
95	30	Denaturation (Step 1)	35
50-70	30	Annealing (Step 2)	
72	60	Elongation (Step 3)	
72	600	Final elongation	1
4	∞	Storage	1

2.2.1.3 Purification of PCR fragments

After a PCR, the solution not only contains the desired PCR product but also the DNA polymerase as well the deoxynucleotides and primers that are usually not used up during a PCR reaction. These contaminants can later on interfere with the restriction digestion of the PCR product and therefore have to be removed beforehand.

Protocol: The PCR product was purified and the contaminants were removed by using the *QIAquick PCR Purification Kit* and by following the manufacturer's instructions. After the procedure, the PCR product was eluted from the silica columns in 30 µl of ddH₂O.

Note: This protocol is not suitable if unspecific PCR fragments are produced during the PCR reaction. In this case it is advisable to perform an agarose gel electrophoresis followed by the gel extraction of the desired PCR product (see Sections 2.2.1.4 and 2.2.1.5).

2.2.1.4 Agarose gel electrophoresis

Agarose gel electrophoresis is a standard method for the separation and purification of DNA molecules such as plasmids or PCR products. The separation itself is hereby mainly dependent on three factors: (i) the molecular weight of the DNA, (ii) the conformation of the DNA (plasmids run faster in the following order and from left to right: relaxed/nicked, linear, supercoiled), and (iii) the density of the agarose gel, which is again dependent on the agarose amount used for casting the gel. To visualize the DNA in the agarose gel matrix after electrophoresis, the gel is stained with ethidiumbromide, a substance that intercalates between two stacked base pairs and shows a strong fluorescence when excited with electromagnetic radiation between 254 and 366 nm (UV light). A valuable asset in agarose gel electrophoresis is a DNA marker. These markers are comprised of DNA fragments of defined lengths that are also separated during electrophoresis and allow for a very exact determination of the molecular weight of DNA fragments.

Protocol: 0.4 - 1 g *StarPure* Agarose were dissolved in 50 ml TAE buffer (Table 2-24) to prepare 0.8 - 2 % (w/v) agarose gels. The agarose TAE suspension was repeatedly heated in a microwave until all agarose was completely dissolved. The clear agarose solution was then cast into an agarose gel casket for cooling and polymerization. After the solution had reached at temperature of about 50 °C during this process, 1µl of ethidiumbromide solution [0.5 µg/ml] was added. When the agarose had completely polymerized, the gel was submerged into an electrophoresis chamber filled with TAE buffer (Table 2-24). DNA samples were diluted in sample buffer (6x *Orange Loading Dye Solution*) and loaded into separate gel pockets. In addition, 7 µl of marker (either *O'GeneRuler 1 kb DNA Ladder* or *O'GeneRuler 100 bp DNA Ladder*) were loaded into a separate gel pocket. The electrophoresis was carried out at 90 V for 60 - 70 min and then analyzed under UV light.

2.2.1.5 Agarose gel extraction

During an agarose gel electrophoresis, DNA molecules are not only separated according to their molecular weights but are also purified from other contaminants such as nucleotides and primers. Each DNA band on an agarose gel therefore corresponds to a purified DNA species. These purified DNA species (e. g. plasmids cut by restriction enzymes or PCR products) are valuable as they will facilitate the subsequent cloning steps such as the restriction digestions of the PCR product or the ligation of this product into the cut vector. The extraction of the purified DNA species from the agarose gel is usually performed by using an agarose gel extraction kit.

Protocol: The DNA species were extracted from agarose gels by using the *QIAquick Gel Extraction Kit* and by following the manufacturer's instructions. The DNA was eluted from the kit-supplied silica columns with 30 µl of ddH₂O.

2.2.1.6 Restriction digestion of plasmids and PCR products

Restriction enzymes (see Section 2.1.3.8) are bacterial type II endonucleases that recognize palindromic sequences in a DNA double helix such as a PCR product or plasmid. In the latter case, diverse recognition sequences for a number of different restriction enzymes usually cluster in a region that is called the multiple cloning site (MCS). Restriction enzymes cleave the DNA backbone by hydrolyzing the phosphodiester bonds between the 3'-OH group of an upstream nucleotide and the 5'-phosphate group of an adjacent downstream nucleotide. Depending on the exact cut locations in both strands, a specific restriction enzyme can give rise to the formation of either sticky ends or blunt ends. DNA fragments with sticky ends feature unpaired nucleotides at their ends, while DNA fragments with blunt ends do not.

Protocol: Restriction digestion samples containing either plasmid or PCR product were prepared according to the pipetting scheme shown in Table 2-29. The samples were incubated at 37 °C for 1 - 24 h, followed by heat inactivation of the restriction enzymes according to the manufacturer’s instructions. The final DNA purification was either performed using a *QIAquick PCR Purification Kit* (see Section 2.2.1.3) or by agarose gel electrophoresis (see Section 2.2.1.4).

Table 2-29. Pipetting scheme for restriction digestions

Components	Stock solution	Composition Plasmid digest	Composition PCR product digest	Volume [μl]
Plasmid	variable	2 μg	-	x
PCR product	variable	-	1 μg	y
Restriction enzyme buffer	10 x	1 x	1 x	2
Restriction enzyme 1	20 u/μl	20 u	20 u	1
Restriction enzyme 2	20 u/μl	20 u	20 u	1
BSA ¹⁾	10 x	1 x	1 x	2
ddH ₂ O	-	-	-	20 - 6 - x or y
Total volume	-	-	-	20

¹⁾ before use, the supplied 100x BSA solution was diluted 10 fold.

2.2.1.7 Ligation

Prior to ligation, a preliminary circular plasmid is formed by filling the gap of a restriction enzyme (see Sections 2.1.3.8 and 2.2.1.6) digested plasmid with a PCR product that was treated with the same set of restriction enzymes. However, to create an intact plasmid it is not enough to simply close the plasmid gap with the PCR product. In addition also the DNA backbone breaks at the contact areas between PCR product and plasmid have to be covalently sealed in a process called ligation. The ligation process itself is mediated by an enzyme that is generally referred to as DNA ligase (most often the T4 ligase; see Section 2.1.3.8), which catalyzes an energy (ATP) consuming and Mg²⁺ dependent reaction that creates a covalent phosphodiester bond between the 3’-OH group of an upstream nucleotide and the 5’-phosphate group of an adjacent downstream nucleotide.

Protocol: Ligation samples were prepared according to the pipetting scheme shown in Table 2-30 and

Equation 2-1, followed by an o/n incubation at 16 °C. The DNA ligase was heat inactivated at 65 °C for 10 min prior to transformation. Alternatively, plasmid and PCR product were ligated using either a *Quick Ligation Kit* or *Rapid DNA Ligation Kit* and by following the manufacturer’s instructions.

Table 2-30. Ligation pipetting scheme

Components	Stock solution	Final solution	Volume [μl]
Plasmid	variable	Equation 2-1	x
PCR product	variable	Equation 2-1	y
Ligase buffer	10 x	1 x	2
T4 ligase	1 u/μl	2 u	2
ATP	50 mM	5 mM	2
ddH ₂ O	-	-	20-6-x-y
Total volume	-	-	20

$$ng(insert) = \frac{3 \cdot ng(plasmid) \cdot bp(insert)}{bp(plasmid)} \quad \text{Equation 2-1}$$

2.2.1.8 Agar plates

Agar plates are used for the growth and selection of bacteria after their transformation with a recombinant plasmid.

Protocol: 25 g of ready-to-use Luria Broth media (Table 2-4) was supplemented with 15 g agar-agar and dissolved in 1 l ddH₂O. The solution was autoclaved and then left to cool down to about 40 °C before antibiotics were added. Then, a lab bench was wiped with 70 % (v/v) ethanol and the neck of the bottle that contained the antibiotic supplemented LB media flamed with a Bunsen burner. The antibiotic supplemented LB media was subsequently poured into Petri dishes and the plates were left at room temperature for several hrs until the agar had polymerized. Finally, agar plates were stored at 4 °C for up to two months.

2.2.1.9 Transformation

Transformation is the process by which *E.coli* cells take up foreign circular DNA (plasmids).

Protocol: 100 μl competent *E.coli* cells were thawed on ice and supplemented with 5 - 10 μl of a ligation mixture. After incubation on ice for 30 min the bacterial suspension was subjected to a heat-shock at 42 °C for 1 min, followed by an additional incubation on ice for 2 min. After incubation, LB medium without antibiotics (preheated at 37 °C) was added to the suspension and the solution incubated at 37 °C for 1 h. Bacteria were then harvested by centrifugation at 800x g for 2 min, and the supernatant (except for 50 μl) discarded. The bacteria were then resuspended in the remaining LB media and plated on agar plates supplemented with the appropriate antibiotics and

then incubated o/n at 37 °C. After 16 hrs, agar plates were evaluated and grown bacterial colonies used for a colony-PCR to test for the presence of the insert.

2.2.1.10 Colony PCR

A colony PCR is based on the same principle as a normal PCR. It is a very useful technique to test high numbers of bacterial colonies in a short time for the presence of the desired DNA insert after transformation. An alternative, but more laborious and time consuming, technique is that of a restriction enzyme test digestion after which the ligated insert is cut out of the plasmid and visualized on an agarose gel.

Protocol: Each bacterial colony was clearly labeled by marking the bottom of the Petri dish directly below the colony with a felt tip pen. The same number was then used to label the corresponding PCR tube. Each PCR tube was filled with 50 µl of the PCR master mix (Table 2-31) and then inoculated with a tiny amount of the bacterial colony, obtained by scratching off a bit of the colony by using a sterile pipet tip. The PCR-tubes were transferred to the *PCR iCycler* and the colony PCR program (Table 2-32) started. Finally, in order to determine the presence of the insert for each colony, the PCR products were analyzed by agarose gel electrophoresis.

Table 2-31. Colony PCR master mix

Components	Stock concentration	Concentrations	Volume per PCR vial [µl]
Bacteria from colony	-	-	Pipet tip
Primer 1 ¹⁾	100 µM	2 µM	1
Primer 2 ¹⁾	100 µM	2 µM	1
dNTPs ²⁾	10 mM	200 µM	1
MgCl ₂	25 mM	1.5 mM	3
Taq LC buffer	10 x	1 x	5
Taq LC polymerase	2.5 u/µl	2.5 u	1
ddH ₂ O	-	-	38
Total volume	-	-	50

¹⁾100µM primer equals 100 pmol/µl

²⁾ total concentration: 10 µM dNTP corresponds to 2.5 µM of each of the four deoxynucleotides

Table 2-32. Colony PCR program

Temp. [°C]	Time [s]	Step	Number of cycles
95	480	Initial denaturation	1
95	30	Denaturation (Step 1)	30
50-70	30	Annealing (Step 2)	
72	60	Elongation (Step 3)	
72	600	Final elongation	1
4	∞	Storage	1

2.2.1.11 Bacterial overnight cultures

Bacterial o/n cultures have the sole purpose of growing bacteria (i) to prepare glycerol stocks for long term storage, (ii) to extract their plasmids for subsequent cloning procedures, or (iii) to inoculate larger amounts of LB media (e. g. for protein expression).

Protocol: 10 ml of LB media supplemented with the necessary antibiotics were inoculated with a single bacterial colony picked from an agar plate, or with a sample of bacteria from stocks that were stored at -80 °C.

2.2.1.12 Glycerol stocks of *E.coli* cells

Glycerol is a very potent cryoprotectant and is therefore widely used to protect (bacterial) cells from being damaged during the freezing and thawing procedures that are necessary for their long term storage.

Protocol: 750 µl of a bacterial o/n culture were mixed with 250 µl of sterile glycerol and frozen at -80 °C.

2.2.1.13 Isolation of plasmid DNA from *E.coli*

The isolation of plasmid DNA from *E.coli* is based on the alkaline lysis of the bacterial cells and the subsequent purification of the plasmids using small spin columns filled with a silica matrix.

Protocol: Plasmid isolation was performed using a *Miniprep Wizard Plus SV Kit* and by following the manufacturer's instructions. The DNA was eluted from the spin column with 30 µl of ddH₂O. For larger plasmid preparations a *Plasmid Maxi Kit* was used. In the latter case the DNA pellet was dissolved in 1 ml of ddH₂O. Although plasmids are fairly stable in water, a higher stability can be achieved by eluting the DNA in 10 mM Tris pH 8.5. DNA sequencing of recombinant plasmids was performed by MWG (Ebersberg, Germany), while the author of this thesis provided the recombinant plasmids as well as the appropriate sequencing primers (Table 2-3).

2.2.1.14 Photometric determination of DNA concentration and purity

Based on the aromatic ring systems of purine and pyrimidine bases, nucleic acids show an absorbance maximum at 260 nm. Due to an effect called hyperchromicity ss DNA absorbs more strongly than ds DNA (Table 2-33).

Table 2-33. Absorbance of ds and ss DNA

	ds DNA [$\mu\text{g/ml}$]	ss DNA [$\mu\text{g/ml}$]
OD ₂₆₀ of 1 equals	40	33

While DNA absorbs strongest at 260 nm, proteins do so at 280 nm. By measuring the absorbance ratio of a DNA sample at these two wavelengths one can estimate the purity of the DNA sample. A 260/280 nm ratio of 1.8 indicates the presence of pure ds DNA (i. e. plasmid), while a ratio below 1.8 can indicate a contamination with protein.

Protocol: DNA stock solutions were diluted in ddH₂O and absorbance values of the solution measured at 260 and 280 nm with a 75 μl quartz cuvette and a spectrophotometer in order to calculate the DNA concentration and to estimate the DNA purity. Alternatively, the absorbance of DNA samples at 260 nm and 280 nm was measured with a *Nanodrop ND-1000* photometer.

2.2.2 Methods in cell biology

2.2.2.1 Cultivation of eukaryotic cells

Cells were grown at 37 °C and 5 % CO₂ in a water-vapor-saturated atmosphere. Cell growth was monitored every second day using a light microscope. Untransfected CHO Lec 3.2.8.1 cells¹⁰⁵ (see Section 2.1.6.2) were grown in CHO Lec culture medium (Table 2-7). CHO Lec 3.2.8.1 cells are adherent cells and were passaged when the culture had reached a confluency of about 90 %. The medium was removed and the cells washed with 37 °C preheated 1x PBS (without Ca²⁺/Mg²⁺). Cells were then overlaid with a thin layer of 37 °C preheated 1x trypsin/EDTA solution and incubated for 3 - 5 minutes, during which the cells detached from the culture flask bottom. The trypsin/EDTA cell suspension was transferred to a conical centrifugation tube (50 ml Falcon type), quickly centrifuged at 300x g for 3 min and the trypsin/EDTA supernatant decanted. The pelleted cells were resuspended and diluted 3 - 10 fold in 37 °C preheated CHO Lec culture medium (Table 2-7), followed by their transfer into new cell culture flasks. After stable transfection with the pcDNA3.1(-) plasmid (see Section 2.1.4.1), CHO Lec 3.2.8.1 cells were grown in CHO Lec

selection medium (Table 2-7) that contained 1.5 mg/ml geneticin (G418) and the protein (UL16-Fc) containing supernatant collected before each passaging of the cells. The supernatant was stored at 4 °C.

The cultivation of CV-1 cells was essentially performed as described in this section for CHO Lec 3.2.8.1 cells. The main difference was the type of basal culture media that was used. CHO Lec 3.2.8.1 cells were grown in α -MEM (Tables 2-7), whereas CV-1 were grown in DMEM (Table 2-8).

2.2.2.2 Roller bottle setup with CHO Lec 3.2.8.1 cells

Roller bottles represent a convenient way to scale up the protein production of adherent eukaryotic cells as they provide a significantly larger growth surface compared to normal cell culture flasks. As a consequence, roller bottles reduce (i) the space requirements to grow the cells, (ii) the material costs and (iii) the amount of labor necessary to maintain the cells and harvest the protein containing supernatant.

Protocol: Four 90 % confluent (175 cm²) cell culture flasks containing CHO Lec 3.2.8.1 cells (see Section 2.1.6.2) were trypsinated as described in Section 2.2.2.1, followed by the transfer of all pelleted cells to a single roller bottle, containing 200 ml of 37 °C preheated CHO Lec culture medium (Table 2-7). The culture medium (instead of selection medium) was used to reduce the stress exerted by the selection marker in order to favor target protein (i. e. UL16-Fc) production. The protein (UL16-Fc) containing medium was collected after 7 days and replaced by fresh CHO Lec 3.2.8.1 culture medium. Roller bottles with cells having grown for 4 weeks were discarded and replaced with new roller bottles containing fresh cells. Cells were grown throughout at 37 °C, 5 % CO₂ in a water-vapor-saturated atmosphere.

2.2.2.3 Cryoconservation of eukaryotic cells

Cells were trypsinated as described in Section 2.2.2.1. However, instead of resuspending the pelleted cells in medium (Tables 2-7 and 2-8), cells were resuspended in 1 ml of cryo solution (Table 2-9) and transferred into a 2 ml cryo tube. The cryo tube was placed inside a 20 °C *Mr. Frosty* freezing container, followed by the transfer of the container into a -80 °C fridge. At -80 °C, the temperature inside the container decreases at 1 °C/min and therefore allows a very gentle freezing of eukaryotic cells. After 4 hrs the cryo tube was removed from the container and stored in liquid nitrogen.

Frozen cells were thawed quickly by placing the frozen cryo tube into a 37 °C water bath for 1 - 2 min and washed with 37 °C preheated PBS to reduce the DMSO concentration. After centrifugation of the PBS solution at 300x g, the cell pellet was resuspended in 25 ml 37 °C preheated medium (Tables 2-7 and 2-8) and transferred to a single (175 cm²) cell culture flask.

2.2.2.4 Counting of living cells

For some applications, such as the transfection of cells (see Section 2.2.2.6) or the generation of kill curves (see Section 2.2.2.5), it is necessary to seed a defined number of cells. In order to do so one needs to know the concentration of living cells in a given cell suspension. This concentration can be determined by staining dead cells with the dye ‘trypan blue’, followed by the subsequent counting of living (non-blue) cells in a Neubauer counting chamber.

Protocol: Adherent cells were trypsinated as described in Section 2.2.2.1 and resuspended in 15 ml of the respective culture medium. 50 µl of this cell suspension were mixed with 50 µl of trypan blue (dilution factor: 2) and pipetted below the cover slip of a Neubauer improved counting chamber. Cells were counted using a light microscope and their concentration calculated using the following equation:

$$c = \frac{N_{av} \cdot F_{tb}}{A \cdot d} \cdot 10^4 \quad \text{Equation 2-2}$$

with

N_{av}	Average number of cells per large square (average based on four large squares)
F_{tb}	Dilution factor of trypan blue (equals 2; see above)
A	Area of one large square (1 mm ²)
d	Depth of the chamber (0.1 mm)
$A \cdot d$	Volume of one large square (0.1 µl)
10^4	$N_{av}/0.1\mu\text{l} \rightarrow N_{av}/1\text{ ml}$
c	Cell concentration [cells/ml]

2.2.2.5 Kill curve

Kill curves are generated to determine the selection marker concentration that is necessary to kill the complete population of a cultured, non-resistant cell line within 5 days. Knowing this kill-concentration of the selection marker will turn out to be very useful when a cell population is selected for cells that were stably transfected with a foreign plasmid during the process of transfection (see Section 2.2.2.6). This is because after transfection basically two subpopulations

exist: Cells belonging to subpopulation I are not resistant against the selection marker as they did not take up the plasmid, which not only contains the gene of interest but also the resistance gene that confers resistance against the selection marker. Cells belonging to subpopulation II however did take up the plasmid and are therefore resistant against the selection marker. Consequently, by exposing a cell population to the kill-concentration of the selection marker, all cells belonging to subpopulation I will be dead after 5 days while the stably transfected cells of subpopulation II will thrive.

Protocol: Untransfected cells were trypsinated as described in Section 2.2.2.1, followed by their resuspension in culture medium without selection marker. Cells were then counted as described in Section 2.2.2.4, followed by the seeding of 50,000 cells in each well of a 6-well plate. In each well a different concentration of selection marker was generated (in case of G418 final concentrations were 0.0, 0.5, 0.8, 1.0, 1.5, and 1.8 mg/ml). After 5 days cells were evaluated with a light microscope and the kill-concentration determined as describe above.

2.2.2.6 Transfection

Transfection generally refers to the transfer of foreign DNA (usually a plasmid carrying the gene of interest, i. e. the gene to be expressed, and a resistance marker gene) into eukaryotic cells. Two different forms of transfection can be distinguished. During a transient transfection, the foreign DNA is only temporarily taken up by the host cell and is usually degraded within a few days. During a stable transfection, however, the foreign DNA is integrated in the host genome and is therefore much more protected against aggregation. To ensure that the foreign DNA remains integrated in the genome, stably transfected host cells have to be kept under the constant selection pressure of the selection marker that is added to the growth medium of the cells.

Protocol: Stable transfections of CHO Lec 3.2.8.1 cells were performed with *Lipofectamine 2000* and by following the manufacturer's instructions. Cells were grown as described in Section 2.2.2.1. In case of UL16-Fc, CHO Lec 3.2.8.1 cells were stably transfected with a recombinant pcDNA3.1(-) vector (see Section 2.1.4.1) containing the genetic information of the UL16-Fc construct described in Section 2.2.3.5.1. After transfection, cells were grown as described in Sections 2.2.2.1 and 2.2.2.2.

2.2.2.7 Subcloning

By performing a stable transfection many cells of the cell population are transfected with the foreign DNA. It is known, however, that individual cells exhibit different phenotypes in terms of protein expression after successful transfection. In other words stably transfected cells, after their selection, can be considered as polyclonal in terms of protein expression. Subcloning is the procedure to isolate individual protein expressing cells (the clones) in order to test them in terms of protein expression. The clone with the highest level of protein expression is then expanded and used to produce the protein. This single clone usually exhibits a much higher protein expression level than the polyclonal culture.

Protocol: CHO Lec 3.2.8.1 cells were trypsinated as described in Section 2.2.2.1. Cells were resuspended in CHO Lec selection medium (see Table 2-7) and the cell concentration counted as described in Section 2.2.2.4. This stock cell suspension was further diluted to a final cell concentration of 1 cell/200 μ l and 200 μ l of this suspension transferred to each well of a flat bottom 96-well plate. This should theoretically result in the seeding of 1 cell (clone) per well. Statistically, however, also wells with zero or more than one clone can also be found. The determination of the clone number per well is performed using a light microscope and by counting the number of cell colonies, each of which is easily visible and based on one seeded cell. Eventually, only the populations that were found in wells with only one colony were expanded and analyzed for protein expression by performing a Protein A/G pull-down assay.

2.2.2.7.1 Protein A/G pull-down assay

The bacterial cell wall proteins Protein A and Protein G were isolated from *Staphylococcus aureus* and group G *streptococci*, respectively. They exhibit molecular weights of 42 (Protein A) and 35 kDa (Protein G) and bind the Fc region of immunoglobulin G (IgG) type antibodies with nanomolar affinity¹⁰⁷. This binding feature is exploited in Protein A and/or Protein G mediated purifications, pull-downs, or immunoprecipitations of IgG subclass antibodies and Fc-tagged proteins, respectively^{107, 108}. Note that recombinant versions of Protein A and Protein G have been developed (e. g. recombinant Protein A/G) in order to improve and/or combine the features of the two individual proteins.

Protocol: Each UL16-Fc expressing CHO Lec 3.2.8.1 cell clone was grown in a separate well of a 6-well plate until a confluency of 100 % was reached. The old medium was exchanged against 2 ml fresh CHO Lec 3.2.8.1 culture medium (Table 2-7), and the cells were incubated for 2 days. 750 μ l

of the UL16-Fc containing supernatant of this culture were removed and filtered using a *Spin-X centrifuge tube filter* (0.22 μm). In addition 10 μl of *Protein A/G UltraLink resin* slurry were transferred to a 1.5 ml reaction vial and washed 3 times with 1 ml of Protein A binding buffer (Table 2-16). During each washing step, Protein A/G beads were resuspended in Protein A binding buffer followed by centrifugation and the disposal of the wash solution in the supernatant. All centrifugation steps using Protein A/G beads were performed at 4,000x g. The Protein A/G pellet obtained after the third centrifugation step was mixed with 750 μl of filtered UL16-Fc supernatant (see above) and the entire suspension incubated for 2 hrs at 4 °C. UL16-Fc bound Protein A/G beads were washed as described above. The UL16-Fc containing Protein A/G pellet obtained after the third centrifugation step was dissolved in 20 μl of reducing SDS sample buffer (Table 2-21), boiled and analyzed by SDS PAGE for UL16-Fc expression that should result in a UL16-Fc protein band at 60 kDa. The clone with the best UL16-Fc expression (indicated by the strongest UL16-Fc band at 60 kDa on the gel) was then used for further UL16-Fc production.

2.2.3 Methods in protein biochemistry

2.2.3.1 Protein concentration determination

Over the years, various assay types have been developed to measure the protein concentration in a given solution. Two of the most common assay types are the Bradford assay and the $\lambda 280$ assay^{109, 110}.

2.2.3.1.1 Bradford assay

The Bradford assay is based on the dye Coomassie Brilliant-Blue G-250. Coomassie binds to arginine and/or hydrophobic amino acid side chains of proteins and absorbs light at 595 nm in this protein-bound state. In general, the Bradford assay is performed in two steps: (Step 1) A number of five to ten BSA samples with distinct, known protein concentrations are prepared and then mixed with Bradford reagent. After a short incubation period the absorption of each sample is measured at 595 nm (A_{595} value) and plotted against the corresponding (known) BSA concentrations in order to derive a standard curve. (Step 2) Then the protein sample with the unknown protein concentration is subjected to the same reaction as the BSA samples and its concentration determined by relating the measured A_{595} value to a protein concentration via the standard curve obtained after step 1.

Protocol: The Bradford assay was performed with the *Bradford Protein Assay* reagent and by following the manufacturer's instructions. The absorption of the sample was measured at 595 nm

(A_{595}) using a spectrophotometer and a 1 ml plastic cuvette. The protein concentration was then calculated based on an equation (Equation 2-3) that had been derived by Pierre Schelling from a BSA calibration curve.

$$c = \frac{A_{595} - 0.004}{0.051 \cdot v} \quad \text{Equation 2-3}$$

with

A_{595} Absorption at 595 nm
 v Sample volume in μl
 c Concentration in mg/ml

2.2.3.1.2 λ 280 assay

By measuring the absorbance of a protein solution at 280 nm (A_{280}) and by knowing the theoretical extinction coefficient of the protein at this wavelength (ϵ_{280}) one can determine the protein concentration in mg/ml using a modified form of Lambert-Beer's law (see Equation 2-4)¹¹¹.

$$c = \frac{A_{280} \cdot M_w}{\epsilon_{280} \cdot d} \quad \text{Equation 2-4}$$

with

A_{280} Absorption at 280 nm
 ϵ_{280} Molar decadic extinction coefficient [$\text{M}^{-1}\text{cm}^{-1}$]
 d Path length of light [cm] = 1
 M_w Molecular weight [g/mol]
 c Concentration [mg/ml]

The extinction coefficient is calculated as shown in Equation 2-5.

$$\epsilon_{280} = n_{\text{Tryptophan}} \cdot 5500 + n_{\text{Tyrosine}} \cdot 1490 + n_{\text{Cystine}} \cdot 125 \quad \text{Equation 2-5}$$

with

n number of the particular amino acid in the protein

The extinction coefficients and molecular weights (without glycosylation) for UL16-Fc, UL16_{mon}, MICB_{short} and MICB_{long} are listed in Table 2-34 and were calculated with the online tool *ProtParam*.

Table 2-34. Extinction coefficients and molecular weights

Components	ϵ_{280} [$M^{-1} \text{ cm}^{-1}$]	Molecular weight [g/mol]
UL16-Fc	62715	44400
UL16 _{mon}	27055	18524
MICB _{short}	25502	23011
MICB _{long}	46535	32702

Protocol: First a reference sample containing pure protein buffer was used to blank the spectrophotometer. Then, 70 μl of protein buffer were added to 5 μl of protein sample followed by the transfer of this mixture into a 75 μl quartz cuvette. The absorption of the sample was measured at 280 nm (A_{280}) and the concentration of the sample determined using Equation 2-4. The obtained concentration value was then corrected by taking the dilution factor into account. Alternatively, the absorbance at 280 nm of undiluted protein samples was measured with a *Nanodrop ND-1,000* photometer.

2.2.3.2 Discontinuous SDS-PAGE

The discontinuous SDS-PAGE is an analytical method to separate proteins based on their molecular weight. Moreover, the use of protein markers, which are comprised of proteins with defined molecular weights, allows for the determination of the molecular weight of the separated proteins. However, to perform this kind of analysis protein samples have to be first diluted in SDS containing sample buffer (Table 2-21) and then boiled for 5-10 minutes. Together, both heat and SDS denature the proteins entirely. The negatively charged SDS also performs a second function. It can bind to the denatured protein (1.4 g SDS per g protein in a 1 % SDS solution) with its hydrophobic portion and thereby coats the protein evenly with the negative charges of its sulphate head group. As a consequence, all proteins will obtain a similar charge/mass ratio, which in turn will lead to the same electrophoretic mobility of each protein during electrophoresis. This explains why proteins during a SDS-PAGE are solely separated by their molecular weight and not by a combination of their molecular weight and native charge.

Protocol: Solutions for the SDS stacking and resolving gels were prepared as described in Table 2-22. Gels were cast using a casting chamber. Protein samples were mixed with the appropriate amount of 4x sample buffer, loaded onto the SDS-Gel and run for 1.5 hrs at 120 V. Gels were transferred to Coomassie staining solution (Table 2-23) and heated in a microwave until the Coomassie solution was hand warm, followed by a staining period of 1 h on an orbital shaker. Gels

were then transferred to destaining solution (Table 2-23), heated again in a microwave until the solution was hand warm, and then left to destain for about 30 min. Finally, gels were evaluated and scanned for documentation.

Note that protein crystals can also be analyzed by SDS PAGE (see Section 3.3.1). However, to detect the nanogram quantities of protein that is present in a protein crystal, a silver stain of the respective SDS gels was performed following a protocol developed by Nesterenko et al.¹¹². The detection limit of a normal Coomassie staining lies in the range between 50 - 100 ng of protein, while the detection limit of a silver stained gel lies between 1 - 10 ng.

2.2.3.3 General aspects of protein purifications performed with FPLC systems

The available chromatographic techniques for the purification of (recombinant) proteins can be divided in four groups based on the protein's properties that are exploited.

- ⇒ Size exclusion chromatography, SEC (also called gel filtration (GF) chromatography) exploits the protein's size but is also dependent on the proteins shape
- ⇒ Ion exchange chromatography (IEX) exploits the protein's charge
- ⇒ Hydrophobic interaction chromatography (HIC) exploits the protein's hydrophobicity
- ⇒ Affinity chromatography (AC) exploits the protein's biorecognition.

In AC methods a ligand (i. e. protein, peptide or small molecule) is coupled to the matrix of a chromatographic column and the solution containing soluble protein is passed over this matrix. Proteins in the solution will recognize their immobilized native ligands, bind to it (biorecognition), and will therefore be retained. However, often this biorecognition property is not a feature of the protein itself but instead conferred by a so-called protein-tag (i. e. a short peptide stretch or a complete protein) that is fused to either the N- or the C-terminus of the protein sequence. These proteins are therefore also called 'fusion proteins'.

In general, protein purifications using one of the chromatographic techniques described above are carried out at 4 °C. The main reason for this is that the cells that produce the recombinant protein also express proteases for their own metabolism. If released during the purification procedure, these proteases can cleave and therefore harm the expressed protein. Lower temperatures combined with protease inhibitors are usually sufficient to minimize protease activity in a cellular lysate or in any other sample that contains protein.

Protocol: All FPLC purification steps were carried out at 4 °C. Buffers were adjusted to the desired pH at 4 °C and were degassed and filtered before use. The FPLC systems that were used in each purification are listed in the respective sections.

2.2.3.4 Constructs, expression, refolding and purification of MICB_{short} and MICB_{long}

2.2.3.4.1 Constructs

The original MICB construct (Figure 2-1) was provided by Prof. Dr. Alexander Steinle and consisted of gene sequences encoding ectodomains $\alpha 1$, $\alpha 2$ and $\alpha 3$ of MICB*002 (named MICB_{long}), which encompassed MICB residues 1-276 that were fused via a two amino acid (Leu-Glu) linker region (not shown in Figure 2-1) to a C-terminal His₆-tag. This construct had been cloned into a pET-21a(+) vector (see Section 2.1.4.2 for details) using the restriction enzymes Nde I and Xho I (see Sections 2.1.3.8 and 2.2.1.6). Using the original MICB construct as a template, a shorter MICB construct (named MICB_{short}) was designed that consisted of gene sequences encoding only ectodomains $\alpha 1$ and $\alpha 2$ (the so-called platform-domain) of MICB (residues 1-184), which were fused to a C-terminal thrombin cleavage site followed by a His₈-tag (Figure 2-1). Like the MICB_{long} construct, also the MICB_{short} construct was cloned into the pET-21a(+) vector using the restriction enzymes Nde I and Xho I (see Sections 2.1.3.8 and 2.2.1.6). The full length sequence of MICB (UniProt: Q29980) is shown in Figure 2-2.

The molecular weights of recombinant proteins MICB_{long}-His₆ and MICB_{short}-His₈ were calculated as 32.7 kDa and 23.0 kDa, respectively.

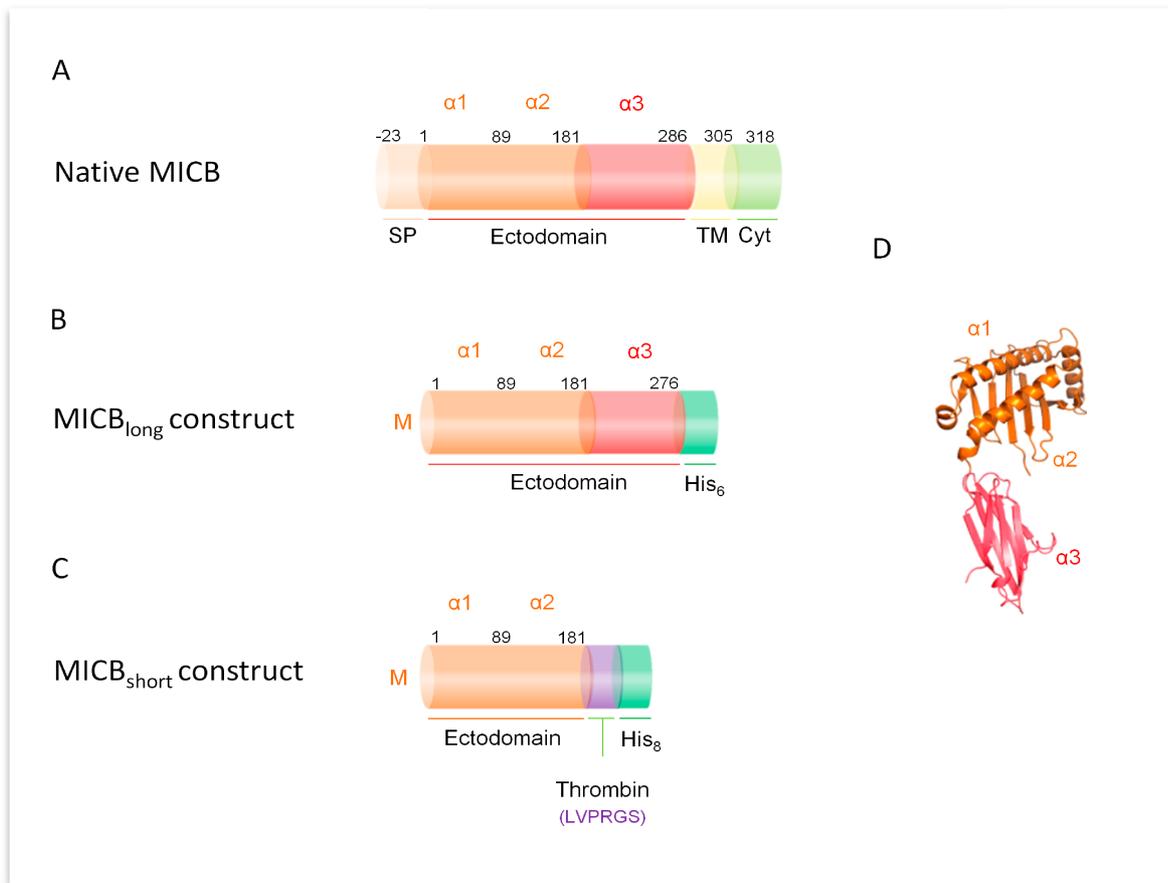


Figure 2-1. MICB constructs. (A-C) Schematic representations of native MICB protein prior to signal peptide cleavage in the ER (panel A) and the designed constructs of MICB_{long} (panel B) and MICB_{short} (panel C) expressed in *E.coli*. Color codes are as follows: ER signal peptide (SP) – bright orange, $\alpha 1\alpha 2$ platform-domain – orange; $\alpha 3$ -domain – red; transmembrane domain (TM) – yellow; cytoplasmic tail (Cyt) – bright green, thrombin cleavage site – purple and His_{8/6}-tags – green. Amino acids in the MICB sequence that mark the beginning and end of the protein regions are indicated above the tubes. The six purple letters in panel C represent the amino acid recognition sequence of thrombin located between the $\alpha 1\alpha 2$ platform-domain of MICB_{short} and its His₈-tag. The capital letter ‘M’ in panels B and C stands for a N-terminal methionine residue in the MICB constructs that is not part of the original MICB sequence. (D) Ribbon tracings of the $\alpha 1\alpha 2$ platform-domain and the membrane proximal $\alpha 3$ domain of MICB (pdb code: 1je6⁶⁴). Domains are colored as in panels A-C.

1	MGLGRVLLFL AVAFPFAPPA AAAEPHSLRY NLMVLSQDGS VQSGFLAEGH	50
51	LDGQPFLRYD RQKRRAKPQG QWAEDVLGAE TWDTEDELT ENGQDLRRTL	100
101	THIKDQKGGI HSLQEIRVCE IHEDSSTRGS RHFYYNGELF LSONLETQES	150
151	TVPQSSRAQT LAMNVTNFWK EDAMKTKTHY RAMQADCLQK LQRYLKSGVA	200
201	IRRTVPPMVN VT SEVSEGN ITVT RASSF YPRNITLTWR QDGVLSHNT	250
251	QQWGDVLPDG NGTYQTWVAT RIRQEEQRF T YMEHSGNH GTHPVPSGKA	300
301	LVLQSQRTEF PYVSAAMPF VIIIIICVPC CKKKTSAEGL PELVSLQVLD	350
351	QHPVGTGDHR DAAQLGFQPL MSATGSTGST EGT	384

Figure 2-2. MICB full length sequence. Shown is the full length sequence of MICB allele 002 prior to signal peptide cleavage in the ER. The five protein regions of MICB (signal peptide, $\alpha 1\alpha 2$ platform-domain, $\alpha 3$ -domain, transmembrane region and cytoplasmic tail) are colored as in Figure 2-1. Cysteines are colored in pink.

2.2.3.4.2 Expression

For decades, *E.coli* has been the first choice for the expression of recombinant proteins. This is owed due to the many beneficial features of *E.coli*: (i) The most widely used *E.coli* strain is the K12 strain, which represents a cultivated strain that has lost its ability to thrive in the human intestine and is therefore very safe to work with. (ii) In addition, foreign genes can be easily introduced into *E.coli* and efficiently selected for with the use of bacterial antibiotics. (iii) Furthermore, *E.coli* not only exhibits exponential growth rates but also high protein expression rates. And last but not least (iv) culturing and storing of *E.coli* cells is usually very straight forward and inexpensive.

Protocol: Both His-tagged MICB_{long} and MICB_{short} constructs were expressed and purified with the same strategy (Figure 2-3): *E.coli* Rosetta 2(DE3) cells (see Section 2.1.6.1) transformed (see Section 2.2.1.9) with the appropriate expression construct were grown in LB medium (Table 2-4), supplemented with 50 µg/ml ampicillin and 34 µg/ml chloramphenicol, to an optical density (OD)₆₀₀ of 0.5 - 0.6. Protein expression was then induced by adding IPTG to the bacterial suspension to make a final concentration of 1 mM, followed by a 4 h incubation period at 37 °C. Bacterial cells were centrifuged at 4,000x g for 20 min and the bacterial pellet resuspended in 150 ml lysis buffer (Table 2-11). Cells were lysed using a high pressure homogenizer followed by the centrifugation of the bacterial lysate at 27,000x g for 20 min. After centrifugation, the bacterial lysate consisted of two distinct solid layers below a clear supernatant. The supernatant was discarded and the soft upper layer carefully removed with a spatula. The hard and brownish lower layer contained the MICB inclusion bodies (IB).

2.2.3.4.3 Refolding

The *E.coli* expression system has many benefits (see Section 2.2.3.4.2). However, the downside of this system is that the production of soluble, natively folded forms of recombinant proteins remains unpredictable. Often instead of soluble, natively folded recombinant proteins, insoluble and misfolded *E.coli* inclusion bodies (IB) are obtained. Although many of the reasons for the formation of IBs remain elusive, clearly (i) the reducing environment of *E.coli*, which prevents the formation of disulfide bonds that are often essential for the proper folding and stability of eukaryotic proteins, and (ii) the inability of *E.coli* to produce post-translational modifications such as glycosylation, contributes to the formation of IBs. Using refolding methods the native form of the recombinant protein can in some cases be reconstituted from the IBs. In a first step, the misfolded proteins are completely unfolded to their random coiled state by strong denaturants such

as urea or guanidine hydrochloride (GuHCl), followed by the stepwise (dialysis approach) or one-step (rapid dilution approach) reduction of the denaturant in refolding buffer, during which the proteins fold back into their native form.

Protocol: IBs were resuspended in Triton X-100 wash buffer (Table 2-12) using a 40 ml Potter homogenizer. The inclusion body suspension was centrifuged at 27,000x g for 20 min. After centrifugation, the supernatant was discarded and the inclusion body pellet again transferred into the Potter homogenizer. This inclusion body cleaning cycle was repeated about 5 times until an almost white inclusion body pellet was obtained. One last cycle was then performed to remove the Triton X-100 detergent with Triton-free wash buffer (Table 2-12). Without disturbing the inclusion body pellet after this last cycle, the centrifuge tube was filled with 25 ml of 8 M urea buffer (Table 2-13), sealed and rotated over night in the cold room. In the morning, the pellet had dissolved completely, and oxidized and reduced glutathione (final concentrations of 0.5 and 5 mM, respectively) were directly added to the MICB containing 8 M urea buffer, followed by another 24 h rotation of the sample at 4 °C. The inclusion bodies were refolded by stepwise arginine/urea dialysis. Briefly, 8 M urea buffer (Table 2-13) containing MICB was dialyzed step by step (each dialysis step taking 24 h) against refolding buffer (Table 2-14). During each of the three steps the concentrations of urea (4M→2M→1M→0M) were decreased (see Table 2-14). In the fourth dialysis step, the arginine concentration was reduced from 400 to 100 mM, followed by a final (fifth) dialysis step during which MICB was dialysed against TBS buffer (pH 8; see Table 2-10). During this last step a significant portion of MICB precipitated and was removed by centrifugation at 27,000x g for 20 min and subsequent filtration (0.22 µm) using a *Steritop Express Plus* filter (0.22 µm). The obtained clear solution contained the soluble MICB and was concentrated using either an *Amicon Ultra 4 or 15* concentrator (MWCO: 5,000 Da). Finally, the concentrated MICB sample was again filtered using a *Steritop ExpressTM Plus* filter (0.22 µm).

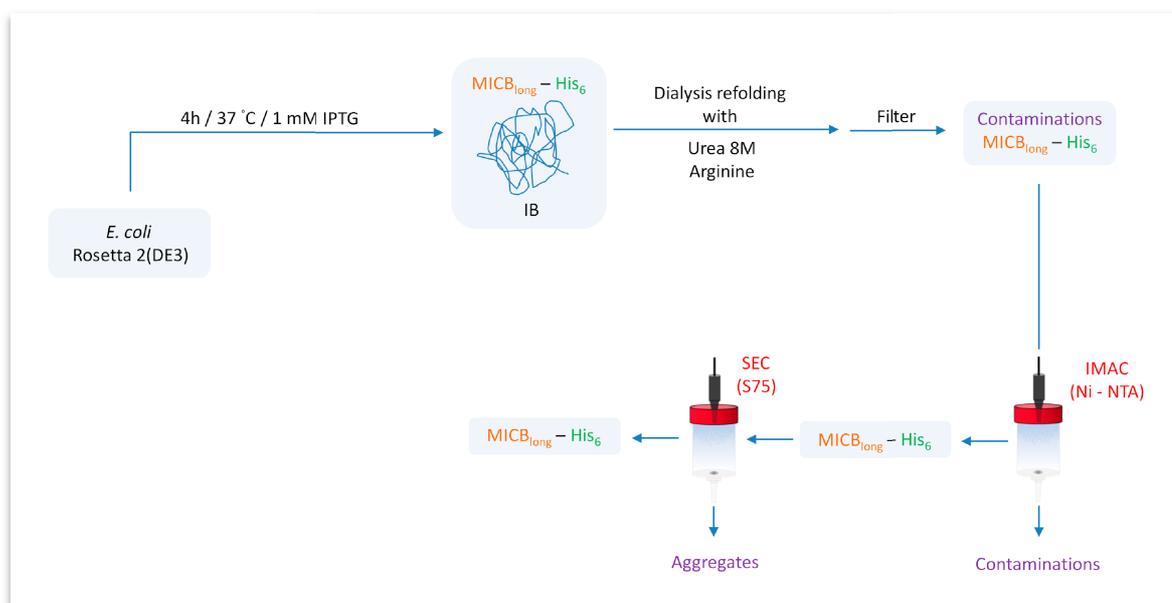


Figure 2-3. MICB expression and purification scheme. Shown are the essential steps of the developed MICB_{long} expression and purification strategy. The same strategy without alteration was also used for MICB_{short}. MICBs were produced by Rosetta 2(DE3) cells in form of insoluble IBs. IBs were refolded and soluble MICB purified by means of IMAC using a *HisTrap* (Ni-NTA) HP column followed by SEC using a *Superdex 75 10/300 GL* column.

2.2.3.4.4 IMAC purification

In immobilized metal affinity chromatography (IMAC), target proteins are fused to a tag consisting of 6-10 histidine (His) residues. Two histidines at a time can then, via their imidazole ring system, establish coordinative bonds with a transition metal ion (i. e. Ni²⁺ or Co²⁺) ligand, which is immobilized to the column matrix via chelating resins such as nitrilo triacetic acid (NTA) or Talon, respectively. The resin itself is covalently bound to the stationary phase of the column. Bound proteins are then eluted with an imidazole gradient, in which the imidazole competes with the bound proteins for metal binding.

Protocol: MICB was loaded onto a *HisTrap* (Ni-NTA) HP column equilibrated with Ni-NTA binding buffer (Table 2-15) and connected to an *Äkta prime* FPLC system. After MICB loading was complete, the column was extensively washed with Ni-NTA binding buffer (Table 2-15), followed by gradient elution of the His-tagged MICB with Ni-NTA elution buffer (length of 0 - 100 % imidazole gradient: 75 ml). Fractions were collected during the entire run and analyzed via SDS PAGE. MICB containing samples were pooled, reequilibrated in TBS buffer (pH 8; see Table 2-10) and concentrated as described in Section 2.2.3.4.3. The entire IMAC run was performed at a flow rate of 1 ml/min and monitored and evaluated using the *PrimeView* software.

2.2.3.4.5 SEC purification

SEC is a very simple, fast and mild form of chromatography that separates proteins according to their differences in size and shape. The stationary phase (the matrix) consists of beads that contain pores of a defined average size that contain the same buffer used in the mobile phase. When a mixture of different molecules is passing through the column, the small molecules can easily penetrate into every part of the pore system. However, when the size of the molecule increases not the whole pore system is accessible to it anymore. At some point the protein size exceeds the size exclusion limit and no further permeation of the molecule into the pore system is permitted. Eventually this translates to the fact that larger molecules will move faster through and elute earlier from the column. However, all molecules (unless unspecificly bound to the matrix) will elute before one total column volume of buffer has passed through the column. An important characteristic of every SEC column is its resolution, which is dependent on many factors such as the packing density, pore size, sample volume, flow rate and column length. In addition, SEC can also be used to determine the size of proteins. To do so, a standard curve is created by plotting the elution time or volume of standard proteins against their size. Having determined the time or volume at which a protein with unknown molecular weight elutes from the column, the weight of the protein can be determined by use of the standard curve.

Protocol: MICB was loaded on a *Superdex 75 10/300 GL* (column volume: 24 ml) equilibrated with TBS (pH 8; Table 2-10) and connected to a *BioLogic DuoFlow* FPLC system. 0.5 ml MICB samples were loaded on the column and eluted with 24 ml of TBS buffer (pH 8; Table 2-10). Fractions of 0.5 ml volume were collected during the entire run and analyzed via SDS-PAGE. MICB containing fractions were pooled, concentrated as described before (see Section 2.2.3.4.3), filtered with a *Spin-X centrifugal tube filter* (0.22 μm) and finally stored at 4 °C. The entire SEC run was performed at a flow rate of 0.5 ml/min and monitored using the *BioLogic Duo Flow* software.

2.2.3.5 Constructs, expression and purification of UL16

2.2.3.5.1 Constructs

The original UL16 construct (see Figure 2-4) was prepared by Dr. Jessica Spreu and provided by Prof. Dr. Alexander Steinle. It was used throughout this thesis without any alteration. Briefly, a recombinant cDNA fragment including the N-terminal signal peptide (residues 1-26) and the ectodomain (residues 27-184) of UL16 was fused to a C-terminal thrombin cleavage site followed

by the human IgG1-Fc sequence and cloned into a pcDNA3.1(-) vector (see Section 2.1.4.1 for details). The full length sequence of UL16 (UniProt: P16757) is shown in Figure 1-6. The four UL16 regions were predicted with *SignalIP 3.0* and *TMHMM*.

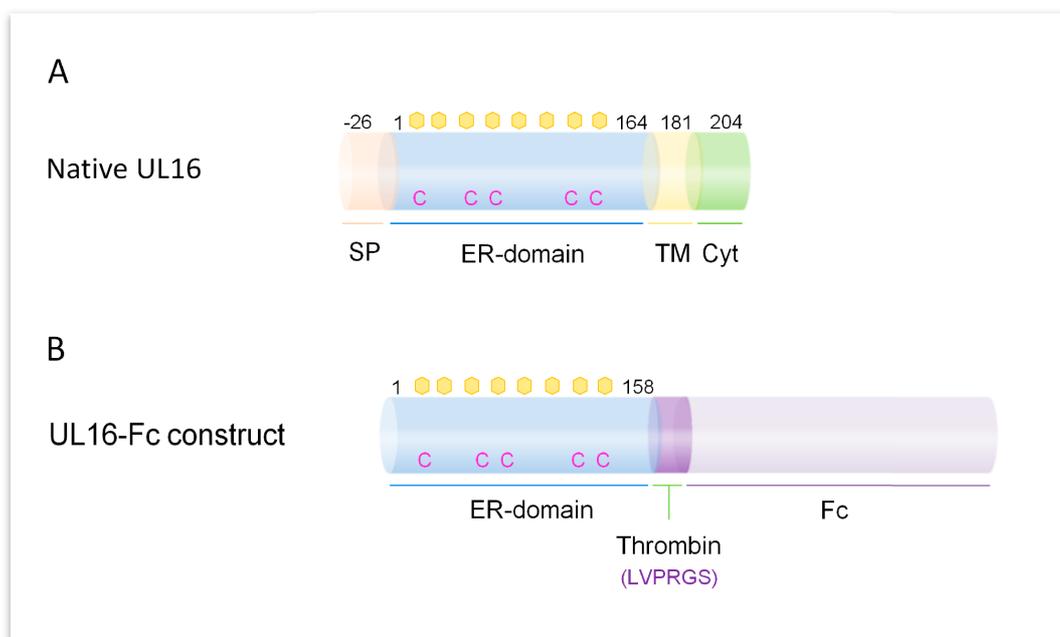


Figure 2-4. UL16 constructs. (A, B) Schematic representations of native UL16 protein prior to signal peptide cleavage in the ER (panel A) and the UL16-Fc construct (panel B) expressed in CV-1 and CHO Lec 3.2.8.1 cells, respectively. The four UL16 protein regions are colored as in Figure 1-6. Amino acids in the UL16 sequence that mark the beginning and end of these protein regions are indicated. The six purple letters in panel B represent the amino acid recognition sequence of thrombin located between the ER-domain of UL16-Fc and its Fc-tag (bright purple). The yellow hexagons represent the eight potential N-linked glycosylation sites in the UL16 ER-domain. The five cysteines present in UL16 are colored in pink (see also Figure 1-6).

2.2.3.5.2 Expression of UL16-Fc

UL16-Fc was expressed in either CV-1 or CHO Lec 3.2.8.1 cells. For details see Sections 2.2.2.1 and 2.2.2.2.

2.2.3.5.3 Protein A purification

The characteristics of Protein A are described in Section 2.2.2.7.1.

Protocol: For each purification, 10 l of cell culture supernatant (from either CHO Lec 3.2.8.1 cells or CV-1 cells; see Section 2.2.2 for details) were filtered through a *Steritop Express™ Plus* filter (0.22 µm), adjusted to pH 9 with 5 M NaOH, and loaded o/n at a flow rate of 5 ml/min onto two serially connected *Protein A HP* columns (5 ml column volume each) using a *Peristaltic EconoPump*. After supernatant loading was finished, the columns were extensively washed with

Protein A binding buffer (Table 2-16). The protein was then eluted in one step (flow rate: 2.5 ml) with 100 % arginine solution (Table 2-16) directly into HEPES reservoir buffer (Table 2-17). Fractions containing UL16-Fc were pooled, dialysed against TBS (pH 8 at 22 °C; Table 2-10), and concentrated using an *Amicon Ultra 15* concentrator with an MWCO of 5,000 Da. Thrombin cleavage was performed with 1 U/mg recombinant protein at 22 °C for 18 h. The cleaved sample was filtered, diluted five-fold in Protein A binding-buffer (Table 2-16), and run at a flow rate of 1 ml/min over two consecutive *Protein A HP* columns (1 ml column volume each) followed by a benzamidine column (1 ml column volume) to remove cleaved Fc tag and thrombin, respectively. Flow-through containing UL16 samples were concentrated and dialyzed against TBS buffer (pH 7.4; Table 2-10) for storage. For a schematic representation of the purification strategy see Figure 2-5.

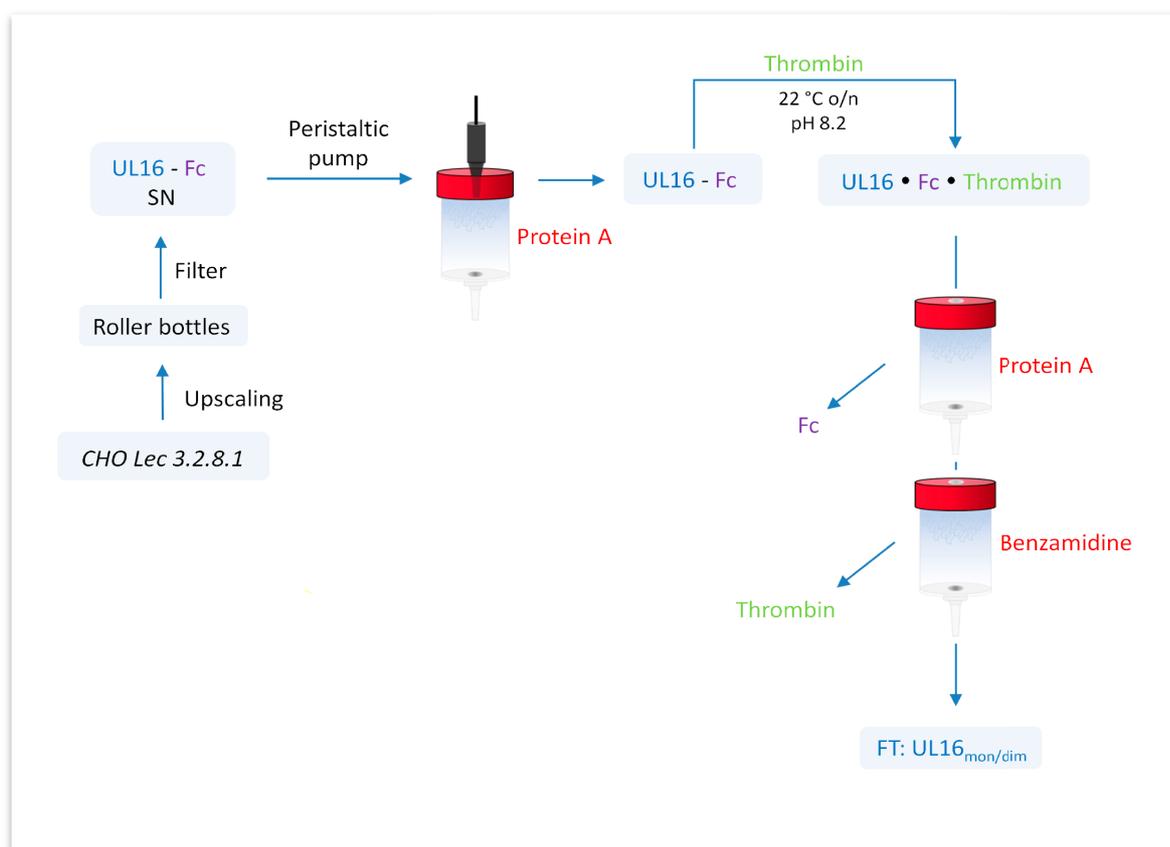


Figure 2-5. UL16 expression and purification scheme. Shown are the essential steps of the developed UL16 expression and purification strategy. UL16-Fc was produced by either CHO Lec 3.2.8.1 or CV-1 cells and purified by passing UL16-Fc containing cell supernatants over a Protein A column. After o/n cleavage of purified UL16-Fc with thrombin, the resulting mixture comprising UL16, Fc-tag and thrombin was passed over a *Protein A HP* column followed directly by a *Benzamidine FF* column in order to remove the cleaved Fc-tag (Fc) and thrombin, respectively. The UL16 however passed through the column and was consequently found in the flow-through (FT). As shown in Section 3.2.7, purified UL16 consists of a monomeric and dimeric species, which is why the purified UL16 in the schematic is named UL16_{mon/dim}.

2.2.3.5.4 IMAC and SEC purification of the UL16_{mon}-MICB_{long} complex

The Ni-NTA (IMAC) purification of the UL16_{mon}-MICB_{long} complex was essentially performed as described in Section 2.2.3.4.4 (IMAC purification of MICB). The SEC step was essentially performed as described in Section 2.2.3.4.5 (SEC purification of MICB). For a schematic representation of the purification strategy see also Figure 2-6.

2.2.3.5.5 IMAC purification of the UL16_{mon}-MICB_{short} complex

The Ni-NTA (IMAC) purification of the UL16_{mon}-MICB_{short} complex was performed exactly as described for the UL16_{mon}-MICB_{long} complex in the previous section.

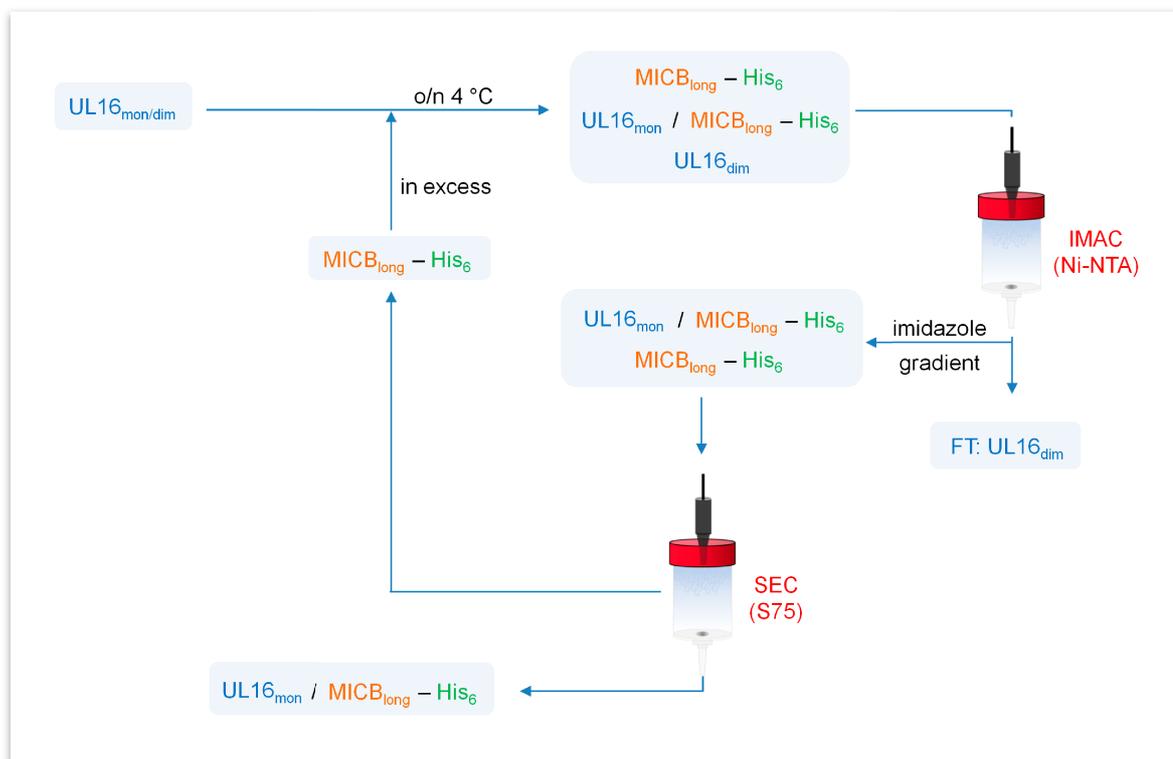


Figure 2-6. UL16_{mon}-MIC_{long} complex purification scheme. Shown are the essential steps of the developed UL16_{mon}-MIC_{long} complex purification strategy. A mixture of UL16 monomer and dimer (UL16_{mon/dim}) was incubated with an excess amount of purified MICB_{long} and the resulting mixture passed over an IMAC *HisTrap* (Ni-NTA) HP column. UL16_{mon}-MIC_{long} complex and MICB_{long} bound to the column and were eluted with an imidazole gradient. The UL16 dimer (UL16_{dim}) however passed through the column and was consequently found in the flow-through (FT). A *Superdex 75 10/300 GL* SEC column was used to separate excess MICB_{long} from the UL16_{mon}-MIC_{long} complex. The separated MICB_{long} was reused in the next purification cycle.

2.2.3.5.6 HIC purification

Hydrophobic interaction chromatography (HIC) exploits the interaction between hydrophobic surface regions of the proteins and the hydrophobic resins of the column matrix. As the hydrophobic regions of various proteins differ in size and hydrophobicity and therefore bind to the hydrophobic resin with different strength, HIC can be used to separate those proteins from each other. Normally, the hydrophobic regions of water-soluble proteins would only weakly interact with the hydrophobic resin (or the hydrophobic regions of a neighboring protein molecule) as their hydrophobic surface regions are shielded in aqueous solution by a cage of highly ordered ('frozen') water molecules. However, by diluting a protein in high-salt buffer the protein's frozen water layer can be removed. The reason for this is that salt ions, as they become strongly solvated in aqueous solution, do not only sequester the free water molecules but also hijack the protein's frozen water molecules once the free water molecules become scarce. The consequence of this so-called 'salting-out' effect is that hydrophobic surface regions of the proteins become exposed to the free water, which facilitates their interaction with the exposed hydrophobic matrix resin. The stepwise or gradient elution of the bound proteins from the column is achieved by stepwise or gradient elution with salt free buffers. Two of the most important variables in HIC (which have to be determined by 'try and error') are (i) the type of hydrophobic resin used and (ii) the type of salt used to dehydrate the proteins in the first place.

Protocol: Briefly, the UL16 monomer/dimer mixture was diluted in HIC-binding buffer, loaded at a flow rate of 2 ml/min onto three serially connected phenyl sepharose columns (5 ml column volume each) and washed extensively with HIC-binding buffer (Table 2-18). Bound UL16_{mon} was eluted in one step with 100 % HIC-elution buffer (Table 2-18) at a flow rate of 2 ml/min and dialyzed against TBS buffer (pH 7.4; Table 2-10). The yield of monomeric UL16_{mon} was 0.2 mg per 1 liter of cell culture supernatant. All chromatographic steps, except the loading of the Protein A columns with cell culture supernatant, were performed with an *Äkta prime* FPLC system and monitored and evaluated using the *PrimeView* software. For a schematic representation of the purification strategy see also Figure 2-7 and compare with Figure 2-5.

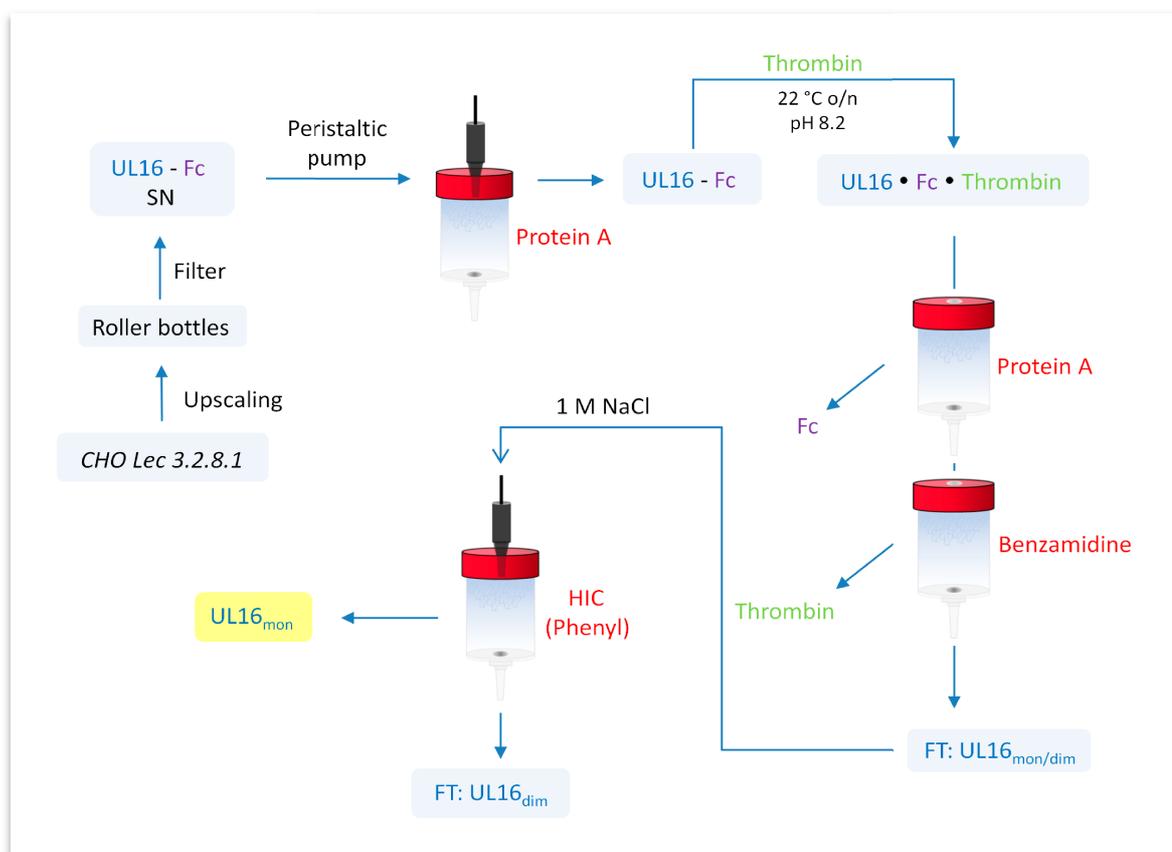


Figure 2-7. UL16_{mon} purification scheme. Shown are the essential steps of the UL16_{mon} purification strategy. UL16_{mon/dim} sample purified as described and shown in Section 3.2.5 and Figure 2-5, respectively, was diluted in 1 M NaCl and then passed over a HIC *Phenyl-HP* column. UL16_{mon} bound to the column and was eluted in one step with 100 % HIC elution buffer (Table 2-18). The UL16 dimer (UL16_{dim}) however passed through the column and was consequently found in the flow-through (FT).

2.2.3.6 Purification of the glycosylated UL16_{mon}-MICB_{short} complex

UL16_{mon} and an excess amount of MICB_{short} were incubated for 16 hrs in TBS buffer (pH 7.4) at 4 °C. The complex was then separated from excess MICB_{short} by performing a SEC, using a *Superdex 75 10/300 GL column*. The SEC run was performed essentially as described in Section 2.2.3.4.5. The UL16_{mon}-MICB_{short} complex sample was further concentrated to 15 mg/ml using a *Microcon YM3* concentrator (MWCO of 3,000 Da), filtered using a *Spin-X centrifuge tube filter* (0.22 μm) and finally used for crystallization. For a schematic representation of the purification strategy see Figure 2-8.

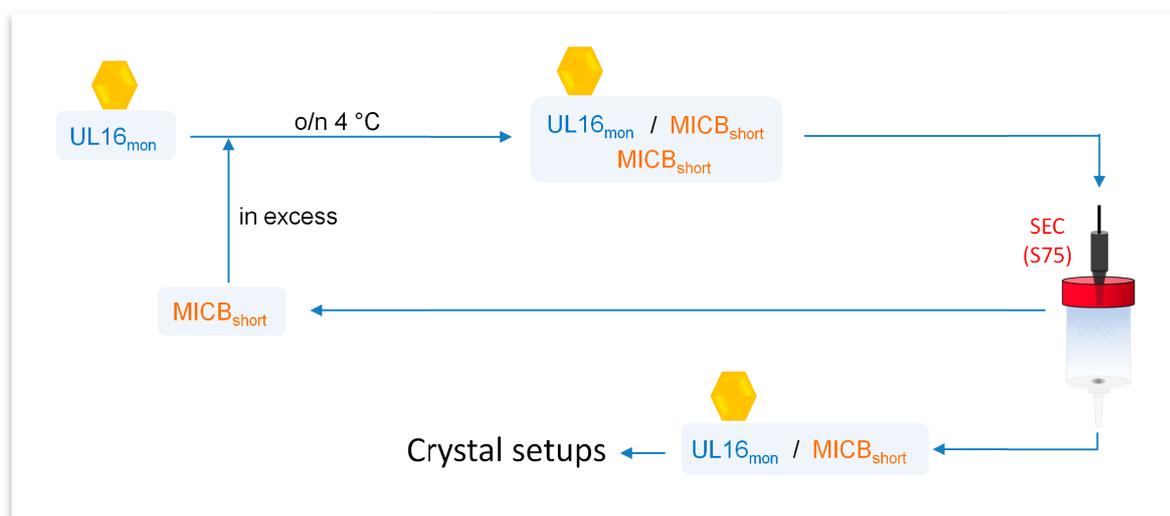


Figure 2-8. Purification of the glycosylated UL16_{mon}-MICB_{short} complex. Purified, glycosylated UL16_{mon} was incubated with an excess amount of purified MICB_{short} and the resulting two component solution passed over a *Superdex 75 10/300 GL* SEC column to separate free excess MICB_{short} from the formed UL16_{mon}-MICB_{short} complex. The orange hexagon represents the relative amount (compare Figure 2-9) of the CHO Lec 3.2.8.1 cell produced glycan moiety on the UL16 protein surface. Purified, glycosylated UL16_{mon}-MICB_{short} complex was used for crystal setups.

2.2.3.7 Purification of deglycosylated UL16_{mon}-MICB_{short} complex

UL16_{mon} and an excess amount of MICB_{short} were incubated for 16 hrs in TBS buffer (pH 7.4) at 4 °C. However, in order to obtain diffracting crystals, the UL16 produced in CHO Lec 3.2.8.1 cells had to be further deglycosylated. This was done by incubating the UL16_{mon}-MICB_{short} complex with Endo H (see Section 2.1.3.8), a glycosidase that cleaves high-mannose and hybrid type glycans in such a way that one N-acetylglucosamine residue (often abbreviated as NAG or GlcNAc) remains covalently connected to the Asn residue of the N-glycosylation site. Briefly, complex was 5 fold diluted in Endo H-buffer (Table 2-19), containing 1 U Endo H per µg complex and incubated for 1 h at 37 °C. To separate UL16_{mon}-MICB_{short} complex from cleaved glycans, Endo H and excess MICB, a SEC was performed using a *Superdex 75 10/300 GL* column. The SEC run was essentially performed as described in Section 2.2.3.4.5. The UL16_{mon}-MICB_{short} complex sample was further concentrated to 15 mg/ml using a *Microcon YM3* concentrator (MWCO of 3,000 Da), filtered using a *Spin-X centrifuge tube filter* (0.22 µm) and finally used for crystallization. For a schematic representation of the purification strategy see Figure 2-9.

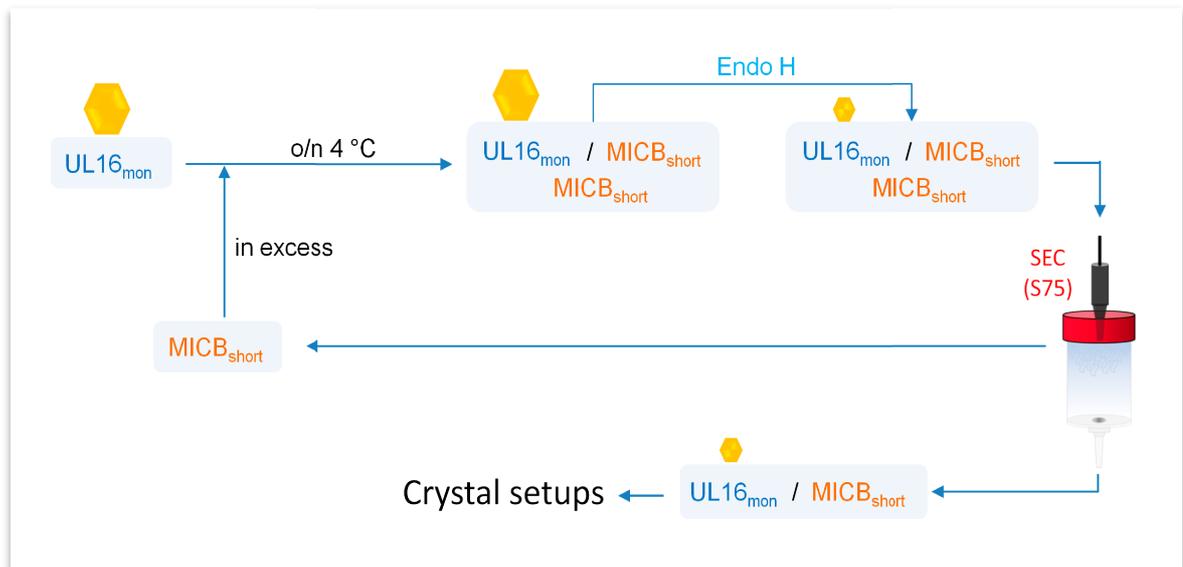


Figure 2-9. Purification of the deglycosylated UL16_{mon}-MICB_{short} complex. Purified, glycosylated UL16_{mon} was incubated with an excess amount of purified MICB_{short} in order to form the UL16_{mon}-MICB_{short} complex. The resulting two component solution consisted of UL16_{mon}-MICB_{short} complex and MICB_{long} and was incubated with Endo H in order to deglycosylate UL16 in its MICB-bound state. The orange hexagon represents the relative amount of the glycan moiety on the UL16 surface before and after the deglycosylation with Endo H (compare Figure 2-8). A *Superdex 75 10/300 GL* SEC was performed to separate cleaved glycans, Endo H and free excess MICB from the deglycosylated UL16_{mon}-MICB_{short} complex, which after its purification, was used for crystal setups.

2.2.3.8 Deglycosylation of CV-1 cell produced UL16

Endo H and PNGase F deglycosylation experiments with CV-1 cell produced UL16 were performed with the chemicals that were supplied with the enzymes and by following the manufacturer's instructions.

2.2.4 Surface Plasmon Resonance

Surface Plasmon Resonance (SPR) is a biosensor chip-based technology to measure biomolecular interactions (e. g. protein-protein, DNA-DNA or Protein-DNA interactions) in real-time. In order to do so, one of the binding partners (the so-called ligand) needs to be covalently (direct assay) or non-covalently (capture assay) immobilized to create the biosensor surface. The second binding partner (the so-called analyte) remains free in solution. First, the soluble analyte is serially diluted in running buffer, and is then injected in series and in a continuous flow manner over the biosensor surface. As analyte molecules from solution bind to the immobilized ligand (i. e. form the complex) a real-time response is generated due to the accumulating mass of analyte. A plot of this response, which is measured in resonance units (RU), against time is called a sensorgram and essentially shows the progress of the interaction (Figure 2-10). Ideally, sensorgrams of a simple 1:1 interaction

of molecules consist of two distinct phases. During the association phase an increase in response is observed due to the formation of the complex. However, at some point the analyte injection is stopped and instead of analyte, running buffer is passed over the biosensor surface. This point marks the beginning of the dissociation phase. During the dissociation phase a decrease in response is observed as complex molecules fall apart and analyte is carried away from the surface in the continuous buffer stream. Binding parameters such as the association and dissociation rate constants (k_a and k_d , respectively) and/or the dissociation constant (K_D) can then be determined by fitting the entire titration series to a predefined binding model with an appropriate software.

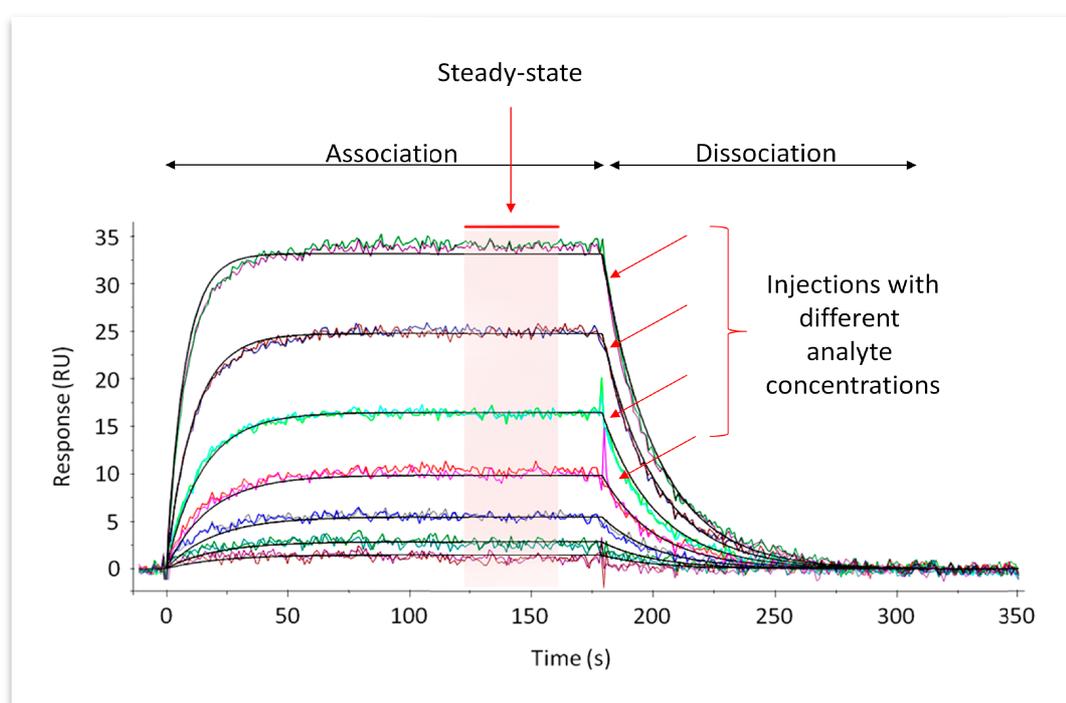


Figure 2-10. SPR data analysis. Shown is a typical SPR data set comprised of seven double-referenced sensorgrams, each of which corresponds to a defined but distinct concentration of injected analyte (in this case UL16_{mon}; see red slanting arrows). Each individual sensorgram is obtained by plotting the SPR response that is caused when the soluble analyte binds the surface immobilized ligand against time (measured usually in seconds). Also indicated are the regions of the individual sensorgrams that are used for the determination of kinetic and steady-state binding parameters, respectively. The association and dissociation phases are used to derive kinetic rate constants (k_a and k_d), while the steady-state phase is used to derive the dissociation constant (K_D). Alternatively, the dissociation constant (K_D) can also be calculated from kinetic rate constants by the relationship $K_D = k_d/k_a$. The figure was modified from Müller *et al.*, 2010¹¹³ (see also Figure 3-25).

Protocol: ‘All SPR experiments were performed and evaluated as described in detail in reference¹¹⁴. Using two consecutive flow cells on a CM5 biosensor chip, MICB_{short} and MICB_{long} ligands, respectively, were each covalently immobilized on the surface of the downstream (experimental) flow cell via amine-coupling chemistry following manufacturer’s instructions, while

the surface of the upstream (reference) flow cell was subjected to the same coupling reaction in the absence of protein. Untagged, monomeric UL16 analyte was serially diluted in running buffer and injected in series over the reference- and experimental (MICB) biosensor surface at 50 $\mu\text{l}/\text{min}$. MICB coated CM5 chips were not regenerated. For the Protein A/G chip preparation, an amount of 3,500 RU (resonance units) of recombinant Protein A/G was covalently immobilized to the upstream (reference) and downstream (experimental) flow cells of a CM5 biosensor chip by amine-coupling chemistry. Fc-tagged ULBP1, ULBP2, ULBP3, ULBP4 and ULBP5 ligands⁸⁹ were diluted in HBS-EP (Table 2-25) and noncovalently bound to the experimental flow cell Protein A/G surface (regarding the used ULBP proteins see also Section 2.1.3.9). After each cycle using a Protein A/G chip, the biosensor surface was regenerated (stripped of any remaining analyte and ligand) with two 1 min injections of 10 mM glycine pH 1.7. Untagged, monomeric UL16 analyte was serially diluted in running buffer and injected in series over the reference- and experimental biosensor surface at 50 $\mu\text{l}/\text{min}$.¹¹³

2.2.5 X-ray crystallography

2.2.5.1 X-rays and data collection

Objects are visible as they diffract electromagnetic radiation. However, in order to be diffracted by the object, the wavelength of the electromagnetic radiation must be of the same size as the object itself. This explains why visible light ($\lambda = 350 - 750 \text{ nm}$) is diffracted by macroscopic and microscopic sized objects, while X-rays ($\lambda = 0.1 - 10 \text{ \AA}$) are diffracted by smaller objects such as atoms, which exhibit a diameter in the range between 1 - 5 \AA . The wavelength of the electromagnetic radiation also determines the so-called resolution limit ($\lambda/2$), which is defined as the shortest distance between two objects at which these objects can still be determined as two singular entities by the observer. If the points are closer together than the resolution limit, they will blur together, making it impossible to distinguish between them. Bonded atoms in macromolecules such as proteins are about 1.5 \AA apart from each other, which explains why protein crystals are commonly exposed to X-ray wavelengths in the range between 0.5 - 2 \AA .

The wavelength used, can vary depending on the X-ray source available and the experimental conditions. X-ray home sources such as in our laboratory make use of the K-shell emission of a rotating copper anode that is positioned in a vacuum compartment and under continuous fire by a beam of accelerated electrons from a heated filament (the cathode) at a potential of 30 to 40 kV. When the accelerated electrons hit the anode, heat is produced. In addition, some of the electrons will also displace electrons from the lowest electron orbital (the K-

shell) of individual copper atoms. In the next step an electron from a higher orbital (the L or M shell) will drop into the empty position of the displaced K-shell electron and thereby emit the excess energy as X-ray photons. Electron transitions from the L→K shell and M→K shell will produce the Cu-K_α radiation ($\lambda = 1.54 \text{ \AA}$) and Cu-K_β radiation ($\lambda = 1.39 \text{ \AA}$), respectively. Since the K_β radiation is less intense than the Cu-K_α radiation and since furthermore it is pivotal to use only monochromatic radiation in an X-ray experiment, a nickel filter is used to essentially eliminate the Cu-K_β radiation.

A different source of X-rays are particle storage rings, which are called synchrotrons that were originally invented to study the physics of subatomic particles. However, synchrotrons have also turned out to be an invaluable asset in the X-ray structural analysis of protein crystals. If compared to X-ray home source systems synchrotrons have many advantages that usually result in better X-ray diffraction data quality: (i) A higher photon flux (i. e. more intense X-ray radiation beams) that often helps to obtain crystal datasets with higher resolution, especially if very small, very thin or very weakly diffracting crystals are measured. (ii) The tuneability of the X-ray wavelength the crystal is exposed to, which allows for the implementation of specific experiments that can help to solve the so-called phase problem. (iii) Detectors that are characterized by reduced signal to noise ratios and shorter data read-out times, which eventually leads to higher data quality and shorter overall measure times, respectively. (iv) And finally a smaller beam size that allows for the testing of different areas of a single crystal.

2.2.5.2 Protein crystal symmetry

Protein crystals consist of a highly ordered three-dimensional array of molecules, which are held together by non-covalent interactions. These non-covalent interactions between neighboring molecules in combination with the high amounts of solvent in a protein crystal make protein crystals very soft and fragile. The smallest identical repeating unit of a crystal is the unit cell, which can be viewed as the smallest building block of the crystal that can generate the entire crystal structure by means of translation operations in three dimensions. The unit cell is defined by six parameters, the lengths of the three cell axes (a , b and c) and the three angles between them ($a/b \rightarrow \gamma$, $a/c \rightarrow \beta$ and $b/c \rightarrow \alpha$). Defined constraints on these axial lengths and interaxial angles of a unit cell give rise to seven crystal systems (triclinic, monoclinic, orthorhombic, tetragonal, trigonal, hexagonal and cubic) with lattice points at each corner of the cell. Besides these primitive (P) unit cells, face (F), body (I) and side (C) centered unit cells can also exist in some crystal systems. Taking these centered unit cells into account, 14 Bravais lattices can be described. During a structural analysis, however, not the unit cell but instead the asymmetric unit (ASU) is at the focus

of interest. This is because the content, i. e. the number and arrangement of the individual molecules in each unit cell (proteins or protein complexes) can always be generated by application of all symmetry operations of the space group to the contents of the ASU. Note, however, that due to the chiral nature of proteins the only symmetry operations allowed and found in protein crystals are translations, rotations, and combinations of both, the screw axes. This is also reflected in the space group symmetry operations, which consist of (i) the abovementioned lattice translations of the ASU (P, I, F, C), (ii) the additional combinations of ASU rotations about the lattice points (point groups) and (iii) the combinations of the ASU rotations and translations about the lines connecting neighboring lattice points (screw axes). Combining the 11 point groups allowed in protein crystals with translations (screw axes) and centering operations (P, I, F, C) gives rise to 65 protein space groups. For a summary of the allowed symmetry operations in crystals with a chiral and non-chiral ASU content, as well as the resulting number of point and space groups see Table 2-35.

Definition – Symmetry operation: An operation that leads to superimposition of an object on itself.

Definition – Symmetry element: An imaginary geometric entity about which a symmetry operation takes place.

Table 2-35. Allowed symmetry operations in crystals with chiral or non-chiral ASU contents

ASU content	Allowed symmetry operations to generate the unit cell	Symmetry element	Number of crystal systems	Number of Bravais lattices	Number of point groups	Number of space groups
Non-chiral	Rotation	Rotation axis	7	14	32	230
	Translation	Translation vector				
	Inversion	Inversion point				
	Reflection	Mirror plane				
	Rotation + Translation	Screw axis				
	Mirror + Translation	Gliding plane				
	Rotation + Inversion	Rotoinversion axis				
Chiral (proteins)	Rotation	Rotation axis	7	14	11	65
	Translation	Translation vector				
	Rotation + Translation	Screw axis				

2.2.5.3 Protein crystal growth

A fundamental requirement to solve the three dimensional structure of a molecule (e. g. of a protein or a protein complex) by means of X-ray crystallography is the availability of protein crystals of sufficient quality and size (ideally 50 - 1,000 μm). The size matters as very small crystals lead to weak signals of the diffracted X-ray beams. The reason for this is that small crystals consist only of

a relatively small number of scatterers (i. e. a relative small number of stacked unit cells). The objective of most crystallization methods is to precipitate proteins from a supersaturated aqueous solution by slowly increasing the concentration of a precipitant and the protein itself without destroying the protein's three-dimensional structure¹¹⁵⁻¹¹⁷. The latter is possible because crystals normally contain high amounts (50 - 70 %) of solvent due to the irregular shape of the proteins. The precipitants are usually salts or water-soluble polymers that, as they become solvated, hijack the highly ordered water molecules from the water cages that automatically form around the hydrophobic (non-polar) portions of a protein in aqueous solution. Therefore, as a consequence of this so-called 'salting-out' effect, hydrophobic surface regions of the proteins become exposed to the free water, which will facilitate their interaction with the exposed hydrophobic regions of other proteins in the solution and thus eventually to the formation of intermolecular precipitates. The most widely used crystallization method that is based on the 'salting-out' principle is vapor diffusion. The course of a typical vapor diffusion crystallization experiment can be roughly divided into three stages that are best illustrated by a phase diagram (Figure 2-11). (Stage 1) At the beginning of a crystallization experiment, the protein solution is mixed with a precipitant containing crystallization solution. The resulting mixture should ideally be undersaturated in terms of protein (red zone in Figure 2-11) and is placed as a hanging or sitting drop in an air-sealed compartment. This compartment also contains a significantly larger volume of the crystallization solution in a spatially separated reservoir. (Stage 2) Then, as a result of the lower overall concentration of solutes in the drop mixture compared to the reservoir solution, water will diffuse from the drop into the reservoir. This vapor diffusion process causes a slow increase in protein and precipitant concentration in the drop up to a point where the solution becomes supersaturated in terms of protein. If during this process the crystal drop mixture reaches the supersaturated precipitation zone (blue zone in Figure 2-11) only amorphous precipitates will be obtained. However, if the drop mixture reaches the moderately supersaturated nucleation zone (yellow zone in Figure 2-11), crystal nuclei will form spontaneously and start growing. (Stage 3) As a consequence of the continuous crystal formation and growth, the concentration of free protein in the crystal drop will decrease, which will consequently move the drop mixture to the less supersaturated metastable zone (green zone in Figure 2-11). In this zone nucleation ceases and only crystal growth continues until an equilibrium between crystal growth and crystal decomposition is reached. At this point the crystal growth eventually stops.

Crystal growth is dependent on many parameters, such as the initial protein and precipitant concentrations, the temperature, the pH of the solution and the presence of additives. Unfortunately there is no way to predict all these parameters and consequently the composition of a successful crystallization condition for a given protein. This forces the crystallographer to systematically test

varying crystallization conditions in order to identify the optimal crystallization condition. Experimentally this is done by using commercial or self-prepared bifactorial grid screens or multifactorial sparse matrix screens. In bifactorial grid screens only two factors, such as precipitant concentration and pH or precipitant concentration, and initial protein concentration are varied. In sparse matrix screens typically 3 - 4 parameters are varied where the composition is based on the analyses of conditions that have led to a successful crystallization of proteins in the past. Sparse matrix screens are usually the starting point in the quest to find the optimal crystallization condition as they allow for a coarse sampling of the crystal condition space. Crystals that may be of low quality can then be improved by refining the crystallization conditions through application of bifactorial grid screens.

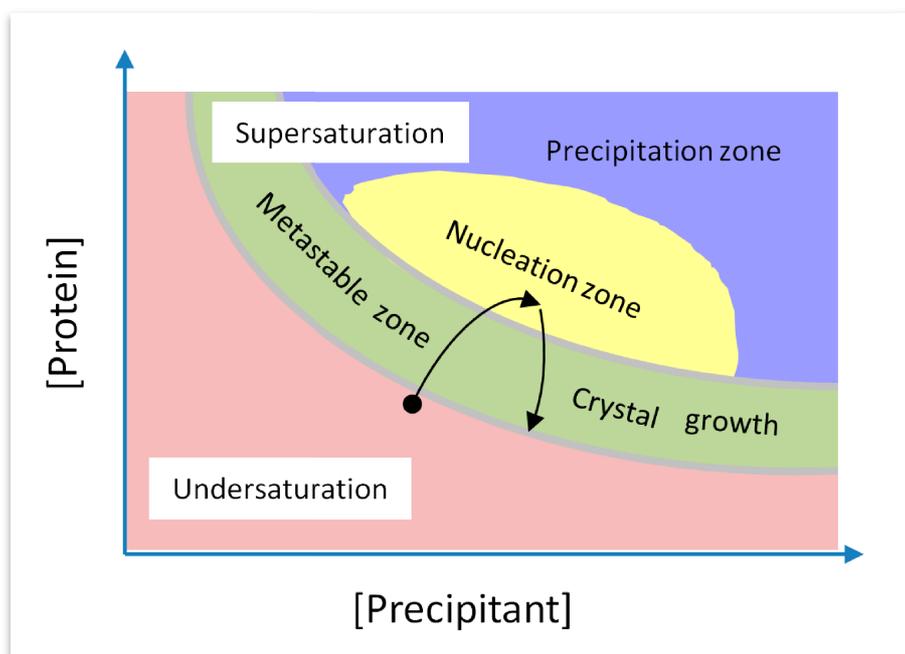


Figure 2-11. The crystallization process in a vapor diffusion experiment. The displayed phase diagram shows the influence of the two crystallization parameters *protein concentration* and *precipitant concentration* on the formation and growth of protein crystals. The black circle marks the starting point of the crystallization experiment (Stage 1), while the arrows indicate the change in both parameters during the processes of vapor diffusion (Stage 2) and crystal growth (Stage 3). The figure was modified from Chayen, 2004¹¹⁵.

Protocol: For initial screening the *Hampton Crystal Screens 1 & 2*, the *Hampton Light Screen*, the *Hampton Peg/Ion 1 Screen*, the *Emerald Wizard 1 & 2 Screens* and the *Hampton Grid Screen Ammonium Sulphate* were used. Initial hits in those screens were then refined through the preparation and implementation of self-made fine screens. The hanging drop vapor diffusion experiments were used when crystal drops were set up manually. 1 μ l of the protein solution was mixed with an equal volume of the crystallization solution (of a grid or sparse matrix screen) and

pipetted onto a siliconized cover slip. The cover slip was turned upside down and placed on top over a single compartment of a 24-well crystallization plate filled with 500 μ l of the crystallization solution. The space between the cover slip and the rim of the compartment was sealed with grease. In cases where a *Freedom Evo 150* robot was used to pipet the crystal drops, the sitting drop method in combination with 96-well crystallization plates was employed. The robot hereby pipets the crystal drops (300 nl protein solution plus 300 nl crystallization solution) on a flat plateau positioned above the reservoir solution (100 μ l). All 96-well plates were sealed at once after the last drop was set by using a self-adhesive transparent foil. Crystal plates were incubated at 4 and 20 $^{\circ}$ C.

2.2.5.4 Protein crystal freezing

Once a protein crystal of suitable size has been obtained through crystallization experiments (see Section 2.2.5.4) it needs to be exposed to X-ray radiation in order to collect the diffraction data from which the structure of the protein can be determined. However, collection of diffraction data may be hampered as protein crystals are sensitive to radiation damage. This radiation damage is caused by the high energy of the X-ray photons and by the way these photons interact with electrons in the exposed matter. Either the primary photon's energy is absorbed by the electron and subsequently released as a secondary photon in a random direction but with the original energy (coherent scattering, also known as Thomson scattering) or the primary photon's energy results in the release of an electron and the emission of a secondary photon in a random direction with reduced energy (incoherent scattering, also known as Compton scattering). Note that the crystallographic structure determination is based on the coherent scattering. Radiation damage is caused by the incoherent scattering of the X-ray photons as the loss of atomic shell electrons results in the formation of free radicals, very reactive chemical species that are very harmful for other macromolecules.

One can reduce the radiation damage caused by free radicals by restricting their movement in the crystal. This can be achieved by performing the X-ray data collection of the protein crystal at very low temperature, which is maintained by a cryostream of nitrogen gas at 100 K. Before a crystal can be placed in the cryostream it has to be protected from ice formation that can severely harm the crystal. Thus, if the crystallization solution does not have cryoprotecting capabilities itself, the protein crystal is first soaked with a cryosolution, consisting of the crystallization solution (also called mother liquor) supplemented with a suitable cryoprotectant such as glycerol or ethylene glycol, followed by mounting it in a fiber loop and flash freezing it in liquid nitrogen (boiling point -196 $^{\circ}$ C) or directly in the cryostream¹¹⁸.

Protocol: Using a quartz capillary (\varnothing 0.5 mm), crystals grown as described in Section 2.2.5.3 were transferred from the crystal drop to a depression well containing 100 μ l of mother liquor in order to clean the crystal of precipitates and protein solution. In the next step the crystal was transferred to a second well that contained 100 μ l of cryosolution, consisting of the mother liquor (Table 2-26) supplemented with cryoprotectant (e. g. polyethylene glycol or glycerol). After a 1 min soaking period in the cryosolution, crystals were fished with an appropriately sized fiber loop, flash frozen in liquid nitrogen and finally mounted on the goniometer head in a cryostream of nitrogen gas. Alternatively, crystals were stored in a liquid nitrogen tank after fiber loop mounting.

2.2.5.5 X-ray diffraction

In order to collect crystallographic data the frozen crystal in its cryoloop is mounted on a goniometer head, that places the crystal directly between an X-ray source and a detector. Once an X-ray beam with a defined wavelength (λ), amplitude (A) and intensity ($I \sim A^2$), hits the crystal, the beam is diffracted into many separate beams that have the same wavelength but altered direction and intensity. For each diffracted beam these two parameters can be determined by recording the positions and intensities of the reflections (spots) that are generated when the diffracted beams hit the detector. Although the reflections appear in two dimensions on the detector, they are actually lattice points of a three-dimensional reciprocal lattice and can be described with the indices h , k and l . The term ‘reciprocal’ refers to the inverse relationship between the lattice translations in the crystal lattice and the lattice translations in the lattice recorded on the detector. Based on this simple mathematical relationship it is possible to use the recorded positions of the reflections in reciprocal space to determine the geometry (i. e. the Bravais lattice) and the dimensions (axes and interaxial angles) of the real unit cell. Then, by applying a Fourier transform to the recorded intensities of the reflections the content of (i. e. the number, arrangement and structures of the molecules in) the real unit cell can be determined. Note here that all atoms in the unit cell contribute to each of the many hkl reflections recorded in a crystallographic data set. This is because all atoms in the unit cell function as X-ray diffraction centers, giving rise to spherical X-ray waves that interfere with each other. The net result of this wave interference is displayed as a diffraction image and strongly dependent on the relative positions of the scattering centers (i. e. the atoms) in the unit cell that produced it. Furthermore, it is important to realize that the diffraction pattern of the crystal, or the content of the myriad of regularly arranged unit cells that make up the crystal lattice, is nothing else than the averaged and amplified diffraction pattern of a single unit cell content sampled at the reciprocal lattice points.

A simplified mathematical description that explains under which conditions the diffraction process by a crystal lattice leads to constructive interference and thus to the presence of a recordable reflection (h, k, l) is given by Bragg's law (Equation 2-6 and Figure 2-12A)¹¹⁹. Bragg's law was developed by Sir William Lawrence Bragg in 1912, who treated the diffraction of the incident beam as reflection from a set of equivalent, parallel real space lattice planes (h, k, l). This explains why the diffraction spots are called 'reflections' and it also explains why both, the set of real space lattice planes and the reciprocal space reflections, share the same set of hkl indices (Figure 2-12A). In detail, Bragg's law states that constructive interference between two diffracted X-rays (i. e. the generation of a reflection) can only be observed if the path difference between the two rays (that impinge upon lattice planes hkl with interplanar lattice spacing d_{hkl} under an angle θ) is a multiple of the utilized wavelength λ .

$$n \cdot \lambda = 2d \cdot \sin\theta \qquad \text{Equation 2-6}$$

with

n	1, 2, 3...
λ	wavelength of the X-ray waves [Å]
d_{hkl}	lattice plane hkl spacing
θ_{hkl}	Bragg angle of the hkl reflection

The sphere of reflection, which is also called the Ewald sphere or Ewald construction, named after its inventor Paul Peter Ewald, is a graphical representation of Bragg's law in reciprocal space (Figure 2-12B)¹²⁰. The crystal is positioned at the center of the Ewald sphere, which features a radius (r) of $1/\lambda$. The point where the incident direct beam exits the Ewald sphere is defined as the origin of the reciprocal lattice and is assigned the hkl indices (0,0,0). A reflection hkl on the detector is observed when a reciprocal lattice point hkl comes in contact with the Ewald sphere. In that case the incident, direct beam is diffracted in such a way that the diffracted beam exits the Ewald sphere through the reciprocal lattice point that is in contact with the sphere. In real space this situation corresponds to the diffraction of the incident, direct beam by the hkl set of real lattice planes, which are indicated in the figure by a single blue plane that slices the crystal. The distance (d^*) between the origin of the reciprocal lattice (0,0,0) and the reciprocal lattice point hkl is connected to the distance of the hkl real lattice planes by an inverse relationship.

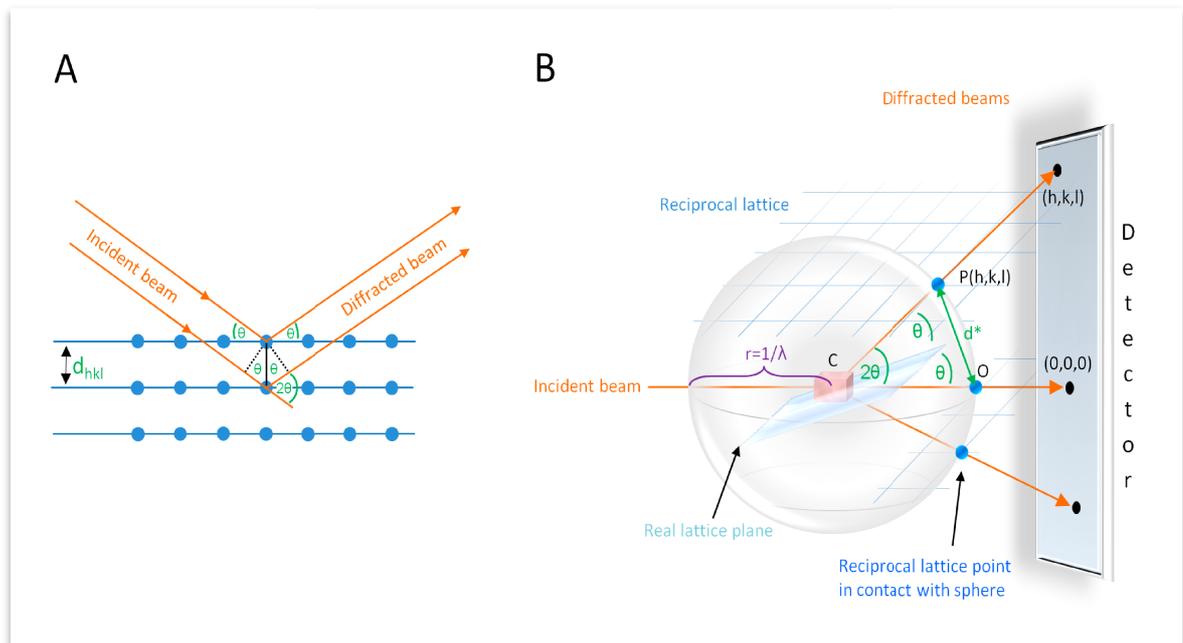


Figure 2-12. Schematic representations of Bragg's law and the Ewald sphere. (A) Bragg's law: λ – X-ray wavelength, d_{hkl} – lattice plane hkl spacing, θ_{hkl} – Bragg angle of the hkl reflection. (B) Ewald sphere: r – radius of the sphere, θ_{hkl} – Bragg angle of the hkl reflection, $d^* = 1/d_{hkl}$, $P(h,k,l)$ – reciprocal lattice point with indices h , k and l , O – origin of the reciprocal lattice. The black circles on the detector represent the reflections caused by reciprocal lattice points (dark blue spheres) that are in contact with the Ewald sphere. Selected areas of the reciprocal lattice are shown in light blue. The arrows mark the incident (bright orange) and diffracted (orange) X-ray beams. C , crystal.

2.2.5.6 Data collection

Detectors used in macromolecular crystallography are area detectors, which can collect many reflections at once in a defined direction. Image plates record X-rays that by hitting a photostimulable phosphor plate (PSP) result in the excitation and storage of electrons in the crystalline phosphor material. The readout is performed by a laser that passes over each spot of the image plate, thereby releasing the energy of the excited electrons in form of blue light that can be measured using a photomultiplier. The recorded intensities of the emitted blue light are proportional to the stored electrons and those are again proportional to the intensities of the X-rays that caused their excitation.

More sophisticated X-ray detectors are charge-coupled devices (CCDs). CCDs consist of pixels (arranged in an array) that accurately count the (X-ray) photons that strike them and that therefore allow for a more precise measurement of a particular reflection's intensity. Each of the photons that hit a CCD pixel in a defined time frame leads to a proportional excitation of electrons (charge) that accumulate in a positively charged region of the pixel. The readout is performed row by row and the charges in the readout pixel row transferred one by one to an amplifier at the edge of the row that converts the charge into a voltage, which is eventually recorded. The greatest

advantage of CCD detectors compared to other detector types such as image plates is their fast readout capability that allows for an extremely efficient data collection. However, the latest advancement in detector technology is represented by two-dimensional hybrid pixel array detectors such as the PILATUS (pixel apparatus for the SLS), which operate in single-photon counting mode and are characterized by several advantages compared to modern CCD detectors, for example a superior signal-to-noise ratio and readout times of only a few milliseconds. For more details see Refs.^{121, 122}.

Protocol: Measurements using synchrotron radiation were performed at the SLS (Villigen, Switzerland) using the PXI (X06SA) and PXIII (X06DA) beamlines that were equipped with a *PILATUS 6M* and a *MAR225* CCD detector, respectively. Crystals measured at synchrotrons were usually tested for diffraction using our in-house X-ray system, comprised of a *MAR345dtb* image plate, a *Rigaku MicroMax-007 HF* rotating anode, and VariMax HF optics. For each X-ray diffraction experiment, a single frozen crystal was mounted on the goniometer head of the X-ray system and positioned in the path of the X-ray source beam in front of the detector. The crystal was exposed to X-ray radiation of defined wavelength, and the generated reflections were recorded by the detector system. During exposure, the crystal was rotated in small angle intervals (usually between 0.25 - 2°) about a fixed axis perpendicular to the X-ray beam. Each interval corresponds to a single recorded diffraction ‘frame’ (or ‘image’) by the detector. A set of recorded test frames was used to guide the strategy of data collection in terms of (i) the optimal crystal to detector distance, (ii) the optimal intensity of the X-ray beam and (iii) the minimal total angle the crystal has to be rotated about in order to record a complete dataset comprised of all the crystal’s unique reflections.

2.2.5.7 Data processing

Data processing of the raw data (recorded images) obtained in Section 2.2.5.6 is performed in three subsequent steps: Indexing, integrating and scaling (see subsequent sections). The objective of all of these steps is to obtain a list of all measured reflections, in which every reflection is properly indexed and is assigned a consistent intensity.

2.2.5.7.1 Indexing

Indexing is performed by using computer programs that search the collected data frames for reflections and then assign hkl values to each one of them. Knowing the positions of the individual reflections as well as the X-ray wavelength and the distance of the crystal from the detector allows

the indexing program to determine the parameters of the unit cell by measuring the distances and angles between the reflections. Therefore, at this stage one can already determine the crystal system but not yet the space group of the crystal as to do so requires at least the raw intensities of the individual reflections.

2.2.5.7.2 Integrating

The objective of the integration process is to assign a defined raw intensity to each recorded reflection, many of which are actually redundant in a given dataset. This happens when, caused by the rotation of the crystal during X-ray data collection, the same lattice point repeatedly hits the Ewald sphere. Moreover, particular reflections may be distributed over more than one frame. These reflections are called ‘partials’, whereas those that contribute to only one frame are called ‘fullies’. Therefore, the first task for the indexing program is to identify the individual frames a single reflection is recorded in. The second task of the indexing program is to analyze the reflections and define their three-dimensional profile. In other words, the program defines which area (size and shape) belongs to a reflection and which doesn’t (i. e. is background). During the actual integration process, all partial and full reflections in each frame are fitted to the determined profile (profile fitting), which means that only the reflection area (no matter if a partial or full reflection is concerned) within the predefined profile is eventually added up to calculate the raw intensity of the individual hkl reflection.

After the integration step a list of the observed hkl reflections (that still includes redundant hkl reflections) and their corresponding raw intensities is obtained that can, in programs such as XDS¹²³, be analyzed in terms of systematic absences, regular patterns of reflections that are missing (absent). Systematic absences are caused by certain internal symmetry elements (screw axes) of the crystal’s unit cell. In combination with the determined Bravais lattice type systematic absences can be used to determine the space group of the unit cell.

2.2.5.7.3 Scaling

The objective of the integration process is to determine the raw intensities of all recorded reflections. Many of the recorded hkl reflections are symmetry related to each other and should therefore theoretically exhibit the same intensity. However, this is usually not the case and instead some variation in intensities of the symmetry related reflections is observed. These variations can be caused by the (i) weakening diffraction of a crystal in the course of the experiment due to radiation damage, (ii) changes in the intensity of the X-ray beam during the experiment, (iii)

different X-ray path lengths through the crystal (crystal shape) and/or (iv) the presence of anisotropy in the crystal itself. In order to correct for these intensity variations the scaling program assigns scaling factors to each hkl reflection, which consequently results in a list of hkl reflections with consistent intensities. However, the list still contains the redundant hkl reflections that were already present in the list of reflections after the integration step. Therefore, as part of the scaling process these redundant hkl reflections are merged to a single hkl reflection. This latter merging step is a data reduction process and therefore sometimes called ‘reducing’. Eventually the output file of the program consists of a table that contains a list of all unique hkl reflections as well as their intensities. Furthermore, tables are generated whose statistical content can be used to define the data quality (see Section 2.2.5.7.4) and the resolution limit of the data.

2.2.5.7.4 Data Quality

Several indicators are used to assess the quality of the data. As the quality of the measured intensities decreases with increasing resolution all of the statistical indicators are calculated for defined resolution bins. The data processing R-factors (reliability factors, also called residual factor) measure quality as the ratio between the mean difference between the magnitudes of values (A and A’) which should be the same (residual value) and the mean magnitude of the measured value (A). The value obtained is a decimal number that is often expressed in % (e. g. 0.3 = 30 %). Symmetry-related reflections should have the same intensity. In this respect R_{sym} (Equation 2-7) assesses data quality by the difference in intensity between these symmetry-related reflections.

$$R_{\text{sym}} = \frac{\sum_{\text{hkl}} \sum_i^n | I_{\text{hkl},i} - \bar{I}_{\text{hkl},i} |}{\sum_{\text{hkl}} \sum_i^n I_{\text{hkl},i}} \quad \text{Equation 2-7}$$

with

- I intensity of the reflection hkl
- \bar{I} mean intensity of symmetry-related reflections of a reflection hkl
- i symmetry-related reflection to hkl
- n number of independent measurements of the reflection hkl (multiplicity)

However, a serious disadvantage of R_{sym} is that its value increases when reflections are measured several times, i. e. when the redundancy of the data is increased. This is counterintuitive as highly

redundant data normally improves the quality of the data. Hence, Diederichs and Karplus introduced the redundancy-independent R-factor, R_{meas} (Equation 2-8)¹²⁴.

$$R_{\text{meas}} = \frac{\sum_{\text{hkl}} \sqrt{\frac{n}{n+1}} \sum_i^n | I_{\text{hkl},i} - \bar{I}_{\text{hkl},i} |}{\sum_{\text{hkl}} \sum_i^n I_{\text{hkl},i}} \quad \text{Equation 2-8}$$

with

- I intensity of the reflection hkl
- \bar{I} mean intensity of symmetry-related reflections of a reflection hkl
- i symmetry-related reflection to hkl
- n number of independent measurements of the reflection hkl (multiplicity)

Another important quality indicator is the ratio between the mean intensity I and the mean standard deviation of intensities σI . The ratio $I/\sigma I$ is an intensity signal to noise ratio and decreases with higher resolution. A resolution shell exhibiting an $I/\sigma I$ value of 2.0 - 3.5 is normally used as the highest resolution shell (i. e. defines the resolution cutoff) unless other indicators such as the R-factors described above argue against this. By plotting the natural logarithm of the mean intensity I of a resolution shells against $\sin^2\theta/\lambda^2$ of the same shell, the latter term being an alternative way of expressing the resolution of a shell, a Wilson plot can be derived¹²⁵. Caused by solvent effects a dip is observed at approximately 5 - 5.5 Å. However, beyond 4 Å the Wilson graph decreases linearly and is only disturbed by secondary structure elements of the protein that can cause a positive peak in the range between 3 - 3.5 Å. The negative slope of the Wilson graph is proportional to the overall temperature factor (B-factor), which is a measure of the thermal motion within the unit cell. Furthermore, the x-value at the intersection of the Wilson graph with the x-axis indicates the approximate resolution limit of the recorded dataset.

Two other statistical parameters that indicate the quality of the data are called completeness and redundancy and are described in Equation 2-9 and Equation 2-10, respectively.

$$\text{Completeness} = \frac{\text{Number of unique reflections measured}}{\text{Total number of unique reflections}} \quad \text{Equation 2-9}$$

$$\text{Redundancy} = \frac{\text{Total number of measured reflections}}{\text{Number of unique reflections measured}} \quad \text{Equation 2-10}$$

2.2.5.7.5 Matthews coefficient

The Matthews coefficient V_M (also called packing parameter) is calculated as shown in Equation 2-11 and describes the volume [\AA^3] within a unit cell that is occupied by protein¹²⁶. While the molecular weight (M) of the crystallized protomers (protein or protein complex) is known even before a crystal is obtained, the unit cell volume (V_{EZ} , calculated from the unit cell parameters) and the number of asymmetric units (n, calculated from the space group) will only be known after the unit cell and the space group have been determined. The only variable in the equation is the number of protomers in the asymmetric unit (z). By trying out different integer numbers for z different V_M values are generated. Based on the comparison of the calculated V_M values to V_M values of already solved crystal structures in a databank, a probability of the most likely z value can be calculated.

$$V_M = \frac{V}{M \cdot z \cdot n} \quad \text{Equation 2-11}$$

with

V_M	Matthews coefficient [$\text{\AA}^3/\text{Da}$]
V	unit cell volume [\AA^3]
M	Molecular weight [Da]
z	Number of protomers in asymmetric unit
n	Number of asymmetric units per unit cell

Once a crystal structure is solved the number of protomers in the ASU can be determined and therefore an accurately defined Matthews coefficient (V_M) be calculated. Knowing the correct V_M value in turn allows one to calculate the solvent content of the crystal (Equation 2-12).

$$V_{\text{solv}} = 1 - \frac{1.23}{V_M} \quad \text{Equation 2-12}$$

with

V_{solv}	Solvent content of the crystal
V_M	Matthews coefficient based on structure determination

2.2.5.8 Phasing

As described in Section 2.2.5.5, all atoms in the unit cell contribute to each of the many hkl reflections recorded in a crystallographic data set. Expressed in mathematical terms, each recorded hkl reflection and therefore each diffracted X-ray beam that causes this reflection, can be described as a complicated periodic function that represents the sum (therefore called Fourier sum, F) of the n superimposed individual periodic wave functions (called Fourier term, f_{hkl}) that originate from each of the n scatterers (i. e. each of the n atoms) in the unit cell. This mathematical relationship was first found by J. B. J. Fourier and is described as a structure-factor equation in crystallography, the mathematical expression of which is shown in Equation 2-13.

$$F_{hkl} = \sum_{j=1}^n f_{hkl} \tag{Equation 2-13}$$

$$f_{hkl} = f_j e^{2\pi i(hx_j + ky_j + lz_j)} \tag{Equation 2-14}$$

with

- F_{hkl} structure factor of reflection hkl (periodic wave generated by superimposition of waves f_{hkl})
- f_{hkl} atomic structure factor (waves created by individual atoms in the unit cell)
- f_j scattering factor of atom j (depends on the element the atom belongs to)
- h, k, l reflection indices (represent three frequencies in each Fourier term)
- i complex number
- n number of all atoms in the unit cell
- x, y, z Coordinates of atom j in the unit cell

In Equation 2-13 a structure factor F_{hkl} is described in terms of the diffraction contribution of each individual atom in the unit cell to reflection hkl. A structure factor can also be described in terms of the individual diffraction contribution of the electron density present in one infinitesimally small volume element p of the unit cell that is centered at position (x, y, z) .

$$F_{hkl} = \int_V p(x, y, z) e^{2\pi i(hx + ky + lz)} dV \tag{Equation 2-15}$$

with

- F_{hkl} structure factor of reflection hkl (periodic wave generated by superimposition of waves f_{hkl})

h,k,l	reflection indices (represent three frequencies in each Fourier term)
i	complex number
x,y,z	Coordinates of volume element p in the unit cell
dV	unit cell volume

Equation 2-15 allows one to describe the structure factors in a crystallographic data set in terms of the electron density of the unit cell (Fourier analysis) and, this also works in the other direction, which allows to describe the electron density of the unit cells in terms of the structure factors (Fourier syntheses; see Equation 2-15). More generally, this reversible mathematical transformation that connects the diffraction pattern (structure factors of the recorded reflections) and the real object (the electron density) is called a Fourier transform.

$$p(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l F_{hkl} e^{-2\pi i(hx+ky+lz)} \quad \text{Equation 2-16}$$

with

$p(x,y,z)$	electron density at position x,y,z
V	the volume of the unit cell
F_{hkl}	the structure factor amplitude of reflection hkl
h,k,l	reflection indices (represent three frequencies in each Fourier term)

Equation 2-17 describes the electron density distribution inside the unit cell as a periodic function. Note that due to the regular arrangement of the unit cells and their content in a crystal, also these electron clouds can (just as the positions of the individual atoms) be described by a periodic function. It is this periodic function of the electron density whose graphical representation is an electron density map that essentially describes the surface properties of the atoms (or molecules) in the unit cell, and therefore the molecular structure of the protein.

Since each structure factor represents a periodic function, one needs to know the (i) amplitude, (ii) the frequency and (iii) the phase of each structure-factor in order to solve Equation 2-16. Unfortunately only the first two parameters can be determined from the obtained diffraction data. The amplitude of the structure factor can be derived from the intensities of the observed reflections (Equation 2-17), while its three frequency terms h, k and l are specified by the position of the observed reflection. The phase information, however, is always lost when X-ray diffraction data is collected on a detector, and therefore indirect methods were developed to solve this so-called phase problem.

$$I_{hkl} = k \cdot F_{hkl}^2 \quad \text{Equation 2-17}$$

with

I_{hkl}	the observed intensity of the reflection hkl
k	constant
F_{hkl}	the structure factor amplitude of the reflection hkl

In the easiest case one can estimate the phases of an unknown structure by using the phases from structure factors of another protein, whose structure is already solved. However, in order for this to work (i. e. for the phases to be similar enough), the model has to be highly homologous to the unknown structure. This known structure is called the ‘phasing model’ or simply ‘model’. In a computer-based two step search called molecular replacement (MR), the known structure is first rotated (3 rotation dimensions) in Patterson space to search for the correct orientation and subsequently translated in space (3 translation dimensions) in order to search for the correct position of the unknown structure in the unit cell. The actual comparison of orientations between model and unknown structure, i. e. the rotation search, is implemented by computing a rotation function that evaluates the correlation between the Patterson map (named after Arthur Lindo Patterson) of the unknown structure and the Patterson maps based on various orientations of the model. A search solution is obtained when the orientation of the model in real space leads to a Patterson map that correlates well with the Patterson map of the unknown structure. As there might be several protomers with several orientations in the ASU also several rotation search solutions may be found in the rotation search. However, to finally be able to superimpose the now properly orientated model protomers correctly onto the unknown structure protomers, i. e. to find the right location of the model protomers, the oriented model protomers have to be translated. This translation search is implemented by computing the R-factor (see Section 2.2.5.9), which compares the overall agreement between the amplitudes of the model structure factors with those of the unknown structure. Again, as there might be several protomers with several locations in the ASU also several translation search solutions may be found in the translation search. Now, distinct sets of the found rotation and translation solutions can be used to superimpose the model structure upon the protomer(s) in the ASU and therefore eventually allow to use the phases of the model structure as the initial phases of the unknown structure in order to obtain the unknown structure’s electron density map.

PHASER is one of the most commonly used computer programs to perform MR operations¹⁰⁰. To judge the quality of the found solution(s), PHASER uses two parameters, the log likelihood gain (LLG) and the TF (translation function) Z-score. A high LLG indicates how much

better the observed data can be described by the model compared to a random distribution of the same atom. TF Z-score values indicate how likely it is that PHASER has found the initial phases to solve the structure. As a guideline, TF Z-scores of 5-6, 6-7, 7-8 and above 8 indicate that is unlikely, possible, probable and very likely that PHASER has solved the structure. If there is more than one solution, i. e. more than one protomer in the ASU, the LLG and the Z-score should increase with every additional solution.

With the initial phases determined by PHASER¹⁰⁰, Equation 2-16 can now be solved in order to calculate the desired electron density map (F_{obs} map) and display it with programs such as Coot¹⁰². To minimize a potential model bias, however, not the F_{obs} map alone but a $F_{\text{obs}}-F_{\text{calc}}$ and a $2F_{\text{obs}}-F_{\text{calc}}$ map are calculated and displayed on the screen. The electron density is model biased because F_{obs} consists of measured intensities and frequencies from the data and phases from the inaccurate model. However, an additional electron density map (F_{calc}) can be calculated from the model by performing a Fourier analysis. Subtracting this model map (F_{calc}) from the model influenced map F_{obs} will result in a so-called ‘difference map’ ($F_{\text{obs}}-F_{\text{calc}}$) that emphasizes the errors (i. e. the differences) between the current model and the true structure. Missing parts in the model appear as positive electron density surface, while areas with misplaced model parts appear as a negative electron density surface. The second map displayed on the screen on top of the difference map ($F_{\text{obs}}-F_{\text{calc}}$) is the $2F_{\text{obs}}-F_{\text{calc}}$ map, which is model biased. The $2F_{\text{obs}}-F_{\text{calc}}$ map basically shows the features of the modeled molecule. For model building the $2F_{\text{obs}}-F_{\text{calc}}$ and $F_{\text{obs}}-F_{\text{calc}}$ maps are usually displayed at contour levels of 1σ and 3σ , respectively, where sigma is the electron background noise.

After initial phasing (e. g. using PHASER¹⁰⁰) the electron density map is often very noisy and difficult to interpret. To improve the initial phases and therefore the electron density map, density modification methods are applied. Solvent flattening is based on the fact that the electron density is generally high in regions where parts of the protein are present, while it should be low at regions of disordered solvent. To perform solvent flattening a mask is generated that separates protein from solvent. The average electron density in the solvent region is then subtracted from the complete map which results in a less noisy electron density map with more detailed features of the protein due to the improved signal to noise ratio. Another density modification is called non-crystallographic symmetry (NCS) averaging. As mentioned before the asymmetric unit (ASU) is the minimal unit of the crystal structure from which one complete unit cell can be generated by application of all symmetry operations defined in the space group of the crystal. Nevertheless, additional symmetry elements (NCS elements) can exist in the asymmetric unit that lead to the superimposition of the protomer on itself. Assuming that the protomer structures are similar, these NCS symmetry elements within the ASU can usually be used to average the molecules in the

asymmetric unit, which like solvent flattening results in a better signal to noise ratio. This is because the protein electron density signal is amplified due to the regular shape of the proteins, while the solvent electron density signal is reduced due the disordered (non-regular shaped) solvent.

2.2.5.9 Structure refinement

Structure refinement is an iterative process of cycles of real- and reciprocal space refinement steps, during which one tries to obtain a more and more chemically and biologically accurate model of the crystallized macromolecule.

The goal of the refinement process is to improve the agreement (i. e. minimize the difference) between the calculated and observed structure factor amplitudes $|F_{calc}|$ and $|F_{obs}|$ by optimizing the parameters of the overall model, which consists of (i) the atomic coordinates, (ii) the occupancies, and (iii) the temperature factors (B factors) of each atom of the macromolecule as well as atoms belonging to other ordered molecules (e. g. ordered water) and/or ions in the ASU. Note that $|F_{obs}|$ is available from the recorded data and that $|F_{calc}|$ is obtained by a Fourier transform of the current model. Mathematically, the refinement process is performed by fitting the experimental data ($|F_{obs}|$) to a model $|F_{calc}|$, whose parameters are adjusted during the fitting process in order to create a fit that deviates as little as possible from the experimental data. More precisely, the fitting procedure relies on an algorithm, which optimizes parameter values by minimizing the sum of the squared residuals (Equation 2-18). The residuals again are the difference between the calculated $|F_{calc}|$ and observed $|F_{obs}|$ structure factor amplitudes.

$$\Phi = \sum_{hkl} \omega_{hkl} (|F_{obs}| - |F_{calc}|)_{hkl}^2 \quad \text{Equation 2-18}$$

with

Φ	sum of the squared residuals (squared residual = $(F_{obs} - F_{calc})^2$)
F_{obs}	observed hkl structure factor amplitude (based on data)
F_{calc}	calculated hkl structure factor amplitude (based on model)
ω	optional weighting factor (reflects reliability of reflection hkl)

The progress of refinement, i. e. the quality of the fit, is monitored by the so-called R (residual or reliability) factor R_{work} (Equation 2-19), which is either expressed as a decimal number or in percent (e. g. 0.5 equals 50 %).

$$R = \frac{\sum ||F_{obs}| - |F_{calc}||}{\sum |F_{obs}|} \quad \text{Equation 2-19}$$

with

R	R factor (R_{work} or R_{free} depending on set of reflections used for F_{obs})
F_{obs}	observed hkl structure factor amplitude (based on data)
F_{calc}	calculated hkl structure factor amplitude (based on model)

Equation 2-19 implies that if the quality of the model is increased during a single refinement cycle (that consists of a real- followed by a reciprocal refinement step), the R_{work} should decrease. Initial models usually result in R_{work} values around 50 %, which is not far away from the theoretical maximum value of 59 % that would be obtained from the same set of model atoms when randomly distributed in the unit cell¹²⁵. Well-refined models on the other hand exhibit R_{work} values between 10 - 25 % depending on the quality of the dataset.

Care must be taken at this point not to overfit the data. Overfitting happens if too many parameters are introduced into a model. This can be viewed as giving the fitting program more and more options to combine parameter values in order to fit the data better. It is important to realize that a fitting program does not care whether a determined parameter value makes sense but instead it is interested in finding a combination of parameter values that (even if meaningless) will allow a better fit of the data to the model (i. e. lead to a lower sum of the squared residuals). What prevents overfitting is usually a high observable (independent reflections) to parameter ratio. While the number of observables will not change for a given dataset the number of parameters depends on the parameterization of the model. To judge at which ratio the overfitting actually starts another R factor, the R_{free} , was introduced that can be calculated in the same fashion as R_{work} (Equation 2-19)¹²⁷. For the calculation of R_{free} a portion of 5 - 10 % of the overall reflections is set aside and flagged as 'free'. These 'free' reflections are not included in the crystallographic refinement process. This means that the fitting program can adapt the model parameters (in number and value) to the 90 - 95 % of the reflections, while the same is not possible with the 'free' reflections that are basically hidden from it. Nevertheless a properly parameterized model after fitting also has to properly predict the 'free' reflections. If after fitting the 'free' reflections are less well described by the model this indicates overfitting. Therefore, the R_{free} is a more unbiased estimate of the model improvement during refinement and has to ideally decrease in the same manner as the R_{work} . In the final model eventually the R_{work} and R_{free} values should ideally not deviate more than 5 %.

To refine the parameters listed above in reciprocal space, different parameterization methods such as rigid body refinement, simulated annealing, individual coordinate refinement,

occupancy refinement and TLS (translation, libration screw) refinement are commonly employed. During rigid body refinement proteins are viewed as entities that are slightly misplaced in orientation and position. The rigid body refinement step essentially rotates and translates this entity and thereby moves the model of the protein from its slightly misplaced to its correct position. Simulated annealing is a mathematical procedure that consists of two steps. In the first step the kinetic energy of the molecules that constitute the refinement model is significantly increased. This allows the molecules or parts of them (e. g. the side chains of a protein or protein loops) to significantly alter their positions and therefore to overcome the energy barriers of the local energy minima they might be trapped in. When the kinetic energy is then reduced during the second step of the simulated annealing procedure there will be an increased chance for the molecules to assume their native conformations, i. e. the conformations with the lowest global energy¹²⁸. The individual coordinate refinement is different from the rigid body refinement as here the positions of misplaced individual atoms and not the entire protein are corrected. The occupancy refinement takes into account that individual atoms do not necessarily occupy a defined space to 100 %. This can for example happen when an amino acid side chain takes up two or more alternative conformations. Consequently, the presence of the side chain at either position over time is less than 100 %. The TLS refinement is a form of B-factor refinement. The B-factor (also called temperature factor) describes the vibrations and oscillations (i. e. the thermal movement) of individual atoms around their central position and therefore accounts for the internal flexibility of the protein. This atomic movement can be described most precisely by assigning anisotropic B factors to each atom. However, this leads to a huge increase in the number of parameters (six parameters per atom) compared to the assignment of isotropic B factors for each atom (one parameter per atom). Therefore only high resolution datasets (<1.5 Å) justify an anisotropic B factor refinement, as only in these datasets the ratio of observables to parameters remains high enough for reliable fitting. However one can simulate the anisotropic movement of individual atoms by TLS refinement, which is not limited to a certain resolution. In TLS refinement segments of the proteins are treated as rigid bodies undergoing TLS (translation, libration, screw) vibrational motion.

The output from reciprocal space refinement consists of a file containing the refined model coordinates, from which $F_{\text{obs}}-F_{\text{calc}}$ and $2F_{\text{obs}}-F_{\text{calc}}$ electron density maps can be calculated. In the next step these model coordinates and electron density maps are visualized by Coot¹⁰², one of the most common programs for real space model refinement. After manual model building in Coot, the modified coordinates are saved in a new coordinate file that is subsequently used in the next round of reciprocal space refinement. The overall refinement progress is monitored by R_{work} and R_{free} and the model building finished when both R-factors have converged.

2.2.5.10 Glycan modeling.

After the structure of the UL16_{mon}-MICB_{short} complex was solved, the GlyProt¹⁰⁴ online server was used to model hybrid and complex glycans linked to the seven Asn residues with observed NAG electron density in order to produce a realistic estimate of size and distribution of the glycan moiety of native UL16.¹¹³

3. Results

3.1 Expression and purification of MICB_{long} and MICB_{short}

The designed MICB constructs MICB_{long} and MICB_{short} (see Section 2.2.3.4.1) were both expressed in *E. coli*. Besides the general benefits of this expression system (see Section 2.2.3.4.2), this choice was mainly based on previous reports that had shown MICB to be successfully expressed as and refolded from *E. coli* produced IBs (see Section 2.2.3.4.3)^{64, 85}.

3.1.1 Bacterial expression of MICB_{long} and MICB_{short}

After induction with 1 mM IPTG, His-tagged MICB_{long} and MICB_{short} proteins were expressed in *E. coli* Rosetta 2(DE3) cells (see Section 2.1.6.1) during a 4 h incubation period at 37 °C (details see Section 2.2.3.4.2). Rosetta 2(DE3) cells supply seven rare tRNAs, which contribute to higher protein expression levels of recombinant proteins that contain codons in their mRNA rarely used in *E. coli*. Indeed, as can be appreciated by comparing the pre IPTG (lane 1) and post IPTG (lane 2) samples analyzed in the A panels of Figures 3-1 and 3-2, *E. coli* Rosetta 2(DE3) cells (Figure 3-2A) showed higher protein expression levels of MICB_{long} as *E. coli* BL21(DE3) cells (Figure 3-1A) that do not supply the rare tRNAs.

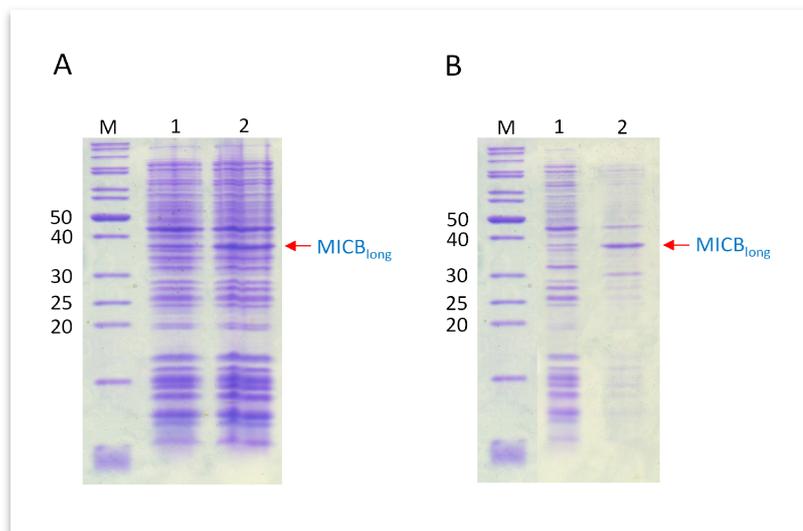


Figure 3-1. MICB_{long} expression in *E. coli* BL21(DE3) cells and MICB_{long} inclusion bodies. (A) Reducing SDS-PAGE showing the expression of MICB_{long} in *E. coli* BL21(DE3) cells before (lane 1) and after (lane 2) induction of protein expression with IPTG. (B) Reducing SDS-PAGE showing the absence and presence of a MICB_{long} protein band in the supernatant (lane 1) and inclusion body pellet (lane 2) of the centrifuged bacterial lysate, respectively. Molecular weight standards (lane M) are marked on the left in units of kDa.

The expression levels of MICB_{long} and MICB_{short} in *E.coli* Rosetta 2(DE3) cells obtained before (pre) and after (post) IPTG induction, respectively, are shown in the A panels of Figures 3-2 and 3-3. In case of MICB_{long} (Figure 3-2A) a dominant protein band is visible in the post IPTG sample (lane 2), whereas the same band is absent in the pre IPTG sample (lane 1). The molecular weight of the protein represented by this band is about 35 kDa, which nicely matches to the calculated molecular weight of MICB_{long} (32.7 kDa) and can therefore be seen as proof of a successful overexpression of MICB_{long}. In case of MICB_{short} however (Figure 3-3A), no clear overexpression was observed as a dominant protein band at or close to the calculated molecular weight of MICB_{short} (23.0 kDa) is absent in the post IPTG sample (lane 2). After MICB_{long} or MICB_{short} expression, bacterial cells were harvested by centrifugation and then lysed using a high pressure homogenizer (details see Section 2.2.3.4.2). The bacterial lysate was centrifuged at 27,000x g, which resulted in a fluid supernatant on top of a solid pellet. A sample of the supernatant was saved and analyzed by SDS-PAGE (see lanes 1 in Figures 3-1B, 3-2B and 3-3B). The pellet of the lysate consisted of two layers. The upper dark brown layer was soft and jelly-like, whereas the light brown lower layer was rather hard and clay-like, the latter features being typical signs of *E.coli* IBs. However, the initial IB pellet is usually contaminated with many impurities and has to be further purified. This was done by performing several washing steps with detergent (Triton X-100) containing buffer (see Section 2.2.3.4.3 and Table 2-12 for details).

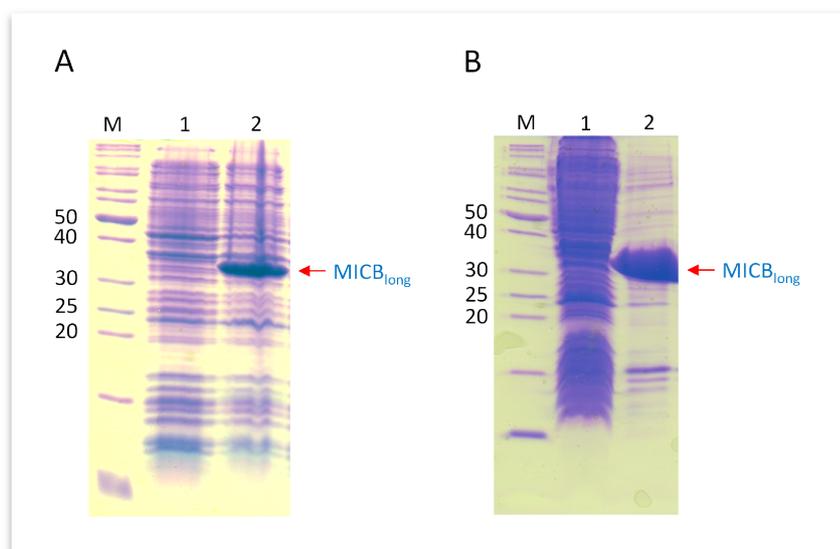


Figure 3-2. MICB_{long} expression in *E.coli* Rosetta 2(DE3) cells and MICB_{long} inclusion bodies. (A) Shown is a reducing SDS-PAGE of the MICB_{long} expression in Rosetta 2(DE3) cells before (lane 1) and after (lane 2) induction of protein production with IPTG. (B) Reducing SDS-PAGE that shows the observed absence and presence of a MICB_{long} protein band in the supernatant (lane 1) and inclusion body pellet (lane 2) of the centrifuged Rosetta 2(DE3) bacterial lysate, respectively. A molecular weight standard (lane M) is marked on the left in units of kDa.

Progress with IB purification was monitored by eye, as each washing step leads to a lighter color of the IB pellet after centrifugation. After a final wash of the IBs with detergent-free buffer (Table 2-12), IBs were analyzed by SDS-PAGE (see lanes 2 in Figures 3-1B, 3-2B and 3-3B). By comparing the supernatant samples (lanes 1) with the IB pellet samples (lanes 2) in Figures 3-2B and 3-3B it becomes quite obvious that both MICB_{long} and MICB_{short} were solely expressed in form of IBs and had to be refolded in order to obtain soluble and natively folded proteins (see Section 2.2.3.4.3). Furthermore, the observation that a MICB_{short} inclusion body pellet was obtained (Figure 3-3B) shows that a missing overexpression band after IPTG induction (lane 2 in Figure 3-3A) does not necessarily mean that there is no expression of the target protein at all.

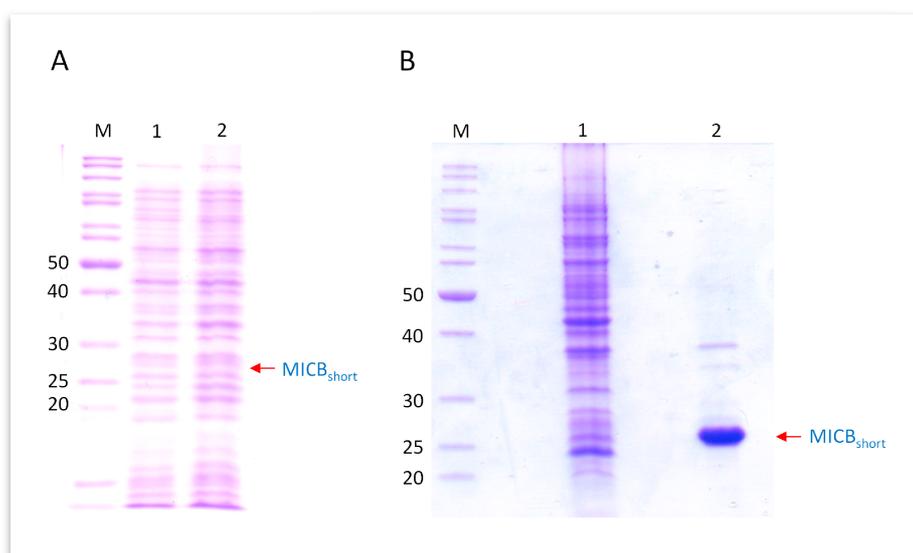


Figure 3-3. MICB_{short} expression in *E.coli* Rosetta 2(DE3) cells and MICB_{short} inclusion bodies. (A) Presented is a reducing SDS-PAGE illustrating the expression of MICB_{short} in Rosetta 2(DE3) cells before (lane 1) and after (lane 2) the induction of protein expression with IPTG. (B) Reducing SDS-PAGE showing the absence and presence of the MICB_{short} protein band in the supernatant (lane 1) and inclusion body pellet (lane 2) of the centrifuged bacterial lysate, respectively. Molecular weight standards (lane M) are marked on the left in units of kDa.

3.1.2 IMAC purification of MICB_{long} and MICB_{short}

Following refolding, MICB_{long} and MICB_{short} were further purified by IMAC (Ni-NTA) affinity chromatography (see Section 2.2.3.4.4). Representative Ni-NTA purifications of MICB_{long} and MICB_{short} are shown in Figures 3-4A and 3-4C, respectively. After loading of the MICB sample, the column was washed with Ni-NTA binding buffer (Table 2-15), and the MICB proteins were subsequently eluted with a gradient of Ni-NTA elution buffer (Table 2-15). Fractions were collected during the complete elution period and those fractions that corresponded to A₂₈₀ elution peaks in the chromatogram (Figures 3-4A and C) were analyzed by SDS-PAGE (Figures 3-4B and D). Figure 3-4B shows the fractions corresponding to the A₂₈₀ elution peak of the MICB_{long}

purification, all of which contained essentially pure MICB_{long}. Figure 3-4D however shows the fractions corresponding to the A₂₈₀ elution peak of the MICB_{short} purification, which for unknown reasons besides MICB_{short} also contained a number of additional protein contaminants. Consequently, all fractions of the MICB_{long} elution peak (Figure 3-4B, fractions 12-24) but only the most pure fractions of the MICB_{short} elution peak (Figure 3-4D, fractions 18-24) were pooled, concentrated and filtered and then used for further purification of MICB_{long} and MICB_{short} by SEC using a *Superdex 75 10/300 GL* column (see Section 2.2.3.4.5).

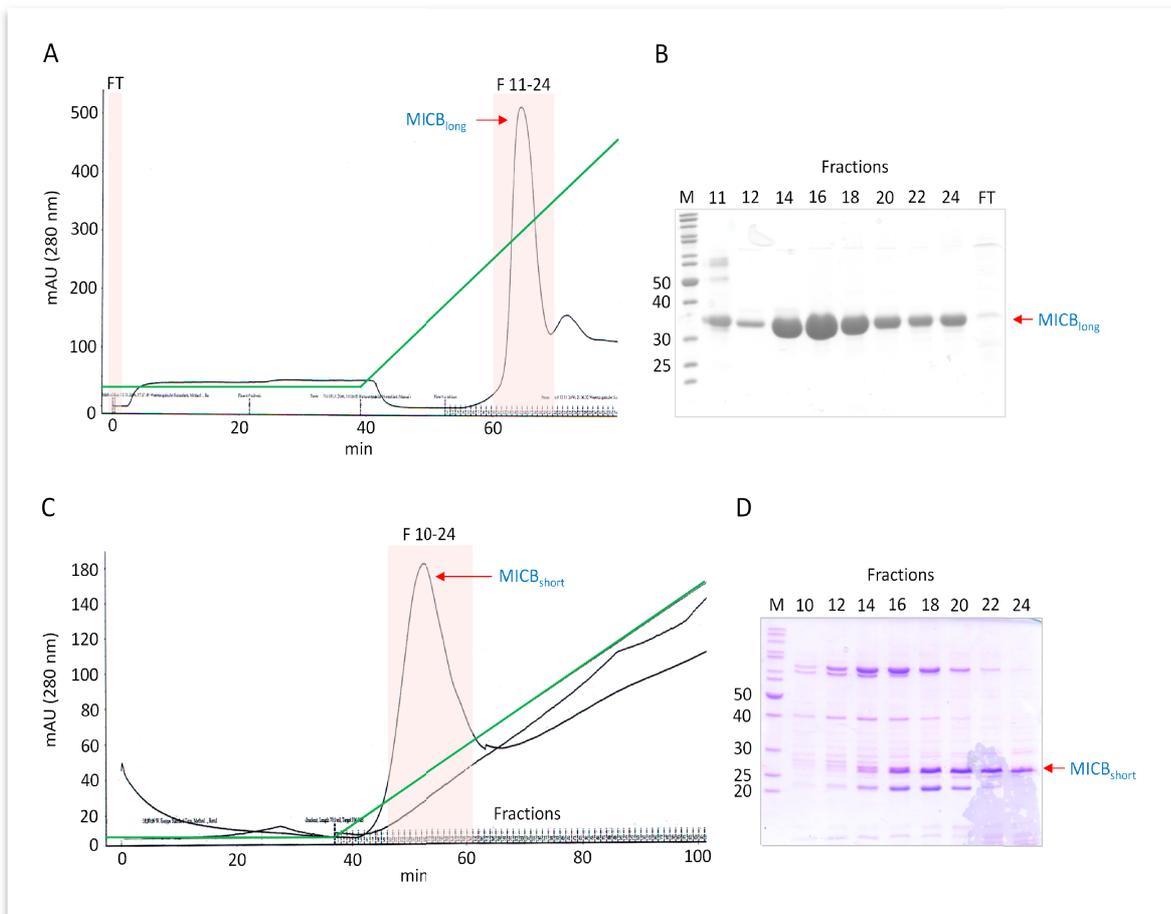


Figure 3-4. IMAC purification of MICB_{long}. (A,C) Chromatograms of the MICB_{long} (panel A) and MICB_{short} (panel C) purification by IMAC using a *HisTrap (Ni-NTA) HP* column. The solid black line shows the A₂₈₀ absorption of samples eluted from the column, while the green line indicates the increase of imidazole in the running buffer. Fractions that were subsequently analyzed by SDS-PAGE are highlighted in red. (B, D) Reducing SDS-PAGEs showing the highlighted fractions from panels A and C, respectively. The label on top of each lane corresponds to the analyzed fraction numbers. Molecular weight standards (lane M) are marked on the left in units of kDa. FT: flow through.

3.1.3 SEC purification of MICB_{long} and MICB_{short}

Representative SEC runs using a *Superdex 75 10/300 GL* column are shown for MICB_{long} and MICB_{short} in Figures 3-5A and 3-5C, respectively. MICB_{long} eluted at 33.5 min from the column, while the smaller MICB_{short}, as expected, was retained longer on the column and eluted after 36 min. The presence of only one major peak for MICB_{long} and MICB_{short} (Figures 3-5A and 3-5C) and the observation that both peaks feature a Gaussian distribution without any significant peak tailing indicates homogeneous monomeric protein species in both cases. Fractions were collected during the entire run and those fractions that corresponded to A₂₈₀ peaks in the chromatogram were analyzed by SDS-PAGE (Figures 3-5B and 3-5D). Fractions 34-38 (Figure 3-5B) and fractions 36-40 (Figure 3-5D) contained very pure MICB_{long} and MICB_{short} protein samples, respectively. This shows that the employed MICB purification strategy consisting of an initial IMAC (Ni-NTA) purification step followed by a SEC (S75) purification step is sufficient to obtain both pure and homogeneous MICB proteins. After SEC the MICB containing fractions were pooled, concentrated and filtered and then used for UL16_{mon}-MICB_{long} (see Section 3.2.10) and UL16_{mon}-MICB_{short} complex formation (see Sections 3.2.14 and 3.2.15) and crystallization (see Sections 3.3.1, 3.3.2 and 3.3.3), respectively.

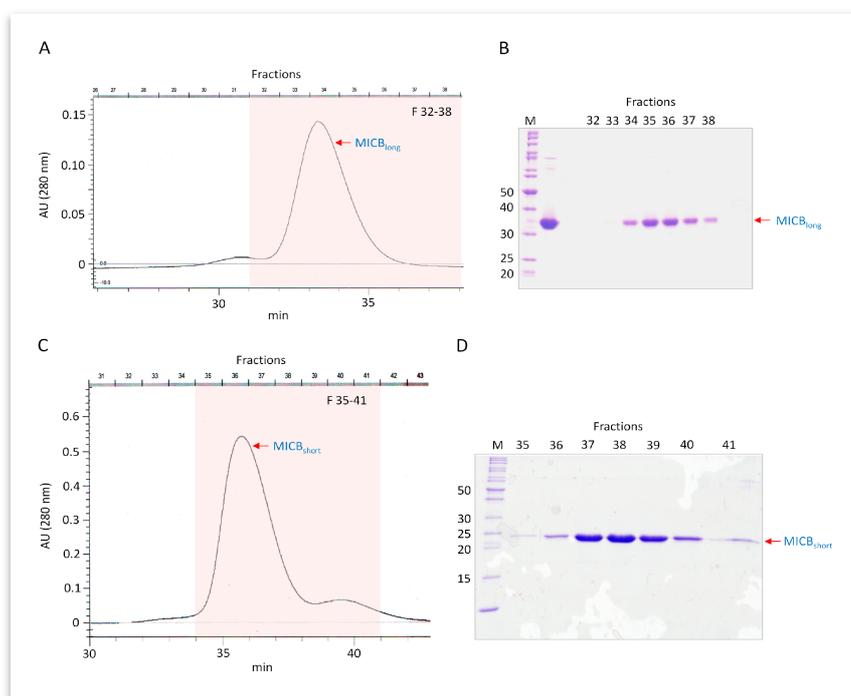


Figure 3-5. SEC purification of MICB_{long}. (A, C) Chromatograms of the MICB_{long} (panel A) and MICB_{short} (panel C) purifications by SEC using a *Superdex 75 10/300 GL* column. The solid black line shows the A₂₈₀ absorption of samples eluted from the column. Fractions that were subsequently analyzed by SDS-PAGE are highlighted in red. (B, D) Reducing SDS-PAGEs showing the highlighted fractions from panels A and C, respectively. The label on top of each lane corresponds to the analyzed fraction number. Molecular weight standards (lane M) are marked on the left in units of kDa.

3.2 Expression and purification of UL16 and the UL16-MICB complex

Efforts to obtain soluble UL16 protein after its expression as His-tagged or GST-tagged fusion proteins in *E.coli* failed. In case of His-tagged UL16, only inclusion bodies were obtained that could not be refolded (data not shown). When fused to a GST-tag, UL16 was initially soluble but precipitated as soon as the GST-tag was cleaved with *PreScission Protease* at 4 °C (data not shown). Taken together, these initial results suggested that the glycosylation of UL16 is important for its solubility and that due to the inability to produce posttranslational modifications of proteins (such as glycosylation), *E.coli* is no suitable expression system for UL16.

3.2.1 Expression of UL16 in CV-1 cells

During the early stages of the project, UL16-Fc (construct see Section 2.2.3.5.1) was successfully expressed in CV-1 cells (see Section 2.2.3.5.2). The cloning of the UL16-Fc construct into pcDNA3.1(-) and the subsequent stable transfection of CV-1 cells with the UL16-Fc containing vector was performed by Dr. Jessica Spreu. Figure 3-6 shows a non-reducing SDS gel of purified UL16 (after Fc-tag cleavage) that was produced in CV-1/UL16-Fc cells. The gel shows two broad, smeary bands that correspond to monomeric and dimeric forms of UL16 (for the monomer/dimer issue see Section 3.2.7). The lower band (monomer, UL16_{mon}) ranges from 35 - 55 kDa, while the upper band ranges from 60 - 120 kDa (dimer, UL16_{dim}). When compared to the deglycosylated UL16 monomer, which should exhibit a molecular weight of 18.5 kDa, it follows that the glycosylated UL16 monomer (Figure 3-6) consists to about 50 - 60 % of glycans. Moreover, the extreme smeariness and broadness of the glycosylated UL16 monomer and dimer bands show that UL16 glycosylation is extremely heterogeneous.

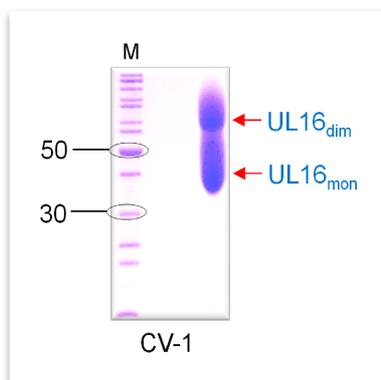


Figure 3-6. Glycosylation state of UL16 expressed by CV-1 cells. Non-reducing SDS-PAGE showing purified, untagged UL16 monomer and dimer species as observed when UL16 is expressed in CV1 cells. Molecular weight standards (lane M) are marked on the left in units of kDa.

3.2.2 Deglycosylation of UL16 produced in CV-1 cells

Since even small glycan heterogeneity (see Sections 3.3.2 and 3.3.3) can turn out to be a major obstacle in the attempts to crystallize a glycoprotein, the strong heterogeneity observed for the CV-1 cell produced UL16 (Figure 3-6) would make it basically impossible to crystallize UL16 in its fully glycosylated state. A common strategy to reduce the glycan heterogeneity of a glycoprotein, which in turn increases the chances to obtain well-diffracting glycoprotein crystals, is the employment of glycosidases such as PNGase F and Endo H. In order to test these two enzymes in terms of their capabilities to efficiently deglycosylate CV-1 cell produced UL16, the latter was separately incubated with both enzymes and the two samples subsequently analyzed by SDS-PAGE.

Lane 1 of the reducing SDS-gel in Figure 3-7 shows CV-1 cell produced UL16 after cleavage with PNGase F, while lane 2 shows the same UL16 after cleavage with EndoH. In case of PNGase F cleavage (Figure 3-7, lane 1) two bands are observed, one at 33 kDa and one at 17 kDa. The 33 kDa band represents PNGase F, while the 17 kDa band represents the completely deglycosylated UL16 protein. In case of Endo H cleavage (Figure 3-7, lane 2) four dominant bands at 27, 25, 20 and 18 kDa are observed. The 27 kDa band represents Endo H, while the 18 kDa band represents the almost completely deglycosylated UL16 since after Endo H cleavage single NAG residues remain attached to the glycosylated asparagine residues of the protein. These NAG residues are the reason why the lowest UL16 band (18 kDa) after Endo H cleavage (lane 2 in Figure 3-7) has a slightly higher molecular weight compared to the UL16 band (17 kDa) after PNGase F cleavage (lane 1 in Figure 3-7) where UL16 is completely stripped of all glycans. More interesting, however, is the presence of the 20 and 25 kDa bands (UL16 glycosylation states) after Endo H cleavage (lane 2 in Figure 3-7), which could indicate the presence of complex-type glycans at some of the UL16 N-linked glycosylation sites that cannot be cleaved by Endo H (see Section 4.1.1).

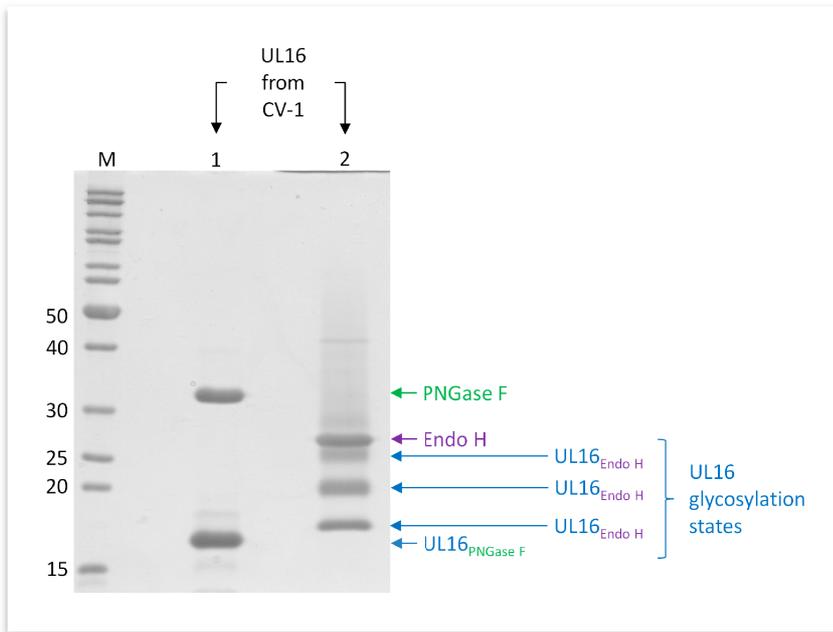


Figure 3-7. Deglycosylation of UL16 by Endo H and PNGase F. Reducing SDS-PAGE showing CV-1 cell produced, purified, untagged UL16 after cleavage with PNGase F (lane 1) and Endo H (lane 2). Protein bands corresponding to PNGase F, Endo H and distinct UL16 glycosylation states are labeled accordingly. Molecular weight standards (lane M) are marked on the left in units of kDa.

3.2.3 Expression of UL16 in CHO Lec 3.2.8.1 cells

An alternative approach to the use of glycosidases in order to reduce the amount and heterogeneity of the glycan shell of a given glycoprotein is the use of specific cell lines such as CHO Lec 3.2.8.1 cells (see Sections 2.1.6.2 and 2.2.2)¹⁰⁵. CHO Lec 3.2.8.1 cells are adherent cells that feature a combination of different lectin (Lec) phenotypes (3, 2, 8 and 1). They are deficient in at least one Golgi glycosyltransferase and two Golgi nucleotide-sugar transporters, and are therefore not able to produce hybrid or complex glycans in their Golgi compartments.

Lec1 mutant: Deficient in N-acetylglucosaminyltransferase 1 (GlcNAc-T1), the enzyme that catalyses the first committed step from oligomannose to hybrid and complex N-linked glycans in the Golgi apparatus.

Lec2 mutant: Exhibits reduced activity of CMP-sialic acid transporters that transport CMP-sialic acid into the Golgi.

Lec3 mutant: Exhibits a similar phenotype as the Lec2 mutants. However, their defect is unknown.

Lec8 mutant: Exhibits a reduced activity of UDP-galactose transporters that transport UDP-galactose reduced transport of UDP-galactose into the Golgi.

However, these mutations do not affect high-mannose glycans, which consist of two NAG residues plus an additional variable number of 5 to 9 mannose residues. Consequently even though CHO Lec 3.2.8.1 cells reduce the amount and heterogeneity of the glycan moiety by not producing hybrid and complex type glycans, some heterogeneity remains due to the variable number of mannose residues of the high-mannose glycans.

CHO Lec 3.2.8.1 cells, stably transfected with the same UL16-Fc construct (see Section 2.2.3.5.1) as the CV-1 cells, were eventually employed to produce a more homogeneous UL16 glycoprotein species. Note that if CHO Lec 3.2.8.1 cells are observed to have an effect on the glycosylation state of a protein, which can be evaluated by comparing the glycosylation state of a protein produced in CV-1 cells (full glycosylation is possible) with the glycosylation state of the same protein produced by CHO Lec 3.2.8.1 cells (missing complex and/or hybrid glycans), this strongly indicates the presence of hybrid and/or complex glycans on the surface of the native protein¹⁰⁵. The monomeric and dimeric forms of UL16 (after Fc-tag cleavage) produced by CHO Lec 3.2.8.1 cells are shown in Figure 3-8 after performing a non-reducing SDS-PAGE. When compared to the monomeric and dimeric forms of UL16 (after Fc-tag cleavage) produced by CV-1 cells (Figure 3-6), it becomes obvious that CHO Lec 3.2.8.1 produced UL16 has narrower and less smeary protein bands on the SDS-gel, which indicates a severely (hybrid and complex type) reduced glycan moiety on the surface of UL16.

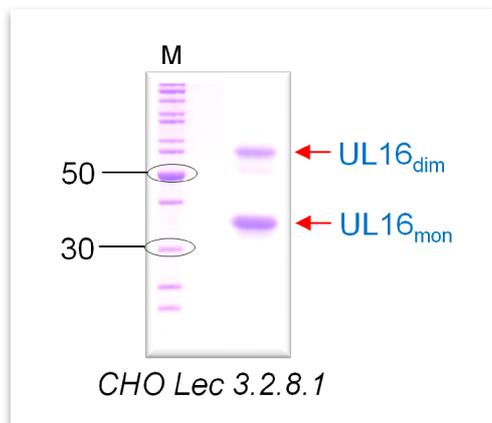


Figure 3-8. Glycosylation state of UL16 expressed by CHO Lec 3.2.8.1 cells. Non-reducing SDS-PAGE showing purified, untagged UL16 monomer and dimer species as observed when UL16 is expressed in CHO Lec 3.2.8.1 cells (compare Figure 3-6). Molecular weight standards (lane M) are marked on the left in units of kDa.

3.2.4 Subcloning of UL16-Fc expressing CHO Lec 3.2.8.1 cells

As individual cells after transfection and initial selection can have different phenotypes in terms of protein expression, it is advisable to isolate individual clones and probe their protein expression levels (see Section 2.2.2.7). Usually this is done by ELISA (enzyme-linked immunosorbent assay) techniques, which allow for an exact quantitative evaluation of the expression level. Since no established Fc-ELISA was available in the laboratory at the time a Protein A/G pull-down assay was performed (see Section 2.2.2.7.1) to qualitatively determine the level of UL16-Fc expression. The outcome of such a pull-down assay can then be easily evaluated by SDS-PAGE. If a clone expresses UL16-Fc, a protein band of about 60 kDa should be found on such a gel under reducing conditions (Mw of glycosylated UL16_{mon} from CHO Lec 3.2.8.1 cells based on SDS-PAGEs: 35 kDa; Mw of single chain Fc: 25 kDa). Indeed, Figure 3-9 shows a 60 kDa band for each of the isolated clones (and the polyclonal cells obtained after initial selection of the stably transfected cells) and therefore indicates a positive expression of UL16-Fc in each case. The second 70 kDa band visible in each lane of the SDS-PAGE in Figure 3-9 is not related to UL16-Fc and rather belongs to a protein of the serum containing cell culture supernatant that binds non-specifically to the Protein A/G beads (data not shown). The intensity of the 70 kDa band was similar for all four tested clones and the polyclonal cells. However, the intensity of the 60 kDa band that corresponds to UL16-Fc differed among the clones. Since clone 05 (lane 05 in Figure 3-9) showed the most prominent UL16-Fc band it was used for the subsequent expression rounds of UL16-Fc.

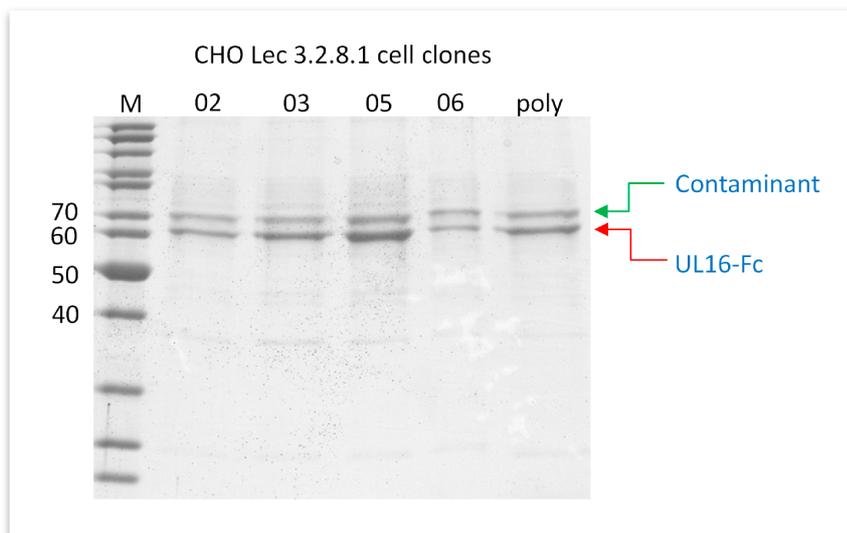


Figure 3-9. Subcloning of UL16-Fc expressing CHO Lec 3.2.8.1 cells. Reducing SDS-PAGE showing the result of a Protein A/G pull down of UL16-Fc from cell culture supernatants produced by distinct CHO Lec 3.2.8.1 clones. The label on top of each lane corresponds to the label of the individual clone analyzed. The 'poly' label refers to a polyclonal population of UL16-Fc expressing CHO Lec 3.2.8.1 cells. Molecular weight standards (lane M) are marked on the left in units of kDa.

3.2.5 UL16-Fc purification by Protein A affinity chromatography

After UL16-Fc clone 05 had been identified as the most promising clone in terms of UL16-Fc expression, its cell population was scaled up significantly in order to set up roller bottles for large scale protein production (see Section 2.2.2.2). After about 5 - 10 liters of UL16-Fc containing cell culture supernatant had been collected, UL16-Fc was purified via Protein A affinity chromatography (see Section 2.2.3.5.3). Figure 3-10A shows a typical chromatographic run of a Protein A-mediated UL16-Fc purification. After loading the Protein A column with UL16-Fc containing cell culture supernatant by means of a peristaltic pump, the column was transferred to an Äkta prime system and washed extensively with Protein A binding buffer (Table 2-16) until the absorbance at 280 nm (A_{280}) reached baseline (Figure 3-10A). The column-bound UL16-Fc was then eluted in one step with 100 % Protein A elution solution (Table 2-16). Fractions were collected during the entire elution period, and those that corresponded to A_{280} peaks in the chromatogram (Figure 3-10A) were analyzed by SDS-PAGE (Figure 3-10B). Fractions 8-14 contained high amounts of relatively pure UL16-Fc and were pooled and concentrated. The concentrated sample was then filtered and subjected to thrombin cleavage in order to remove the Fc-tag (see Section 2.2.3.5.3).

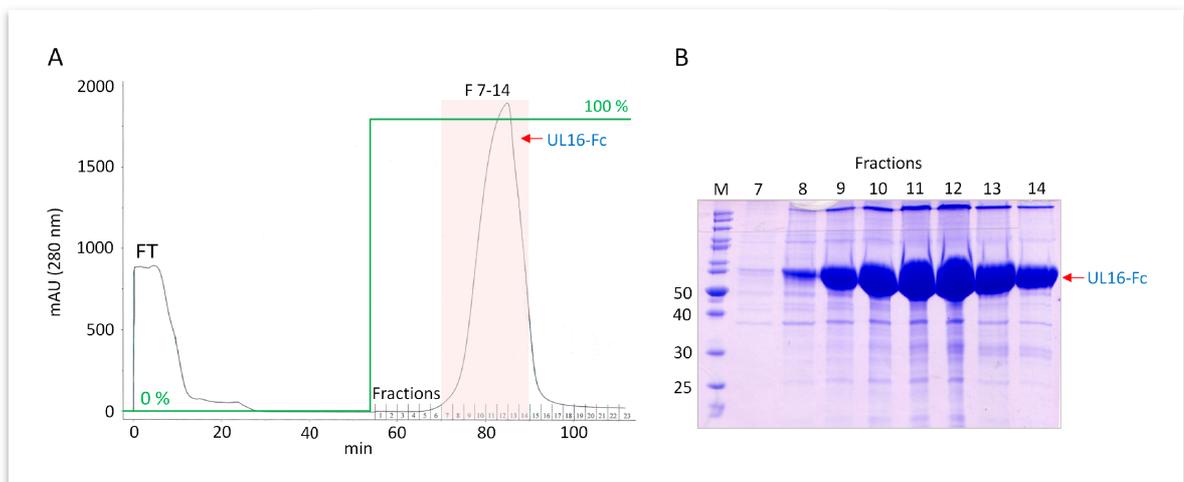


Figure 3-10. AC purification of UL16-Fc. (A) Chromatogram of the UL16-Fc (produced by CHO Lec 3.2.8.1 cells) purification by AC using a *Protein A HP* column. The solid black line shows the A_{280} absorption of samples eluted from the column, while the green line indicates the percentage of Protein A elution solution in the running buffer. Fractions that were subsequently analyzed by SDS PAGE are highlighted in red. FT: flow through. (B) Reducing SDS-PAGE showing the highlighted fractions from panel A. The label on top of each lane corresponds to the analyzed fraction number. Molecular weight standards (lane M) are marked on the left in units of kDa. The UL16-Fc purification and the SDS PAGE shown in this figure were performed by Irmgard Hähnlein-Schick.

3.2.6 UL16 purification by Protein A and benzamidine affinity chromatography

After thrombin cleavage, the protein solution consisted of three components (Figure 2-7): the UL16 protein, the cleaved Fc-tag, and thrombin. To remove the latter two components, the solution was diluted in Protein A binding buffer (Table 2-16) and passed over a Protein A column that was directly followed by a benzamidine column. Figure 3-11A shows a representative chromatographic run of this type. Fractions were collected during the entire run and those fractions that corresponded to A_{280} peaks in the chromatogram (Figure 3-11A) were analyzed by SDS-PAGE (Figure 3-11B). As UL16 lacks an Fc-tag it cannot bind to the *Protein A HP* column and was consequently found in the flow-through (see lane 1 in Figure 3-11B). Once the A_{280} flow-through peak (corresponding to UL16) had again reached baseline, the Fc-tag and thrombin were eluted together from the *Protein A HP* and *Benzamidine FF* columns in one step (see lane 2 in Figure 3-11B) with 100 % Protein A elution solution (Table 2-16). A non-reducing SDS-PAGE of the pooled, concentrated and filtered flow-through and eluate samples shows that this purification strategy efficiently separates UL16 from the cleaved Fc-tag (compare lane 1 and 2 in Figure 3-11B) and thrombin and is therefore suitable to obtain highly purified UL16 (shown in lane 1 in Figure 3-11B). The fact that a thrombin band (M_w of bovine thrombin protease: 37 kDa) is not visible in lane 2 (eluate sample) of Figure 3-11B is most likely due to the relatively small amounts of thrombin used for UL16-Fc digestion.

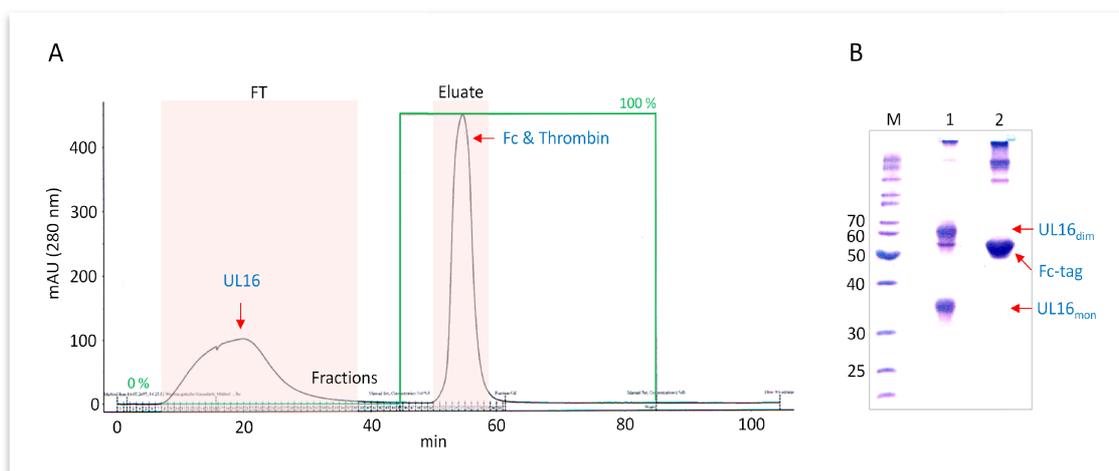


Figure 3-11. AC purification of UL16. (A) Chromatogram of the UL16 purification after thrombin cleavage of CHO Lec 3.2.8.1 cell produced UL16-Fc by affinity chromatography (AC) using a *Protein A HP* column in combination with a *Benzamidine FF* column. The solid black line shows the A_{280} absorption of samples eluted from the columns, while the green line indicates the percentage of Protein A elution solution in the running buffer. Fractions that were subsequently analyzed by SDS PAGE are highlighted in red. FT: flow through. (B) Non-reducing SDS-PAGE of the pooled FT (lane 1) and eluate (lane 2) fractions highlighted in panel A. UL16 containing FT samples were pooled, concentrated and filtered. Molecular weight standards (lane M) are marked on the left in units of kDa.

3.2.7 Monomeric and dimeric forms of UL16

As is readily visible from lane 1 of the non-reducing SDS-gel shown in Figure 3-11B, purified UL16 produced in CHO Lec 3.2.8.1 cells consists to equal parts of a monomeric (UL16_{mon}) and dimeric (UL16_{dim}) protein species that exhibit molecular weights of 35 and 60 kDa, respectively. Interestingly, the dimer band completely disappears when the UL16 sample is incubated with reducing (i. e. β -ME containing) SDS sample buffer. This observation is also shown in Figure 3-12, where lane 1 of the SDS-gel shows one UL16 band (UL16_{mon}) at 35 kDa after incubation with reducing SDS sample buffer, while lane 2 shows two UL16 bands (UL16_{mon} and UL16_{dim}) at 35 and 60 kDa, respectively, after incubation with non-reducing SDS sample buffer. The observation that the UL16 dimer only exists under non-reducing conditions strongly suggests that the UL16 dimer consists of two disulfide linked UL16 monomers. These observations, of course, give rise to the question if both UL16 species are active in terms of MICB binding.

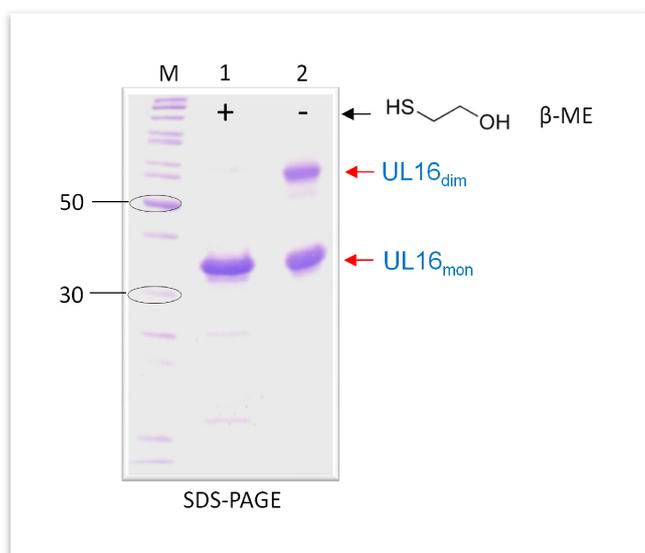


Figure 3-12. Monomeric and dimeric forms of UL16. SDS-PAGE gel showing the same sample of a UL16 monomer and dimer mixture (UL16_{mon/dim}) after incubation with (lane 1) and without (lane 2) β -ME containing SDS sample buffer (Table 2-21). Molecular weight standards (lane M) are marked on the left in units of kDa.

3.2.8 The monomer represents the active form of UL16

To determine which UL16 species is active in terms of MICB binding, samples comprising (i) a mixture of UL16_{mon} and UL16_{dim}, (ii) MICB_{long} and (iii) a mixture of all three proteins were analyzed by separately applying each of the three samples to a *Superdex 75 PC 3.2/30* SEC column connected to a SMART FPLC system. The chromatographic runs and corresponding non-reducing

SDS-gels are shown in Figure 3-13. The chromatogram of the UL16_{mon/dim} sample (black color code in Figure 3-13) shows two peaks with maxima at 102 min (see also fractions 30 and 31 in panel A) and 108 min (see also fractions 34 and 35 in panel A). The peak at 102 min corresponds to UL16_{dim} while the peak at 108 min corresponds to UL16_{mon}. Although the theoretical size of MICB_{long} (Mw: 32.7 kDa) equals the size of the glycosylated UL16 monomer produced in CHO Lec 3.2.8.1 cells (Mw: 35 kDa; Figure 3-13), MICB_{long} elutes (red color code in Figure 3-13) at 120 min (see also fractions 40 and 41 in panel C) and therefore much later than UL16_{mon} (t=108 min). A possible explanation for this observation could be that MICB_{long} has a shape that, compared to glycosylated UL16_{mon}, allows it to better access the pores of the SEC matrix, which in turn would lead to MICB_{long} being retained longer on its way through the column. Alternatively it is also possible that MICB_{long} interacts non-specifically with the column matrix, which would also result in elevated elution times. However, when MICB_{long} was incubated together with a mixture of both UL16 species (green color code in Figure 3-13) the MICB_{long} peak at 120 min disappeared (compare red and green chromatograms in Figure 3-13). Instead, MICB_{long} was now found in fractions 31-34 (panel D in Figure 3-13), which correspond to an elution time of about 105 min that lies between the elution times of UL16_{mon} (t=108) and UL16_{dim} (t=102). These results strongly indicate that MICB_{long} only bound the UL16 monomer as otherwise a more significant shift of the complex peak (green chromatogram in Figure 3-13) and MICB_{long} itself (panel D in Figure 3-13) to shorter elution times should have been observed. Therefore, one can conclude that only the UL16 monomer is active in terms of MICB_{long} binding. However, the questions remain what causes the dimerization of the UL16 monomers and why the dimer is unable to bind NKG2D ligands such as MICB.

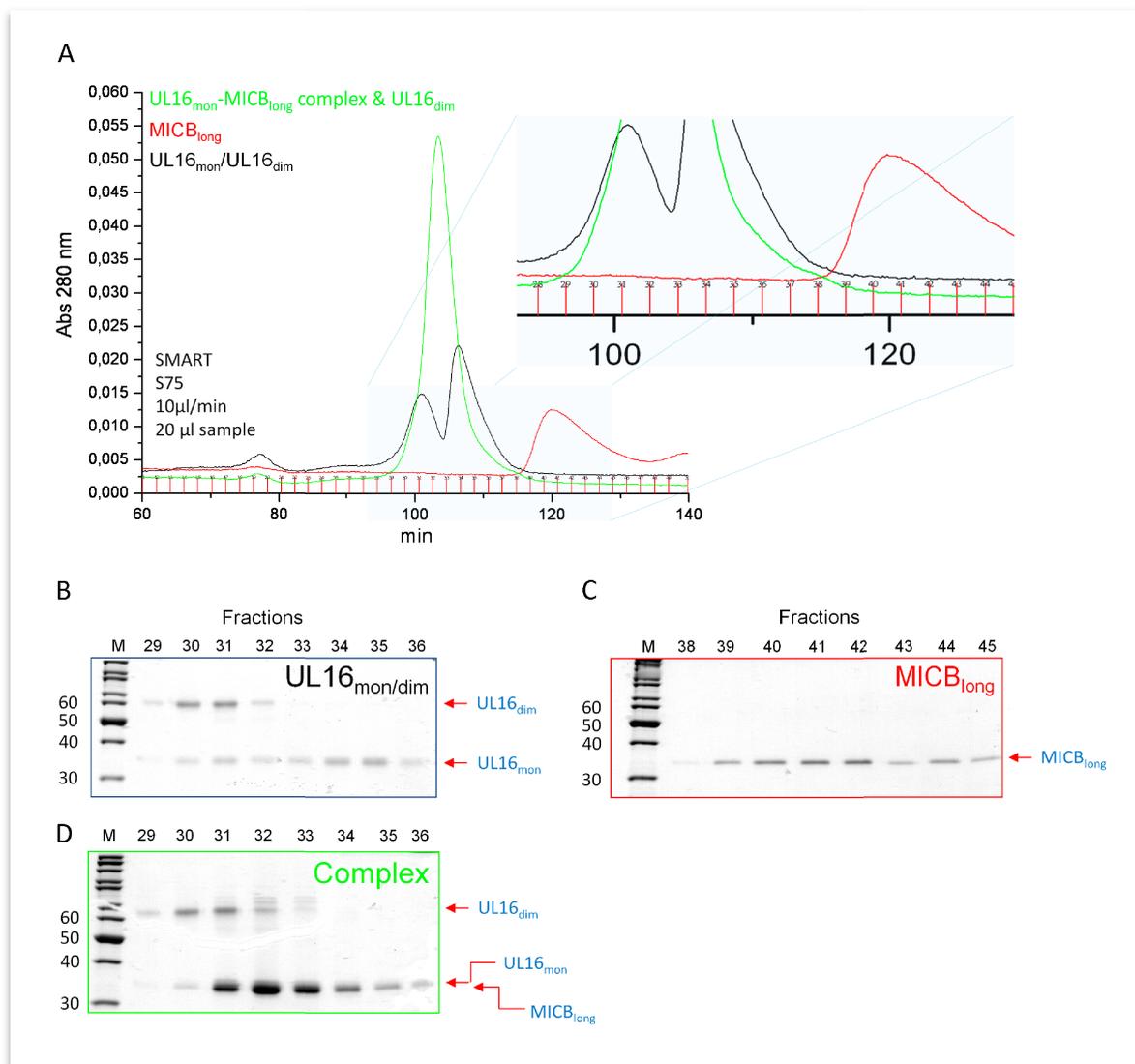


Figure 3-13. UL16_{mon}-MICB_{long} complex formation monitored by SEC. (A) Chromatograms of samples comprised of (i) UL16_{mon}-MICB_{long} complex plus UL16_{dim} (green curve), (ii) UL16_{mon/dim} (black curve) and (iii) MICB_{long} (red curve) when analyzed by analytical SEC using a *Superdex 75 PC 3.2/30* column connected to a *SMART FPLC* system. In each case, the curves show the A₂₈₀ absorption of samples eluted from the column. The inlay shows an enlarged area of the presented chromatograms. (B-D) A₂₈₀ peak fractions of the chromatograms shown in panel A, analyzed by non-reducing SDS-PAGE. The gel frames are colored according to the color code of the chromatograms shown in panel A. The label on top of each lane corresponds to the analyzed fraction number. Molecular weight standards (lane M) are marked on the left in units of kDa.

3.2.9 Standard methods fail to separate the UL16 monomer from the dimer

In order to crystallize a protein species (i. e. a single protein or a protein complex) it should be as homogeneous as possible to facilitate crystal growth and proper crystal packing. It follows that the presence of significant amounts of inactive UL16 dimers in a UL16 monomer or a UL16_{mon}-MICB_{long} complex sample would most likely hamper the formation of well-diffracting

crystals. Therefore it was tried with different methods to separate the inactive UL16 dimer from the active UL16 monomer. As described in the previous section, Figure 3-13 shows a SEC run using a *Superdex 75 PC 3.2/30* column connected to a *SMART FPLC* system. As the flow rate was set to 10 $\mu\text{l}/\text{min}$ this setup represented the highest analytical SEC resolution that could be reached in our laboratory at the time of the experiment. The red curve represented the chromatogram of $\text{MICB}_{\text{long}}$, while the black and green curves represented the chromatograms of a $\text{UL16}_{\text{mon/dim}}$ sample and a $\text{UL16}_{\text{mon/dim}}$ sample incubated with $\text{MICB}_{\text{long}}$ (Figure 3-13), respectively. One can easily see that, even when an analytical high resolution SEC column in combination with an extremely low flow rate was used, UL16 monomer and dimer (black chromatogram in Figure 3-13) could not be completely separated. Therefore, it is more than likely that a preparative SEC to separate the UL16 monomer from the dimer will not be very successful. Indeed, when this was tried by means of a preparative *Superdex 75 10/300 GL* SEC column only one peak, comprised of two overlapping peaks from the UL16 monomer and dimer, was obtained (data not shown). Since UL16 monomer and dimer are not directly separable in this way, one could envision a SEC that would separate the $\text{UL16}_{\text{mon}}\text{-MICB}_{\text{long}}$ complex from UL16_{dim} after incubation of the $\text{UL16}_{\text{mon/dim}}$ mixture with $\text{MICB}_{\text{long}}$. However, as can be seen from the single peak of the green chromatogram in Figure 3-13, which corresponds to a mixture of $\text{UL16}_{\text{mon}}\text{-MICB}_{\text{long}}$ complex and UL16_{dim} , also this attempt fails as the UL16 dimer and the $\text{UL16}_{\text{mon}}\text{-MICB}_{\text{long}}$ complex are too similar in size and therefore elute from the SEC column at the same time. A similar result was obtained with the $\text{UL16}_{\text{mon}}\text{-MICB}_{\text{short}}$ complex (Figure 3-14). Therefore one can conclude that SEC is not a suitable method to separate (i) UL16_{mon} , (ii) the $\text{UL16}_{\text{mon}}\text{-MICB}_{\text{long}}$ complex or (iii) the $\text{UL16}_{\text{mon}}\text{-MICB}_{\text{short}}$ from the UL16 dimer.

In addition to SEC, also cation and anion exchange chromatography (IEX) failed to separate the UL16 monomer from the dimer (data not shown). Further attempts aimed at reducing the intermolecular disulfide bond between the UL16 dimer protomers with reducing agents such as DTT, $\beta\text{-ME}$ or 2-mercaptoethanolamine (2-MEA) were also not successful as under all reducing conditions that were tested not only the intermolecular disulfide bond of the UL16 dimer but also the intramolecular disulfide bonds of the already existent and dimer-derived UL16 monomers were apparently reduced (data not shown).

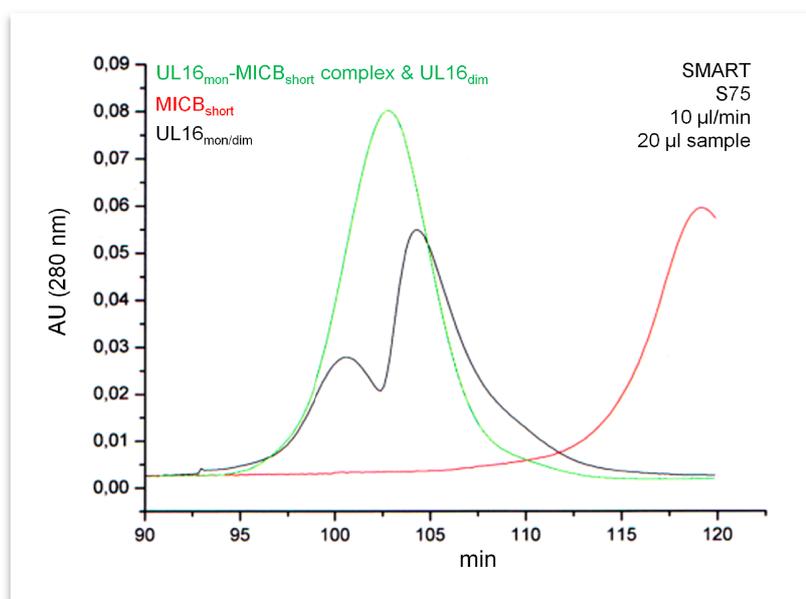


Figure 3-14. UL16_{mon}-MICB_{short} complex formation monitored by SEC. (A) Chromatograms of samples comprised of (i) UL16_{mon}-MICB_{short} complex plus UL16_{dim} (green curve), (ii) UL16_{mon/dim} (black curve) and (iii) MICB_{short} (red curve) when analyzed by analytical SEC using a *Superdex 75 PC 3.2/30* column connected to a SMART FPLC system. In each case the curves show the A₂₈₀ absorption of samples eluted from the column.

3.2.10 Ni-NTA purification of the UL16_{mon}-MICB_{long} complex

Chromatographic methods such as SEC and IEX had failed to separate the UL16 dimer from the UL16 monomer or from the UL16_{mon}-MICB_{long} complex, respectively (see Section 3.2.9). Therefore, a new strategy was devised to separate the UL16 dimer from the UL16_{mon}-MICB_{long} complex based on the ability of MICB_{long} to bind to a Ni-NTA column by means of its N-terminally fused His₆-tag (see Section 2.2.3.5.4 and Figure 2-6). Briefly, UL16_{mon/dim} mixture was incubated with an excess amount of purified MICB_{long} (see Sections 2.2.3.4 and 3.1), which resulted in a three component solution comprised of (i) UL16_{mon}-MICB_{long} complex, (ii) non-UL16-bound MICB_{long} ('free MICB_{long}'), and (iii) UL16_{dim} (see Figure 2-6). This solution was diluted in Ni-NTA binding buffer (Table 2-15) and then passed over a *HisTrap (Ni-NTA) HP* column (Figure 3-15A). Since UL16_{dim} does not have a His-tag that would allow it to bind to the *HisTrap (Ni-NTA) HP* column, UL16_{dim} was found in the flow-through accordingly (Figure 3-15A, lane 13 in Figure 3-15B and lane 1 in Figure 3-15C). This was not the case for the UL16_{mon}-MICB_{long} complex and the excess amounts of free His-tagged MICB_{long}, both of which bound tightly to the *HisTrap (Ni-NTA) HP* column via the His₆-tag provided by MICB_{long} and which had to be eluted from the *HisTrap (Ni-NTA) HP* column with elevated concentrations of imidazole (Figure 3-15A, lanes 43, 45, 47 in Figure 3-15B and lane 2 in Figure 3-15C). It was also observed that the protein band on the SDS-

gel corresponding to UL16_{mon} was rather faint when samples had been incubated with non-reducing SDS sample buffer while the band was better defined when samples had been incubated with non-reducing SDS sample buffer (compare lanes 43, 45, 47 in Figure 3-15B with lane 2 in Figure 3-15C).

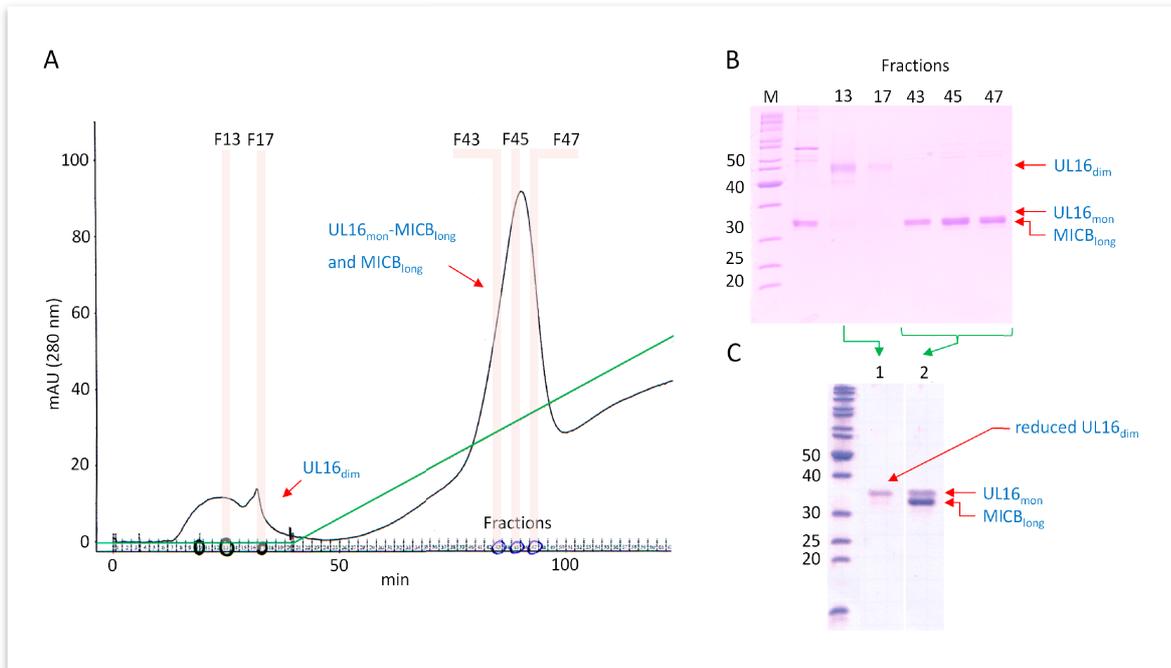


Figure 3-15. IMAC purification of the UL16_{mon}-MICB_{long} complex. (A) Chromatogram of the UL16_{mon}-MICB_{long} purification by IMAC using a *HisTrap (Ni-NTA) HP* column. The solid black line shows the A₂₈₀ absorption of samples eluted from the column, while the green line indicates the increase of imidazole in the running buffer. Fractions that were subsequently analyzed by SDS-PAGE are highlighted in red with the corresponding fraction numbers labeled. (B) Reducing SDS-PAGE of the fractions highlighted in panel A. The label on top of each lane corresponds to the analyzed fraction number. (C) Non-reducing SDS-PAGE of the same samples shown in panel B (indicated by green arrows). UL16_{mon}-MIC_{long} complex containing fractions 39 - 50 were pooled and concentrated. Molecular weight standards (lane M) are marked on the left in units of kDa.

3.2.11 SEC purification of the glycosylated UL16_{mon}-MICB_{long} complex

To separate the excess amount of free MICB_{long} from the UL16_{mon}-MICB_{long} complex, an additional SEC step using a *Superdex 75 10/300 GL* column was employed. The resulting chromatogram is shown in Figure 3-16A and shows two almost completely separated peaks at 29.5 min (Peak-1) and 32.5 min (Peak-2). Fractions were collected during the entire run. As can be seen from the reducing and non-reducing gels shown in Figure 3-16B, Peak-1 corresponded to the UL16_{mon}-MICB_{long} complex, while Peak-2 corresponded to free MICB_{long}. It was also observed again that the UL16_{mon} band is poorly visible under non-reducing conditions (left gel in Figure 3-16B), while it was clearly

visible under reducing conditions (right gel in Figure 3-16B). The UL16_{mon}-MICB_{long} complex samples were pooled, concentrated and filtered before they were used for crystal setups. MICB_{long} samples were treated similarly but instead of being used for crystallization trials they were stored at 4 °C and later used in a new round of UL16_{mon}-MICB_{long} complex formation and purification.

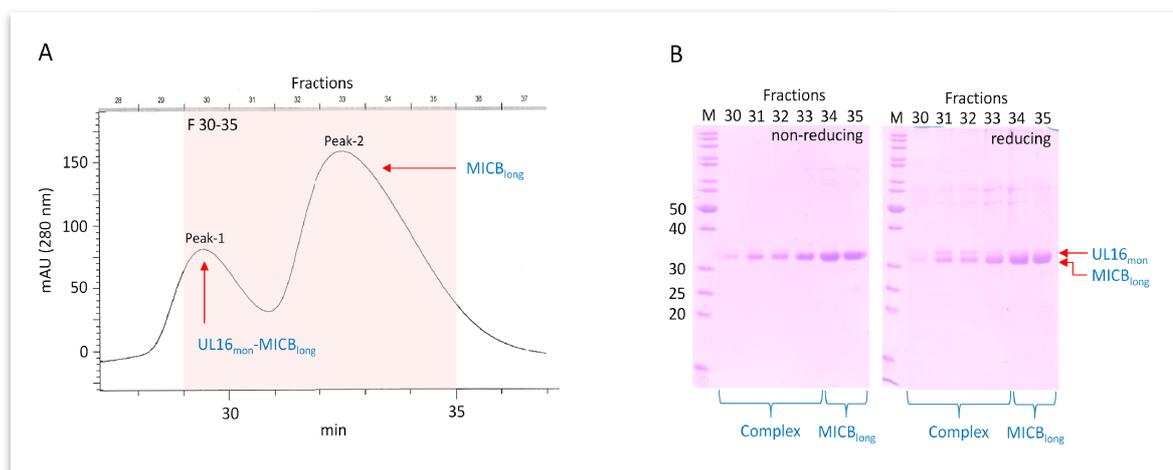


Figure 3-16. SEC purification of the UL16_{mon}-MICB_{long} complex. (A) Chromatogram of the UL16_{mon}-MICB_{long} complex purification by SEC using a *Superdex 75 10/300 GL* column. The solid black line shows the A_{280} absorption of samples eluted from the column. Fractions of the chromatogram that were subsequently analyzed by SDS-PAGE are highlighted in red with the corresponding fraction numbers labeled. (B) Non-reducing (left) and reducing (right) SDS-PAGE of the highlighted fractions in panel A. The label on top of each lane corresponds to the analyzed fraction number. UL16_{mon}-MICB_{long} complex containing fractions are indicated below the gels. Molecular weight standards (lane M) are marked on the left in units of kDa.

Figure 3-17A shows an analytical SEC run of the purified UL16_{mon}-MICB_{long} complex (Figure 3-16) performed with a *Superdex 200 PC 3.2/30* column connected to a *SMART FPLC* system. The SEC run resulted in a single symmetric A_{280} peak at 30.5 min (Figure 3-17A). Fractions of the peak were analyzed by reducing SDS-PAGE and found to contain pure, homogeneous UL16_{mon}-MICB_{long} complex (Figure 3-17B). This proved that the abovementioned chromatographic strategy to purify the UL16_{mon}-MICB_{long} complex by means of a Ni-NTA column followed by a Superdex 75 column was successful.

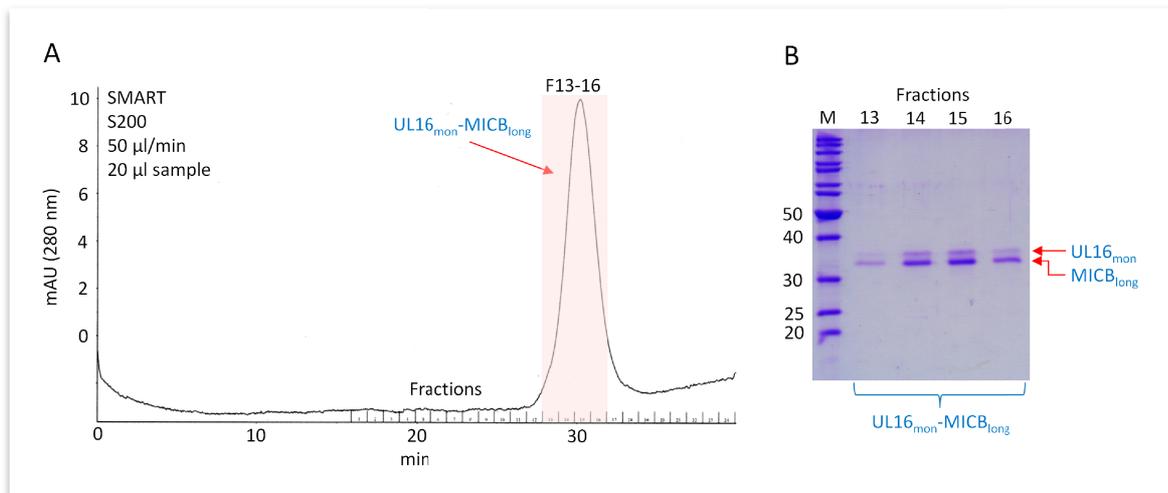


Figure 3-17. Analytical SEC of the UL16_{mon}-MICB_{long} complex. (A) Chromatogram of a UL16_{mon}-MICB_{long} complex sample that was purified using the strategy presented in Figure 3-16 when analyzed by analytical SEC using a *Superdex 200 PC 3.2/30* column connected to a *SMART* FPLC system. Fractions of the chromatogram that were subsequently analyzed by SDS-PAGE are highlighted in red with the corresponding fraction numbers labeled. (B) Non-reducing SDS-PAGE of the highlighted fractions in panel A. The label on top of each lane corresponds to the analyzed fraction number. UL16_{mon}-MICB_{long} complex containing fractions are indicated below the gel. Molecular weight standards (lane M) are marked on the left in units of kDa.

3.2.12 Ni-NTA purification of the UL16_{mon}-MICB_{short} complex

As described in detail in Section 3.3.1, crystallization trials with the UL16_{mon}-MICB_{long} complex (purification see previous section) revealed that it was in principle possible to produce UL16-MICB complex crystals comprised of UL16_{mon} and the MHC class I-like $\alpha 1\alpha 2$ platform-domain of MICB (termed MICB _{$\alpha 1\alpha 2$}). To reproduce the coincidentally obtained UL16_{mon}-MICB _{$\alpha 1\alpha 2$} crystals, a new MICB construct, MICB_{short} (see Section 2.2.3.4.1 and Figure 2-1), was designed that consisted only of the MHC class I-like $\alpha 1\alpha 2$ platform-domain.

Expression and purification protocols that had been developed for MICB_{long} turned out to work also perfectly for MICB_{short}, which is not surprising if the high similarity of both constructs is considered. However, attempts to purify the UL16_{mon}-MICB_{short} complex with the same strategy as devised for UL16_{mon}-MICB_{long} were surprisingly not successful as the Ni-NTA affinity chromatography failed completely (Figure 3-18A). UL16_{mon/dim} and an excess amount of MICB_{short} had been incubated to form the three component solution consisting of (i) UL16_{dim}, (ii) UL16_{mon}-MICB_{short} and (iii) free, non-UL16-bound MICB. When this solution was passed over the Ni-NTA column, a large peak was observed in the flow through, while only a small peak was observed when the column was eluted with an imidazole gradient (Figure 3-18A). This is the complete opposite of what was observed during the UL16_{mon}-MICB_{long} Ni-NTA purification step (compare chromatograms in Figures 3-15A and 3-18A). As can be seen from the reducing SDS-gel

in Figure 3-18B the flow-through peak contained not only the UL16_{dim} but also UL16_{mon} and MICB_{short} proteins, while the elution peak contained a small amount of free non UL16-bound MICB_{short}. The most likely explanation of the observed results was that the UL16_{mon} and MICB_{short} proteins that were found alongside UL16_{dim} in the flow-through actually represented the UL16_{mon}-MICB_{short} complex, which was for some reason unable to bind to the Ni-NTA column.

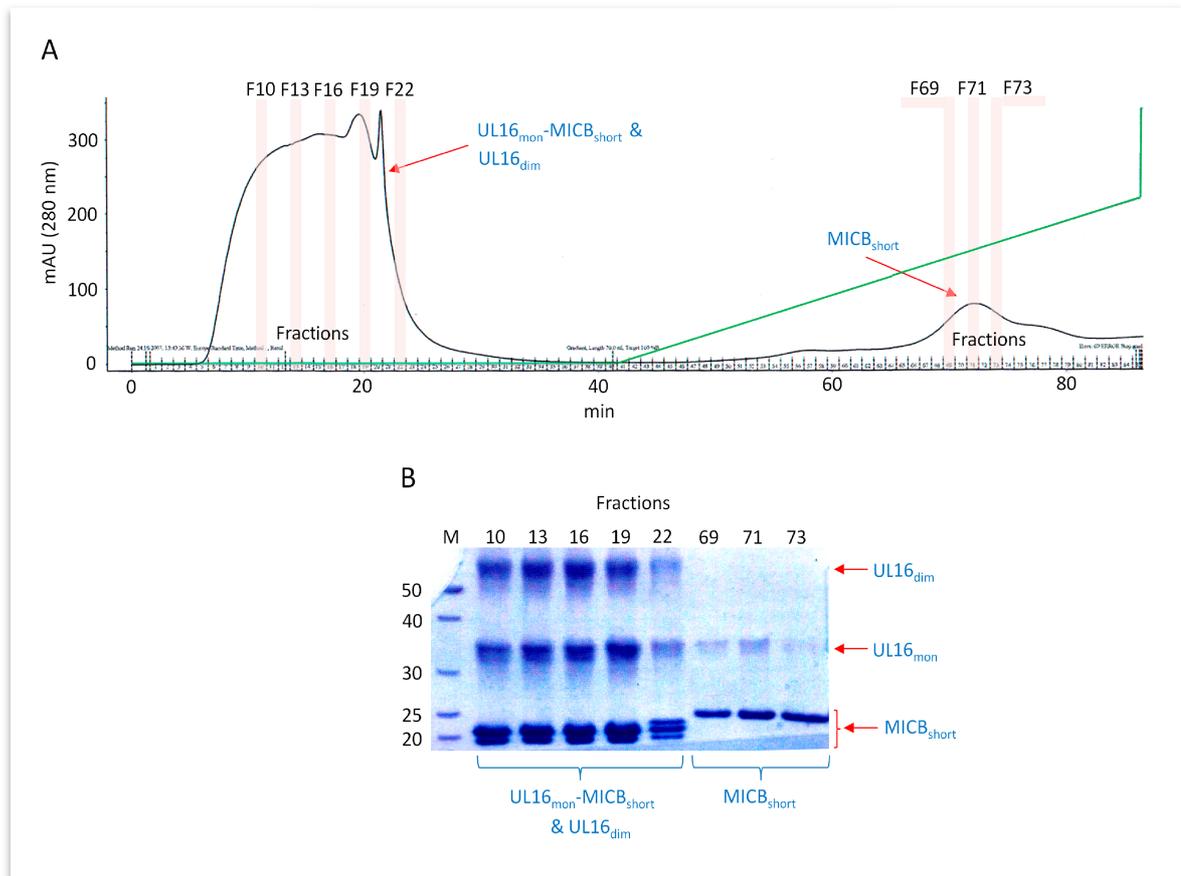


Figure 3-18. IMAC purification of the UL16_{mon}-MICB_{short} complex. (A) Chromatogram of the UL16_{mon}-MICB_{short} purification by IMAC using a *HisTrap* (Ni-NTA) *HP* column. The solid black line shows the A₂₈₀ absorption of samples eluted from the column, while the green line indicates the increase of imidazole in the running buffer. Fractions that were subsequently analyzed by SDS-PAGE are highlighted in red with fraction numbers indicated. (B) Non-reducing SDS-PAGE of the fractions highlighted in panel A. The label on top of each lane corresponds to the analyzed fraction number. Fractions containing a mixture of UL16_{mon}-MICB_{short} complex and UL16_{dim} and those containing only MICB_{long} are indicated below the gel. Molecular weight standards (lane M) are marked on the left in units of kDa.

3.2.13 HIC purification of the UL16 monomer

The strategy that finally allowed for the separation of the active UL16 monomer from the inactive UL16 dimer involved the use of *Phenyl-HP* hydrophobic interaction columns (HIC; see Section 2.2.3.5.6). Figure 3-19A shows a representative purification of the UL16 monomer using this type

of chromatography that is based on the so-called ‘salting-out effect’ of proteins (see Section 2.2.3.5.6 for details). The UL16 dimer species did not bind to the phenyl matrix and was found in the flow-through (see lane 1 in Figure 3-19B). On the other hand, the UL16 monomer species bound tightly and was eluted in one step with 100 % of HIC elution buffer (Table 2-18; see lane 2 in Figure 3-19B).

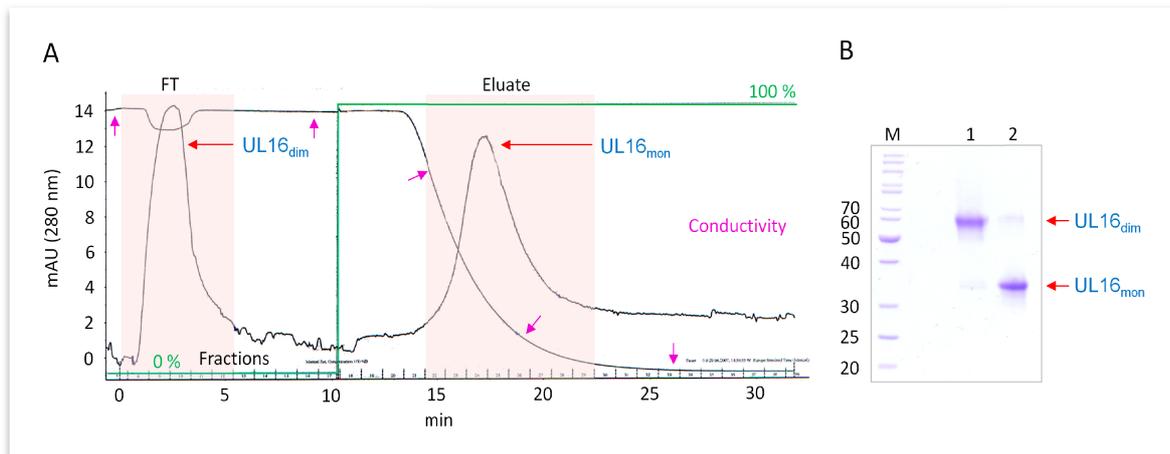


Figure 3-19. HIC purification of the UL16 monomer. (A) Chromatogram of the UL16 monomer (UL16_{mon}) separation from the dimer (UL16_{dimer}) by HIC using a *Phenyl HP* column. The solid black line shows the A₂₈₀ absorption of samples eluted from the column, while the green line indicates the percentage of HIC elution buffer in the running buffer. The pink arrows mark the path of the conductivity curve. Flow-through (FT) and eluate fractions (see labels on top of the chromatogram) that were pooled and subsequently analyzed by SDS-PAGE are highlighted in red. (B) Non-reducing SDS-PAGE of the pooled FT (lane 1) and eluate (lane 2) fractions highlighted in panel A. The label on top of each lane corresponds to the analyzed fraction number (see panel A). Molecular weight standards (lane M) are marked on the left in units of kDa.

3.2.14 SEC purification of glycosylated UL16_{mon}-MICB_{short} complex

To obtain the UL16_{mon}-MICB_{short} complex, purified UL16 monomer was first mixed and incubated with an excess amount of MICB_{short} and the complex then separated from unbound MICB_{short} by SEC (see Section 2.2.3.6). Figure 3-20A shows a representative SEC run of the UL16_{mon}-MICB_{short} complex purification. Two protein peaks are clearly visible and almost entirely separated from each other. The non-reducing gel in Figure 3-20B shows that the early protein peak (Peak-1: 31 min) in the chromatogram represents the UL16_{mon}-MICB_{short} complex, while the late peak (Peak-2: 36 min) represents MICB_{short}. The UL16_{mon} band is observed at 32 kDa on the gel, while the MICB_{short} band is running at about 25 kDa. Additional bands above or below the UL16_{mon} and MICB_{short} protein band in lane 1 of Figure 3-20B were most likely caused by a gel artifact. The complex samples (Peak-1) were pooled, concentrated to 15 mg/ml and after 0.2 μm filtration used for crystallization trials (see Section 3.3.2). The samples containing free MICB_{short} (Peak-2), were also pooled and

concentrated and the obtained pure MICB_{short} used in subsequent cycles of UL16_{mon}-MICB_{short} complex formation and purification.

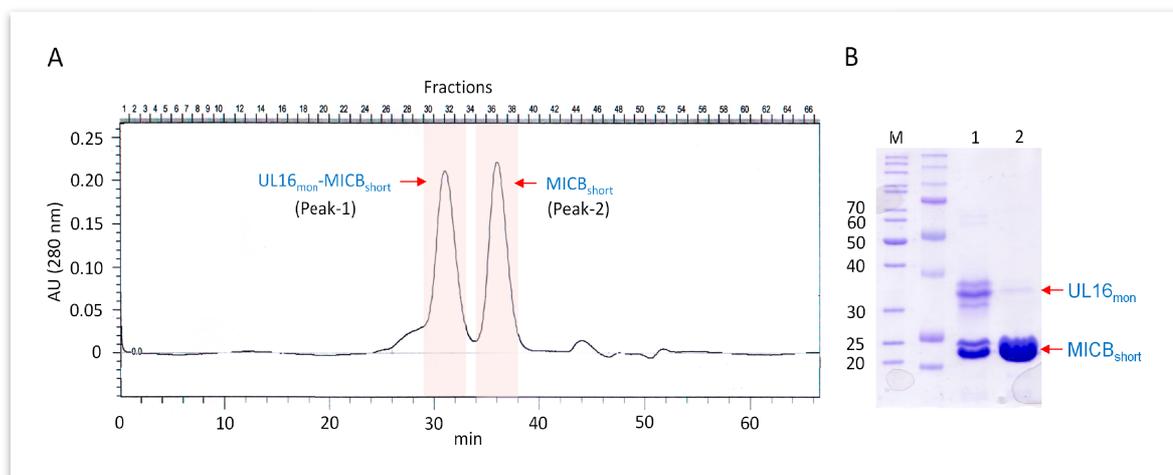


Figure 3-20. Purification of glycosylated UL16_{mon}-MICB_{short} complex. (A) Chromatogram of the UL16_{mon}-MICB_{short} complex purification by SEC using a *Superdex 75 10/300 GL* column. The solid black line shows the A₂₈₀ absorption of samples eluted from the column. The red highlighted fractions of each peak were pooled and subsequently analyzed by SDS-PAGE. (B) Non-reducing SDS-PAGE of the pooled peak fractions highlighted in panel A. Peak-1 and Peak-2 were analyzed in lanes 1 and 2, respectively. Molecular weight standards (lane M) are marked on the left in units of kDa.

3.2.15 SEC purification of deglycosylated UL16_{mon}-MICB_{short} complex

To obtain the deglycosylated UL16_{mon}-MICB_{short} complex, purified UL16 monomer was first mixed and incubated with an excess amount of MICB_{short}, followed by its deglycosylation with Endo H (see Section 2.2.3.7). The deglycosylated complex was then separated from excess MICB_{short} by SEC of which a representative run is shown in Figure 3-21A. As shown on the non-reducing gel in Figure 3-21B, the main peak in the chromatogram (Figure 3-21A) represents the deglycosylated UL16_{mon}-MICB_{short} complex, while the shoulder that elutes after the main peak represents unbound MICB_{short} excess. The protein band that corresponds to the left shoulder of the main peak could not be clearly identified. The deglycosylated UL16_{mon}-MICB_{short} complex samples (fractions 32-38) were pooled, concentrated to 15 mg/ml and, after 0.2 μm filtration, used for crystallization trials. For details about the crystallization of the deglycosylated UL16_{mon}-MICB_{short} complex see Section 3.3.3.

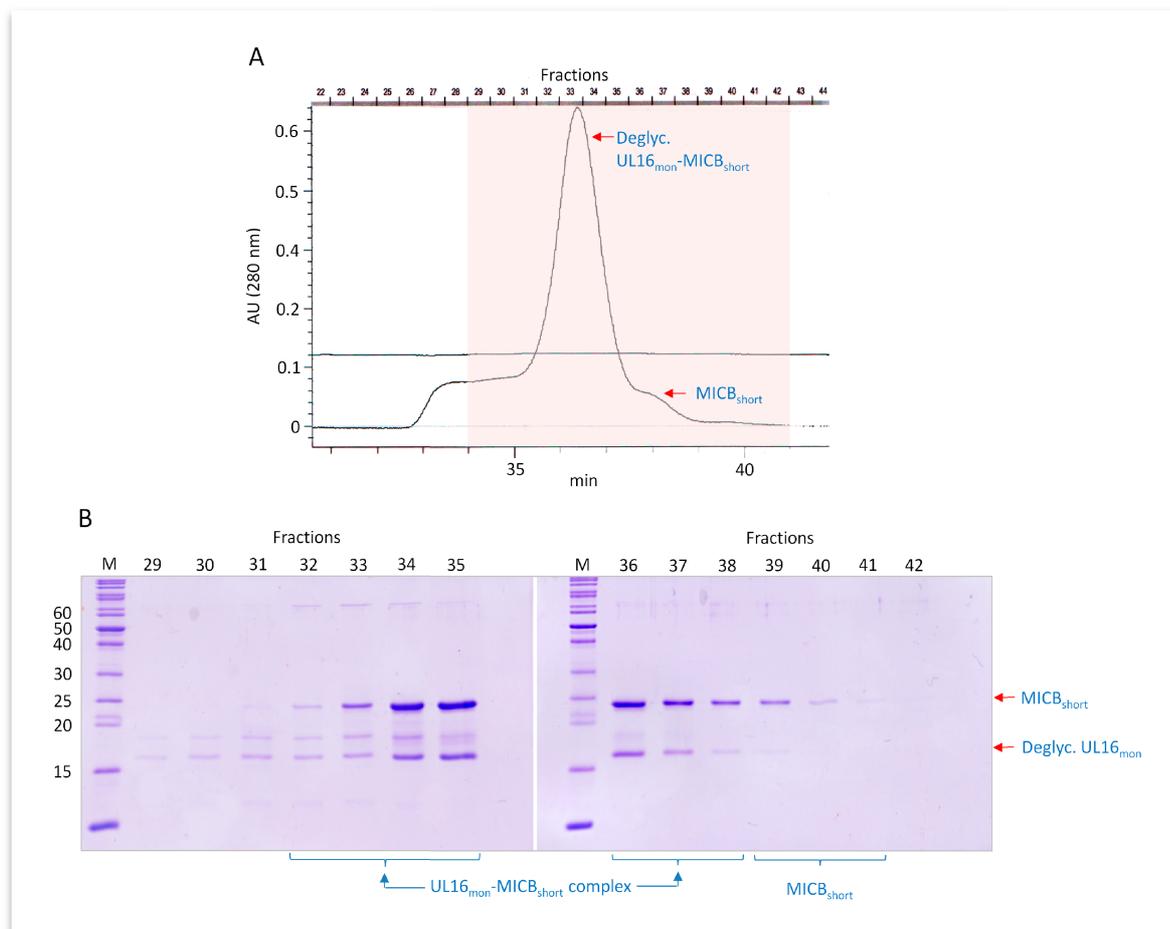


Figure 3-21. Purification of Endo H deglycosylated UL16_{mon}-MICB_{short} complex. (A) Chromatogram of the Endo H deglycosylated UL16_{mon}-MICB_{short} complex purification by SEC using a *Superdex 75 10/300 GL* column. The solid black line shows the A₂₈₀ absorption of samples eluted from the column. Fractions that were subsequently analyzed by SDS-PAGE are highlighted in red. (B) Non-reducing SDS-PAGE of fractions highlighted in panel A. The label on top of each lane corresponds to the analyzed fraction number. Molecular weight standards (lane M) are marked on the left in units of kDa.

3.3 Crystallization of the UL16-MICB complex

3.3.1 Glycosylated UL16_{mon}-MICB_{long} complex

Initial crystals of the glycosylated UL16_{mon}-MICB_{long} complex (see Section 3.2.11) were obtained by mixing a 15 mg/ml complex solution 1:1 with crystallization solutions provided by the Hampton Light Screen (conditions 15 and 28). Crystals grew at 20 °C over a period of 1 month using the hanging drop vapor diffusion method (i. e. crystal drops were pipetted by hand). Condition 15 comprised 0.2 M ammonium sulphate, 0.1 M sodium cacodylate trihydrate pH 6.5 and 15 % (w/v) PEG 8000, while condition 28 comprised 0.2 M sodium acetate trihydrate, 0.1 M sodium cacodylate trihydrate pH 6.5 and 15 % (w/v) PEG 8000. Glycosylated UL16_{mon}-MICB_{long} complex

crystals were fished, soaked in reservoir solution enriched with 25 % glycerol, and then flash frozen in liquid nitrogen prior to their X-ray exposure (see Section 2.2.5.4). Unfortunately, when tested at our X-ray home source, neither a protein nor a salt diffraction pattern could be observed. Therefore, to make sure the obtained crystals really consisted of UL16_{mon}-MICB_{long} complex a large crystal, which during handling broke into pieces (Figure 3-22), was extensively washed with mother liquor, diluted in reducing SDS sample buffer, and finally analyzed by SDS-PAGE.

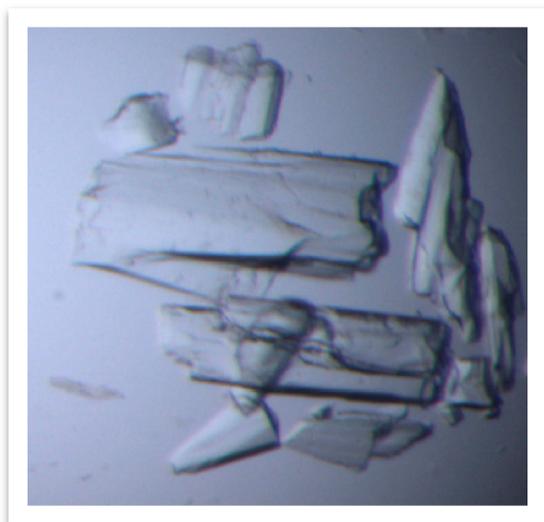


Figure 3-22. Crystal obtained from a protein solution containing UL16_{mon}-MICB_{long} complex.

In total, three samples were analyzed by SDS-PAGE under reducing conditions (Figure 3-23): A purified UL16_{mon/dim} sample of glycosylated UL16 produced in CHO Lec 3.2.8.1 cells (lane 1), a purified MICB_{long} sample (lane 3) and the dissolved crystal of the glycosylated UL16_{mon}-MICB_{short} complex (lane 2). Since only tiny amounts of protein are usually present in a crystal, instead of a standard Coomassie staining a silver staining (see Section 2.1.5.5) of the SDS-gel shown in Figure 3-23 was performed. The protein bands visible in lane 2 (Crystal) proved that the obtained crystals did indeed contain protein. By comparing the protein bands from lane 2 with the positive controls shown in lanes 1 (UL16_{mon}) and 3 (MICB_{long}) it became obvious however, that only UL16 (35 kDa) was present in the crystal, while a band of MICB_{long} (32 kDa) was missing. Instead two additional bands (purple labels) at 50 (labeled MS-1) and 20 kDa (labeled MS-2) were found that seemed unrelated to the UL16_{mon} and MICB_{long} proteins used for setting up the crystals. In order to determine which proteins were represented by the two additional protein bands, both bands were cut out of the gel and sent for mass spectrometric analysis. The latter was performed externally (TOPLAB GmbH, Martinsried). While the identification of the protein corresponding to the MS-1 band failed, the protein that corresponded to the MS-2 band was successfully identified as the

MICB $\alpha 1\alpha 2$ platform-domain (Figure 2-1). This means, that what had crystallized was not the expected $UL16_{mon}$ -MICB $_{long}$ complex but a complex consisting of $UL16_{mon}$ and a MICB that had been deprived of its $\alpha 3$ -domain (Figure 2-1).

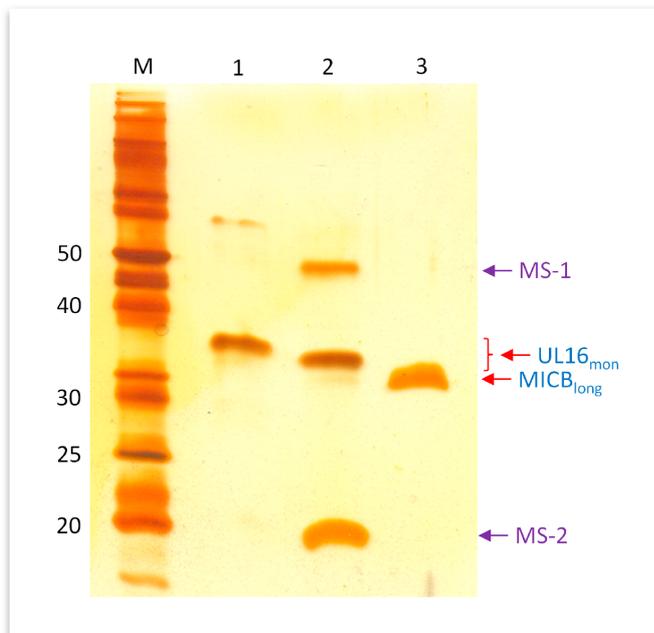


Figure 3-23. Crystal composition of the potential $UL16_{mon}$ -MICB $_{long}$ crystal. (A) Reducing SDS-PAGE of samples comprised of (i) a mixture of purified $UL16$ monomer and dimer (lane 1), (ii) fragments of the apparent $UL16_{mon}$ -MICB $_{long}$ crystal (lane 2) and (iii) purified MICB $_{long}$ (lane 3). Protein bands labeled MS-1 and MS-2 were analyzed by mass spectrometry. Molecular weight standards (lane M) are marked on the left in units of kDa. The gel was silver stained (see Section 2.2.3.2).

3.3.2 Glycosylated $UL16_{mon}$ -MICB $_{short}$ complex

Crystals of the (glycosylated) $UL16_{mon}$ -MICB $_{short}$ complex (see Section 3.2.14) were obtained by mixing a 15 mg/ml complex solution 1:1 with crystallization solutions provided by the Hampton Crystal Screens 1 (condition 15) & 2 (condition 26) and the Hampton Light Screen (conditions 15 and 28). Crystals grew at 20 °C over a period of 1 month using the sitting drop vapor diffusion method (i. e. crystal drops were pipetted by crystallization robot). As expected, all identified conditions were extremely similar to those initially identified in Section 3.3.1. Based on these results, conditions were further refined with self-made fine screens. All identified conditions essentially comprised 0.1 M of buffer at pH 6.5 (either sodium cacodylate or MES monohydrate), 0.2 M of salt (either ammonium sulphate or sodium acetate) and 15 - 30 % of precipitant (PEG 8000 or PEG monomethylether 5000). Glycosylated $UL16_{mon}$ -MICB $_{short}$ complex crystals were fished, soaked in reservoir solution enriched with cryoprotectant (various types were tested), and then flash frozen in liquid nitrogen prior to data collection (see Section 2.2.5.4). A picture of the

obtained crystals is shown in Figure 3-24. Crystals diffracted at our X-ray home source to about 8 - 10 Å, a resolution that could not be increased even when crystals were exposed to high intensity synchrotron radiation. Attempts to improve the resolution by dehydrating the crystals (performed essentially as described in Ref.¹²⁹) were not successful.

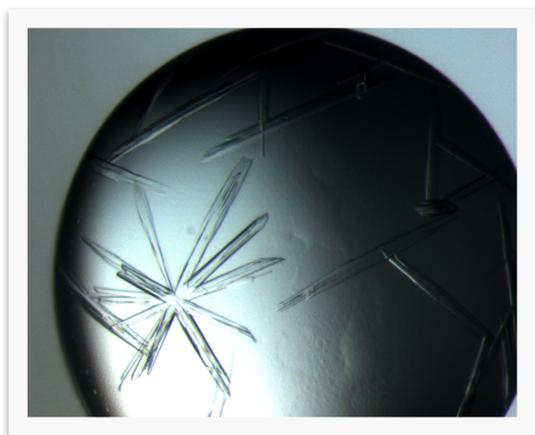


Figure 3-24. Crystals of the UL16_{mon}-MICB_{short} complex. The crystal shown was grown from a UL16_{mon}-MICB_{short} complex protein solution.

3.3.3 Deglycosylated UL16_{mon}-MICB_{short} complex

A pivotal step in obtaining well-diffracting UL16_{mon}-MICB_{short} complex crystals was the deglycosylation of UL16_{mon}, performed after complex formation with MICB_{short}. Crystals of the deglycosylated UL16_{mon}-MICB_{short} complex were obtained under essentially the same conditions as described in Section 3.3.2. The best diffracting UL16_{mon}-MICB_{short} crystals grew after mixing 15 mg/ml deglycosylated complex solution 1:1 with a crystallization solution containing 0.2 M ammonium sulphate, 0.1 M sodium cacodylate pH 6.5, and 25 % PEG 8000. Deglycosylated UL16_{mon}-MICB_{short} crystals grew at 4 °C over a time period of 4 months using the sitting drop vapor diffusion method (i. e. crystal drops were pipetted by crystallization robot). They were fished, soaked in reservoir solution enriched with 15 % ethylene glycol, and then flash frozen in liquid nitrogen prior to data collection (see Section 2.2.5.4). Deglycosylated UL16_{mon}-MICB_{short} crystals diffracted to 1.8 Å using synchrotron radiation. However, no deglycosylated UL16_{mon}-MICB_{short} crystals were obtained at 20 °C.

3.4 SPR measurements of UL16-NKG2D ligand interactions

SPR experiments were performed to obtain the kinetic and steady-state binding parameters for different UL16_{mon}-NKG2D ligand interactions (Table 3-1). These experiments revealed an average dissociation constant (K_D) for the UL16_{mon}-MICB_{long} complex of 67 nM (Figure 3-25A and C), while the average K_D value for the UL16_{mon}-MICB_{short} complex was determined to be 66 nM (Figure 3-25B and D), respectively. These K_D values are virtually identical and therefore indicate that the $\alpha 3$ -domain of MICB is not involved in the binding to UL16. Furthermore, also various ULBP molecules were tested for their interaction with the UL16 monomer. The average affinities for the UL16_{mon}-ULBP1 (Figure 3-26B) and UL16_{mon}-ULBP2 (Figure 3-26A and C) complexes were shown to be 12 and 43 nM, respectively, whereas no interaction was observed between the UL16 monomer and NKG2D ligands ULBP3, ULBP4 or ULBP5. The highest UL16 monomer (analyte) concentration tested was 100 μ M. One may wonder at this point why no steady-state analysis was performed for the interaction between UL16 monomer and ULBP1 (Table 3-1 and Figure 3-26B). This is because the low UL16 monomer concentrations did not reach the chemical equilibrium during association, which is an important requirement to obtain reliable steady-state K_D values that can be subsequently compared to the K_D values obtained from kinetic SPR experiments.

Table 3-1. Kinetic and affinity data determined by SPR¹¹³

Experimental setup		Steady-State-Analysis	Kinetic Analysis			Average Affinity
Immobilized (ligand)	in solution (analyte)	K_D [M]	K_D [M]	k_{on} [$M^{-1}s^{-1}$]	k_{off} [s^{-1}]	K_D [M]
MICB _{short} ($\alpha 1-2$)	UL16	$66 \cdot 10^{-9}$	$66 \cdot 10^{-9}$	$0.82 \cdot 10^6$	0.054	$66 \cdot 10^{-9}$
MICB _{long} ($\alpha 1-3$)		$68 \cdot 10^{-9}$	$66 \cdot 10^{-9}$	$0.68 \cdot 10^6$	0.045	$67 \cdot 10^{-9}$
ULBP1-Fc		n.d.	$12 \cdot 10^{-9}$	$1.31 \cdot 10^6$	0.016	$12 \cdot 10^{-9}$
ULBP2-Fc		$43 \cdot 10^{-9}$	$42 \cdot 10^{-9}$	$1.71 \cdot 10^6$	0.072	$43 \cdot 10^{-9}$
ULBP3-Fc		No SPR response was observed upon injection of 100 μ M UL16				
ULBP4-Fc						
ULBP5-Fc						

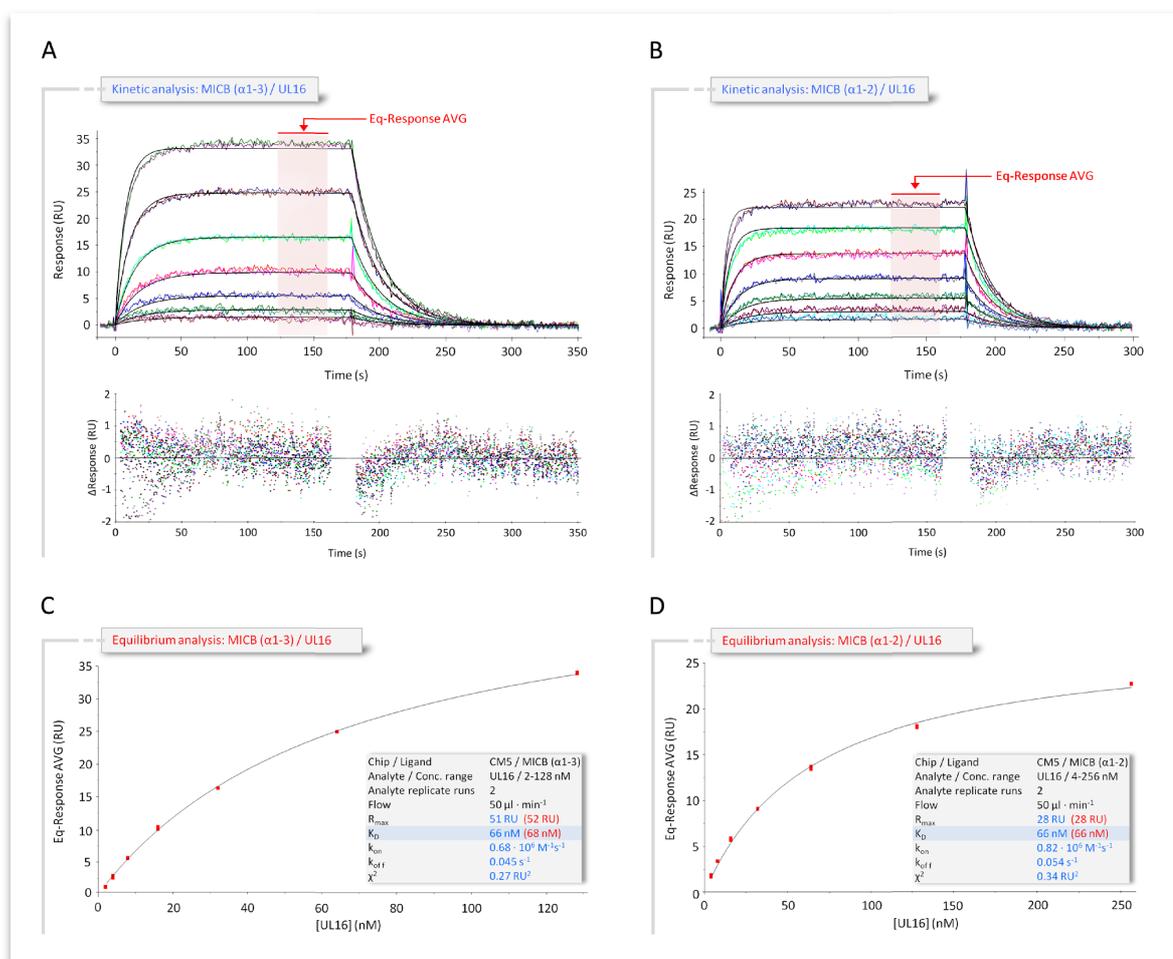


Figure 3-25. ‘Kinetic and equilibrium SPR analyses of UL16 interactions with MICB. (A,B) Kinetic analyses of UL16 binding to covalently immobilized MICB proteins comprising domains α 1 and α 2 only (MICB_{short}) (A), and domains α 1, α 2 and α 3 (MICB_{long}). (B), Each individual analyte concentration was injected twice and data are representative of at least two separate experiments with similar results. Double-referenced sensorgrams (shown in color) are overlaid with fits of a ‘1:1 binding with mass transfer’ model (black lines). Corresponding residual plots below the sensorgrams show the kinetic-fit range and absolute deviation (Δ) of data points from curve fit values. The red arrow and the red highlighted area of the sensorgram series indicate data used to determine averaged (AVG) equilibrium (Eq) response values (Eq-Response AVG) for equilibrium analysis. (C,D) Equilibrium analysis of UL16 binding to MICB_{short} (C) and MICB_{long} (D). Averaged equilibrium response values (red squares) are plotted against injected UL16 concentrations and fitted to a ‘1:1 Langmuir isotherm’ model (black line). The shaded boxes contain additional information about setup details (black font) and measured parameters from kinetic (blue font) and equilibrium analysis (red font).’¹¹³

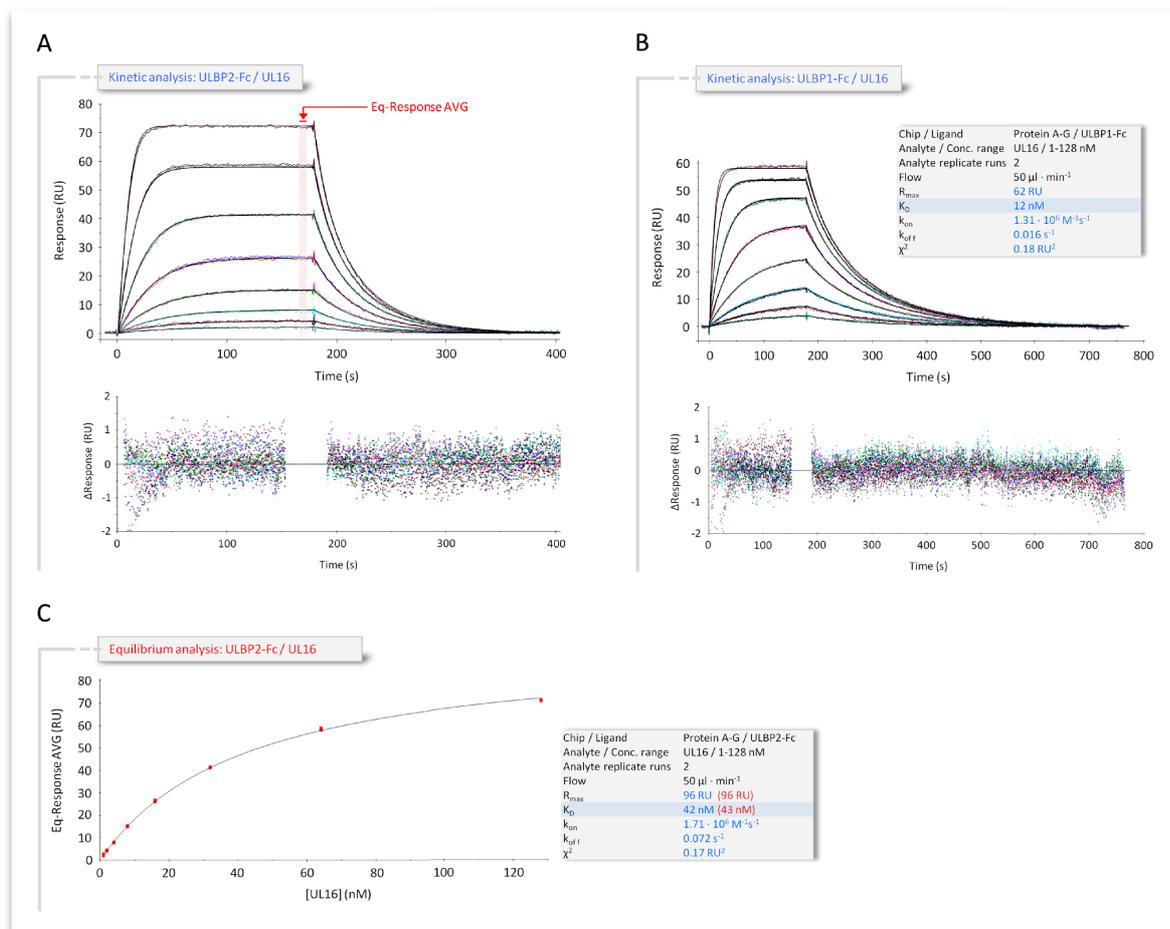


Figure 3-26. ‘Kinetic and equilibrium SPR analyses of UL16 interactions with ULBP1 and ULBP2. (A, B) Kinetic analyses of UL16 binding to Protein A-G captured (A) ULBP2-Fc and (B) ULBP1-Fc ectodomains. Each individual analyte concentration was injected twice and data are representative of at least two separate experiments with similar results. Double-referenced sensorgrams (shown in color) are overlaid with fits of a ‘1:1 binding with mass transfer’ model (black lines). Corresponding residual plots below the sensorgrams show the kinetic-fit range and absolute deviation (Δ) of data points from curve fit values. The red arrow and the red highlighted area of the sensorgram series indicate data used to determine averaged (AVG) equilibrium (Eq) response values (Eq-Response AVG) for equilibrium analysis. (C), Equilibrium analysis of UL16 binding to the ULBP2-Fc ectodomain. Averaged equilibrium response values (red squares) are plotted against injected UL16 concentrations and fitted to a ‘1:1 Langmuir isotherm’ model (black line). The shaded boxes contain additional information about setup details (black font) and results from kinetic (blue font) and equilibrium analysis (red font).’¹¹³

3.5 Structure determination of the UL16-MICB complex

3.5.1 Data collection

The data used for solving the crystal structure of the deglycosylated UL16_{mon}-MICB_{short} complex were collected at the Swiss Light Source (SLS) in Villigen-Switzerland using the PX (also called X06SA) beamline and the PILATUS 6M detector. Setup details are shown in Table 3-2.

Table 3-2. Data collection parameters

Parameter	Value
Crystal to detector distance	325 mm
Wavelength	1.0013 Å
Oscillation angle	0.25 °
Beam intensity	55 %
Frames per set	360 frames
Number of data sets ¹⁾	3 sets
Number of crystals used	1 crystal

¹⁾ Each data set was recorded from a different part of the crystal

3.5.2 Data processing

The three subsets of data collected from the single UL16_{mon}-MICB_{short} crystal were separately indexed and integrated with XDS¹²³ (X-ray Detector Software). In each case, XDS suggested a number of Bravais lattice types that were consistent with the observed locations of the diffraction spots (marked with an asterisk in Table 3-3).

Table 3-3. Bravais lattice types consistent with observed reflections

LATTICE- CHARACTER	BRAVAIS- LATTICE	QUALITY OF FIT	UNIT CELL CONSTANTS (ANGSTROM & DEGREES)					
			a	b	c	alpha	beta	gamma
* 31	aP	0.0	57.7	103.4	145.3	89.9	89.7	90.0
* 44	aP	0.2	57.7	103.4	145.3	90.1	90.3	90.0
* 33	mP	3.3	57.7	103.4	145.3	90.1	90.3	90.0
* 35	mP	5.9	103.4	57.7	145.3	90.3	90.1	90.0
* 34	mP	8.7	57.7	145.3	103.4	90.1	90.0	90.3
* 32	oP	8.9	57.7	103.4	145.3	90.1	90.3	90.0
37	mC	244.8	296.0	57.7	103.4	90.0	90.1	79.0
28	mC	247.3	57.7	296.0	103.4	89.9	90.0	79.0
36	oC	247.8	57.7	296.0	103.4	89.9	90.0	101.0
29	mC	250.0	57.7	214.7	145.3	90.0	90.3	74.4
39	mC	255.9	214.7	57.7	145.3	90.3	90.1	74.4
38	oC	258.9	57.7	214.7	145.3	89.9	90.3	105.6
27	mC	494.3	214.7	57.7	178.2	90.2	124.0	74.4
26	oF	616.3	57.7	214.7	296.0	87.0	101.0	105.6
41	mC	798.7	308.3	103.4	57.7	90.0	90.2	70.5
30	mC	803.8	103.4	308.3	57.7	89.8	90.0	70.5
40	oC	804.2	103.4	308.3	57.7	89.8	90.0	109.5
⋮								

Further analytic cycles with XDS¹²³ favored a primitive orthorhombic lattice (oP) and space group 19 (i. e. space group P2₁2₁2₁), as the reflections that should be systematically absent in space group P2₁2₁2₁ were indeed not found (or had negligible intensities) among the observed reflections (see CORRECT.LP output in Table 3-4). Under this latter assumption the three XDS¹²³ data-output hkl-files (one from each data set) were scaled together using XSCALE¹²³.

Table 3-4. Systematic absences.

REFLECTIONS OF TYPE H,0,0 0,K,0 0,0,L OR EXPECTED TO BE ABSENT (*)							
H	K	L	RESOLUTION	INTENSITY	SIGMA	INTENSITY/SIGMA	#OBSERVED
0	0	3	48.928	0.2863E+02	0.6522E+01	4.39	1*
0	0	4	36.696	0.3731E+05	0.1141E+04	32.70	2
0	0	5	29.357	0.5458E+01	0.6047E+01	0.90	2*
0	0	6	24.464	0.7440E+04	0.2290E+03	32.48	2
0	0	7	20.969	-0.1202E+02	0.9362E+01	-1.28	2*
0	0	8	18.348	0.1606E+05	0.4928E+03	32.60	2
0	0	9	16.309	0.1389E+02	0.1095E+02	1.27	2*
0	0	10	14.678	0.5355E+05	0.1639E+04	32.68	2
0	0	11	13.344	0.1579E+02	0.1451E+02	1.09	2*
0	0	12	12.232	0.1097E+04	0.3957E+02	27.72	2
0	0	13	11.291	0.2212E+02	0.2175E+02	1.02	1*
0	0	14	10.485	0.6743E+05	0.2918E+04	23.11	1
0	0	15	9.786	0.7422E+00	0.1863E+02	0.04	2*
0	0	16	9.174	0.3059E+04	0.9934E+02	30.79	2
0	0	17	8.634	0.2110E+02	0.2028E+02	1.04	2*
0	0	18	8.155	0.1998E+03	0.2253E+02	8.87	2
0	0	19	7.725	-0.1075E+02	0.2938E+02	-0.37	1*
0	0	21	6.990	0.4773E+01	0.2314E+02	0.21	2*
0	0	22	6.672	0.3177E+04	0.1048E+03	30.32	2
0	0	23	6.382	-0.7702E+01	0.2408E+02	-0.32	2*
0	0	24	6.116	0.1214E+05	0.3775E+03	32.16	2
0	0	25	5.871	-0.3122E+02	0.2740E+02	-1.14	2*
0	0	26	5.646	0.3035E+04	0.1027E+03	29.56	2
0	0	27	5.436	0.4786E+01	0.2887E+02	0.17	2*
0	0	28	5.242	0.4622E+05	0.1418E+04	32.59	2
0	0	29	5.061	0.7062E+02	0.5149E+02	1.37	1*
0	0	30	4.893	0.3292E+04	0.1598E+03	20.60	1
0	0	31	4.735	0.1026E+02	0.3290E+02	0.31	2*
0	0	32	4.587	0.1091E+05	0.3429E+03	31.83	2
0	0	33	4.448	0.2713E+01	0.3952E+02	0.07	2*
0	0	34	4.317	0.9524E+04	0.4246E+03	22.43	1
0	0	35	4.194	-0.1265E+03	0.5208E+02	-2.43	1*
0	0	36	4.077	0.1835E+05	0.5701E+03	32.18	2
0	0	37	3.967	-0.8119E+02	0.4278E+02	-1.90	2*
0	0	38	3.863	0.5971E+03	0.5208E+02	11.46	2
0	0	39	3.764	-0.3828E+02	0.4732E+02	-0.81	2*
0	0	41	3.580	-0.2514E+01	0.5145E+02	-0.05	2*
0	0	42	3.495	0.7103E+04	0.2331E+03	30.47	2
0	0	43	3.414	-0.2473E+02	0.5264E+02	-0.47	2*
0	0	44	3.336	0.1964E+05	0.6126E+03	32.06	2
0	0	45	3.262	0.2194E+02	0.9504E+02	0.23	1*
0	0	46	3.191	0.1628E+05	0.5102E+03	31.91	2
0	0	47	3.123	0.1302E+02	0.5205E+02	0.25	2*
0	0	48	3.058	0.1016E+05	0.3242E+03	31.34	2
0	0	49	2.996	0.1557E+02	0.4611E+02	0.34	2*
0	0	50	2.936	0.3496E+05	0.1524E+04	22.94	1
0	0	51	2.878	-0.3498E+02	0.4777E+02	-0.73	2*
0	0	53	2.769	0.9020E+02	0.4898E+02	1.84	2*
0	0	54	2.718	0.2453E+03	0.4978E+02	4.93	2
0	0	55	2.669	-0.3857E+02	0.4675E+02	-0.82	2*

Table 3-5 shows an excerpt of the overall statistical output from XSCALE¹²³, revealing the redundancy (based on number of observed and possible unique reflections), the completeness, the observed R-factor and the I/σI for each resolution shell and the complete data set. Based on the critical evaluation of these factors the resolution limit of the dataset was set to 1.8 Å.

Table 3-5. XSCALE output.

SUBSET OF RESOLUTION LIMIT	INTENSITY DATA OBSERVED	NUMBER OF REFLECTIONS UNIQUE	SIGNAL/NOISE POSSIBLE	>= -3.0 AS FUNCTION OF DATA	R-FACTOR observed	R-FACTOR expected	I/SIGMA	R-meas	Rmerge-F	Anomal Corr	SigAno	Nano	
8.05	9201	1056	1071	98.6%	4.3%	6.1%	9198	34.51	4.6%	1.9%	-8%	0.562	720
5.69	16102	1792	1799	99.7%	5.0%	6.3%	16090	32.85	5.3%	2.1%	-12%	0.668	1419
4.65	20903	2289	2298	99.6%	4.7%	6.2%	20898	34.08	5.0%	2.0%	-9%	0.654	1895
4.02	23435	2672	2680	99.7%	5.0%	6.2%	23425	33.36	5.3%	2.1%	-10%	0.700	2268
3.60	27544	2994	3012	99.4%	5.5%	6.3%	27540	32.88	5.8%	2.3%	-17%	0.699	2586
3.29	31312	3305	3324	99.4%	5.8%	6.5%	31311	31.56	6.1%	2.6%	-12%	0.745	2909
3.04	32652	3548	3574	99.3%	6.4%	6.9%	32649	28.21	6.8%	3.1%	-9%	0.790	3133
2.85	35522	3840	3863	99.4%	7.5%	7.7%	35515	25.32	7.9%	3.8%	-9%	0.800	3440
2.68	39714	4047	4072	99.4%	8.3%	8.4%	39708	23.94	8.7%	4.1%	-5%	0.808	3652
2.55	43058	4265	4295	99.3%	9.8%	9.9%	43049	21.16	10.4%	5.0%	-6%	0.790	3862
2.43	43066	4446	4491	99.0%	11.5%	11.5%	43062	18.39	12.1%	6.5%	-3%	0.805	4029
2.32	46506	4680	4732	98.9%	12.7%	12.6%	46503	17.32	13.4%	6.9%	-3%	0.803	4265
2.23	47094	4842	4900	98.8%	14.8%	14.7%	47087	15.37	15.6%	8.1%	-1%	0.818	4422
2.15	51568	5055	5124	98.7%	17.6%	17.9%	51561	13.66	18.5%	9.3%	-1%	0.807	4654
2.08	53675	5175	5243	98.7%	20.7%	21.4%	53670	11.91	21.7%	11.3%	1%	0.802	4774
2.01	53671	5370	5461	98.3%	25.1%	26.0%	53665	10.02	26.5%	14.8%	2%	0.798	4962
1.95	41021	5498	5519	97.8%	26.8%	28.0%	40986	7.67	28.8%	20.2%	1%	0.767	4723
1.90	37097	5658	5782	97.9%	33.0%	35.3%	37062	5.81	35.8%	28.5%	4%	0.759	4822
1.85	39653	5824	5969	97.6%	45.4%	49.0%	39616	4.43	49.0%	38.0%	5%	0.749	4896
1.80	40545	5916	6066	97.5%	58.5%	64.1%	40508	3.44	63.1%	49.4%	6%	0.729	4934
total	733339	82272	83374	98.7%	7.6%	8.5%	733103	16.93	8.0%	7.2%	-2%	0.771	72364

A summary of the data collection statistics from Table 3-5, including the space group and the XSCALE-refined unit cell parameters is shown in Table 3-6.

Table 3-6. Data collection statistics¹¹³

Statistic	Value
Space group	P2 ₁ 2 ₁ 2 ₁
No. of complexes in asymmetric unit	2
Unit cell dimension	
a, b, c (Å)	58.1, 104.2, 146.8
α, β, γ (°)	90, 90, 90
Resolution (Å)	50-1.8
R _{merge}	7.2 (49.4) ¹⁾
I/σI	16.9 (3.4)
Completeness (%)	98.7 (97.5)
Redundancy	8.9 (6.9)
Wilson Factor [Å ²]	24.1

¹⁾ Highest resolution shell is shown in parenthesis

Based on the XSCALE statistical output also a Wilson plot (Figure 3-27) was generated (performed by Dr. Georg Zocher) that (i) supported the resolution cutoff at about 1.8 Å (see Section 2.2.5.7.3) and (ii) returned an overall B-factor of 25.4 Å² (Figure 3-27)¹²⁵.

XSCALE was also used to define 5 % of the total number of reflections in the data set for the calculation of the R_{free} factor that is used as a quality control during structure refinement (see Section 2.2.5.9). Finally, the XSCALE data-output hkl-file was converted to an mtz-file using the programs XDSCONV and f2mtz.

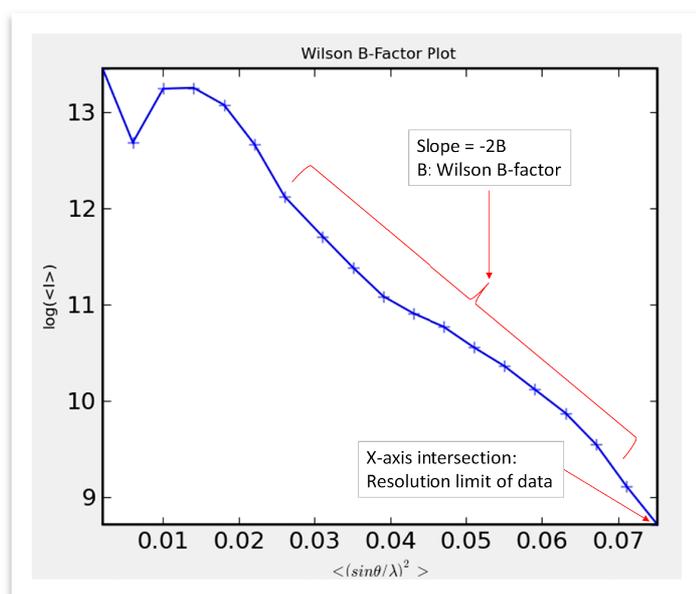


Figure 3-27. Wilson plot of the UL16_{mon}-MICB_{short} dataset. The Wilson plot shown in this figure was generated by Dr. Georg Zocher (using the XSCALE output and a self-written program) and shows the average intensities (y-axis) of reflections recorded at various defined resolutions (x-axis).

3.5.3 Matthews coefficient and solvent content

The use of (i) the Matthews probability calculator (see Section 2.1.2.6), (ii) the data shown in Table 3-6 (unit cell parameters and space group) and (iii) the molecular weight of the UL16_{mon}-MICB_{short} protomers (Mw: 44,000 Da) allowed for the calculation (see Section 2.2.5.7.5) of the most likely number of protomers (z) in the asymmetric unit (ASU). The calculation indicated that two UL16_{mon}-MICB_{short} molecules share one ASU. With two protomers/ASU, the Matthews coefficient was calculated to be $2.7 \text{ \AA}^3/\text{Da}$, which corresponds to a crystal solvent content of 54 %.

3.5.4 Structure determination

The structure was solved by molecular replacement (MR) as implemented in PHASER¹⁰⁰ using the unliganded MICB structure⁶⁴ (PDB code: 1je6) as search model. The search parameters used in PHASER¹⁰⁰ are listed in Table 3-7.

Table 3-7. PHASER input parameters

Parameter	Value
Resolution for search	50 - 2.5
Number of (MICB) copies to search for	2
Mass of protomer (UL16 _{mon} -MICB _{short})	44,000

In total, PHASER¹⁰⁰ found two solutions with satisfactory LLG, RF Z-score and TF Z-score values as listed in Table 3-8.

Table 3-8. PHASER solutions

Parameter	Solution 1	Solution 1 + Solution 2
LLG	214	671
RF Z-core	6.3	5.7
TF T-score	14.9	25.4

In addition to statistical parameters, the PHASER¹⁰⁰ output also consisted of a coordinate pdb-file that contained the positions (and therefore the phases) of the two correctly placed MICB molecules in the ASU. Together with the mtz-file obtained after XDSCONV/cad/f2mtz (see Section 3.5.2), an initial electron density map could be displayed in Coot¹⁰².

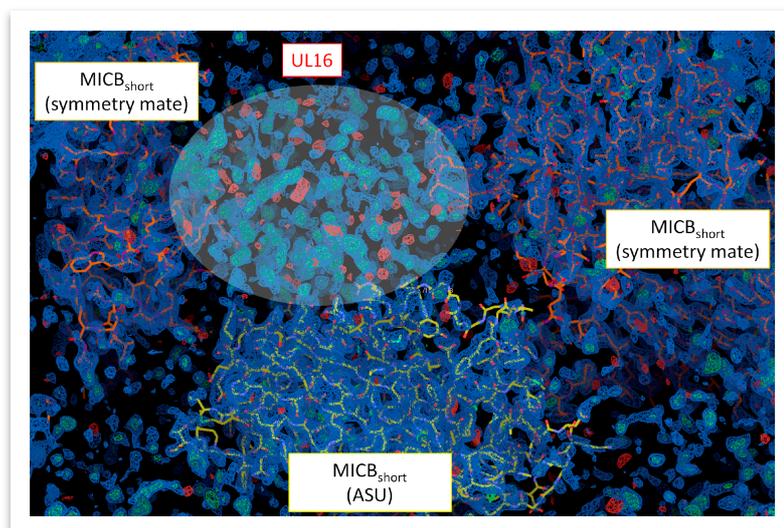


Figure 3-28. Starting model of the UL16_{mon}-MICB_{short} complex after PHASER¹⁰⁰. Model of one out of two MICB_{short} (yellow sticks) molecules present in the ASU overlaid with the corresponding $2F_{\text{obs}}-F_{\text{calc}}$ (blue mesh) and $F_{\text{obs}}-F_{\text{calc}}$ (green and red mesh) electron density maps at contour levels of about 1.5 and 3.5 σ , respectively. Also shown are the models of two MICB_{short} symmetry mates (orange sticks). The large patch of positive electron density (green mesh, see highlighted area) on top of MICB_{short} indicates the position of UL16_{mon}. Residual electron density is observed for other areas surrounding the yellow colored MICB_{short} molecule. The figure was created with Coot¹⁰².

The initial density map (Figure 3-28) already clearly showed the approximate location of the two UL16 molecules as positive electron density in areas adjacent to the UL16_{mon}-MICB_{short} interface. It also revealed that both UL16_{mon}-MICB_{short} protomers in the ASU are related by a two-fold noncrystallographic symmetry (NCS). After phases were improved by solvent flattening and NCS averaging using RESOLVE¹³⁰ (performed by Dr. Georg Zocher), the Ca chains of the two UL16

molecules in close proximity to the UL16_{mon}-MICB_{short} interface were defined well enough (Figure 3-29) to start building short stretches of glycine residues (performed by Dr. Georg Zocher).

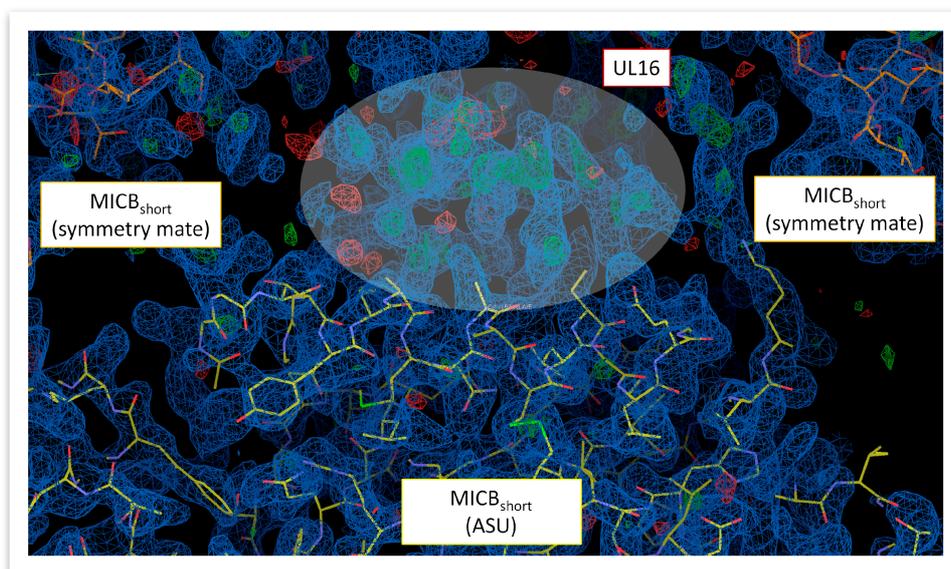


Figure 3-29. Starting model of the UL16_{mon}-MICB_{short} complex after RESOLVE¹³⁰. Model of one out of two MICB_{short} (yellow sticks) molecules present in the ASU overlaid with the corresponding $2F_{\text{obs}}-F_{\text{calc}}$ (blue mesh) and $F_{\text{obs}}-F_{\text{calc}}$ (green and red mesh) electron density maps at contour levels of about 1.5 and 3.5 σ , respectively. Also shown are the models of two MICB_{short} symmetry mates (orange sticks). The large patch of unoccupied electron density (blue mesh, see highlighted area) on top of MICB_{short} indicates the position of UL16_{mon}. Almost no electron density (black regions) is observed for other areas surrounding the yellow colored MICB_{short} molecule. The figure was created with Coot¹⁰².

After one round of refinement with Refmac, ArpWrp¹³¹ (Version 7.0) was used for automated (glycine) model building and refinement (performed by Dr. Georg Zocher). The glycine model produced by ArpWrp¹³¹ is shown in Figure 3-30. In subsequent rounds of manual model building and refinement using Coot¹⁰² and PHENIX¹⁰¹, respectively (see Section 3.5.5), the stretches of glycine residues were extended, and the glycines were exchanged against the correct amino acid residues as defined by the UL16 primary sequence (Figure 3-31).

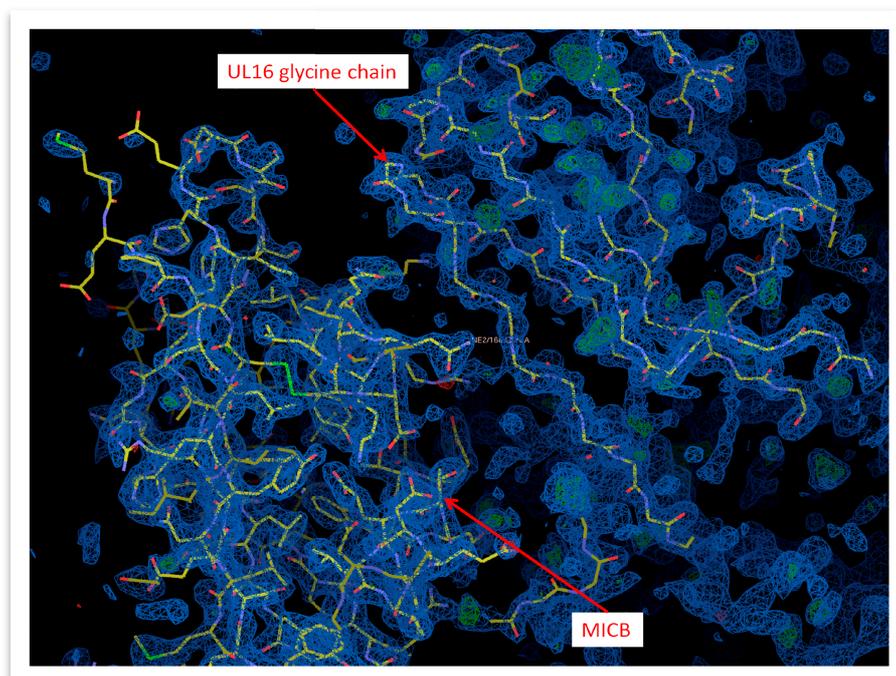


Figure 3-30. Glycine model of UL16_{mon} obtained with ArpWrp¹³¹. Model of the UL16_{mon} glycine chain (top right corner) built by ArpWrp overlaid with the corresponding $2F_{obs}-F_{calc}$ (blue mesh) and $F_{obs}-F_{calc}$ (green and red mesh) electron density maps at contour levels of about 1.5 and 3.5 σ , respectively. Also shown is the model of MICB_{short} in the bottom left corner. The figure was created with Coot¹⁰².

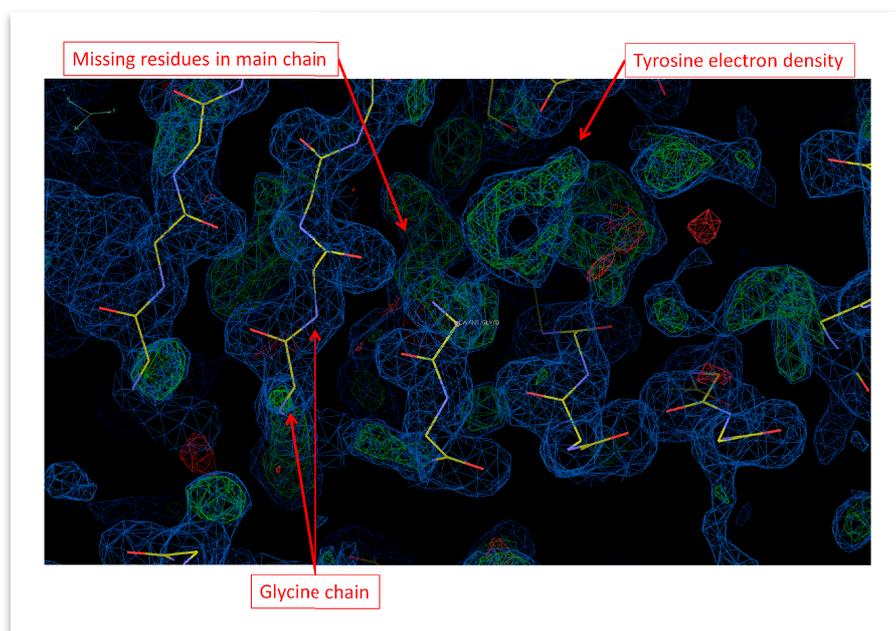


Figure 3-31. Details of the initial UL16_{mon} glycine model obtained with ArpWrp¹³¹. Magnified portion of the UL16 glycine model (see Figure 3-30) overlaid with the corresponding $2F_{obs}-F_{calc}$ (blue mesh) and $F_{obs}-F_{calc}$ (green and red mesh) electron density maps at contour levels of 1 and 3 σ , respectively. Stretches of already built glycine residues, missing amino acid residues and missing amino acid side chains (e. g. of a tyrosine residue) are marked with red arrows. The figure was created with Coot¹⁰².

3.5.5 Structural refinement

Reciprocal space refinement was performed with PHENIX¹⁰¹, and included rigid body refinement, simulated annealing, individual coordinate refinement, occupancy refinement, and TLS refinement of 26 TLS groups assigned by the TLSMD Server¹³². Real space refinement (i. e. model building) was done with Coot¹⁰². The refinement progress was monitored throughout by the decrease of R_{work} and R_{free} and by ensuring that the deviation between both R-factors would not exceed 5 %. The final model had R_{work} and R_{free} values of 17.7 % and 21.3 %, respectively, and was validated using PROCHECK¹⁰⁰ and WHAT_IF¹³³. Secondary structure elements were assigned with DSSP¹³⁴. Ramachandran statistics were obtained with Coot¹⁰² version 0.5. Structural figures were created with PyMOL¹⁰³. The complete refinement statistics are shown in Table 3-9.

Table 3-9. Refinement statistics¹¹³

Statistic	Value
Resolution (Å)	50-1.8
No. reflections	
measured	733339
unique	82272
R_{work}	17.7
R_{free} (test set of 5%)	21.3
No. of non-H atoms	5980
Protein	5084
Carbohydrates (NAGs)	196
PEG8000	49
Acetate	12
Water	639
Average isotropic B factor (Å ²)	
Protein main chain	28.0
Protein side chain	34.1
Carbohydrates (NAGs)	41.5
PEG8000	55.6
Acetate	65.0
Water	39.1
R.m.s deviations	
Bond lengths (Å)	0.006
Bond angles (°)	1.057
Ramachandran regions ¹⁾	
most favorable (%)	97.4
allowed	2.6
outliers	0

¹⁾ Determined with Coot¹⁰² version 0.5

3.6 The crystal structure of UL16

‘The UL16 ectodomain folds into a modified version of the immunoglobulin (Ig)-like domain (Figure 3-32A). The presence of nine β -strands, arranged in two antiparallel β -sheets (formed by β -strands A, G, F, C, C', C'' and β -strands D, E, B, respectively) and a central disulfide bond linking β -strands B and F clearly classifies it as a variable (V-type) Ig-like domain^{32, 33, 35}. In contrast to classical V-type Ig domains, however, UL16 also has an additional N-terminal ‘plug’ (amino acids 27-50), formed by a two-stranded antiparallel β -sheet (β -strands X1 and X2) and a short 3_{10} -helix (Figure 3-32A). The plug covers the concave side of the AGFCC'C'' β -sheet and is covalently linked to the Ig-like core with a disulfide bond between β -strands X2 and F. The UL16_{mon}-MICB_{short} complex was partially deglycosylated prior to crystallization, leaving only single NAG molecules attached to glycosylation sites. While there is no evidence for O-linked glycosylation, the electron density maps provide clear evidence for the presence of NAGs at seven out of eight putative N-glycosylation sites (asparagines 35, 41, 68, 84, 95, 101 and 132). Six of these seven asparagine residues (Asn41, 68, 84, 95, 101, 132) carry NAG residues that are clearly defined by electron density (Figure 3-32A). While extra density is present at the seventh residue, Asn35, this density is not well defined, and consequently no NAG residue was built at this location. No extra electron density is observed in the vicinity of the eighth asparagine, Asn145, and thus this residue is either not glycosylated or carries an especially flexible glycan moiety. Modeling experiments show that native glycosylation would effectively shield much of the UL16 surface from solvent (Figure 3-32B). In particular, the outward-facing AGFCC'C'' β -sheet and the N-terminal plug are expected to be mostly covered with glycans in the fully glycosylated protein. By contrast, the solvent-exposed face of the DEB β -sheet is devoid of glycans and available for interactions with other proteins.’¹¹³

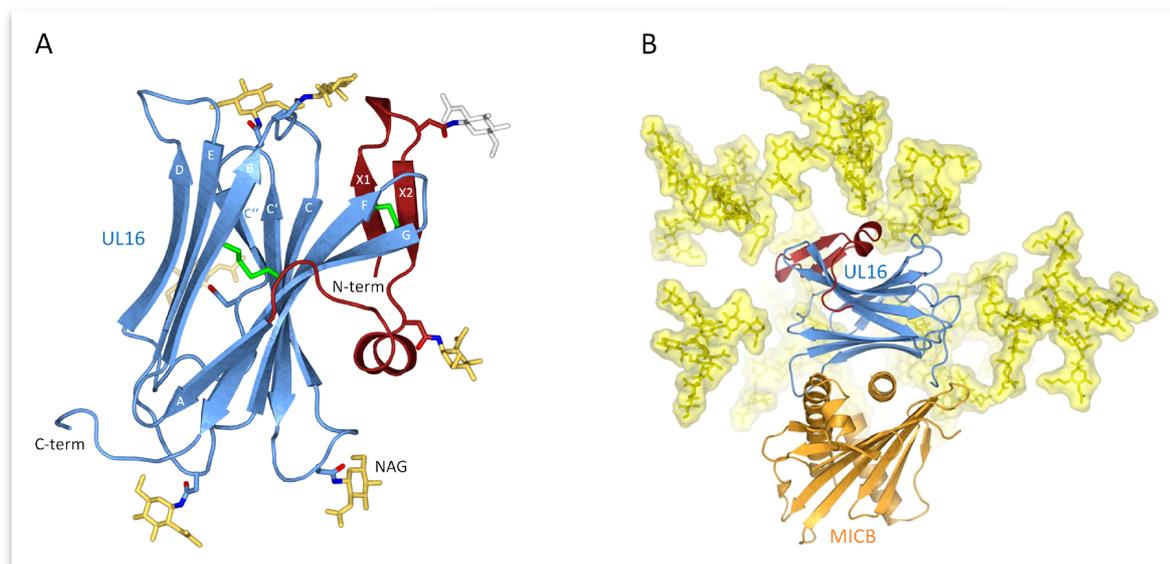


Figure 3-32. ‘Structure of UL16 in complex with MICB_{short}. (A), Ribbon tracing of the structure of UL16. The portions of UL16 belonging to the V-type IgSF fold the N-terminal ‘plug’ are colored blue and red, respectively³⁵. Glycosylated asparagines (nitrogen atoms dark blue, oxygen atoms red) and attached NAG residues (yellow) are shown as ball-and-stick models. The grey NAG residue attached to Asn35 has high temperature factors and was therefore not included in the refinement. Disulfide bonds are shown in green. (B) Structure of the UL16_{mon}-MICB_{short} complex. UL16 is colored as in panel A, MICB_{short} is shown in orange. In order to visualize the native glycosylation of UL16, modeled glycans are shown in yellow as ball-and-stick models with a semitransparent surface.’¹¹³

3.7 The crystal structure of MICB_{short}

‘The extracellular region of MICB consists of two structural domains, the $\alpha 1\alpha 2$ -platform domain (MICB_{short}) and the C-type Ig-like $\alpha 3$ -domain⁶⁴. The $\alpha 3$ -domain is present only in the MIC family members of NKG2D ligands, but not among members of the ULBP family^{19, 20, 61, 63, 64}. As previously reported for the unliganded MICB⁶⁴, MICB_{short} folds into a structure that closely resembles MHC class I molecules, with two long parallel α -helices, contributed by domains $\alpha 1$ and $\alpha 2$, arranged above an eight-stranded antiparallel β -sheet (Figure 3-32B; for nomenclature of domains and secondary structure elements see Figure 3-33). Comparison of MICB_{short} with the structure of the unliganded MICB ectodomain⁶⁴ shows that the platform domain remains essentially unchanged upon engagement of UL16 (root-mean-square deviation of 1.4 Å for 172 common C α atoms). Although minor differences are seen within three surface-exposed loops and a short N-terminal helix ($\alpha 0$), the residues in these regions have elevated temperature factors and do not contact UL16.’¹¹³

3.8 The UL16-MICB_{short} interface

'UL16 primarily engages MICB_{short} via a predominantly hydrophobic, glycan-free (see also Ref.⁸⁸) surface comprised of its DEB β -sheet and the adjacent β -strand A, with additional contacts provided by the DE-loop (connecting β -strands D and E) and four amino acids (aa 160-163) at the C-terminus (Figure 3-33B and Figure 3-34).'¹¹³

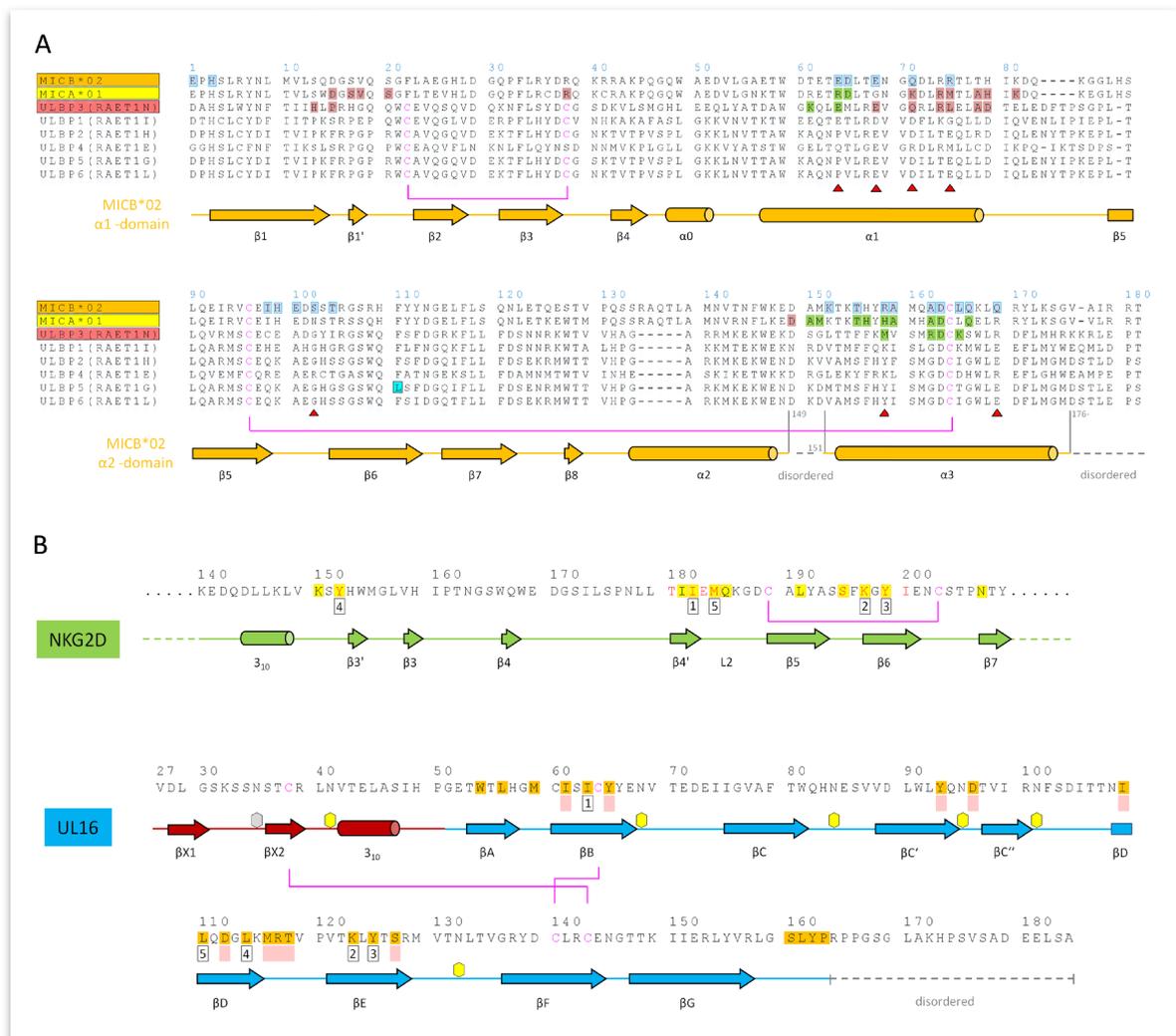


Figure 3-33. 'Amino acid sequences of NKG2D ligands, NKG2D, and UL16. (A) Sequence alignment of NKG2D ligands. Sequences of the α 1 α 2-platform domains of NKG2D ligands MICA*001, MICB*002, ULBP1, ULBP2, ULBP3, ULBP4, ULBP5 and ULBP6 are included in the alignment. The alternative RAET nomenclature of ULBP proteins is indicated. Secondary structure elements as observed in the structure of MICB in complex with UL16 were assigned by DSSP¹³⁴ and are represented with cylinders (helices) and arrows (β -strands) below the alignment. Helices are named as described in⁶¹. Residues shaded in blue contact UL16 in the UL16_{mon}-MICB_{short} complex. Residues shaded in salmon contact the salmon-colored NKG2D monomer (see Discussion and Figure 4-4A and B) in the MICA-NKG2D and ULBP3-NKG2D complex structures^{20, 61}. Residues shaded in green contact the green NKG2D monomer (Figure 4-4A and B) in the MICA-NKG2D and ULBP3-NKG2D complex structures. Residues marked with a red triangle indicate substitutions between MICA and MICB in regions that contact UL16 in the UL16_{mon}-MICB_{short} complex. The ULBP5 residue boxed in cyan was recently shown to be the major determinant for diminished binding to

NKG2D and UL16¹³⁵. Disulfide bridges and corresponding cysteines are represented with magenta lines. Gaps are indicated by (-). (B) Structural mimicry of UL16. Shown are relevant portions of the sequences of the green human NKG2D monomer^{20, 61} (Figure 4-4A and B) and UL16. Secondary structure elements as observed in the structure of MICB in complex with UL16 and MICA in complex with NKG2D²⁰, respectively, were assigned and represented as described in panel A. The five residues marked with numbered black boxes below the sequence define the central binding motif that engages MICB or, in the case of NKG2D, MICA²⁰, in a similar manner (Figure 4-4C). Residues with the same number superimpose in space, although they are located in different regions in the protein sequences. Residues shaded in yellow and orange form contacts with MICA in the case of NKG2D²⁰ and with MICB in the case of UL16, respectively. NKG2D residues in red contact ULBP3 in the ULBP3-NKG2D complex⁶¹. Residues that augment the central binding motif, performing similar functions in the UL16-MICB and NKG2D-MICA complexes are marked with filled light red boxes below the sequence. An example is shown in Figure 4-4C. Disulfide bridges are represented with magenta lines. Hexagons mark the seven UL16 asparagine residues linked with glycans as observed in the UL16_{mon}-MICB_{short} complex.¹¹³

‘This surface interacts with the two long parallel helices at the top of the MICB $\alpha 1\alpha 2$ platform-domain and the $\beta 5\beta 6$ -loop connecting β -strands $\beta 5$ and $\beta 6$ of MICB (Figure 3-33A and Figure 3-34), shielding an area of 2194 Å² from solvent. With the exception of the MICB region that corresponds to the peptide-binding groove in MHC class I proteins, the contact area contains few interfacial solvent-filled cavities. The complex features good surface complementarity ($Sc = 0.77$) and is highly curved (planarity = 4.0)^{136, 137}. Its overall organization resembles a saddle with two stirrups (UL16) that is mounted on horseback (MICB) (Figure 3-34A, see also Figure 3-32B). The saddle is formed by the DEB β -sheet, whereas the stirrups are contributed by the DE-loop and the C-terminus on either side of the sheet. To facilitate discussion of interactions, the three regions of the UL16-MICB interface were divided into three regions (A, B and C, Figure 3-34). **Contact region A**, which is located at the center of the complex and mostly hydrophobic in nature, contributes 54 % of the total contact area. Interactions predominantly involve residues within the DEB β -sheet and β -strand A of UL16. Eight UL16 residues (Trp54, Leu56, Met59, Ile61, Ile63, Tyr125, Leu110 and Leu114) define a compact hydrophobic face that interacts with non-polar regions of MICB residues in its central $\alpha 3$ -helix. These interactions are augmented with a salt bridge between UL16 Asp112 and MICB Lys152 and a number of mostly water-mediated hydrogen bonds (Figure 3-34B). **Contact region B**, with 23 % of the total contact area, is located at one end of the DEB β -sheet and within the DE-loop of UL16. UL16 residues in this region contact several acidic residues (Glu64, Asp65 and Glu68) in the $\alpha 1$ -helix of MICB, mostly via polar interactions (Figure 3-34C). **Contact region C**, which contributes 23 % to the total contact area, is located on the other side of the UL16 saddle. Here, the C-terminus of the UL16 ectodomain interacts with the $\beta 5\beta 6$ -loop and the N-terminus of MICB_{short} via a mixture of hydrophilic and hydrophobic contacts (Figure 3-34D). The overall architecture of the complex, with its large contact area and substantial number of interactions between contacting residues, indicates tight

binding, which is in agreement with the obtained SPR data that place the affinity of the UL16 monomer for MICB_{short} at 66 nM (Figure 3-25 and Table 3-1).¹¹³

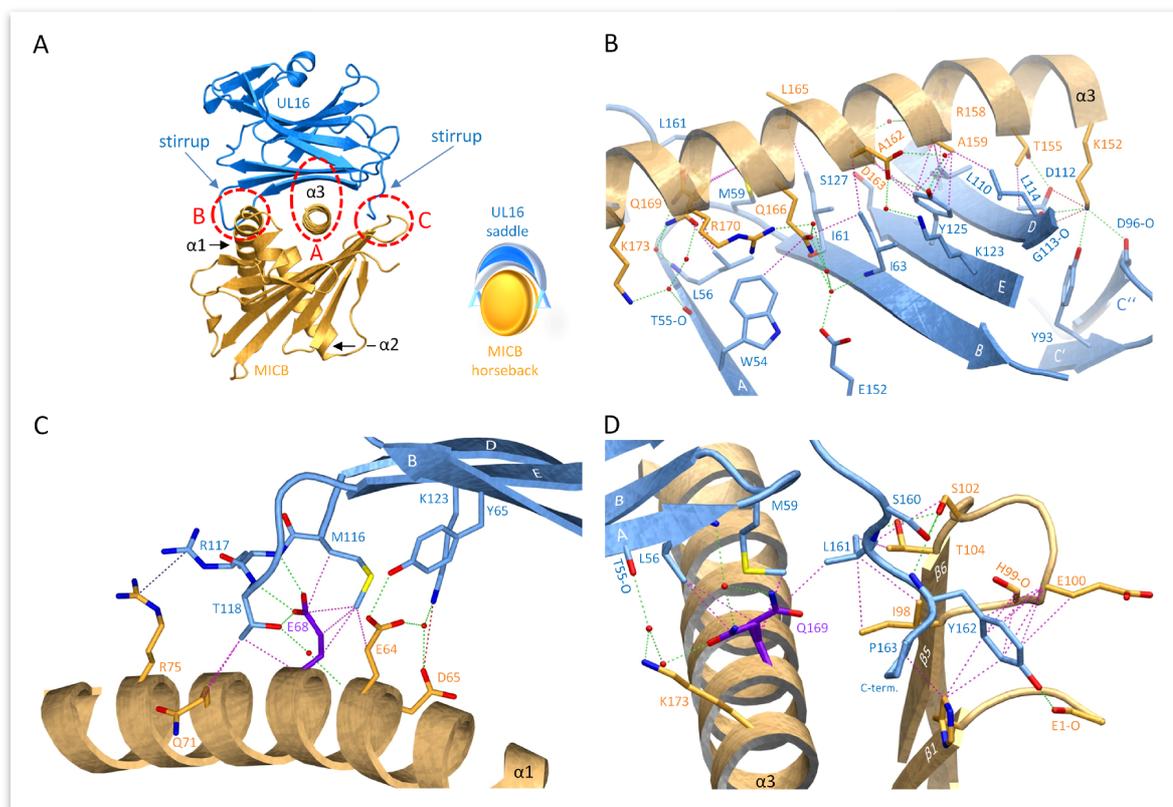


Figure 3-34. ‘Interaction between UL16 and MICB. (A), Ribbon tracing of the complex using the color code from Figure 3-32. Also shown in the lower right corner of panel A is a schematic representation of the ‘saddle on horseback’ arrangement between UL16 and MICB. (B-D), The three major contact regions A, B and C of the complex. Nitrogen, oxygen and sulfur atoms are colored blue, red, and yellow, respectively. Hydrogen bonds and salt bridges are represented with dashed green and red lines, respectively, and hydrophobic contacts (distance <math>< 4.0 \text{ \AA}</math>) are shown as dashed magenta lines. The dashed blue line indicates π - π interactions of two arginine guanidinium groups. Water molecules are shown as red spheres.’¹¹³

4. Discussion

4.1 The UL16-MICB complex - purification and crystallization obstacles

4.1.1 CV-1 cell produced UL16 is not suitable for crystallization

Deglycosylation is often a pivotal step in the process of obtaining high quality crystals of glycoproteins. This is due to the observation that the glycans (as shown in Figure 3-6) are usually extremely heterogeneous. Glycan heterogeneity means that the very same asparagine (Asn) residue of a defined N-glycosylation recognition sequence (Asn-X-Ser/Thr) in the glycoprotein's primary sequence can carry glycan trees that vary for instance in (i) the glycan type (high-mannose, hybrid or complex glycans), (ii) the presence of a core fucose and (iii) the number and length of the polysaccharide branches.

A heterogeneous glycoprotein mixture (which is usually obtained after purification) can be imagined as consisting of a common invariable core (the protein) and a highly variable shell (the heterogeneous glycans). In order for a protein or glycoprotein to lead to the formation of high quality, i. e. well-diffracting crystals, the individual molecules have to be regularly packed against each other. Regular packing however is only possible if the surfaces of the molecules that are to be crystallized are very similar and therefore allow for the regular formation of stable crystal contacts when the individual molecules are arranged in the three-dimensional crystal lattice. It follows that this regular packing and the formation of regular and stable crystal contacts will not be possible for glycoproteins with their highly heterogeneous glycan shells. This can be a possible reason when only poorly diffracting crystals (if any crystals at all) of glycoproteins are obtained. To remove the heterogeneity, glycosidases are usually employed. Two of the most common glycosidases are Endo H (Mw: 29 kDa) and PNGase F (Mw: 36 kDa). PNGase F virtually removes all N-linked oligosaccharides (high mannose, complex and hybrid glycans) from glycoproteins, whereas Endo H only removes high-mannose and hybrid N-linked glycans. Therefore the pattern of protein bands that is visualized on a SDS-gel after a glycosylated protein has been cleaved with both glycosidases under denaturing conditions can give valuable information about the presence of complex type N-linked glycans. This kind of assay was performed with UL16 produced in CV-1 cells (Figure 3-6). As CV-1 cells were derived from monkeys, i. e. from close relatives of humans, one can assume that the UL16 produced in this cell type carries a similar set of glycans as one would expect from a human cell infected with HCMV.

Although the UL16 heterogeneity can be reduced by Endo H (see lane 2 in Figure 3-6) at least three UL16 glycoprotein species remain that will continue to make crystallization difficult and would most likely require additional purification steps in order to isolate a single glycoform for crystallization. The question arises now, why UL16 was not deglycosylated with PNGase F, as in that case only one UL16 species (without any glycans at all) would have been obtained (see lane 1 in Figure 3-6).

This is because PNGase F hydrolyses the bond between the carbonyl group of the asparagine side chain and the neighboring nitrogen atom that creates the glycosidic bond with the first NAG residue of the glycan tree. Therefore the PNGase F enzymatic reaction always results in the transformation of asparagine to aspartate residues when the glycans are cleaved off. Due to the introduction of a negative charge by the aspartate, the structure of the protein may (locally or even globally) change, which of course leads to the question of how much the protein core of a PNGase F cleaved glycoprotein actually still resembles the native fold of the protein. Therefore, Endo H is generally preferred for the deglycosylation of glycoproteins since after Endo H cleavage one NAG residue remains attached to the N-linked asparagine residues of the protein, which not only avoids possible structural changes in the protein but also allows for the direct identification of asparagine sites that are actually glycosylated. In conclusion, due to the extreme and heterogeneous glycosylation and the difficulties to effectively deglycosylate the protein with Endo H, CV-1 cells are not very suitable to express UL16 protein that is supposed to be used for crystallization.

As a consequence and as outlined in Section 3.2.3, CHO Lec 3.2.8.1 cells, which are deficient in a number of glycosyltransferases and Golgi nucleotide-sugar transporters, were eventually employed to produce a more homogeneous UL16 glycoprotein species that was used throughout in the attempts to crystallize the UL16-MICB complex¹⁰⁵ (see Section 3.6, 3.7 and 3.8). Note however, that the CHO Lec 3.2.8.1 produced UL16 monomer still exhibits a molecular weight of 35 kDa (Mw of UL16 without glycans: 18.5 kDa), which shows that still about 50 % of the CHO Lec 3.2.8.1 produced UL16 glycoprotein consists of glycans (Figure 3-8). These remaining glycans are all of the high-mannose type as CHO Lec 3.2.8.1 cells only disable the synthesis of hybrid- and complex type glycans (see Section 3.2.3). This and the results obtained from the deglycosylation experiments of CV-1 produced UL16 (see above) suggest that when natively expressed in human cells after viral infection, the UL16 protein is surrounded by a glycan layer consisting of high-mannose type-, complex type- and likely also hybrid type glycans.

4.1.2 The UL16 dimer is likely formed by locally misfolded UL16 monomers

After its purification (see Sections 3.2.5 and 3.2.6) CHO Lec 3.2.8.1 cell expressed UL16 consists to equal parts of a monomeric (UL16_{mon}) and a disulfide-linked dimeric (UL16_{dim}) protein species (see Section 3.2.7). Hence, the question was raised (see Section 3.2.8) what causes the dimerization of the UL16 monomers. In order to provide an answer to this question, the UL16 structure, as observed in the structure of the UL16_{mon}-MICB_{short} complex, was analyzed in terms of cysteine residues (Figure 4-1; see also Figure 3-33). As became apparent from the crystal structure of the UL16_{mon}-MICB_{short} complex (see Section 3.6), UL16 features a V-type Ig-like fold (Figure 4-1A), which consists of two antiparallel β -sheets (the DEB-sheet and the AGFCC'C''-sheet) that together form a β -sandwich. The surface exposed side of the DEB-sheet (Figures 4-1A and B) is comprised of a set of predominantly hydrophobic amino acid side chains and can therefore be considered as a relatively large hydrophobic patch on the molecular surface of the UL16 monomer. Furthermore, five cysteines are present in the Ig-like domain of the UL16 monomer, four of which contribute to the formation of two disulfide bonds (green bonds in Figure 4-1B; see also Figure 3-33). It follows, that the only cysteine residue that remains non-disulfide bonded in the UL16 monomer is Cys60, which is located at the N-terminus of β -strand *b* that together with β -strands *d* and *e* comprises the hydrophobic DEB-sheet (labeled in Figure 4-1). Cys60 is very well defined (i. e. does not show multiple side chain conformers) in the electron density of the UL16_{mon}-MICB_{short} crystal structure (Figure 4-2A) and points with its free SH-group towards the hydrophobic core of the Ig-like fold (Figures 4-1B and 4-2B). In other words, the SH-group of Cys60 is not surface exposed in natively folded UL16 monomers but instead buried in the core of the Ig-like domain (Figure 4-2B). As a consequence, Cys60 should not be able to participate in any disulfide bond formation that could result in the disulfide-bridged UL16 dimer. It follows further that the disulfide bridge formation between two UL16 monomers can only be explained if the SH-group of Cys60 becomes surface exposed, i. e. if it becomes part of the outer surface of the DEB-sheet. This in turn would suggest that the UL16 monomers (those that participate in the UL16 dimer formation) are locally misfolded at or close to the Cys60 residues. Note that also the possibility remains that the Ig-like domain of UL16 monomer is not only locally but globally misfolded. In the latter case also other cysteines than Cys60 might be involved in the UL16 dimer formation. Assuming that the UL16 monomers are only locally misfolded, a possible reason why the Cys60 residues become surface exposed can again be given based on the obtained structure of the UL16_{mon}-MICB_{short} complex. In the latter, Cys60 is stabilized in the Ig-like core of UL16 by a very weak hydrogen bond (distance of 4 Å) that is formed between the SH-group of Cys60 and the OH-group of Tyr-138 (Figure 4-2). Otherwise Cys60 is mainly surrounded by the hydrophobic portions of the surrounding amino acid

residues Val-156, Tyr-138, Thr-126, Arg-128 and Val-130 (Figure 4-2B). One can therefore speculate that the position of the rather hydrophilic Cys60 in natively folded UL16 monomers is energetically not very favorable. This could be an explanation why a certain percentage of the UL16 monomers are locally misfolded around Cys60 and why they present this cysteine side chain on the surface exposed side of the DEB sheet, which then leads to the formation of the disulfide linked UL16 dimers.

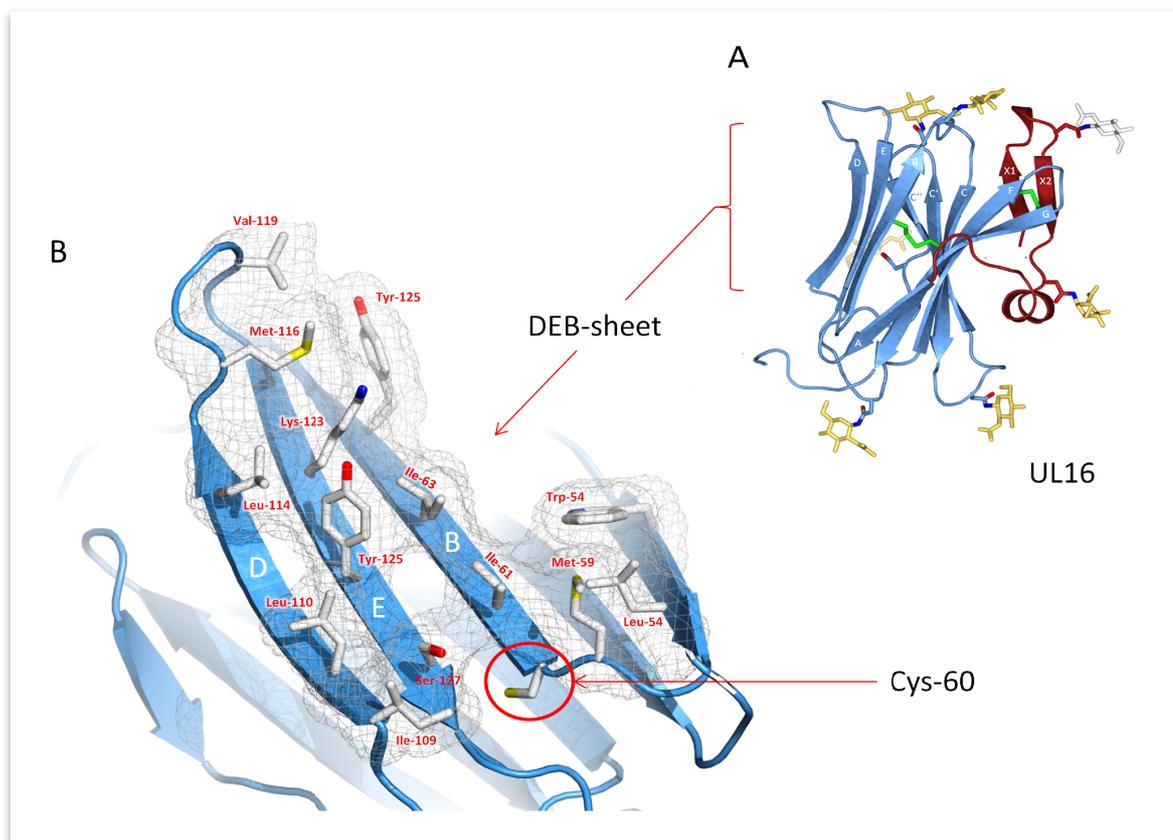


Figure 4-1. The role of Cys60 in UL16 dimer formation. (A) Ribbon drawing of the UL16 structure showing the DEB and AGFCC'C'' β -sheet sandwich of the Ig-like fold of UL16 (for details see Section 3.6 and Figure 3-32A). (B) Close-up view of the UL16 DEB sheet (see panel A) showing the predominantly hydrophobic side chains that comprise this sheet as well as the position (red circle) and orientation of the Cys60 side chain. The gray mesh represents the surface of the depicted side chains. Atoms are colored as in Figure 3-34.

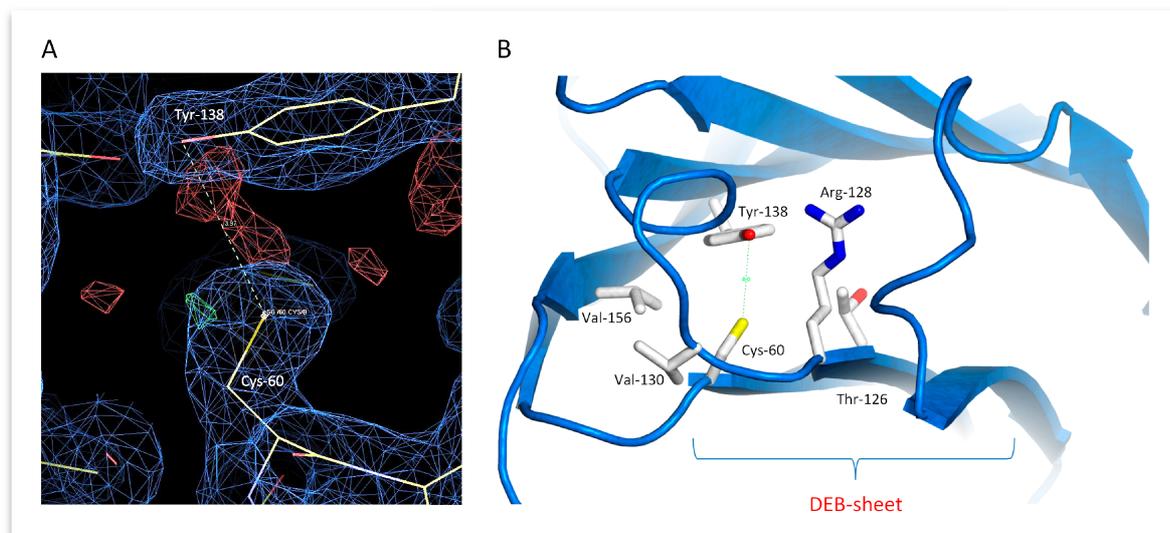


Figure 4-2. Environment of the Cys60 side chain of UL16. (A) Model of the UL16 side chains Cys60 and Tyr-138 (shown as sticks) as observed in the crystal structure of the UL16_{mon}-MICB_{short} complex overlaid with the corresponding $2F_{\text{obs}}-F_{\text{calc}}$ (blue mesh) and $F_{\text{obs}}-F_{\text{calc}}$ (green and red mesh) electron density maps at contour levels of 1 and 3 σ , respectively. The dotted line represents a weak hydrogen bond with a distance of 4 Å and unfavorable angle, formed between the SH-group of Cys60 and the side chain OH-group of Tyr-138. Oxygen and sulfur atoms are colored red and yellow, respectively. (B) Ribbon drawing of the DEB and AGFCC'C'' β -sheet sandwich of UL16 with focus on the Ig-like core residues (shown as sticks) of the UL16 core residues and Cys60 side chain of UL16. The hydrophobic parts of the UL16 core residues and Cys60 are colored white. Nitrogen, oxygen and sulfur atoms are colored blue, red, and yellow, respectively.

4.1.3 MICB binding sites are inaccessible in the UL16 dimer

Based on the results presented in Section 3.2.8 it was concluded that only the UL16 monomer can interact with MICB_{long}, whereas the UL16 dimer obviously cannot. However, the question may arise if this specific binding behavior only applies to MICB_{long}. In other words, could maybe MICB_{short} still engage the UL16 dimer? Functional SPR data of the UL16 monomer interactions with MICB_{short} and MICB_{long} (see Section 3.43.7) suggests otherwise. This data clearly shows that both MICB_{short} and MICB_{long} bind the UL16 monomer in the very same fashion by means of their identical $\alpha1\alpha2$ -platform domains, whereas the $\alpha3$ -domain of MICB_{long} is not involved in UL16 binding. Based on this identical binding mode of both MICB's in case of the UL16 monomer it seems rather unlikely that MICB_{short} and MICB_{long} would feature distinct binding modes in case of the UL16 dimer, unless there was steric interference of the UL16_{dim}/MICB_{long} interaction caused by the $\alpha3$ -domain of MICB_{long}.

A more fundamental question is, why the UL16 dimer is unable to bind NKG2D ligands such as MICB in the first place (see Section 3.2.8). Again, possible answers to this question can be presented in light of the UL16_{mon}-MICB_{short} complex structure. UL16 mainly contacts MICB_{short} (and MICB_{long}) via the predominantly hydrophobic residues on the surface exposed side of the DEB-sheet (see contact region A in Figure 3-34). The Cys60 residue that likely is responsible for the UL16 dimer formation (see previous section) lies within this DEB sheet and therefore in the vicinity of the UL16 binding site of MICB. It follows that the formation of a disulfide bond between the Cys60 residues of two UL16 monomers that form the dimer will most likely result in the complete occlusion of the two monomer binding sites. This conclusion is not only supported by structural evidence but also by distinct binding behavior of both UL16 species in terms of Phenyl HP columns (discussed in Section 4.1.6).

When it was shown that the UL16 dimer is completely inactive in terms of MICB_{long} binding (see Section 3.2.8), attempts were made to separate either the UL16 monomer or the UL16_{mon}-MICB_{long} complex from the inactive and therefore contaminating UL16 dimer. However, standard chromatographic methods such as SEC or IEX failed to provide an efficient separation strategy (see Section 3.2.9). It was also tried to reduce the disulfide-linked dimer to the monomer by means of different reducing agents but also this approach remained unsuccessful (see Section 3.2.9). However, as outlined in the previous section, the intermolecular disulfide bond(s) are most likely caused by either locally or globally misfolded UL16 monomers. Therefore it has to be doubted that a successful reduction of the UL16 dimer would lead to natively folded UL16 monomers.

4.1.4 Purification and crystallization of the UL16_{mon}-MICB_{long} complex

It was shown that purified UL16 in solution (see Sections 3.2.5 and 3.2.6), consists to equal amounts of UL16 monomers and dimers (see Section 3.2.7). It was further shown that (i) the dimer is comprised of two disulfide bridged UL16 monomers (see Section 3.2.7), that (ii) only the monomer is active, while the dimer is not (see Section 3.2.8) and that (iii) standard chromatographic methods fail to separate the UL16 dimer from the UL16 monomer as well as from the UL16_{mon}-MICB_{long} complex (Section 3.2.9).

Eventually, the only purification strategy that led to the efficient disposal of the UL16 dimer was the separation of the UL16 dimer from a mixture comprised of UL16_{mon}-MICB_{long} complex and excess amounts of MICB_{long} by means of IMAC AC (see Section 3.2.10). After separation from excess MICB_{long} by SEC (see Section 3.2.11), the obtained UL16_{mon}-MICB_{long} complex was used in crystallization trials that successfully led to the growth of, unfortunately, non-

diffracting crystals (see Section 3.3.1). However, when these crystals were analyzed by SDS-PAGE it turned out that what had crystallized was not the expected UL16_{mon}-MICB_{long} complex but a complex comprised of UL16 monomer and the MHC class I-like $\alpha 1\alpha 2$ platform-domain of MICB (named MICB _{$\alpha 1\alpha 2$}). In other words, MICB_{long} had been deprived of its $\alpha 3$ -domain (see Section 3.3.1). The most likely explanation of this totally unexpected result is a contamination of the UL16_{mon}-MICB_{long} complex with trypsin (a serine protease that is frequently used for various applications in our laboratory) prior to the setup of the crystal drops. The analysis of the full length amino acid sequence of MICB with the online tool *PeptideCutter* revealed two potential trypsin cleavage sites at position 179 and position 180 of MICB, two residues at the very N-terminal end of the MICB $\alpha 2$ -domain (see Figure 3-33). Nevertheless, the coincidental cleavage of MICB_{long} showed that it was in principle possible to crystallize a UL16_{mon}-MICB _{$\alpha 1\alpha 2$} complex under the conditions described in Section 3.3.1.

In order to reproduce these UL16_{mon}-MICB _{$\alpha 1\alpha 2$} crystals, a shorter MICB construct, MICB_{short}, was designed (see section 2.2.3.4) that consisted only of the MICB $\alpha 1\alpha 2$ platform-domain followed by an N-terminal thrombin cleavage site and a His₈-tag. This construct was then successfully expressed and purified by following the protocols that had been devised for the expression and purification of MICB_{long} (see Sections 3.1).

4.1.5 Separation of UL16_{dim} from UL16_{mon}-MICB_{short} complex by Ni-NTA AC

Unfortunately however, attempts to separate the UL16 dimer from the UL16_{mon}-MICB_{short} complex by means of IMAC AC repeatedly failed (see Section 3.2.12). This was very surprising since, as described in Section 3.2.10, the same strategy had worked perfectly well for the separation of the UL16 dimer from the UL16_{mon}-MICB_{long} complex (see Section 3.2.10). Two possible reasons for these observations may exist. The first explanation could be that the His₈-tag of MICB_{short} was cleaved off by thrombin protease. This clearly is a possibility since the Fc-tag of UL16-Fc is cleaved by thrombin as part of the UL16 purification strategy (see Section 3.2.6). Although benzamidine columns are specifically used in this strategy to trap the utilized thrombin, it cannot be ruled out that some thrombin molecules pass through these columns and consequently contaminate the obtained UL16_{mon/dim} sample. It follows that by mixing these potentially contaminated UL16_{mon/dim} samples with MICB_{short} in order to set up the UL16_{mon}-MICB_{short} complex, thrombin gains access to the thrombin cleavage site of MICB_{short}. This explanation is supported by the fact that differently from MICB_{short}, the His₆-tag of MICB_{long} is not cleavable by thrombin (see Section 2.2.3.4.1 and Figure 2-1) and by the observation that the MICB_{short} which was eluted from the Ni-NTA column with imidazole (lanes 69, 71 and 73 in Figure 3-18B) exhibited a slightly higher

molecular weight than the MICB_{short} found in the flow through samples (lanes 10, 13, 16, 19 and 22 in Figure 3-18B). However, a second explanation could be that the His₈-tag of MICB_{short} was covered upon binding to the UL16 monomer. That this explanation is a possibility too can be illustrated by comparing the structure of the UL16_{mon}-MICB_{short} complex (pdb code: 2wy3; Figure 4-3A) with the structure of the UL16_{mon}-MICB_{long} complex (Figure 4-3B), which was obtained by the superimposition of the UL16_{mon}-MICB_{short} structure with the structure of MICB_{long} (pdb code: 1je6⁶⁴). In the UL16_{mon}-MICB_{long} structure, the His-tag is positioned at the N-terminus of the MICB α 3-domain and therefore located at a great distance to the UL16 binding site (Figure 4-3B). The opposite however is observed in the case of the UL16_{mon}-MICB_{short} complex, in which the accessibility of the His-tag might be severely restricted due to its close proximity to the C-terminal end of the UL16 Ig-like domain (Figure 4-3A).

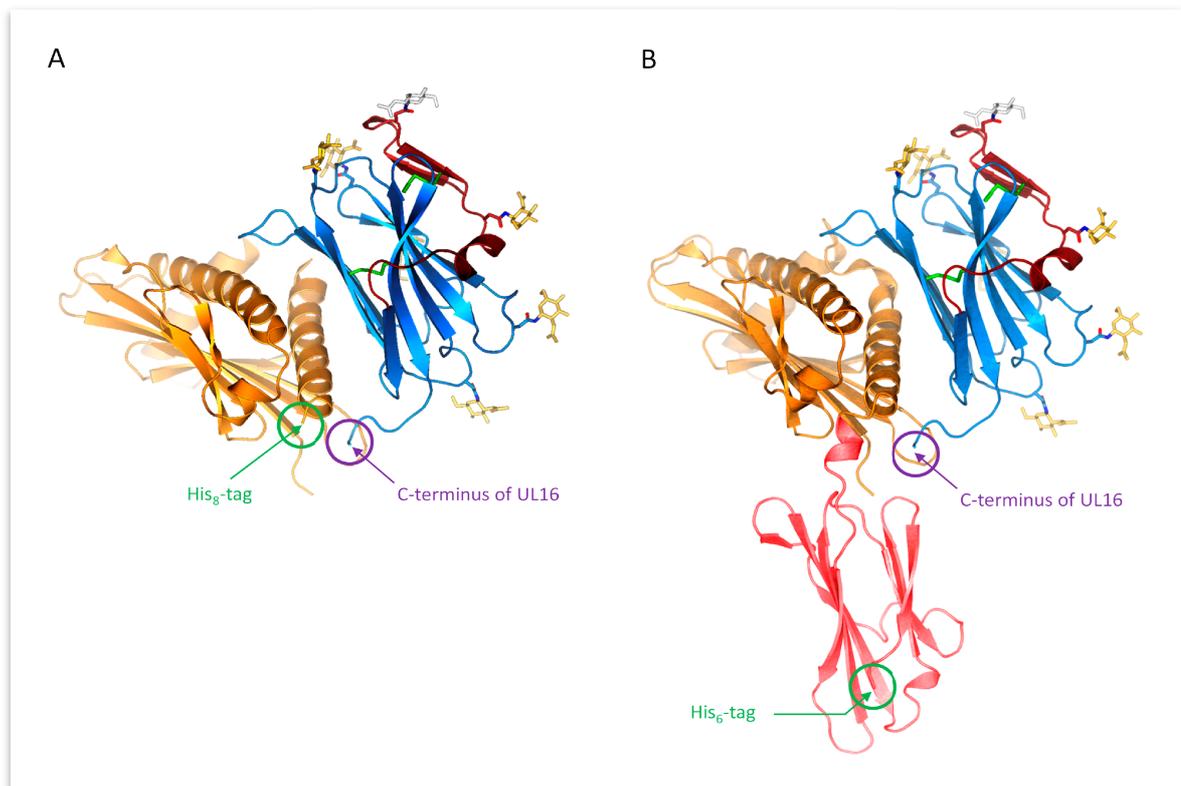


Figure 4-3. Positions of the His-tag in MICB_{short} and MICB_{long}. (A) Ribbon tracing of the UL16_{mon}-MICB_{short} structure (pdb code: 2wy3¹¹³) with the position of the N-terminal His-tag of MICB_{short} indicated by a green arrow. Also shown (purple arrow) is the C-terminus of UL16. (B) Ribbon tracing of a UL16_{mon}-MICB_{long} structure, generated from the structures of the UL16_{mon}-MICB_{short} complex and MICB_{long} (pdb code: 1je6⁶⁴) by superimposing the α 1 α 2-platform domains (orange in both panels) of both MICB structures. The ribbon tracing of the MICB_{short} α 1 α 2-platform domain is not shown for reasons of clarity. The positions of the N-terminal His-tag of MICB_{long} and the C-terminus of UL16 are indicated as in panel A.

In conclusion, the inability to separate the UL16 dimer from the UL16_{mon}-MICB_{short} complex with the strategy that had been devised and successfully been used for the purification of the UL16_{mon}-MICB_{long} complex, marked a major setback in the project.

4.1.6 Purification of the UL16 monomer

The failure to separate the UL16 dimer from the UL16_{mon}-MICB_{short} complex made it necessary to refocus on the goal of finding a purification strategy that would reliably separate the UL16 monomers from the dimers. The strategy that was finally successful in this respect involved the use of HIC columns, a column type that separates proteins according to their hydrophobic surface properties (see Sections 2.2.3.3 and Section 3.2.13). When a mixture of UL16 monomer and UL16 dimer was passed over a *Phenyl HP* column, the monomer was retained, whereas the dimer passed through the column and was accordingly found in the flow-through (see Section 3.2.13). Of course, one may wonder what the reason behind this distinct binding behavior of the two UL16 species in terms of the *Phenyl HP* column might be.

It was shown in Section 3.6 that the outward-facing AGFCC'C" β -sheet and the N-terminal plug of the UL16 monomer are covalently linked to eight NAG residues that represent what is left of the UL16 glycan trees after Endo H cleavage. In contrast, the solvent-exposed side of the DEB β -sheet was shown to constitute an extensive hydrophobic face (Figure 4-1) on the surface of UL16 (comprised of UL16 residues Trp54, Leu56, Met59, Ile61, Ile63, Tyr125, Leu110 and Leu114) that is not only devoid of glycans but also represents the major binding site of MICB (54 % of the total contact area). Its accessibility and its hydrophobic properties also make the outer side of this DEB β -sheet especially well suited to mediate the contacts between the UL16 monomer and the hydrophobic *Phenyl HP* column matrix. It follows that any kind of effective shielding of this hydrophobic UL16 surface region would not only lead to the inability of the UL16 molecules to interact with the hydrophobic column matrix but also to their failure to contact MICB. However, this could be exactly what happens when the UL16 dimer should indeed be formed by two locally misfolded UL16 monomers via their exposed cysteine residues at position 60 (see Section 4.1.2). The reason for this is that the Cys60 residues themselves are part of the DEB β -sheets of UL16 and that a disulfide bond between them would cause the UL16 dimer-forming monomers to mutually shield their DEB β -sheets and in this way their binding site for both the *Phenyl HP* matrix and MICB. It needs to be pointed out, however, that the formation of the UL16 dimer by two globally misfolded UL16 monomers (as discussed in Section 4.1.2), could be an alternative explanation for the abovementioned binding characteristics of the UL16 dimer as the global misfolding would most

likely lead to severe alterations in the DEB β -sheet of UL16 that also could, even if the DEB β -sheet was surface exposed, lead to the inability of UL16 to bind the phenyl matrix or MICB.

4.1.7 Surface plasmon resonance measurements

The HIC purification strategy that allowed to separate the UL16 monomer from the UL16 dimer in a MICB independent fashion marked a major step forward on the way towards the crystal structure of the UL16_{mon}-MICB_{short} complex. Furthermore, the UL16 monomer could also now reliably be used as the analyte (the soluble binding partner) in SPR experiments (see Section 3.4). Knowing the exact concentration of the analyte is of pivotal importance in SPR experiments as the concentration of the analyte directly influences the determination of the binding parameters. It is therefore obvious that a mixture consisting of an unknown ratio of UL16 monomer and dimer is not suitable for SPR applications because the concentration of the active UL16 monomer in this mixture cannot be reliably determined with standard protein determination assays such as the Bradford assay or the λ_{280} assay.

The SPR data obtained for the interactions between purified UL16 monomer and the two MICB proteins, MICB_{long} and MICB_{short}, revealed affinity values of 67 and 66 nM, respectively, which can be considered as being virtually identical (see Section 3.4 Table 3-4 and Table 3-1). However, the affinity is only an indirect measure for the stability of a complex. The direct parameter that characterizes the complex stability is the off-rate constant k_d , which can be obtained by performing kinetic SPR experiments^{138, 139}. The k_d values determined for the UL16_{mon}-MICB_{long} and UL16_{mon}-MICB_{short} interactions were 0.045 and 0.054 s⁻¹ (see Section 3.4 and Table 3-1), respectively. This high similarity can be interpreted in the way that the α 3-domain of MICB_{long} is not involved in binding to UL16, an information that was not available from the structure of the UL16_{mon}-MICB_{short} complex and that is in good agreement with a previous report published by Spreu *et al.*⁸⁸. Also the interactions of other NKG2D ligands such as ULBP1, ULBP2, ULBP3, ULBP4 and ULBP5 with the UL16 monomer were tested (see Section 3.4 and Table 3-1). The obtained data showed that MICA, ULBP3, ULBP4 and ULBP5 do not interact with the UL16 monomer, while ULBP1 and ULBP2 bind tightly. These results are in good agreement with the current literature^{65, 84-90}. In detail, the UL16_{mon}-ULBP1 and UL16_{mon}-ULBP2 complexes exhibit k_d values of 0.016 and 0.072 s⁻¹, respectively. This places the complex stability of the UL16_{mon}-ULBP2 complex close to the one observed for the UL16_{mon}-MICB interaction and also shows that ULBP1 is the NKG2D ligand that forms the most stable complex with UL16 (see Section 3.4 and Table 3-1). The reasons for the UL16 monomer failure to engage other NKG2D ligands such as MICA or ULBP3 will be discussed in detail in Section 4.2.3.

4.1.8 Purification and crystallization of the UL16_{mon}-MICB_{short} complex

In addition to SPR studies, purified UL16 monomer was also used to obtain the UL16_{mon}-MICB_{short} complex for crystallization attempts. Briefly, UL16_{mon} and MICB_{short} were mixed and incubated and the formed complex purified by SEC. The crystallization attempts that followed were eventually successful (see Section 3.3.2) and resulted in crystals that were similar to those that had been obtained coincidentally during crystallization experiments with the UL16_{mon}-MICB_{long} complex (see Section 3.3.1). Also the crystallization conditions were quite similar (see Sections 3.3.1 and 3.3.2), which is likely attributed to the high structural similarity between the UL16_{mon}-MICB_{short} and the UL16_{mon}-MICB_{long} complexes. Unfortunately, crystals of the UL16_{mon}-MICB_{short} complex diffracted only to about 8 - 10 Å. A possible reason for this result could be the remaining heterogeneity of the high-mannose glycan trees that still cover the surface of the UL16 monomer after its expression by CHO Lec 3.2.8.1 cells. To test this hypothesis the UL16_{mon}-MICB_{short} complex was further deglycosylated enzymatically by use of the glycosidase Endo H (see next section) and the crystallization attempts repeated with a deglycosylated UL16_{mon}-MICB_{short} complex. The reasons why Endo H and not PNGase F was used to deglycosylate the UL16_{mon}-MICB_{short} complex are described in Section 4.1.1.

4.1.9 Purification and crystallization of the deglycosylated UL16_{mon}-MICB_{short} complex

Attempts to crystallize the glycosylated UL16 monomer in its non-MICB-bound state never returned any promising results. Furthermore, attempts to crystallize deglycosylated UL16 monomer were hampered due to the observation that UL16 monomer, after its deglycosylation with Endo H, started to precipitate. This indicates that the glycan shell of UL16, besides its other possible functions (see Section 3.6), prevents UL16 from aggregating, possibly by increasing its overall solubility. This would make also sense considering that the outer face of the DEB β -sheet, and therefore an extensive area of the overall UL16 surface, is extremely hydrophobic (Figure 4-1). However, when deglycosylation of UL16 was performed in its MICB-bound state, no UL16 aggregation or complex aggregation was observed. Therefore, the strategy was adjusted in a way that the deglycosylation step with Endo H followed after the formation of the UL16_{mon}-MICB_{short} complex (see Section 3.2.15). Briefly, purified UL16 monomer was first mixed and incubated with MICB_{short} and the formed complex further purified by SEC (see Section 3.2.14). Then, in a second step, the UL16_{mon}-MICB_{short} complex was deglycosylated with Endo H and, again, further purified by SEC (see Section 3.2.15).

Crystallization trials with the purified, deglycosylated UL16_{mon}-MICB_{short} complex showed that this deglycosylated complex could be successfully crystallized under the exact same conditions as the glycosylated UL16_{mon}-MICB_{short} complex (see Section 3.3.3). The obtained crystals diffracted to about 1.8 Å and a full crystallographic dataset was recorded, which was subsequently used to solve and analyze the crystal structure of the deglycosylated UL16_{mon}-MICB_{short} complex (see Sections 3.5 - 3.8).

Taken together, these results clearly show how severely the resolution of X-ray diffraction data can depend on the glycosylation state of the crystallized proteins (see Sections 3.3.2 and 3.3.3). When CHO Lec 3.2.8.1 cell produced UL16 was used (where each N-glycosylation site was most likely connected to a glycan tree of 7 - 11 sugar residues), the resolution of the UL16_{mon}-MICB_{short} complex was in the range between 8 - 9 Å, whereas when UL16 was further deglycosylated with Endo H (where each N-glycosylation site is only connected to a single NAG residue), the resolution of the UL16_{mon}-MICB_{short} complex was dramatically increased to 1.8 Å.

The 1.8 Å dataset was eventually used to solve the structure of the UL16_{mon}-MICB_{short} complex, which to the best of the author's knowledge is the first structure of a viral immunoevasin in complex with a stimulatory NK cell receptor ligand. The results of the structural analyses of the complex were presented in Sections 3.6 (UL16_{mon}), 3.7 (MICB_{short}) and 3.8 (UL16_{mon}-MICB_{short} interface). The structural insights gained from the UL16_{mon}-MICB_{short} structure will be discussed in the subsequent sections, starting with the comparison between the UL16_{mon}-MICB_{short} complex and the NKG2D-MICA complex.

4.2 The UL16-MICB complex - structural insights into NKG2D ligand binding by UL16 and NKG2D

4.2.1 Comparison between the UL16-MICB and the NKG2D-MICA complex

A crystal structure of the NKG2D homodimer bound to MICB is unavailable. However, the NKG2D structure in complex with the highly homologous MICA protein shows that both NKG2D monomers make extensive contacts with the long helices at the top of the MICA $\alpha 1\alpha 2$ -platform domain. The NKG2D-MICA complex buries a surface area of 2170 Å², which is almost exactly the same area buried in the UL16_{mon}-MICB_{short} complex. A superimposition of the two complexes demonstrates that contacts formed by UL16 overlap substantially with those made by one NKG2D monomer (Figure 4-4A and B). One could therefore envision a scenario in which UL16 acts as a direct competitor for NKG2D⁶⁵, perhaps even displacing it from its ligands. While the higher affinity of UL16 for MICB_{short} and ULBP1 (K_D values of 66 and 12 nM, respectively) (Figures 3-25

and 3-26 and Table 3-1) compared with the respective affinities of NKG2D for the same ligands (K_D values of 0.8 and 1 μ M, respectively)⁶⁰ would support this scenario, most reports to date indicate that UL16 acts inside the cell and is therefore unlikely to compete with NKG2D for ligand binding^{14, 19, 85, 86, 113}.

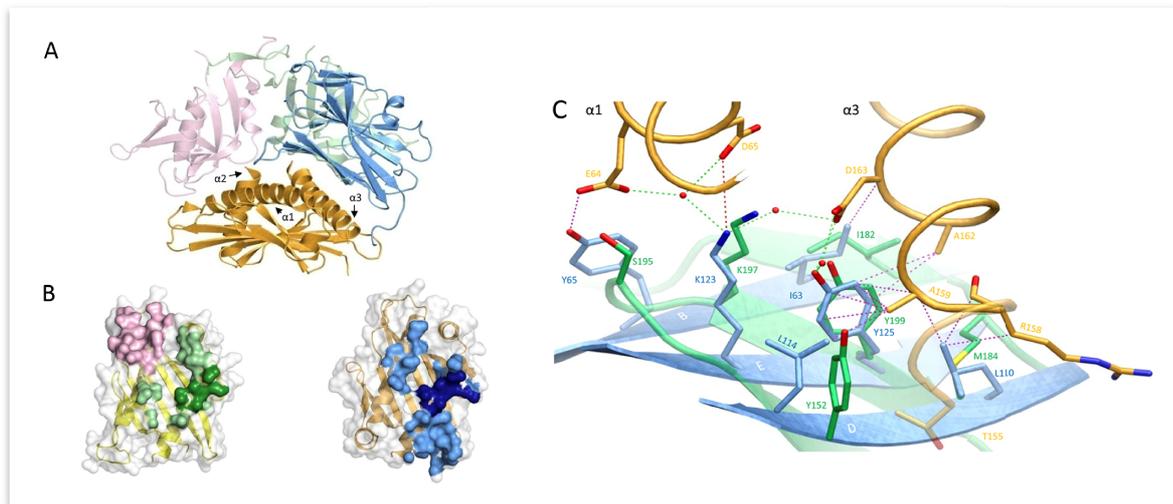


Figure 4-4. ‘Comparison of the UL16_{mon}-MICB_{short} and NKG2D-MICA complex structures. In all panels, the two NKG2D monomers are shown in salmon and green, whereas UL16 and MICB are colored blue and orange, respectively. (A) Superposition of the UL16_{mon}-MICB_{short} complex onto the MICA-NKG2D complex²⁰. MICA, which is very similar to MICB, is not shown for clarity. (B) Ribbon drawings of the $\alpha1\alpha2$ -platform domains of MICA (left side, yellow) and MICB (right side, orange), with their molecular surfaces outlined in grey. Surface-exposed areas of residues that are buried upon complex formation with NKG2D and UL16, respectively, are colored using the color scheme of panel (A). MICB/MICA residues 155, 158, 159, 162 and 163, which contact both UL16 and NKG2D in a similar manner are shown in darker green and blue shading, respectively. (C) Structural mimicry of UL16. Close-up view of the core region of the structures shown in panel (A) with UL16 residues Ile63, Lys123, Tyr125, Leu110, Leu114 that superimpose with chemically equivalent NKG2D residues Ile182, Lys197, Tyr199, Met184 and Tyr152. Side chain atoms, UL16-MICB contacts, and water molecules are colored as described in the legend to Figure 3-34.’¹¹³

‘The detailed comparison of the central contact regions in each case reveals that, despite having entirely different folds, NKG2D and UL16 use an almost identical pattern of amino acid side chains to engage their ligands (Figure 4-4C, see also Figure 3-33 and Figure 4-6). In UL16, this pattern includes the MICB-contacting residues Ile63, Lys123 and Tyr125, while NKG2D uses an identical pattern of residues, Ile182, Lys197, and Tyr199, to form very similar contacts with MICA. Remarkably, although the three side chains are contributed by different structural elements in each case, their position in space overlaps closely (Figure 4-4C). This is also true for two additional UL16 residues, Leu110 and Leu114, which are hydrophobic in nature and overlap with chemically related NKG2D residues Met184 and Tyr152 (Figure 4-4C). Together, the five residues constitute a predominantly hydrophobic binding motif that is common to NKG2D and UL16 (Figure 4-4B, C), and that forms the center of the interaction with the MIC molecules. This central

binding motif is augmented by additional contacts, such as those mediated by UL16 residue Tyr65 and NKG2D residue Ser195, that perform similar functions in the UL16_{mon}-MICB_{short} and NKG2D-MICA²⁰ complexes (see Figure 3-33B and Figure 4-4C). Since all MICA and MICB residues contacted by this central binding motif are identical, and since the structures of MICA and MICB superimpose well in this region, one can conclude that UL16 mimics a key structural motif of NKG2D with an entirely different fold in order to engage MICB. Furthermore, it can be considered likely that the central binding motif of the UL16 monomer also plays an important role in the recognition of other NKG2D ligands, for which structures of complexes with UL16 are not yet available.’¹¹³

4.2.2 The promiscuous binding mode of UL16

‘Bacterial and viral pathogens often interfere with cellular activities and immunosurveillance processes to enhance their survival and effectiveness¹⁴⁰. This is typically achieved by virulence factors, which imitate the function of a host protein by mimicking its key structural features. In the majority of such cases, pathogens first hijack and then manipulate host genes to produce structurally homologous versions of host proteins^{82, 140-143}. Thus, virulence factors and host proteins are derived from the same origin and arise from divergent evolution. However, structural mimics can also be generated through convergent evolution. Although differing in evolutionary origin and three-dimensional structure, the virulence factors have in this case evolved to mimic key structural features of cellular proteins. Examples for the latter strategy, which can only be revealed through structural analyses, are still exceedingly rare and are limited to a small number of virulence factors^{140, 144, 145}. The comparison of HCMV UL16 with human NKG2D, reveals a striking example of convergent evolution¹⁴⁰. A set of five predominantly hydrophobic core residues on the UL16 surface precisely mimics a set of five equivalent residues in the central region of the interface used by the structurally unrelated immunoreceptor NKG2D to interact with its ligands.’¹¹³

‘As this central binding motif represents only a portion of the total interface between NKG2D and its ligands (Figure 4-4), one may wonder why UL16 mimics just this particular structural motif of NKG2D. McFarland *et al.* reported that residues constituting this motif (Tyr152, Met184 and Tyr199) form the basis for the highly degenerate ligand recognition mode of NKG2D^{59, 60}. They proposed a ‘rigid adaptation’ mechanism, in which a rigid binding site on NKG2D uses the same set of predominantly hydrophobic core residues to make diverse interactions with a series of chemically and structurally distinct ligand residues. As an example, Tyr199 and Tyr152 of NKG2D can accommodate residues as diverse as Ala, Met or Phe at ligand position 159⁶⁰ (Figure 3-33A and Figure 4-6). Mimicry of these core residues likely enables UL16

to employ this binding mechanism of NKG2D to contact a similar set of ligands. The ‘rigid adaptation’ concept is furthermore supported by the finding that UL16 engages its ligands via a rigid β -sheet, which does not allow for much conformational flexibility. The ligand residues contacted by NKG2D and UL16 in MICA and MICB, respectively, are Asp65, Thr155, Ala159, Ala162, Asp163 and the hydrophobic portions of the Arg/His158 side chain (Figure 3-33A, Figure 4-4C, and Figure 4-6)^{20, 59, 60}. Since NKG2D and UL16 both evolved the same central binding motif in order to contact this specific set of ligand residues, the latter likely represent binding hot spots in MICA and MICB⁵⁹. Furthermore, these residues probably are also of major importance for interactions with ULBP molecules (Figure 3-33A and Figure 4-6). Note for instance that (i) based on the ‘rigid adaptation’ concept the amino acid at ligand position 159 can be quite variable in size and chemical nature, (ii) Asp163 is conserved in all NKG2D ligands, and (iii) alanine and glycine dominate at position 162.’¹¹³

4.2.3 Key determinants of NKG2D-ligand binding to UL16

‘Unlike NKG2D, UL16 engages only MICB, ULBP1, ULBP2 and ULBP6, but not MICA, ULBP3, ULBP4 and ULBP5^{65, 86-88, 90, 135}. The SPR measurements show that UL16 binds MICB with high affinity, whereas the affinity of UL16 for MICA is negligible (Table 3-1), in line with earlier studies^{14, 85, 88}. Given the high degree of similarity between MICA and MICB at the sequence and structural level, the inability of UL16 to engage MICA is puzzling. In order to better understand the structural parameters that guide UL16 binding to MICB vs. MICA, Spreu *et al.*⁸⁸ assayed binding of soluble UL16-Fc to MICB chimeras in which they had exchanged domains, subdomains and single amino acids of MICB against equivalent regions of MICA. These experiments clearly demonstrated that recognition by UL16 is linked to residues projecting from the helical structures in the MICB α 2-domain. However, the molecular mechanism by which these residues confer selectivity remained unclear.’¹¹³

‘The crystal structure of the UL16_{mon}-MICB_{short} complex now allows us to identify the key determinants of NKG2D-ligand binding to UL16. The performed structural alignment of MICA and MICB identifies only seven MICB residues that contact UL16 in the complex and that are replaced by other amino acids in MICA (Figure 3-33). Residues at positions 64, 71, 75, 102 and 158 can assume alternate conformations that would not interfere with binding, and could in some cases even mediate favorable contacts with UL16. Therefore, their effect on UL16 binding is likely to be negligible (see also Spreu *et al.*⁸⁸). Replacement of α 1-domain Glu68 with glycine (Figure 3-34C) in MICA would eliminate several hydrophobic contacts and three hydrogen bonds with UL16 residues 117 and 118, and could therefore conceivably have a negative effect on UL16

binding. However, as complete replacement of the α 1-domain of MICB by MICA (including residue Glu68) did not significantly affect UL16 binding⁸⁸, residue 68 is probably not a key determinant of UL16 binding.¹¹³

‘On the other hand, however, Gln169 in the α 2-domain of MICB is likely to be critical. The UL16-MICB structure shows that substitution of Gln169 with arginine, which is present at this position in MICA, would lead to steric clashes with UL16 residues Met59 and Leu161 (Figure 4-5A) that would prevent binding. This is in perfect agreement with previous experiments demonstrating that MICB carrying a Gln169Arg substitution no longer bound UL16⁸⁸. Based on the following reasons, it is in fact likely that the side chain at position 169 is not only the key determinant of selective UL16 binding to the MIC molecules but all NKG2D ligands. (i) All NKG2D ligands that carry a glutamine or glutamate at position 169, i. e. MICB, ULBP1, ULBP2 and ULBP6, bind UL16, while all ligands that have an arginine at this position, i. e. MICA, ULBP3 and ULBP4, do not bind UL16 (Figure 3-33). Although ULBP5 also carries a glutamate at position 169 and should therefore bind UL16, Wittenbrink *et al.* demonstrated by mutational studies that a substitution in the α 2-domain, which is unique among all NKG2D ligands (Figure 3-33), prevents binding of ULBP5 to UL16¹³⁵. (ii) Arg169 has a similar conformation, stabilized by contacts with surrounding hydrophobic residues, in the unliganded⁶³ and liganded²⁰ MICA structures (Figure 4-5A). In this orientation, however, the Arg169 side chain would clash with UL16 residues. Modeling suggests that the arginine side chain could adopt only a single rotamer conformation, sandwiched between the hydrophobic side chain regions of Leu172 and Lys173, that would not result in steric clashes with UL16 (Figure 4-5A). However, such a rotamer is only seen in 2 % of all observed arginines¹⁰². (iii) The conformation of Arg169 in the ULBP3 structure⁶¹, which is held in place by a salt bridge to Asp170, would also clash with UL16 (Figure 4-5A). A similar arrangement of Arg169 can be expected for ULBP4, where Asp170 is replaced with glutamate (Figure 3-33). It is also important to note that Arg169 is not located near the NKG2D binding site and therefore does not play a role in the interaction of either MICA or ULBP3 with NKG2D.’¹¹³

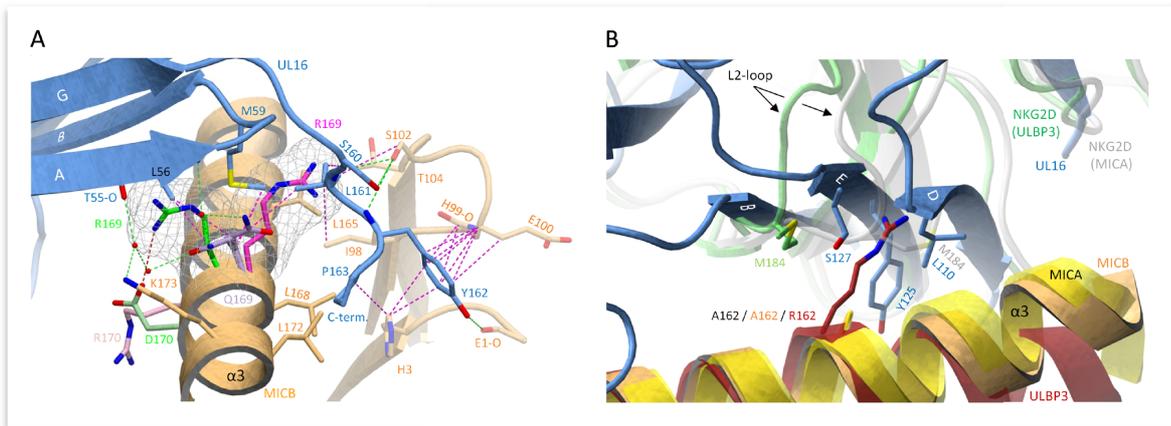


Figure 4-5. ‘Selectivity of NKG2D-ligand binding by UL16. (A) The $\alpha 1\alpha 2$ -platform domain of NKG2D-bound MICA²⁰ was superimposed onto MICB_{short}, but only the MICA side chains Arg170 (pink) and Arg169 (magenta) are shown. The $\alpha 1\alpha 2$ -domain of NKG2D-bound ULBP3⁶¹ was also superimposed onto MICB_{short}, and only the ULBP3 side chains of Arg169 (green) and Asp170 (light green) are shown. Cages surrounding the two arginines of MICA and ULBP3 at position 169 depict the area that these side chains would require in a space-filling model. In both cases, the arginine side chains would clash with UL16 residues. (B) The $\alpha 1\alpha 2$ -platform domain of NKG2D-bound MICA²⁰ (yellow) and ULBP3⁶¹ (red), respectively, was superimposed onto MICB_{short} (orange). The side chains of alanine (present in MICB and MICA) and arginine (present in ULBP3) at position 162 are shown. Also shown are the Met184 side chains of both the MICA-bound (white) and ULBP3-bound (green) NKG2D monomers, both of which correspond to the green NKG2D monomer in Figure 4-4A and B. Conformational changes of the L2-loop of MICA-bound NKG2D displaces Met184 and allows for the accommodation of Arg162 in ULBP3-bound NKG2D. In UL16, the rigid DEB sheet does not allow for a similar conformational adjustment, and ULBP3 residue Arg162 would therefore clash with UL16 residues.’¹¹³

‘A second important requirement for binding of NKG2D ligands to UL16 is the presence of a small hydrophobic side chain at position 162. In the UL16_{mon}-MICB_{short} complex, Ala162 faces towards Tyr125, a UL16 footprint residue (Figure 4-4C and Figure 4-5B). With the exception of ULBP3, which has an arginine at this position, all other NKG2D ligands have either an alanine or a glycine at position 162 (Figure 3-33A Figure 4-6). The long and positively charged Arg side chain of ULBP3 would clash with several UL16 residues (Figure 4-5B), likely contributing to the failure of UL16 to bind ULBP3^{14, 19, 65, 85} (Table 3-1). Interestingly, Arg162 would also clash with Met184 of the L2-loop of NKG2D in its MICA-liganded form. To allow for ULBP3 binding, NKG2D undergoes a conformational adjustment in which the L2-loop displaces Met184, resulting in sufficient space for the accommodation of Arg162 (Figure 4-5B). However, the rigid DEB β -sheet of UL16, which would not allow for such larger conformational adjustments, is unlikely to accommodate Arg162.’¹¹³

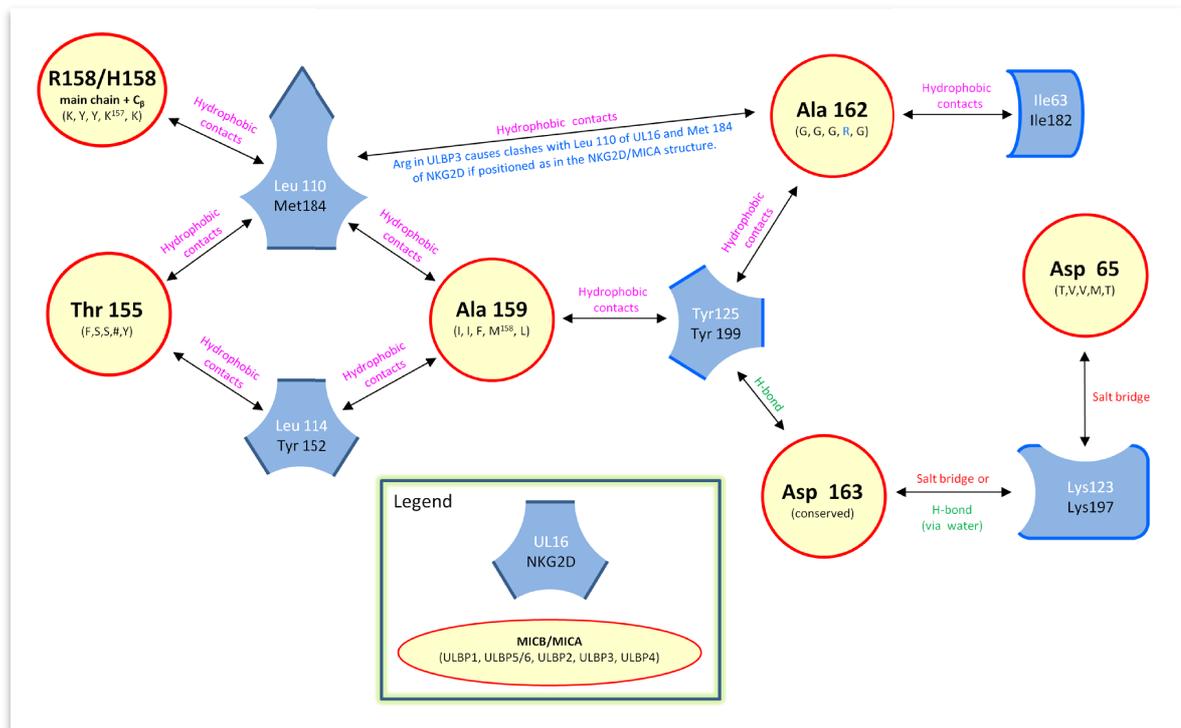


Figure 4-6. ‘Schematic view of the structural mimicry of UL16. The blue regions highlight the five UL16 and NKG2D footprint residues participating in structural mimicry. UL16 residues are shown in white, the corresponding NKG2D residues are shown in black. MICA and MICB residues that are contacted by the footprint are placed in yellow circles, at the approximate position of interaction. Also shown are the amino acids at corresponding positions in ULBP1, ULBP5/6, ULBP2, ULBP3 and ULBP4. In ULBP3⁶¹, a kink in the α -3-helix starting at position 162 (Figure 4-5B) causes a one-residue shift towards the N-terminus. In these cases, the shifted ULBP3 residue that corresponds to the MICB residue is given by a superscript number following the ULBP3 one letter code. As an example, ULBP3 position Met168 and not Val169 corresponds to MICB position Ala159. Also as a result of the helix kink, no ULBP3 residue corresponds in space to the MICB residue in position 155, indicated by (#). Interactions between residues are represented with arrows, accompanied by green text for hydrogen bonds, red text for salt bridges, and magenta text for hydrophobic contacts; the blue text indicates the clash of ULBP3 Arg162 (Figure 3-33A) with Leu100 of UL16 or Met184 of NKG2D as observed in the MICA/NKG2D complex structure²⁰ (Figure 4-5B).’¹¹³

4.2.4 UL16 and the evolutionary diversification of NKG2D ligands

‘Taken together, these analyses suggest that some NKG2D ligands apparently bypass intracellular retention by UL16 through alteration of a small number of key residues at strategic locations of their potential UL16 binding interface. It is therefore quite likely that the selective pressure exerted by UL16 contributed to drive the diversification of NKG2D ligands, which eventually may have led to the emergence of non-UL16 binding ligands such as MICA and ULBP3^{14, 19, 38, 56, 65}. Further support for an HCMV-driven diversification of NKG2D ligands comes from studies by Cosman and colleagues showing that the HCMV immunoevasin UL142 targets most MICA allelic variants except MICA*008^{14, 38, 91, 93}. Intriguingly, MICA*008 contains a truncated cytoplasmic domain and is by far the most frequent MICA variant in many populations^{91, 93}. As yet, no direct interaction of

UL142 and MICA has been shown and the molecular mechanisms of MICA sequestration by UL142 are unknown. In contrast to UL16, UL142 and MCMV-encoded immunoevasins m145, m152, and m155 that suppress surface expression of mouse NKG2D ligands MULT-1, RAE-1, and H60, respectively, are predicted to have an MHC class I-like fold^{14, 56, 91, 146-150}. It will be of great interest to determine the structural basis of NKG2D-ligand engagement by MHC class I-like HCMV immunoevasins and to compare these interactions of two MHC class I-like molecules to those of the NKG2D-like ligand binding mode of UL16.¹¹³

'NK receptors binding to MHC class I or class I-like molecules belong to two structurally distinct families, the Ig superfamily and the C-type lectin superfamily¹⁵¹. While NKG2D belongs to the latter group, the structural analysis presented here shows that UL16 assumes an Ig-like fold. Therefore, one may ask whether UL16 is related to the Ig-like NK receptors that bind MHC class I molecules, such as the leukocyte Ig-like receptors (LIRs) or the killer immunoglobulin-like receptors (KIRs). Structures of LIR-1 in complex with HLA-A2¹⁵¹ and with the HCMV MHC class I decoy UL18⁸² show that, in both cases, LIR-1 contacts β 2-microglobulin and the α 3-domain of the HLA-A2 and UL18 ligands via loops located at the interdomain hinge region of its two tandem Ig domains. In contrast KIRs, like UL16, engage the α -helical parts of the platform domain of MHC class I molecules, but, similar to LIRs, employ loops at the interdomain hinge region of their Ig domains for this interaction¹⁵¹. Therefore, LIRs and KIRs exhibit an MHC class I-binding mode that is distinct from that used by UL16. Since there is also no obvious sequence homology between these Ig-like NK receptors and UL16, this supports the view that UL16 evolved independently, mimicking a central binding motif of the structurally unrelated NKG2D immunoreceptor.'¹¹³

4.2.5 Accession numbers

Atomic coordinates and structure factors of the UL16_{mon}-MICB_{short} complex have been deposited with the Protein Data Bank under accession code 2wy3.¹¹³

5. Conclusions

The activating immunoreceptor NKG2D stimulates anti-viral and anti-tumor effector responses of human cytotoxic lymphocytes such as natural killer (NK) cells, CD8 $\alpha\beta$ T cells and $\gamma\delta$ T cells. In contrast to many other immunoreceptors, NKG2D is fairly promiscuous in its ability to recognize a diverse array of at least eight human MHC class I-related self-ligands. These NKG2D ligands are upregulated in a cell stress-associated manner and thereby allow for immunorecognition of infected or transformed cells. Several diverse and independent immune evasive strategies are employed by HCMV to avert an NKG2D-mediated immunorecognition, underscoring the importance of this particular surveillance system. One of the most well-known and extensively studied strategies of HCMV to subvert NKG2D-mediated immune responses is mediated by the HCMV glycoprotein UL16. UL16 counteracts NKG2D-mediated antiviral immune responses by intracellular retention of the distantly related NKG2D ligands MICB, ULBP1, ULBP2 and ULBP6. However, other NKG2D ligands such as the MICB-homolog MICA, the ULBP2-homolog RAET1G, and ULBPs ULBP3 and ULBP4 are not targeted by UL16. These findings contributed to the hypothesis that diversification of NKG2D ligands has been driven by viral immunoevasins rather than by transformed cells.

The work presented in this thesis has led to the crystal structure determination of the HCMV immunoevasin UL16 in complex with the NKG2D immunoreceptor ligand MICB at 1.8 Å resolution. This is, to the best of the author's knowledge, the first structure of a viral immunoevasin in complex with a stimulatory NK cell receptor ligand. Furthermore, the affinities and kinetics of the various UL16–NKG2D-ligand interactions were determined by surface plasmon resonance (SPR). Major new findings based on the obtained structural and functional data are listed on the subsequent two pages in respect to the aims formulated at the onset of this work.

What is the three-dimensional structure of UL16?

- ⇒ **Determination of the crystal structure of UL16 in complex with MICB at 1.8 Å resolution.** UL16 adopts an immunoglobulin-like (Ig-like) fold and, similarly to the C-type lectin-like receptor NKG2D, binds across the α -helices of the MICB $\alpha 1\alpha 2$ -platform domain. Although Ig-like proteins share a common structural fold, their sequence identities are often extremely low. This explains why the Ig-like fold of UL16 could not be predicted from its primary sequence. Comparison of the UL16-MICB complex with known crystal structures of MICA and ULBP3 in complex with NKG2D allowed for an insightful comparative analyses of interaction modes by NKG2D and UL16, respectively.

What are the structural features allowing UL16 to (partially) recapitulate the promiscuous engagement of highly diverse ligands by NKG2D?

- ⇒ **In a rare example of structural mimicry (presumably due to convergent evolution) Ig-like UL16 uses a set of residues at the center of its BDE β -sheet to mimic a central binding motif employed by the structurally unrelated lectin-like NKG2D immunoreceptor to interact with its diverse MHC class I-related ligands.** This result may be also of more general interest because it strongly suggests that HCMV and humans have independently evolved two structurally distinct receptors that nevertheless share the same central ligand binding motif in order to achieve promiscuous binding to MIC and ULBP molecules. Whereas the great majority of structurally defined virulence factors simply constitute structurally homologous versions of host proteins that were acquired through hijacking of host genes, only two structurally validated examples of viral proteins have been described that, similarly to UL16, likely are products of convergent evolution.

What are the binding parameters (kinetic rate constants and affinities) of the individual UL16-NKG2D ligand interactions?

- ⇒ **Surface plasmon resonance measurements showed that UL16 binds MICB, ULBP1, and ULBP2 with similar affinities that lie in the nanomolar range (12 - 66 nM).** The off-rate constants, which represent a more direct measure of the complex stability, ranged between 0.016 s^{-1} for ULBP1 and 0.054 s^{-2} for MICB_{short}. No binding was observed between UL16 and NKG2D ligands MICA, ULBP3, ULBP4 and ULBP5. ULBP6 was not tested.

What is the structural basis for UL16 binding to a subset of diverse NKG2D ligands (MICB, ULBP1 and ULBP2), and for its inability to bind the remaining NKG2D ligands (MICA, ULBP3 and RAET1G)?

⇒ **Comparison of UL16-MICB with the NKG2D-MICA and NKG2D-ULBP3 complexes reveals that substitutions at a few central positions of MICB that are key to engaging UL16 (e. g. Q169R) allow MICA and ULBP3 to bypass UL16 engagement without significantly affecting NKG2D binding.**

Can structural data provide support for the hypothesis that the diversification of NKG2D ligands resulted from selective pressure exerted by viral immunoevasins?

⇒ **The finding that UL16 mimics a central binding motif of the NKG2D immunoreceptor in order to engage MICB suggests that the UL16 escape variants among the NKG2D ligands have evolved through alteration of key residues at strategic locations. Assuming that ULBP1 and ULBP2 engage UL16 in a manner similar to MICB, this strongly supports a scenario in which selective pressure exerted by viral immunoevasins such as UL16 was and probably still is a major driving force of the NKG2D ligand diversification.**

Taken together this work provides new insights into the structural basis of the immunological arms race between a persistent human pathogen (HCMV) and cellular surveillance systems (NK cells) of the human immune system, exemplified by the promiscuous binding mode of the HCMV immunoevasin UL16 and the diversification of NKG2D ligands.

6. Zusammenfassung

Der aktivierende Immunrezeptor NKG2D stimuliert antivirale und antitumorale Immunantworten humaner zytotoxischer Lymphozyten wie etwa Natürlicher Killerzellen (NK Zellen), CD8 $\alpha\beta$ T-Zellen und $\gamma\delta$ T-Zellen. NKG2D ist in der Lage mindestens acht in ihrer Sequenz zum Teil sehr unterschiedliche MHC Klasse I-ähnliche, körpereigene Liganden zu erkennen und zeigt somit ein im Gegensatz zu vielen anderen Immunrezeptoren sehr promiskuitives Ligandenbindungsverhalten. Zellulärer Stress wie beispielsweise eine virale Infektion oder eine maligne Transformation führt zu einer Erhöhung der NKG2D Ligandendichte auf der Oberfläche der gestressten Zellen und ermöglicht auf diese Weise die NKG2D-vermittelte Erkennung virusinfizierter bzw. bösartiger Zellen. Die besondere Bedeutung der NKG2D-vermittelten Immunüberwachung wird unter anderem durch den humanen Zytomegalovirus (HCMV) verdeutlicht, der zahlreiche Evasionsmechanismen entwickelt hat, um einer Erkennung durch NKG2D zu entgehen. Eine der in diesem Sinne bekanntesten und bestuntersuchten Evasionsstrategien basiert auf dem HCMV Glykoprotein UL16. UL16 wirkt einer NKG2D-vermittelten antiviralen Immunantwort entgegen, indem es die in ihrer Sequenz zum Teil sehr unterschiedlichen NKG2D Liganden MICB, ULBP1, ULBP2 und ULBP6 (RAET1L) im Inneren der Zelle zurückhält und somit eine Erhöhung der Oberflächenexpression dieser Liganden verhindert. Interessanterweise existieren aber auch NKG2D Liganden wie beispielsweise das zu MICB homologe MICA, das zu ULBP2 homologe ULBP5 (RAET1G) sowie ULBP3 und ULBP4, die nicht von UL16 erkannt und in der Zelle zurückgehalten werden. Basierend auf diesen Ergebnissen wurde die Hypothese formuliert, dass die starke Diversifikation der NKG2D Liganden ihre Ursache möglicherweise weniger in einem durch bösartige Zellen, als vielmehr in einem durch virale Immunevasine hervorgerufenen Selektionsdruck hat.

Die hier präsentierte Arbeit beschreibt den Weg zur Lösung der Kristallstruktur des HCMV Immunevasins UL16 im Komplex mit dem NKG2D Immunrezeptorliganden MICB. Die Auflösung der Struktur beträgt 1.8 Å und ist, nach bestem Wissen des Autors dieser Arbeit, die erste eines viralen Immunevasins im Komplex mit einem Liganden eines stimulierenden NK Rezeptors. Zudem wurden die Affinitäten und Kinetiken der verschiedenen UL16-NKG2D Ligandeninteraktionen mittels Oberflächenplasmonresonanz (SPR) ermittelt. Die auf Basis dieser strukturellen und funktionellen Daten gewonnenen neuen Erkenntnisse werden auf den nachfolgenden Seiten im einzelnen aufgeführt und insbesondere in Bezug zu den am Anfang dieser Arbeit formulierten Ziele gesetzt.

Wie sieht die dreidimensionale Struktur von UL16 aus?

- ⇒ **Bestimmung der Kristallstruktur des UL16-MICB Komplexes mit einer Auflösung von 1.8 Å.** UL16 zeigt eine immunglobulinähnliche (Ig-ähnliche) Faltung und bindet, ähnlich wie der C-Typ lektinähnliche Rezeptor NKG2D, die α -Helices der MICB $\alpha 1\alpha 2$ Plattformdomäne. Trotz ihrer sehr ähnlichen strukturellen Faltung weisen Ig-ähnliche Proteine oft größere Unterschiede in ihrer Aminosäuresequenz auf. Dies wiederum macht deutlich warum die Ig-ähnliche Faltung des UL16 Proteins nicht aufgrund seiner Aminosäuresequenz vorhergesagt werden konnte. Zudem führte ein Vergleich des UL16-MICB Komplexes mit den ebenfalls bekannten Kristallstrukturen von MICA und ULBP3 im Komplex mit NKG2D zu einer aufschlussreichen vergleichenden Analyse der Interaktionsmodi von NKG2D bzw. UL16.

Welche strukturellen Eigenschaften erlauben es UL16 den promiskuitiven Ligandenbindemodus von NKG2D partiell zu kopieren?

- ⇒ **In einem seltenen Beispiel von struktureller Mimikry (vermutlich das Resultat einer konvergenten Evolution) imitiert UL16 mittels eines Satzes von Aminosäuren im Zentrum seines BDE β -Faltblatts ein zentrales Bindemotiv des zu ihm strukturell völlig unterschiedlichen NKG2D Immunrezeptors, um so mit dessen unterschiedlichen MHC Klasse I-ähnlichen Liganden interagieren zu können.** Dieses Ergebnis ist von allgemeinem Interesse, denn es zeigt, dass HCMV und der Mensch vollständig unabhängig voneinander zwei strukturell völlig unterschiedliche Rezeptoren entwickelt haben, die jedoch das gleiche zentrale Bindemotiv teilen, um so die promiskuitive Bindung an die MIC und ULBP Moleküle zu gewährleisten. Während die große Mehrheit der bislang strukturell bekannten Virulenzfaktoren lediglich homologe Versionen von Wirtsproteinen darstellen, die vom Virus durch die Übernahme von Wirtsgenen erlangt wurden, sind nur zwei strukturell belegte Beispiele bekannt bei denen die viralen Proteine, ähnlich wie UL16, wahrscheinlich das Produkt einer konvergenten Evolution darstellen.

Wie unterscheiden sich die Bindeparameter (kinetische Ratenkonstanten und Affinitäten) der verschiedenen UL16-NKG2D Ligandeninteraktionen?

- ⇒ **Oberflächenplasmonresonanzmessungen haben gezeigt, dass UL16 die NKG2D Liganden MICB, ULBP1 und ULBP2 mit ähnlichen, im nanomolaren Bereich (12-66 nM) angesiedelten, Affinitäten bindet.** Die Dissoziationsratenkonstanten, die ein direktes Maß für die Stabilität eines Komplexes darstellen, lagen im Bereich zwischen $0,016 \text{ s}^{-1}$ für ULBP1 und $0,054 \text{ s}^{-1}$ für MICB_{short}. Keine Bindung wurde zwischen UL16 und den NKG2D Liganden MICA, ULBP3, ULBP4 und ULBP5 beobachtet. ULBP6 wurde nicht getestet.

Was ist die strukturelle Basis, die es UL16 erlaubt bestimmte NKG2D Liganden wie MICB, ULBP1 und ULBP2 zu binden, während eine Bindung an andere NKG2D Liganden wie MICA, ULBP3 und RAET1G nicht möglich ist?

- ⇒ **Ein Vergleich des UL16-MICB Komplexes mit den NKG2D-MICA und NKG2D-ULBP3 Komplexen zeigt, dass der Austausch einiger weniger MICB Aminosäuren an Positionen die von zentraler Bedeutung für die UL16 Bindung sind (z. B. Q169R) es MICA und ULBP3 erlauben einer UL16 Bindung zu entgehen ohne gleichzeitig negative Auswirkungen auf deren Bindung an NKG2D zu haben.**

Können strukturelle Daten die Hypothese untermauern, dass die Diversifikation der NKG2D Liganden das Resultat eines Selektionsdruckes seitens viraler Immunevasine ist?

- ⇒ **Die Tatsache dass UL16 ein zentrales Bindemotiv des NKG2D Immunrezeptors nachahmt um MICB binden zu können deutet darauf hin, dass sich UL16 „escape“ Varianten unter den NKG2D Liganden durch den Austausch von Schlüsselaminosäureresten an für die Bindung an UL16 strategisch wichtigen Positionen entwickelt haben. Unter der Annahme, dass ULBP1 und ULBP2 UL16 in ähnlicher Weise binden wie MICB, kann man davon ausgehen, dass selektiver Druck durch virale Immunevasine wie UL16 die treibende Kraft der NKG2D Ligandendiversifikation war und vermutlich immer noch ist.**

Zusammengefasst liefert die vorliegende Arbeit am Beispiel des promiskuitiven Bindemodus des HCMV Immunevasins UL16 und der Diversifikation der NKG2D Liganden neue Einblicke in die molekulare Basis des immunologischen Wettlaufs zwischen einem persistierenden humanen Pathogen (HCMV) und zellulären Überwachungssystemen (NK Zellen) des menschlichen Immunsystems.

7. References

1. Dranoff, G. Cytokines in cancer pathogenesis and cancer therapy. *Nat Rev Cancer* **4**, 11-22 (2004).
2. Blum, K.S. & Pabst, R. Lymphocyte numbers and subsets in the human blood. Do they mirror the situation in all organs? *Immunol Lett* **108**, 45-51 (2007).
3. Trepel, F. Number and distribution of lymphocytes in man. A critical analysis. *Klin Wochenschr* **52**, 511-515 (1974).
4. Market, E. & Papavasiliou, F.N. V(D)J recombination and the evolution of the adaptive immune system. *PLoS Biol* **1**, E16 (2003).
5. Silverstein, A.M. The Clonal Selection Theory: what it really is and why modern challenges are misplaced. *Nat Immunol* **3**, 793-796 (2002).
6. Kiessling, R., Klein, E. & Wigzell, H. "Natural" killer cells in the mouse. I. Cytotoxic cells with specificity for mouse Moloney leukemia cells. Specificity and distribution according to genotype. *Eur J Immunol* **5**, 112-117 (1975).
7. Herberman, R.B., Nunn, M.E. & Lavrin, D.H. Natural cytotoxic reactivity of mouse lymphoid cells against syngeneic acid allogeneic tumors. I. Distribution of reactivity and specificity. *Int J Cancer* **16**, 216-229 (1975).
8. Yokoyama, W.M. Mistaken notions about natural killer cells. *Nat Immunol* **9**, 481-485 (2008).
9. Cooper, M.A., Fehniger, T.A. & Caligiuri, M.A. The biology of human natural killer-cell subsets. *Trends Immunol* **22**, 633-640 (2001).
10. Farag, S.S. & Caligiuri, M.A. Human natural killer cell development and biology. *Blood Rev* **20**, 123-137 (2006).
11. Caligiuri, M.A. Human natural killer cells. *Blood* **112**, 461-469 (2008).
12. Cerwenka, A. & Lanier, L.L. Natural killer cells, viruses and cancer. *Nat Rev Immunol* **1**, 41-49 (2001).
13. Lodoen, M.B. & Lanier, L.L. Natural killer cells as an initial defense against pathogens. *Curr Opin Immunol* **18**, 391-398 (2006).
14. Lanier, L.L. Evolutionary struggles between NK cells and viruses. *Nat Rev Immunol* **8**, 259-268 (2008).
15. Zamai, L. *et al.* NK cells and cancer. *J Immunol* **178**, 4011-4016 (2007).
16. Waldhauer, I. & Steinle, A. NK cells and cancer immunosurveillance. *Oncogene* **27**, 5932-5943 (2008).
17. Yokoyama, W.M. & Plougastel, B.F. Immune functions encoded by the natural killer gene complex. *Nat Rev Immunol* **3**, 304-316 (2003).
18. Lanier, L.L. NK cell recognition. *Annu Rev Immunol* **23**, 225-274 (2005).
19. Eagle, R.A. & Trowsdale, J. Promiscuity and the single receptor: NKG2D. *Nat Rev Immunol* **7**, 737-744 (2007).
20. Li, P. *et al.* Complex structure of the activating immunoreceptor NKG2D and its MHC class I-like ligand MICA. *Nat Immunol* **2**, 443-451 (2001).
21. Barten, R., Torkar, M., Haude, A., Trowsdale, J. & Wilson, M.J. Divergent and convergent evolution of NK-cell receptors. **22**, 52-57 (2001).
22. Sun, P.D. Structure and function of natural-killer-cell receptors. *Immunol Res* **27**, 539-548 (2003).
23. Radaev, S. & Sun, P.D. Structure and function of natural killer cell surface receptors. *Annu Rev Biophys Biomol Struct* **32**, 93-114 (2003).
24. Moretta, L. & Moretta, A. Killer immunoglobulin-like receptors. *Current Opinion in Immunology* **16**, 626-633 (2004).
25. Drickamer, K. C-type lectin-like domains. *Curr Opin Struct Biol* **9**, 585-590 (1999).

26. Weis, W.I., Taylor, M.E. & Drickamer, K. The C-type lectin superfamily in the immune system. *Immunol Rev* **163**, 19-34 (1998).
27. Zelensky, A.N. & Gready, J.E. The C-type lectin-like domain superfamily. *FEBS J* **272**, 6179-6217 (2005).
28. Zelensky, A.N. & Gready, J.E. Comparative analysis of structural properties of the C-type-lectin-like domain (CTLD). *Proteins* **52**, 466-477 (2003).
29. Weis, W.I., Kahn, R., Fourme, R., Drickamer, K. & Hendrickson, W.A. Structure of the calcium-dependent lectin domain from a rat mannose-binding protein determined by MAD phasing. *Science* **254**, 1608-1615 (1991).
30. Llera, A.S., Viedma, F., Sanchez-Madrid, F. & Tormo, J. Crystal structure of the C-type lectin-like domain from the human hematopoietic cell receptor CD69. *J Biol Chem* **276**, 7312-7319 (2001).
31. Halaby, D.M. & Mornon, J.P. The immunoglobulin superfamily: an insight on its tissular, species, and functional diversity. *J Mol Evol* **46**, 389-400 (1998).
32. Harpaz, Y. & Chothia, C. Many of the immunoglobulin superfamily domains in cell adhesion molecules and surface receptors belong to a new structural set which is close to that containing variable domains. *J Mol Biol* **238**, 528-539 (1994).
33. Halaby, D.M., Poupon, A. & Mornon, J. The immunoglobulin fold family: sequence analysis and 3D structure comparisons. *Protein Eng* **12**, 563-571 (1999).
34. Barclay, A.N. Ig-like domains: evolution from simple interaction molecules to sophisticated antigen recognition. *Proc Natl Acad Sci U S A* **96**, 14672-14674 (1999).
35. Bork, P., Holm, L. & Sander, C. The immunoglobulin fold. Structural classification, sequence patterns and common core. *J Mol Biol* **242**, 309-320 (1994).
36. Williams, A.F. & Barclay, A.N. The immunoglobulin superfamily--domains for cell surface recognition. *Annu Rev Immunol* **6**, 381-405 (1988).
37. Kelley, J. & Trowsdale, J. Features of MHC and NK gene clusters. *Transplant Immunology* **14**, 129-134 (2005).
38. Jonjic, S., Babic, M., Polic, B. & Krmpotic, A. Immune evasion of natural killer cells by viruses. *Curr Opin Immunol* **20**, 30-38 (2008).
39. Lodoen, M.B. & Lanier, L.L. Viral modulation of NK cell immunity. *Nat Rev Microbiol* **3**, 59-69 (2005).
40. Ljunggren, H.G. & Karre, K. In search of the 'missing self': MHC molecules and NK cell recognition. *Immunol Today* **11**, 237-244 (1990).
41. Lanier, L.L. Up on the tightrope: natural killer cell activation and inhibition. *Nat Immunol* **9**, 495-502 (2008).
42. Lopez-Larrea, C., Suarez-Alvarez, B., Lopez-Soto, A., Lopez-Vazquez, A. & Gonzalez, S. The NKG2D receptor: sensing stressed cells. *Trends Mol Med* **14**, 179-189 (2008).
43. Gonzalez, S., Groh, V. & Spies, T. Immunobiology of human NKG2D and its ligands. *Curr Top Microbiol Immunol* **298**, 121-138 (2006).
44. Vivier, E., Tomasello, E., Baratin, M., Walzer, T. & Ugolini, S. Functions of natural killer cells. *Nat Immunol* **9**, 503-510 (2008).
45. Long, E.O. *et al.* Killer cell inhibitory receptors: diversity, specificity, and function. *Immunological Reviews* **155**, 135-144 (1997).
46. Diefenbach, A. & Raulet, D.H. Strategies for target cell recognition by natural killer cells. *Immunological Reviews* **181**, 170-184 (2001).
47. Raulet, D.H. Roles of the NKG2D immunoreceptor and its ligands. *Nat Rev Immunol* **3**, 781-790 (2003).
48. Lanier, L.L. NKG2D in innate and adaptive immunity. *Adv Exp Med Biol* **560**, 51-56 (2005).
49. Bauer, S. *et al.* Activation of NK cells and T cells by NKG2D, a receptor for stress-inducible MICA. *Science* **285**, 727-729 (1999).

50. Kubin, M. *et al.* ULBP1, 2, 3: novel MHC class I-related molecules that bind to human cytomegalovirus glycoprotein UL16, activate NK cells. *Eur J Immunol* **31**, 1428-1437 (2001).
51. Roberts, A.I. *et al.* NKG2D receptors induced by IL-15 costimulate CD28-negative effector CTL in the tissue microenvironment. *J Immunol* **167**, 5527-5530 (2001).
52. Sutherland, C.L. *et al.* UL16-binding proteins, novel MHC class I-related proteins, bind to NKG2D and activate multiple signaling pathways in primary NK cells. *J Immunol* **168**, 671-679 (2002).
53. Castriconi, R. *et al.* Transforming growth factor beta 1 inhibits expression of NKp30 and NKG2D receptors: consequences for the NK-mediated killing of dendritic cells. *Proc Natl Acad Sci U S A* **100**, 4120-4125 (2003).
54. Friese, M.A. *et al.* RNA interference targeting transforming growth factor-beta enhances NKG2D-mediated antiglioma immune response, inhibits glioma cell migration and invasiveness, and abrogates tumorigenicity in vivo. *Cancer Res* **64**, 7596-7603 (2004).
55. Wu, J. *et al.* An activating immunoreceptor complex formed by NKG2D and DAP10. *Science* **285**, 730-732 (1999).
56. Jonjic, S., Polic, B. & Krmpotic, A. Viral inhibitors of NKG2D ligands: friends or foes of immune surveillance? *Eur J Immunol* **38**, 2952-2956 (2008).
57. Mistry, A.R. & O'Callaghan, C.A. Regulation of ligands for the activating receptor NKG2D. *Immunology* **121**, 439-447 (2007).
58. Houchins, J.P., Yabe, T., McSherry, C. & Bach, F.H. DNA sequence analysis of NKG2, a family of related cDNA clones encoding type II integral membrane proteins on human natural killer cells. *J Exp Med* **173**, 1017-1020 (1991).
59. McFarland, B.J., Kortemme, T., Yu, S.F., Baker, D. & Strong, R.K. Symmetry recognizing asymmetry: analysis of the interactions between the C-type lectin-like immunoreceptor NKG2D and MHC class I-like ligands. *Structure* **11**, 411-422 (2003).
60. McFarland, B.J. & Strong, R.K. Thermodynamic analysis of degenerate recognition by the NKG2D immunoreceptor: not induced fit but rigid adaptation. *Immunity* **19**, 803-812 (2003).
61. Radaev, S., Rostro, B., Brooks, A.G., Colonna, M. & Sun, P.D. Conformational plasticity revealed by the cocrystal structure of NKG2D and its class I MHC-like ligand ULBP3. *Immunity* **15**, 1039-1049 (2001).
62. Eagle, R.A., Jafferji, I. & Barrow, A.D. Beyond Stressed Self: Evidence for NKG2D Ligand Expression on Healthy Cells. *Curr Immunol Rev* **5**, 22-34 (2009).
63. Li, P. *et al.* Crystal structure of the MHC class I homolog MIC-A, a gammadelta T cell ligand. *Immunity* **10**, 577-584 (1999).
64. Holmes, M.A., Li, P., Petersdorf, E.W. & Strong, R.K. Structural studies of allelic diversity of the MHC class I homolog MIC-B, a stress-inducible ligand for the activating immunoreceptor NKG2D. *J Immunol* **169**, 1395-1400 (2002).
65. Cosman, D. *et al.* ULBPs, novel MHC class I-related molecules, bind to CMV glycoprotein UL16 and stimulate NK cytotoxicity through the NKG2D receptor. *Immunity* **14**, 123-133 (2001).
66. Gonzalez, S., Lopez-Soto, A., Suarez-Alvarez, B., Lopez-Vazquez, A. & Lopez-Larrea, C. NKG2D ligands: key targets of the immune response. *Trends Immunol* **29**, 397-403 (2008).
67. Cao, W. & He, W. UL16 binding proteins. *Immunobiology* **209**, 283-290 (2004).
68. Stephens, H.A. MICA and MICB genes: can the enigma of their polymorphism be resolved? *Trends Immunol* **22**, 378-385 (2001).
69. Ohashi, M., Eagle, R.A. & Trowsdale, J. Post-translational modification of the NKG2D ligand RAET1G leads to cell surface expression of a GPI-linked isoform. *J Biol Chem* (2010).
70. Romphruk, A.V. *et al.* Polymorphisms of NKG2D ligands: diverse RAET1/ULBP genes in northeastern Thais. *Immunogenetics* **61**, 611-617 (2009).

71. Davison, A.J. *et al.* The order Herpesvirales. *Arch Virol* **154**, 171-177 (2009).
72. Davison, A.J. Herpesvirus systematics. *Vet Microbiol* (2010).
73. Davison, A.J. Evolution of the herpesviruses. *Vet Microbiol* **86**, 69-88 (2002).
74. Reddehase, M.J. Antigens and immunoevasins: opponents in cytomegalovirus immune surveillance. *Nat Rev Immunol* **2**, 831-844 (2002).
75. Huang, E.S. & Johnson, R.A. Human cytomegalovirus - no longer just a DNA virus. *Nat Med* **6**, 863-864 (2000).
76. Fannin Rider, P.J., Dunn, W., Yang, E. & Liu, F. Human cytomegalovirus microRNAs. *Curr Top Microbiol Immunol* **325**, 21-39 (2008).
77. Murphy, E., Rigoutsos, I., Shibuya, T. & Shenk, T.E. Reevaluation of human cytomegalovirus coding potential. *Proc Natl Acad Sci U S A* **100**, 13585-13590 (2003).
78. Powers, C., DeFilippis, V., Malouli, D. & Fruh, K. Cytomegalovirus immune evasion. *Curr Top Microbiol Immunol* **325**, 333-359 (2008).
79. Wilkinson, G.W. *et al.* Modulation of natural killer cells by human cytomegalovirus. *J Clin Virol* **41**, 206-212 (2008).
80. Farrell, H.E., Degli-Esposti, M.A. & Davis-Poynter, N.J. Cytomegalovirus evasion of natural killer cell responses. *Immunol Rev* **168**, 187-197 (1999).
81. Wiertz, E.J., Devlin, R., Collins, H.L. & Rensing, M.E. Herpesvirus interference with major histocompatibility complex class II-restricted T-cell activation. *J Virol* **81**, 4389-4396 (2007).
82. Yang, Z. & Bjorkman, P.J. Structure of UL18, a peptide-binding viral MHC mimic, bound to a host inhibitory receptor. *Proc Natl Acad Sci U S A* **105**, 10095-10100 (2008).
83. Chapman, T.L., Heikeman, A.P. & Bjorkman, P.J. The inhibitory receptor LIR-1 uses a common binding interaction to recognize class I MHC molecules and the viral homolog UL18. *Immunity* **11**, 603-613 (1999).
84. Wu, J. *et al.* Intracellular retention of the MHC class I-related chain B ligand of NKG2D by the human cytomegalovirus UL16 glycoprotein. *J Immunol* **170**, 4196-4200 (2003).
85. Welte, S.A. *et al.* Selective intracellular retention of virally induced NKG2D ligands by the human cytomegalovirus UL16 glycoprotein. *Eur J Immunol* **33**, 194-203 (2003).
86. Dunn, C. *et al.* Human cytomegalovirus glycoprotein UL16 causes intracellular sequestration of NKG2D ligands, protecting against natural killer cell cytotoxicity. *J Exp Med* **197**, 1427-1439 (2003).
87. Eagle, R.A., Traherne, J.A., Hair, J.R., Jafferji, I. & Trowsdale, J. ULBP6/RAET1L is an additional human NKG2D ligand. *Eur J Immunol* (2009).
88. Spreu, J., Stehle, T. & Steinle, A. Human cytomegalovirus-encoded UL16 discriminates MIC molecules by their alpha2 domains. *J Immunol* **177**, 3143-3149 (2006).
89. Wittenbrink, M., Spreu, J. & Steinle, A. Differential NKG2D binding to highly related human NKG2D ligands ULBP2 and RAET1G is determined by a single amino acid in the alpha2 domain. *Eur J Immunol* **39**, 1642-1651 (2009).
90. Chalupny, N.J., Sutherland, C.L., Lawrence, W.A., Rein-Weston, A. & Cosman, D. ULBP4 is a novel ligand for human NKG2D. *Biochem Biophys Res Commun* **305**, 129-135 (2003).
91. Chalupny, N.J., Rein-Weston, A., Dosch, S. & Cosman, D. Down-regulation of the NKG2D ligand MICA by the human cytomegalovirus glycoprotein UL142. *Biochem Biophys Res Commun* **346**, 175-181 (2006).
92. Stern-Ginossar, N. *et al.* Host immune system gene targeting by a viral miRNA. *Science* **317**, 376-381 (2007).
93. Ashiru, O. *et al.* NKG2D ligand MICA is retained in the cis-Golgi apparatus by human cytomegalovirus protein UL142. *J Virol* **83**, 12345-12354 (2009).
94. Pfeffer, S. *et al.* Identification of microRNAs of the herpesvirus family. *Nat Meth* **2**, 269-276 (2005).
95. Cullen, B.R. Immunology. Outwitted by viral RNAs. *Science* **317**, 329-330 (2007).

96. Cullen, B.R. Viruses and microRNAs. *Nat Genet* (2006).
97. Kaye, J., Browne, H., Stoffel, M. & Minson, T. The UL16 gene of human cytomegalovirus encodes a glycoprotein that is dispensable for growth in vitro. *J Virol* **66**, 6609-6615 (1992).
98. Vales-Gomez, M. & Reyburn, H.T. Intracellular trafficking of the HCMV immunoevasin UL16 depends on elements present in both its cytoplasmic and transmembrane domains. *J Mol Biol* **363**, 908-917 (2006).
99. Vales-Gomez, M. *et al.* The human cytomegalovirus glycoprotein UL16 traffics through the plasma membrane and the nuclear envelope. *Cell Microbiol* **8**, 581-590 (2006).
100. Collaborative Computational Project, N. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50**, 760-763 (1994).
101. Adams, P.D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* **58**, 1948-1954 (2002).
102. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132 (2004).
103. DeLano, W.L. The Pymol Molecular Graphics System. (2002).
104. Bohne-Lang, A. & von der Lieth, C.W. GlyProt: in silico glycosylation of proteins. *Nucleic Acids Res* **33**, W214-219 (2005).
105. Stanley, P. Chinese hamster ovary cell mutants with multiple glycosylation defects for production of glycoproteins with minimal carbohydrate heterogeneity. *Mol Cell Biol* **9**, 377-383 (1989).
106. Casasnovas, J.M., Larvie, M. & Stehle, T. Crystal structure of two CD46 domains reveals an extended measles virus-binding surface. *EMBO J* **18**, 2911-2922 (1999).
107. Akerstrom, B. & Bjorck, L. A physicochemical study of protein G, a molecule with unique immunoglobulin G-binding properties. *J Biol Chem* **261**, 10240-10247 (1986).
108. Sjoquist, J., Meloun, B. & Hjelm, H. Protein A isolated from *Staphylococcus aureus* after digestion with lysostaphin. *Eur J Biochem* **29**, 572-578 (1972).
109. Bradford, M.M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **72**, 248-254 (1976).
110. Stoscheck, C.M. Quantitation of protein. *Methods Enzymol* **182**, 50-68 (1990).
111. Gill, S.C. & von Hippel, P.H. Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem* **182**, 319-326 (1989).
112. Nesterenko, M.V., Tilley, M. & Upton, S.J. A simple modification of Blum's silver stain method allows for 30 minute detection of proteins in polyacrylamide gels. *J Biochem Biophys Methods* **28**, 239-242 (1994).
113. Muller, S., Zocher, G., Steinle, A. & Stehle, T. Structure of the HCMV UL16-MICB complex elucidates select binding of a viral immunoevasin to diverse NKG2D ligands. *PLoS Pathog* **6**, e1000723 (2010).
114. Persson, B.D. *et al.* An arginine switch in the species B adenovirus knob determines high-affinity engagement of cellular receptor CD46. *J Virol* **83**, 673-686 (2009).
115. Chayen, N.E. Turning protein crystallisation from an art into a science. *Curr Opin Struct Biol* **14**, 577-583 (2004).
116. Chayen, N.E. & Saridakis, E. Protein crystallization: from purified protein to diffraction-quality crystal. *Nat Methods* **5**, 147-153 (2008).
117. Saridakis, E. & Chayen, N.E. Towards a 'universal' nucleant for protein crystallization. *Trends Biotechnol* **27**, 99-106 (2009).
118. Rodgers, D.W. Cryocrystallography. *Structure* **2**, 1135-1140 (1994).
119. Bragg, W.L. The Diffraction of Short Electromagnetic Waves by a Crystal. *Proceedings of the Cambridge Philosophical Society* **17**, 43-57 (1913).
120. Ewald, P.P. Das "reziproke Gitter" in der Strukturtheorie. *Z. Kristallogr.* **56**, 129-156 (1921).

121. Kraft, P. *et al.* Performance of single-photon-counting PILATUS detector modules. *Journal of Synchrotron Radiation* **16**, 368-375 (2009).
122. Hulsen, G., Broennimann, C., Eikenberry, E.F. & Wagner, A. Protein crystallography with a novel large-area pixel detector. *Journal of Applied Crystallography* **39**, 550-557 (2006).
123. Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *Journal of Applied Crystallography* **26**, 795-800 (1993).
124. Diederichs, K. & Karplus, P.A. Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nat Struct Biol* **4**, 269-275 (1997).
125. Wilson, A. The probability distribution of X-ray intensities. *Acta Crystallographica* **2**, 318-321 (1949).
126. Matthews, B.W. Solvent content of protein crystals. *J Mol Biol* **33**, 491-497 (1968).
127. Brunger, A.T. Assessment of phase accuracy by cross validation: the free R value. Methods and applications. *Acta Crystallogr D Biol Crystallogr* **49**, 24-36 (1993).
128. Brunger, A.T. & Rice, L.M. Crystallographic refinement by simulated annealing: methods and applications. *Methods Enzymol* **277**, 243-269 (1997).
129. Heras, B. *et al.* Dehydration converts DsbG crystal diffraction from low to high resolution. *Structure* **11**, 139-145 (2003).
130. Terwilliger, T.C. SOLVE and RESOLVE: automated structure solution and density modification. *Methods Enzymol* **374**, 22-37 (2003).
131. Morris, R.J., Perrakis, A. & Lamzin, V.S. ARP/wARP and automatic interpretation of protein electron density maps. *Methods Enzymol* **374**, 229-244 (2003).
132. Painter, J. & Merritt, E.A. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr D Biol Crystallogr* **62**, 439-450 (2006).
133. Vriend, G. WHAT IF: a molecular modeling and drug design program. *J Mol Graph* **8**, 52-56, 29 (1990).
134. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577-2637 (1983).
135. Wittenbrink, M., Spreu, J. & Steinle, A. Differential NKG2D binding to highly related human NKG2D ligands ULBP2 and RAET1G is determined by a single amino acid in the alpha2 domain. *Eur J Immunol* (2009).
136. Bahadur, R.P. & Zacharias, M. The interface of protein-protein complexes: analysis of contacts and prediction of interactions. *Cell Mol Life Sci* **65**, 1059-1072 (2008).
137. Lawrence, M.C. & Colman, P.M. Shape complementarity at protein/protein interfaces. *J Mol Biol* **234**, 946-950 (1993).
138. Williams, A.F. Cellular interactions. Out of equilibrium. *Nature* **352**, 473-474 (1991).
139. Copeland, R.A., Pompliano, D.L. & Meek, T.D. Drug-target residence time and its implications for lead optimization. *Nat Rev Drug Discov* **5**, 730-739 (2006).
140. Stebbins, C.E. & Galan, J.E. Structural mimicry in bacterial virulence. *Nature* **412**, 701-705 (2001).
141. Alcami, A. Viral mimicry of cytokines, chemokines and their receptors. *Nat Rev Immunol* **3**, 36-50 (2003).
142. Lambris, J.D., Ricklin, D. & Geisbrecht, B.V. Complement evasion by human pathogens. *Nat Rev Microbiol* **6**, 132-142 (2008).
143. van Cleef, K.W., Smit, M.J., Bruggeman, C.A. & Vink, C. Cytomegalovirus-encoded homologs of G protein-coupled receptors and chemokines. *J Clin Virol* **35**, 343-348 (2006).
144. Alexander, J.M. *et al.* Structural basis of chemokine sequestration by a herpesvirus decoy receptor. *Cell* **111**, 343-356 (2002).
145. Carfi, A., Smith, C.A., Smolak, P.J., McGrew, J. & Wiley, D.C. Structure of a soluble secreted chemokine inhibitor vCCI (p35) from cowpox virus. *Proc Natl Acad Sci U S A* **96**, 12379-12383 (1999).

146. Mans, J. *et al.* Structure and function of murine cytomegalovirus MHC-I-like molecules: how the virus turned the host defense to its advantage. *Immunol Res* **43**, 264-279 (2009).
147. Smith, H.R. *et al.* Recognition of a virus-encoded ligand by a natural killer cell activation receptor. *Proc Natl Acad Sci U S A* **99**, 8826-8831 (2002).
148. Krmpotic, A. *et al.* NK cell activation through the NKG2D ligand MULT-1 is selectively prevented by the glycoprotein encoded by mouse cytomegalovirus gene m145. *J Exp Med* **201**, 211-220 (2005).
149. Lodoen, M.B. *et al.* The cytomegalovirus m155 gene product subverts natural killer cell antiviral protection by disruption of H60-NKG2D interactions. *J Exp Med* **200**, 1075-1081 (2004).
150. Lenac, T., Arapovic, J., Traven, L., Krmpotic, A. & Jonjic, S. Murine cytomegalovirus regulation of NKG2D ligands. *Med Microbiol Immunol* **197**, 159-166 (2008).
151. Deng, L. & Mariuzza, R.A. Structural basis for recognition of MHC and MHC-like ligands by natural killer cell receptors. *Semin Immunol* **18**, 159-166 (2006).

8. Publications

The following list contains all publications the author of this thesis contributed to.

Scientific Publications

2005

1. Hunter DJ, Roberts GA, Ost TW, White JH, Muller S, et al. (2005) Analysis of the domain properties of the novel cytochrome P450 RhF. *FEBS Lett* 579: 2215-2220.

2009

2. Persson BD, Muller S, Reiter DM, Schmitt BB, Marttila M, et al. (2009) An arginine switch in the species B adenovirus knob determines high-affinity engagement of cellular receptor CD46. *J Virol* 83: 673-686.

2010

3. Muller S, Zocher G, Steinle A, Stehle T (2010) Structure of the HCMV UL16-MICB complex elucidates select binding of a viral immunoevasin to diverse NKG2D ligands. *PLoS Pathog* 6: e1000723.
4. Cupelli K, Muller S, Persson BD, Jost M, Arnberg N, et al. (2010) Structure of adenovirus type 21 knob in complex with CD46 reveals key differences in receptor contacts among species B adenoviruses. *J Virol* 84: 3189-3200.

9. Abbreviations

A	absorbance
Å	Angstrom (1Å = 0.1 nm)
aa	amino acid
Ab	antibody
AC	affinity chromatography
ADCC	antibody dependent cellular cytotoxicity
Amp	ampicillin
APC	antigen presenting cell
APS	ammoniumperoxodisulphate
ASU	asymmetric unit
ATP	adenosine triphosphate
BCA	bicinchoninic acid
BCR	B cell receptor
bp	base pair
BSA	bovine serum albumin
BTP	bis-tris propane
Cam	chloramphenicol
CCD	charge-coupled device
CHO	Chinese hamster ovary
CID	cytomegalic inclusion disease
conc.	Concentration
CRD	carbohydrate recognition domain
CRD	carbohydrate recognition domain
CTL	cytotoxic T lymphocyte
CTLD	C-type lectin-like domain
CTLDcps	C-type lectin-like domain containing proteins
CTLR	C-type lectin-like receptor
Cyt	cytoplasmic tail
Da	Dalton
dd	double distilled
DNA	deoxyribonucleic acid
dNTP	2'-deoxynucleotide triphosphate
DPBS	Dulbecco's PBS
ds	double-stranded
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	ethylenediaminetetraacetic acid
ELISA	enzyme-linked immunosorbent assay
EndoH	endoglycosidase H
Eq	equilibrium
ER	endoplasmic reticulum
FCS	fetal calf serum
Fig.	figure
FPLC	fast Protein Liquid Chromatography
FT	flow through
g	earth gravity (9.81 m/s ²)
GF	gel filtration
GlcNAc	N-acetylglucosamine
GSH	glutathione, reduced

GSSG	glutathione, oxidized
GST	glutathione S-transferase
h	hour
HEPES	N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid
His-tag	histidine tag
HLA	human leukocyte antigen
hrs	hours
IB	inclusion bodies
IEX	ionexchange chromatography
Ig	immunoglobulin
IL	interleukin
IMAC	immobilized metal affinity chromatography
IPTG	isopropyl β -D-1-thiogalactopyranoside
ITAM	immunoreceptor tyrosine-based activation motif
ITIM	immunoreceptor tyrosine-based inhibitory motif
K	Kelvin (0 K = - 273.15 °C)
k_a	dissociation rate constant (also off rate constant)
Kan	kanamycin
k_d	association rate constant (also on rate constant)
K_D	dissociation constant
LB	Luria Bertani
Lec	lectin
LLG	log likelihood gain
Man	mannose residue
MBL	mannose-binding lectin
MCS	multiple cloning site
MEA	mercaptoethylamine
MHC	major histocompatibility complex
MIC	MHC class I chain-related molecule
miRNA (miR)	microRNA
MOPS	3-(N-morpholino)propanesulfonic acid
MR	molecular replacement
Mw	molecular weight
MWCO	molecular weight cut off
n	nucleotide
NAG	N-acetylglucosamine
NCS	non-crystallographic symmetry
NKc	natural killer cells
NKC	natural killer gene complex
NKD	natural killer (NK) domain
NKG2D	natural-killer group 2, member D
NKT	natural killer T cells
NTA	nitilotriacetic acid
o/n	overnight
OD	optical density
ORF	open reading frame
P20	polysorbate 20
PAGE	polyacrylamide gel electrophoresis
PAMP	pathogen associated molecular patterns
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PEG	polyethylene glycol

Abbreviations

Pfu	<i>Pyrococcus furiosus</i>
PI-3K	phosphatidylinositol-3-kinase
PMSF	phenylmethylsulfonyl fluoride
PRR	pattern recognition receptor
PSP	photostimulable phosphor plate
PVR	poliovirus receptor
r. m. s. d.	root mean square deviation
RNA	ribonucleic acid
rpm	revolutions per minute
RU	resonance unit
s	seconds
S 75 / 200 / 300	superdex 75 / 200 / 300
SDS	sodium dodecyl sulphate
SEC	size exclusion chromatography
SLS	Swiss Light Source
SOB	super optimal broth
SP	signal peptide
SPR	surface plasmon resonance
ST	stalk region
SV40	simian virus 40
Tab.	table
TAE	tris-acetate EDTA
Taq	<i>Thermus aquaticus</i>
TCR	T cell receptor
TEMED	N,N,N',N'-tetramethylethylenediamine
Temp.	temperature
Tet	tetracycline
Tfb	transformation buffer
T _m	melting temperature
TM	transmembrane domain
TM	transmembrane region/domain
Tris	tris(hydroxymethyl)aminomethane
UL	unique long
ULBP	UL16-binding protein
US	unique short
UV	ultraviolet
V	volts
v/v	volume per volume
vol.	volume
w/v	weight per volume
wt	wild-type
β2m	β ₂ -microglobulin
β-ME	2-mercaptoethanol
ε	extinction coefficient (M ⁻¹ cm ⁻¹)
λ	wavelength (nm)

10. Acknowledgements

Prof. Dr. Thilo Stehle danke ich sehr herzlich dafür, dass ich meine Doktorarbeit unter seiner Anleitung anfertigen konnte, als auch für sein fortwährendes Interesse an meiner Arbeit und seine ausgesprochen intensive Förderung. Zudem möchte ich mich für sein Vertrauen in mich und meine Arbeit bedanken und nicht zuletzt für die enormen Freiheiten die er mir bei der Umsetzung der von mir bearbeiteten Projekte gegeben hat.

Prof. Dr. Alexander Steinle danke ich sehr herzlich für die kontinuierliche Unterstützung und erstklassige Zusammenarbeit im Rahmen unseres UL16 Kooperationsprojektes sowie für seinen wertvollen Beitrag und sein Engagement beim Verfassen der UL16 Publikation.

Für seine generell große Hilfsbereitschaft, eine tolle Zusammenarbeit im Labor und seine großartige Hilfe bei der Lösung und Analyse der UL16-MICB Struktur möchte ich mich bei Georg Zocher bedanken.

Irmgard Hähnlein-Schick danke ich für die tolle Zusammenarbeit und ihre große Unterstützung im Rahmen des UL16 Projektes. Vielmehr aber noch möchte ich mich bei Irmgard dafür bedanken, dass sie immer ein offenes Ohr für mich hatte, mir in allen Dingen (beruflichen wie privaten) immer mit Rat und Tat zur Seite gestanden hat und mir immer ehrlich ihre Meinung gesagt hat.

Ein weiterer Dank geht an Jessica Spreu, deren wissenschaftliche Vorarbeiten eine wichtige Grundlage für das UL16 Projekt geschaffen haben.

Alexandra Thor hat mir sehr viel Arbeit in der Zellkultur abgenommen und auch dafür möchte ich mich natürlich bedanken.

Desweiteren möchte ich mich bei David Persson, Shawn Clark und Zaigham Mahmood Khan für die vielen hilfreichen wissenschaftlichen Diskussionen, die tolle Zusammenarbeit, viel Aufmunterung und Unterstützung und viel Spaß während und nach der Arbeit bedanken.

Gerne möchte ich mich auch bei allen anderen Mitarbeitern des Arbeitskreis Stehle für ihre Hilfsbereitschaft und die angenehme Arbeitsatmosphäre bedanken.

Benny und Patrick: Wahre Freunde wie euch findet man nur ganz selten.

Bedanken möchte ich mich im Besonderen bei meinem Patenonkel Andreas und seiner Frau Beate, die immer ein offenes Ohr für mich haben und immer da sind wenn ich sie brauche. Auch meinen Schwestern Annika und Anna-Lena, meiner Großmutter Christa und dem ganzen Rest meiner Familie möchte ich dafür danken, dass ich immer auf sie zählen kann.

Ein besonderer Dank gilt meiner Freundin Mareike, die immer zu mir hält, immer fest an mich glaubt und mich so nimmt wie ich bin. Auch Mareikes Eltern, Theo und Renate, möchte ich ganz herzlich für ihre freundliche Unterstützung in den vergangenen Jahren danken.

Den größten Dank aber möchte ich meinen Eltern, Elke und Erhard, aussprechen, die mich zeit meines Lebens immer liebevoll unterstützt und mir in allen Lagen den Rücken freigehalten haben. Bessere Eltern kann man sich nicht wünschen.

11. Academic teachers

UNIVERSITY OF TÜBINGEN, GERMANY

Prof. Ammermann, Prof. Bardele, Dr. Bayer, Prof. Bisswanger, Prof. Bock, Prof. Bohley, Prof. Christen, Prof. Dodt, Prof. Dringen, Prof. Duszenko, Prof. Eisele, Prof. Gauglitz, Prof. Grossmann, Dr. Günzl, Prof. Häfeling, Prof. Hamprecht, Prof. Jäger, Dr. Kalbacher, Prof. Klein, Prof. Krug, Prof. Lindner, Prof. Madeo, PD Dr. Maier, Prof. Mayer, Prof. Mecke, Prof. Ninnemann, Prof. Oberhammer, Prof. Pommer, Prof. Probst, Prof. Rammensee, Dr. Reinecke, Prof. Reutter, Dr. Sarrazin, Prof. Schild, Prof. Schott, Prof. Schwarz, Prof. Stehle, Prof. Steinle, PD Dr. Stoeva, Prof. Strähle, PD Dr. Verleysdonk, Prof. Voelter, Prof. Wagner, Prof. Weber, Prof. Weser, Prof. Wohlleben, Prof. Zeller.

UNIVERSITY OF EDINBURGH, UNITED KINGDOM, SCOTLAND

Prof. Chapman

12. Curriculum vitae

PERSONAL DETAILS

Name: Steffen Müller
Date of birth: June 13, 1979
Place of birth: Göttingen, Germany
Nationality: German

HIGHER EDUCATION

12/2005-05/2010 PhD thesis at the Interfaculty Institute for Biochemistry, University of Tübingen, Germany, supervised by Prof. Dr. Thilo Stehle.
Title: Structural and functional characterization of the protein-protein interaction between the HCMV immunoevasin UL16 and the NKG2D ligand MICB

10/2005 Diploma degree in Biochemistry

02/2005-10/2005 Diploma thesis at the Interfaculty Institute for Cell Biology, Department of Immunology, University of Tübingen, Germany, supervised by Prof. Dr. Hans-Georg Rammensee und Prof. Dr. Alexander Steinle.
Title: Identification of intracellular signal molecules of the NK receptor NKp80

10/2002-06/2003 Studies in Biochemistry at the Department of Chemistry, University of Edinburgh, United Kingdom, Scotland, supervised by Prof. Dr. Steve K. Chapman

10/1999-10/2005 Studies in Biochemistry, University of Tübingen, Germany

08/1998-08/1999 Alternative military service, Evangelic Hospital Weende e. V., Göttingen, Germany

SECONDARY SCHOOL EDUCATION

1995-1998 Gymnasium, Hainberg-Gymnasium Göttingen, Germany

1991-1995 Middle school, Voigt-Realschule Göttingen, Germany

13. Lebenslauf

PERSÖNLICHE DATEN

Name: Steffen Müller
 Geburtsdatum: 13 Juni 1979
 Geburtsort: Göttingen, Deutschland
 Nationalität: Deutsch

HOCHSCHULAUSBILDUNG

12/2005-05/2010 Anfertigung der Doktorarbeit am Interfakultären Institut für Biochemie, Universität Tübingen, Deutschland, unter Anleitung von Prof. Dr. Thilo Stehle.
 Titel: Strukturelle und funktionelle Charakterisierung der Protein-Protein Interaktion zwischen dem HCMV Immunevasin UL16 und dem NKG2D Liganden MICB

10/2005 Diplom in Biochemie

02/2005-10/2005 Anfertigung der Diplomarbeit am Interfakultären Institut für Zellbiologie, Abteilung Immunologie, Universität Tübingen, Deutschland, unter Anleitung von Prof. Dr. Hans-Georg Rammensee und Prof. Dr. Alexander Steinle.
 Titel: Identifikation intrazellulärer Signalmoleküle des NK Rezeptors NKp80

10/2002-06/2003 Studium der Biochemie, Institut für Chemie, Universität Edinburg, Vereinigtes Königreich, Schottland, unter Anleitung von Prof. Dr. Steve K. Chapman

10/1999-10/2005 Studium der Biochemie, Universität Tübingen, Deutschland

08/1998-08/1999 Wehersatzdienst, Evangelisches Krankenhaus Weende e. V., Göttingen, Deutschland

WEITERFÜHRENDE SCHULBILDUNG

1995-1998 Gymnasium, Hainberg-Gymnasium Göttingen, Deutschland
 Abschluss: Abitur

1991-1995 Realschule, Voigt-Realschule Göttingen, Deutschland
 Abschluss: Erweiterter Sekundarabschluss I